

SECOND EDITION

CHEMICAL ENGINEERING

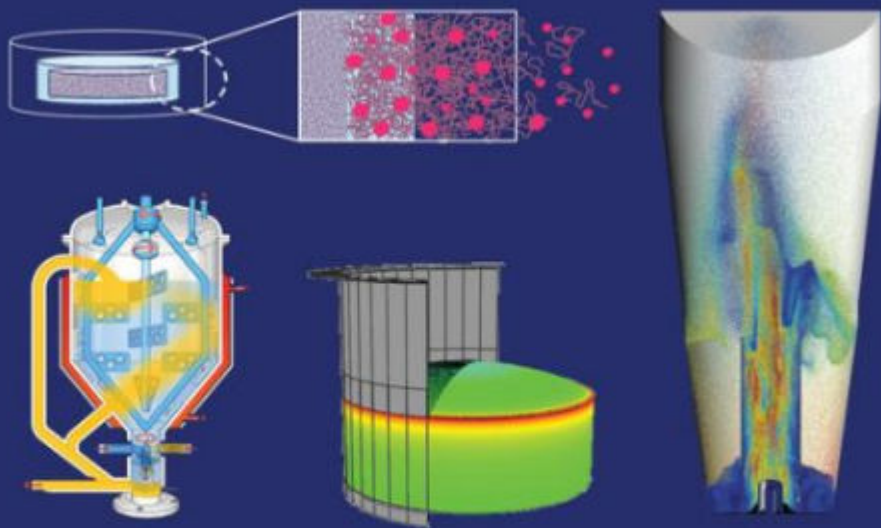
IN THE PHARMACEUTICAL INDUSTRY

DRUG PRODUCT DESIGN, DEVELOPMENT AND MODELING

EDITED BY

MARY T. AM ENDE

DAVID J. AM ENDE



CHEMICAL ENGINEERING IN THE PHARMACEUTICAL INDUSTRY



<https://t.me/ebookgate>

CHEMICAL ENGINEERING IN THE PHARMACEUTICAL INDUSTRY

Drug Product Design, Development, and Modeling

Second Edition

Edited by

MARY T. AM ENDE

Lyndra Therapeutics, Inc., Watertown, MA, USA

DAVID J. AM ENDE

Nalas Engineering Services, Inc., Centerbrook, CT, USA

WILEY

AIChE 
The Global Home of Chemical Engineers

This edition first published 2019
© 2019 John Wiley & Sons, Inc.

A Joint Publication of the American Institute of Chemical Engineers and John Wiley & Sons, Inc.

Edition History

Chemical Engineering in the Pharmaceutical Industry, First edition, John Wiley & Sons, 2011.

All rights reserved. No part of this publication may be reproduced, stored in a retrieval system, or transmitted, in any form or by any means, electronic, mechanical, photocopying, recording or otherwise, except as permitted by law. Advice on how to obtain permission to reuse material from this title is available at <http://www.wiley.com/go/permissions>.

The right of Mary T. am Ende and David J. am Ende to be identified as the authors of the editorial material in this work has been asserted in accordance with law.

Registered Office

John Wiley & Sons, Inc., 111 River Street, Hoboken, NJ 07030, USA

Editorial Office

111 River Street, Hoboken, NJ 07030, USA

For details of our global editorial offices, customer services, and more information about Wiley products visit us at www.wiley.com.

Wiley also publishes its books in a variety of electronic formats and by print-on-demand. Some content that appears in standard print versions of this book may not be available in other formats.

Limit of Liability/Disclaimer of Warranty

In view of ongoing research, equipment modifications, changes in governmental regulations, and the constant flow of information relating to the use of experimental reagents, equipment, and devices, the reader is urged to review and evaluate the information provided in the package insert or instructions for each chemical, piece of equipment, reagent, or device for, among other things, any changes in the instructions or indication of usage and for added warnings and precautions. While the publisher and authors have used their best efforts in preparing this work, they make no representations or warranties with respect to the accuracy or completeness of the contents of this work and specifically disclaim all warranties, including without limitation any implied warranties of merchantability or fitness for a particular purpose. No warranty may be created or extended by sales representatives, written sales materials or promotional statements for this work. The fact that an organization, website, or product is referred to in this work as a citation and/or potential source of further information does not mean that the publisher and authors endorse the information or services the organization, website, or product may provide or recommendations it may make. This work is sold with the understanding that the publisher is not engaged in rendering professional services. The advice and strategies contained herein may not be suitable for your situation. You should consult with a specialist where appropriate. Further, readers should be aware that websites listed in this work may have changed or disappeared between when this work was written and when it is read. Neither the publisher nor authors shall be liable for any loss of profit or any other commercial damages, including but not limited to special, incidental, consequential, or other damages.

Library of Congress Cataloging-in-Publication Data

Names: Ende, Mary Tanya am, editor. | Ende, David J. am, editor.

Title: Chemical engineering in the pharmaceutical industry : drug product design, development and modeling / edited by Mary Tanya am Ende, Global R&D, Pfizer, Inc., Groton, CT, USA, David J. am Ende, Nalas Engineering Services, Inc., Centerbrook, CT, USA.

Description: Second edition. | Hoboken, NJ, USA : Wiley, 2019. | Includes bibliographical references and index. |

Identifiers: LCCN 2018054741 (print) | LCCN 2018055054 (ebook) | ISBN 9781119285519 (AdobePDF) | ISBN 9781119285502 (ePub) | ISBN 9781119285496 (hardcover)

Subjects: LCSH: Pharmaceutical technology. | Biochemical engineering. | Chemical engineering.

Classification: LCC RS192 (ebook) | LCC RS192 .C525 2019 (print) | DDC 615.1/9–dc23

LC record available at <https://lccn.loc.gov/2018054741>

Cover Design: Wiley

Cover Images: Courtesy of Johannes G. Khinast, © ProXES, Inc.,

© Pfizer Inc., Courtesy of Xiao Yu Wu

Set in 10/12pt Times by SPi Global, Pondicherry, India

Printed in the United States of America

10 9 8 7 6 5 4 3 2 1

CONTENTS

LIST OF CONTRIBUTORS	ix
PREFACE	xv
UNIT CONVERSIONS	xvii
PART I INTRODUCTION	1
1 Chemical Engineering in the Pharmaceutical Industry: An Introduction	3
<i>David J. am Ende and Mary T. am Ende</i>	
PART II DRUG PRODUCT DESIGN, DEVELOPMENT, AND MODELING	19
2 Design of Solid Dosage Formulations	21
<i>Kevin J. Bittorf, Tapan Sanghvi, and Jeffrey P. Katstra</i>	
3 Powder Process Challenges and Solutions	53
<i>Thomas Baxter and James Prescott</i>	
4 Design and Scale-up of Dry Granulation Processes	81
<i>Howard J. Stamato and Omar L. Sprockel</i>	
5 Model-based Development of Roller Compaction Processes	119
<i>Gavin Reynolds</i>	
6 Wet Granulation Processes	147
<i>Karen P. Hapgood and James D. Litster</i>	
7 Toward a Generic Model for Twin-screw Wet Granulation	173
<i>Daan Van Hauwermeiren, Maxim Verstraeten, Michaël Ghijs, Kai Lee, Neil Turnbull, Mary T. am Ende, Pankaj Doshi, David Wilsdon, Thomas De Beer, and Ingmar Nopens</i>	

8	Modeling a Dosator Capsule Filling Process for Hard-shell Capsules	187
	<i>Peter Loidolt, Eva Faulhammer, and Johannes G. Khinast</i>	
9	Powder Compaction: Process Design and Understanding	203
	<i>David Wilson, Ron Roberts, and John Blyth</i>	
10	Punch Sticking: Factors and Solutions	227
	<i>Daryl M. Simmons</i>	
11	Spray Atomization Modeling for Tablet Film Coating Processes	245
	<i>Alfred Berchielli, Pankaj Doshi, Alberto Aliseda, and Juan C. Lasheras</i>	
12	Spray Drying and Amorphous Dispersions	267
	<i>Kristin J.M. Ploeger, Pavithra Sundararajan, Pedro C. Valente, Kenneth J. Rosenberg, João G. Henriques, and Paige Adack</i>	
13	The Freeze Drying Process: The Use of Mathematical Modeling in Process Design, Understanding, and Scale-up	293
	<i>Venkat Koganti, Sumit Luthra, and Michael J. Pikal</i>	
14	Sterilization Processes in the Pharmaceutical Industry	311
	<i>Piero M. Armenante and Otute Akiti</i>	
15	Controlled Release Technology and Design of Oral Controlled Release Dosage Forms	381
	<i>Avinash G. Thombre, Xiao Yu (Shirley) Wu, and Mary T. am Ende</i>	
16	Process Design and Development for Novel Pharmaceutical Dosage Forms	409
	<i>Leah Appel, Joshua Shockey, Matthew Shaffer, and Jennifer Chu</i>	
17	Multiscale Modeling of a Pharmaceutical Fluid Bed Coating Process Using CFD/DEM and Population Balance Models to Predict Coating Uniformity	419
	<i>Avik Sarkar, Dalibor Jajcevic, Peter Böhlring, Peter Toson, Matej Zadravec, Brian Shoemaker, Pankaj Doshi, Johannes Khinast, and Mary T. am Ende</i>	
18	Process Design of Topical Semisolids: Application of Fundamental Concepts in Pharmaceutical Engineering to PEG Ointment Development	451
	<i>Amanda Samuel, Thean Yeoh, Rolf Larsen, and Avik Sarkar</i>	
19	Achieving a Hot Melt Extrusion Design Space for the Production of Solid Solutions	469
	<i>Luke Schenck, Gregory M. Troup, Mike Lowinger, Li Li, and Craig Mckelvey</i>	
20	Drug Product Process Modeling	489
	<i>Mary T. am Ende, William Ketterhagen, Andrew Prpich, Pankaj Doshi, Salvador García-Muñoz, and Rahul Bharadwajh</i>	
PART III	CONTINUOUS MANUFACTURING	527
21	Continuous Manufacturing in Secondary Production	529
	<i>Martin Warman</i>	

22	Continuous Direct Compression Using Portable Continuous Miniature Modular & Manufacturing (PCM&M)	547
	<i>Daniel O. Blackwood, Alexandre Bonmassieux, and Giuseppe Cogoni</i>	
23	Process Control Levels for Continuous Pharmaceutical Tablet Manufacturing	561
	<i>Niels Nicolai, Ingmar Nopens, Maxim Verstraeten, and Thomas De Beer</i>	
PART IV APPLIED STATISTICS AND REGULATORY ENVIRONMENT		585
24	Multivariate Analysis for Pharmaceutical and Medical Device Development	587
	<i>Frederick H. Long</i>	
25	Pharmaceutical Manufacturing: The Role of Multivariate Analysis in Design Space, Control Strategy, Process Understanding, Troubleshooting, and Optimization	601
	<i>Theodora Kourti</i>	
26	Quality by Design: Pilot to Reality-The Honeymoon Phase to the Stormy Years	631
	<i>Mary T. am Ende and Christine B. Seymour</i>	
	INDEX	645

LIST OF CONTRIBUTORS

Paige Adack

Senior Scientist
Pharmaceutical Commercialization Technology
Merck & Co., Inc.
West Point, PA, USA

Otute Akiti, PhD

Head of CMC
BlackThorn Therapeutics
San Francisco, CA, USA

Alberto Aliseda, PhD

Assistant Professor
Department of Mechanical Engineering
University of Washington
Seattle, WA, USA

David J. am Ende, PhD

President
Nalas Engineering Services, Inc.
Centerbrook, CT, USA

Mary T. am Ende, PhD

Research Fellow
Process Modeling & Engineering Technology Group
Pfizer, Inc.
Groton, CT, USA
Current address: Lyndra Therapeutics, Watertown,
MA, USA

Leah Appel, PhD

Managing Partner
Green Ridge Consulting
Bend, OR, USA

Piero M. Armenante, PhD

Distinguished Professor of Chemical Engineering
Otto H. York Department of Chemical and Materials
Engineering
New Jersey Institute of Technology
University Heights
Newark, NJ, USA

Thomas Baxter

Director
Jenike & Johanson, Inc.
Tyngsboro, MA, USA

Alfred Berchielli

Senior Principal Scientist
Drug Product Design
Pfizer, Inc.
Groton, CT, USA

Rahul Bharadwajh, PhD

Vice President, Engineering and Business Development
Engineering Simulation and Scientific Software (ESSS),
Rocky DEM, Woburn, MA, USA

Kevin J. Bittorf, PhD, MBA, PEng

Formulation Development
Vertex Pharmaceuticals
Cambridge, MA, USA
Current affiliation: Principal Consultant Simulation LLC
Boston, MA, USA

Daniel O. Blackwood

Research Fellow
Drug Product Design
Pfizer, Inc.
Groton, CT, USA

John Blyth

Senior Scientist
AstraZeneca
Macclesfield, UK

Peter Böhling

Scientist
Research Center Pharmaceutical Engineering GmbH
Graz, Austria

Alexandre Bonnassieux

Scientist
Drug Product Design
Pfizer, Inc.
Groton, CT, USA

Jennifer Chu, PhD

Technology Development Lead
FreeThink Technologies, Inc.
Branford, CT, USA

Giuseppe Cogoni

Chemometrician/Data Analyst
Analytical R&D
Pfizer, Inc.
Groton, CT, USA

Thomas De Beer, PhD

Professor
Laboratory of Pharmaceutical Process Analytical Technology
Ghent University
Ghent, Belgium

Pankaj Doshi, PhD

Head of Process Modeling & Engineering Technology Group
Pfizer, Inc.
Groton, CT, USA

Eva Faulhammer, PhD

Senior Scientist
Research Center Pharmaceutical Engineering GmbH
Graz, Austria

Salvador García-Muñoz, PhD

Small Molecule Design and Development
Eli Lilly and Company
Indianapolis, IN, USA

Michaël Ghijs

BIOMATH
Ghent University
Ghent, Belgium

Karen P. Hapgood, PhD

Professor of Engineering
School of Engineering
Deakin University
Geelong, Victoria, Australia

João G. Henriques

Team Leader
R&D Drug Product Development
Hovione Farmacênciã SA
Loures, Portugal

Dalibor Jajcevic, PhD

Scientific Coordinator
Research Center Pharmaceutical Engineering GmbH
Graz, Austria

Jeffrey P. Katstra, MS

Formulation Development
Vertex Pharmaceuticals
Cambridge, MA, USA
Current affiliation: Associate Director
Agius Pharmaceuticals
Cambridge, MA, USA

William Ketterhagen, PhD

Process Modeling & Engineering Technology Group
Pfizer, Inc.
Groton, CT, USA
Current address: Drug Product Development, Research
and Development, AbbVie Inc., North Chicago,
IL, USA

Johannes G. Khinast, PhD Univ.-Prof.

CEO
Research Center Pharmaceutical Engineering GmbH
and
Head of the Institute for Process and
Particle Engineering
Graz University of Technology
Graz, Austria

Venkat Koganti, PhD

Celgene Corporation
Summit, NJ, USA

Theodora Kourti, PhD

Department Chemical Engineering
McMaster University
Hamilton, Ontario, Canada

Rolf Larsen

Senior Principal Scientist
Formulation Process and Design Group
Pfizer, Inc.
Groton, CT, USA

Juan C. Lasheras, PhD

Stanford and Beverly Penner Professor of
Applied Sciences
Distinguished Professor of Mechanical & Aerospace
Engineering and Bioengineering
Jacobs School of Engineering
University of California San Diego
La Jolla, CA, USA

Kai Lee, PhD

Process Scientist/Chemical Engineer
Drug Product Design
Pfizer Ltd
Sandwich, UK

Li Li, PhD

Principal Scientist
Merck & Co., Inc.
West Point, PA, USA

James D. Litster, PhD

Professor of Chemical Engineering
Department of Chemical and Biological Engineering
The University of Sheffield
Sheffield, UK

Peter Loidolt, PhD

Junior Researcher
Institute for Process and Particle Engineering
Graz University of Technology
Graz, Austria

Frederick H. Long, PhD

President
Spectroscopic Solutions, LLC
Randolph, NJ, USA

Mike Lowinger

Principal Scientist
Merck & Co., Inc.
Rahway, NJ, USA

Sumit Luthra, PhD

Principal Scientist
Pfizer Worldwide Research and Development
Pfizer, Inc.
Andover, MA, USA

Craig McKelvey, PhD

Distinguished Investigator
Merck & Co., Inc.
West Point, PA, USA

Niels Nicolai

BIOMATH
Ghent University
Ghent, Belgium

Ingmar Nopens, PhD

Professor
BIOMATH
Ghent University
Ghent, Belgium

Michael J. Pikal, PhD (deceased)

Former Pfizer Distinguished Endowed Chair in
Pharmaceutical Technology
Professor of Pharmaceutical Sciences
Pharmaceutics
School of Pharmacy
University of Connecticut
Storrs, CT, USA

Kristin J.M. Ploeger, PhD

Principal Scientist
Pharmaceutical Commercialization
Technology
Merck & Co., Inc.
West Point, PA, USA

James Prescott

Vice President
Jenike & Johanson, Inc.
Tyngsboro, MA, USA

Andrew Prpich, MS

Senior Scientist
Process Modeling & Engineering Technology Group
Pfizer, Inc.
Groton, CT, USA

Gavin Reynolds, PhD, CEng, FIChemE

Principal Scientist
Pharmaceutical Technology & Development
AstraZeneca
Macclesfield, UK

Ron Roberts

Associate Principal Scientist
AstraZeneca
Macclesfield, UK

Kenneth J. Rosenberg, PhD

Associate Principal Scientist
Center for Materials Science and Engineering
Merck & Co., Inc.
West Point, PA, USA

Amanda Samuel, PhD

Principal Scientist
Formulation Process and Design Group
Pfizer, Inc.
Groton, CT, USA

Tapan Sanghvi, PhD, MBA

Senior Scientist
Formulation Development
Vertex Pharmaceuticals
Cambridge, MA, USA

Avik Sarkar, PhD

Principal Scientist
Process Modeling & Engineering Technology Group
Pfizer, Inc.
Groton, CT, USA

Luke Schenck

Principal Scientist
Merck & Co., Inc.
Rahway, NJ, USA

Christine B. Seymour, PhD

Director
Global Regulatory Affairs CMC
Pfizer, Inc.
Groton, CT, USA

Matthew Shaffer

Lonza, Inc.
Bend, OR, USA

Joshua Shockey, PE

Partner
Green Ridge Consulting
Bend, OR, USA

Brian Shoemaker

Process Modeling & Engineering Technology Group
Pfizer, Inc.
Groton, CT, USA
Current address: Department of Chemical
& Environmental Engineering,
Yale University, New Haven, CT, USA

Daryl M. Simmons

Manager
Eurofins Lancaster Laboratories PSS
Lancaster, PA, USA
Current address: Three Rivers Community College
Norwich, CT USA

Omar L. Sprockel, PhD

Head, Engineering Technologies
Product Development
Bristol-Myers Squibb
New Brunswick, NJ, USA

Howard J. Stamato

Associate Director
Global Regulatory, Safety and Biometrics, Research
and Development
Bristol-Myers Squibb
Hopewell, NJ, USA

Pavithra Sundararajan, PhD

Associate Principal Scientist
Formulation Sciences
Merck & Co., Inc.
West Point, PA, USA

Avinash G. Thombre, PhD

Research Fellow
Drug Product Design
Pfizer, Inc.
Groton, CT, USA

Peter Toson, PhD

Senior Scientist
Research Center Pharmaceutical Engineering GmbH
Graz, Austria

Gregory M. Troup, PhD

Sr. Principal Scientist
Merck & Co., Inc.
West Point, PA, USA

Neil Turnbull

Associate Research Fellow
Drug Product Design
Pfizer Ltd
Sandwich, UK

Pedro C. Valente, PhD

Senior Scientist, Team Leader
R&D Drug Product Development
Hovione Farmaciência SA
Loures, Portugal

Daan Van Hauwermeiren
BIOMATH, Ghent University
Ghent, Belgium

Maxim Verstraeten
Laboratory of Pharmaceutical Process Analytical
Technology
Ghent University
Ghent, Belgium

Martin Warman, PhD
Martin Warman Consultancy Ltd
Chestfield, UK

David Wilsdon, PhD
Principal Scientist
Analytical R&D
Pfizer Ltd
Sandwich, UK

David Wilson, PhD
Associate Principal Scientist
AstraZeneca
Macclesfield, UK

Xiao Yu (Shirley) Wu, PhD, FAAPS
Director of Advanced Pharmaceuticals and
Drug Delivery Laboratory
Leslie Dan Faculty of Pharmacy
University of Toronto
Toronto, Ontario, Canada

Thean Yeoh, PhD
Associate Research Fellow
Formulation Process and Design
Pfizer, Inc.
Groton, CT, USA

Matej Zadavec, PhD
Senior Scientist
Research Center Pharmaceutical Engineering GmbH
Graz, Austria

PREFACE

Pharmaceutical research and development is unique to traditional chemical engineering curricula, which has focused intensively on the chemical industry. This book is intended to be used as a professional reference and as a textbook reference for undergraduate or graduate studies in engineering and pharmaceutical sciences. Many of the experimental methods related to drug product design and process development are learned on the job. This book is intended to provide many of those important concepts that R&D engineers and manufacturing engineers should know and be familiar in those roles. Formulation design and process development of drug products will be treated from the engineer's perspective and span from solid to semisolid and lyophilized drug products and sterilization. Technology transfer and scale-up of batch processes will be exemplified experimentally and computationally, including *in silico* process modeling tools that streamline experimental screening approaches. The emerging field of continuous drug product manufacturing will also be discussed by skilled professionals. Although continuous manufacturing is in the mainstream for chemical engineers, it is unique in the pharmaceutical industry with regard to the range of scales and the complex economics associated with transforming existing batch plant capacity. Basic analytics for quantitation of drug product quality attributes, such as potency, purity, content uniformity, and dissolution, will be addressed with consideration of the applied statistics, process analytical technology (PAT), and process control. In addition, contemporary methods of data analysis will be introduced, and these concepts extended into quality by design strategies for regulatory filings. Advances in the drug product pharmaceutical R&D are now being strongly supported by precompetitive consortia. Finally, technical chapters on commonly used software tools with examples are an important part of this book.

This book deals with the elements of chemical engineering science unique to drug product development and commercialization specifically related to the successful formulation design and process development of the active pharmaceutical ingredient (API) into the desired dosage form. It emphasizes the need for scientific integration of chemical engineering and pharmaceutical sciences during R&D, as well as with manufacturing engineers, analytical chemists, and other scientific disciplines necessary to deliver pharmaceuticals to the market place. As part of a multidisciplinary team in R&D, engineering contributes to process design, process understanding, and process development, which ultimately enables improvements in quality, reduces cost, and ensures safe, robust processes are transferred to manufacturing. As cost and time pressures increase, engineers play an important role in leveraging process modeling tools that can help focus the experimental work more rapidly with techniques to ensure the desired formulation and manufacturing process will scale as planned – so as to avoid surprises on scale-up. This book covers the basic chemical engineering theories with its emphasis toward providing experimental methods, analysis, and contemporary process modeling methods in chemical engineering. This book provides guidance on analytical methods for engineers in R&D as well as manufacturing. In addition, emphasis is given on experimental techniques and considerations necessary to address scale-up issues and approach general process design-related challenges to pharmaceutical process R&D. As a professional reference it is intended to be part “text” book and part “how-to” book and includes many worked examples related to problem solving via experimental and modeling methods. The book is organized to provide a foundational introduction on challenges and opportunities for chemical engineers in this industry in Part I. In Part II, chemistry and engineering activities

related to drug product design, development, and modeling are presented. Part III is focused on drug product continuous manufacturing. Finally, Part IV is focused on applied statistics and regulatory environment, with examples of their applications to pharmaceutical products.

I am grateful to all the contributing authors for making this book possible. I would also like to thank my supervisors and leadership team for their long-standing support of the important role chemical engineers play at Pfizer. A special note of gratitude to Lyndra Therapeutics for making my next career endeavor an inspiring one. Thank you to my graduate advisor, Professor Nicholas Peppas, for all of your amazing support in my academic development and opportunity to pursue chemical engineering in the pharmaceutical field.

I would also like to state a special note of gratitude to my ever-supportive family (David, Nathan, Noah, and Brianna) for encouraging me to pursue this opportunity to serve as editor. It is a pleasure to work with my strongest advocate in my career and life, who is also the best chemical engineer I know – my husband David. Finally, I am ever grateful to my family (James, Donna, Tami, Kevin, Jaime, Miles, and Michele) for their unwavering belief in me to pursue degrees in chemical engineering at the University of Iowa (BS 1988) and Purdue University (PhD 1993).

*MARY T. AM ENDE, PHD
Vice President
Process Development
Lyndra Therapeutics, Inc.
Watertown, MA, USA*

UNIT CONVERSIONS

Quantity	Equivalent Values
Length	$1 \text{ m} = 100 \text{ cm} = 1000 \text{ mm} = 10^6 \mu\text{m} = 10^{10} \text{ \AA}$ $1 \text{ m} = 39.37 \text{ in} = 3.2808 \text{ ft} = 1.0936 \text{ yards} = 0.0006214 \text{ mile}$ $1 \text{ ft} = 12 \text{ in} = 0.3048 \text{ m} = 1/3 \text{ yard} = 30.48 \text{ cm}$
Area	$1 \text{ m}^2 = 10.76 \text{ ft}^2 = 1550 \text{ in}^2 = 10\,000 \text{ cm}^2$ $1 \text{ in}^2 = 6.4516 \text{ cm}^2 = 645.16 \text{ mm}^2 = 0.006\,94 \text{ ft}^2$ $1 \text{ ft}^2 = 929.03 \text{ cm}^2 = 0.092\,903 \text{ m}^2$ Example: cross sectional area of 1/4" ID tube: $\frac{\pi d^2}{4} = \frac{\pi 0.25^2}{4} = 0.0491 \text{ in}^2 = 0.3167 \text{ cm}^2$
Volume	$1 \text{ m}^3 = 1000 \text{ L} = 10^6 \text{ cm}^3 \text{ (ml)} = 1000 \text{ dm}^3$ $1 \text{ m}^3 = 35.3145 \text{ ft}^3 = 220.83 \text{ imperial gallons} = 264.17 \text{ gal (U)}$ $1 \text{ ft}^3 = 1728 \text{ in}^3 = 7.4805 \text{ gal (US)} = 0.028\,317 \text{ m}^3 = 28.317 \text{ L}$ $1 \text{ gal (US)} = 3.785 \text{ L} = 0.1337 \text{ ft}^3 = 231 \text{ in}^3 = 4 \text{ quart} = 8 \text{ pints}$ $1 \text{ L} = 0.264 \text{ gal} = 1.0567 \text{ quart} = 2.113 \text{ pint} = 4.2267 \text{ cup} = 202.88 \text{ teaspoon (US)} = 0.035\,31 \text{ ft}^3 = 61.02 \text{ in}^3$
Mass	$1 \text{ kg} = 1000 \text{ g} = 0.001 \text{ metric ton (MT)} = 2.204\,62 \text{ lb}_m = 35.273\,92 \text{ oz}$ $1 \text{ lb}_m = 16 \text{ oz} = 453.593 \text{ g} = 0.453\,593 \text{ kg}$ $1 \text{ ton (metric)} = 1000 \text{ kg} = 2204.6 \text{ lb}_m$
Pressure	$1 \text{ atm} = 1.013\,25 \text{ bar} = 1.013\,25 \times 10^5 \frac{\text{N}}{\text{m}^2} \text{ (Pa)} = 0.101\,325 \text{ MPa}$ $= 101.325 \text{ kPa} = 1.013\,25 \times 10^6 \frac{\text{dynes}}{\text{cm}^2} = 1.033 \frac{\text{kg}_f}{\text{cm}^2}$ $= 760 \text{ mm Hg at } 0^\circ\text{C (torr)} = 10.333 \text{ m H}_2\text{O at } 4^\circ\text{C}$ $= 14.696 \frac{\text{lb}_f}{\text{in}^2} \text{ (psi)} = 33.9 \text{ ft H}_2\text{O at } 4^\circ\text{C} = 2116 \frac{\text{lb}_f}{\text{ft}^2}$ $= 29.921 \text{ in Hg at } 0^\circ\text{C}$ $1 \text{ MPa} = 9.869 \text{ atm} = 10 \text{ bar} = 145.04 \text{ psi}$ $1 \text{ psi} = 2.31 \text{ ft H}_2\text{O} = 0.0680 \text{ atm}$ $= 703.1 \frac{\text{kg}_f}{\text{m}^2} = 0.070\,31 \frac{\text{kg}_f}{\text{cm}^2} = 51.71 \text{ mm Hg}$ <i>Note:</i> $P_{\text{absolute}} = P_{\text{gauge}} + P_{\text{atmospheric}}$ For example, if a pressure gauge reads 30 psig, then the absolute pressure is 44.7 psia, <i>i.e.</i> $P_{\text{absolute}} = P_{\text{gauge}} + P_{\text{atmospheric}} = 30 + 14.7 = 44.7 \text{ psia}$

Quantity	Equivalent Values
Pressure (<i>continued</i>)	<p>Vacuum: A vacuum gauge may have range from 0 to -30 in Hg. So if the vacuum gauge is reading -25.0 in Hg then the absolute pressure is $P_{\text{absolute}} = P_{\text{gauge}} + P_{\text{atmospheric}} = -25.0 + 29.921 = +4.921 \text{ in Hg} = 0.16 \text{ atm} = 2.42 \text{ psia}$ (assuming that the atmospheric pressure is taken at sea level equivalent to 1.0 atm = 29.921 in Hg)</p>
Temperature	<p>$^{\circ}\text{C} = 5/9 (^{\circ}\text{F} - 32)$ $^{\circ}\text{F} = 9/5 ^{\circ}\text{C} + 32$ $\text{K} = ^{\circ}\text{C} + 273.15 = 5/9 ^{\circ}\text{R}$ $^{\circ}\text{R} = ^{\circ}\text{F} + 459.67$ Freezing point H₂O = 0 °C or 32 °F or 273.15 K Boiling point H₂O = 100 °C or 212 °F or 373.15 K</p>
Density (ρ)	<p>$1 \frac{\text{g}}{\text{cm}^3} = 1 \frac{\text{g}}{\text{ml}} = 1 \frac{\text{kg}}{\text{L}} = 62.43 \frac{\text{lb}_m}{\text{ft}^3} = 1000 \frac{\text{kg}}{\text{m}^3} = 8.345 \frac{\text{lb}_m}{\text{U.S.gal}}$ $100 \frac{\text{lb}_m}{\text{ft}^3} = 1601.85 \frac{\text{kg}}{\text{m}^3} = 1.602 \frac{\text{g}}{\text{cm}^3}$ $\rho(\text{H}_2\text{O}, 20^{\circ}\text{C}) = 998.2 \frac{\text{kg}}{\text{m}^3} = 0.9982 \frac{\text{g}}{\text{cm}^3}$ $\rho(\text{H}_2\text{SO}_4, 25^{\circ}\text{C}, 95 \text{ wt } \%) = 1.84 \frac{\text{g}}{\text{cm}^3}$</p>
Force	<p>$1 \text{ N} = 1 \frac{\text{kg} \cdot \text{m}}{\text{sec}^2} = 10^5 \text{ dynes} = 10^5 \text{ g} \frac{\text{g} \cdot \text{cm}}{\text{sec}^2} = 0.224 81 \text{ lb}_f = 0.102 \text{ kg}_f$ $1 \text{ lb}_f = 32.174 \frac{\text{lb}_m \cdot \text{ft}}{\text{sec}^2} = 4.4482 \text{ N} = 4.4482 \times 10^5 \text{ dynes}$ $1 \text{ kg}_f = 1 \text{ kg} \cdot 9.806 65 \frac{\text{m}}{\text{sec}^2} = 9.806 65 \frac{\text{kg} \cdot \text{m}}{\text{sec}^2} = 9.806 65 \text{ N} = 2.205 \text{ lb}_f$</p>
Energy	<p>$1 \text{ W} = \text{J/sec}$ 1 calorie = 4.184 J (thermochemical) $1 \text{ J} = 1 \text{ N} \cdot \text{m} = 1 \text{ W} \cdot \text{sec} = 0.239 01 \text{ cal} = 10^7 \text{ ergs} = 10^7 \text{ dyne} \cdot \text{cm}$ $1 \text{ J} = 2.778 \times 10^{-7} \text{ kW} \cdot \text{h} = 0.7376 \text{ ft} \cdot \text{lb}_f = 0.00094845 \text{ Btu}$ $1 \text{ Btu} = 1054.35 \text{ J} = 1.054 \text{ kJ} = 251.996 \text{ cal}$ $= 0.2929 \text{ W} \cdot \text{h} = 1054.35 \text{ N} \cdot \text{m}$ $1 \text{ kWh} = 3.6 \text{ MJ}$</p>
Heat generation rate	<p>$1 \frac{\text{Btu}}{\text{lb}_m \cdot \text{hr}} = 0.64612 \frac{\text{W}}{\text{kg}}$</p>
Heat transfer coefficient (U_o, h_o)	<p>$1 \frac{\text{W}}{(\text{m}^2 \text{K})} = 0.1761 \frac{\text{Btu}}{(\text{hr ft}^2 ^{\circ}\text{F})}$ $1 \frac{\text{Btu}}{(\text{hr ft}^2 ^{\circ}\text{F})} = 5.678 \frac{\text{W}}{(\text{m}^2 \text{K})} = 4.882 \frac{\text{kcal}}{(\text{hr m}^2 ^{\circ}\text{C})}$</p>
Nusselt Number (Nu)	<p style="text-align: center;">$\text{Nu} \equiv \frac{hD}{k} \equiv \frac{\text{conduction} + \text{convection}}{\text{conduction}}$</p> <p>where h is the heat transfer coefficient, D is the pipe diameter, and k is thermal conductivity. For purely <i>laminar</i> and fully developed pipe flow (conduction dominates) limiting cases:</p> <ul style="list-style-type: none"> • <i>Case of uniform heat flux or constant temperature difference:</i> $\text{Nu}_{\infty} = \frac{h_{\infty} D}{k} \approx 4.364$ • <i>Case of constant wall temperature:</i> $\text{Nu}_{\infty} = \frac{h_{\infty} D}{k} \approx 3.656$

Quantity	Equivalent Values
Nusselt Number (Nu) (continued)	<p>For <i>turbulent</i> pipe flow: Dittus-Boelter Equation:</p> $\text{Nu} \equiv \frac{hD}{k} = 0.023 \text{Re}^{0.8} \text{Pr}^n$ $n = \begin{pmatrix} 0.4 \text{ for heating} \\ 0.3 \text{ for cooling} \end{pmatrix}$ <p>Dittus-Boelter valid for: $10\,000 < \text{Re} < 120\,000$, $0.7 < \text{Pr} < 120$, $L/D > 60$ (ie fully developed) and when the pipe temperature is within 10°F for liquids and 100°F for gases.</p> <p>Sieder-Tate Equation:</p> $\text{Nu} \equiv \frac{hD}{k} = 0.023 \text{Re}^{0.8} \text{Pr}^{1/3} \left(\frac{\mu_b}{\mu_{\text{wall}}} \right)^{0.14}$ <p>Valid for $\text{Re} > 10\,000$, $L/D > 60$, higher Prandtl numbers $0.7 < \text{Pr} < 16\,700$, and larger temperature differences between bulk and wall. Properties evaluated at bulk temperature except for μ_{wall} which is evaluated at the wall temperature.</p> <p>Source: From Pitts and Sissom [1].</p>
Prandtl Number (Pr)	$\text{Pr} = \frac{C_p \mu}{k} = \frac{\text{viscous diffusion rate}}{\text{heat conduction rate}}$ <p>Prandtl number is a characteristic of the fluid. Liquids in general have high Prandtl numbers.</p> <ul style="list-style-type: none"> • Ethylene glycol $0^\circ\text{C} = 615$ • Water at $20^\circ\text{C} = 7.02$ • Water at $80^\circ\text{C} = 2.22$ • Steam $107^\circ\text{C} \approx 1.06$ • Gases ≈ 0.7
Latent heat	$1 \frac{\text{Btu}}{\text{lb}_m} = 2.326 \frac{\text{kJ}}{\text{kg}}$ $1 \frac{\text{J}}{\text{g}} = 0.23901 \frac{\text{cal}}{\text{g}}$ <p>H_2O: $\Delta H_{\text{melting}} = \Delta H_{\text{fusion}} = 6.01 \text{ kJ/mol}$ or 334 J/g H_2O: $\Delta H_{\text{vaporization}} \text{H}_2\text{O} = 2230 \text{ J/g} = 40.65 \text{ kJ/mol}$</p>
Power	$1 \text{ W} = 1 \frac{\text{J}}{\text{sec}} = 1 \frac{\text{kg} \cdot \text{m}^2}{\text{sec}^3} = 1 \frac{\text{Nm}}{\text{sec}} = 0.23901 \frac{\text{cal}}{\text{sec}}$ $= 0.7376 \frac{\text{ft} \cdot \text{lb}_f}{\text{sec}} = 0.0009485 \frac{\text{Btu}}{\text{sec}}$ $= 3.414 \frac{\text{Btu}}{\text{hr}} = 0.001341 \text{ hp}$
Power/volume	$1 \frac{\text{W}}{\text{L}} = \frac{\text{kW}}{\text{m}^3} = 0.03798 \frac{\text{hp}}{\text{ft}^3} = 96.67 \frac{\text{Btu}}{\text{hr} \cdot \text{ft}^3} = 12.9235 \frac{\text{Btu}}{\text{hr} \cdot \text{gal}}$
Specific heat (C_p)	$1 \frac{\text{kJ}}{(\text{kg} \cdot \text{K})} = \frac{\text{J}}{(\text{g} \cdot \text{K})} = 0.23901 \frac{\text{kcal}}{(\text{kg} \cdot ^\circ\text{C})} = 0.23901 \frac{\text{Btu}}{(\text{lb}_m \cdot ^\circ\text{F})}$ $1 \frac{\text{Btu}}{(\text{lb}_m \cdot ^\circ\text{F})} = 1 \frac{\text{cal}}{(\text{g} \cdot ^\circ\text{C})} = 4184 \frac{\text{J}}{\text{kg} \cdot \text{K}}$

Quantity	Equivalent Values
Specific heat (C_p) (continued)	For water (20 °C): $C_p = 4184 \frac{\text{J}}{\text{kg} \cdot \text{K}} = 1 \frac{\text{cal}}{\text{gm} \cdot ^\circ\text{C}} = 1 \frac{\text{Btu}}{(\text{lb} \cdot ^\circ\text{F})}$ For air (20 °C): $C_p = 1013 \frac{\text{J}}{\text{kg} \cdot \text{K}} = 29.29 \frac{\text{J}}{\text{mol} \cdot \text{K}}$ $= 0.24 \frac{\text{cal}}{\text{gm} \cdot ^\circ\text{C}} = 7 \frac{\text{cal}}{\text{mol} \cdot ^\circ\text{C}}$
Thermal conductivity (k)	$1 \frac{\text{Btu}}{(\text{hr} \cdot \text{ft} \cdot ^\circ\text{F})} = 1.7307 \frac{\text{W}}{(\text{m} \cdot \text{K})} = 0.00413 \frac{\text{cal}}{(\text{sec} \cdot \text{cm} \cdot \text{K})}$ $1 \frac{\text{W}}{(\text{m} \cdot \text{K})} = 0.5779 \frac{\text{Btu}}{(\text{hr} \cdot \text{ft} \cdot ^\circ\text{F})} = 0.85984 \frac{\text{kcal}}{(\text{hr} \cdot \text{m} \cdot ^\circ\text{C})}$ $1 \frac{\text{W}}{(\text{m} \cdot \text{K})} = 0.00239 \frac{\text{cal}}{\text{sec} \cdot \text{cm} \cdot ^\circ\text{C}} = 0.578 \frac{\text{Btu}}{\text{hr} \cdot \text{ft} \cdot ^\circ\text{F}}$ Thermal conductivity $k \cong$ independent of pressure For water (20 °C): $k = 0.597 \frac{\text{W}}{\text{m} \cdot \text{K}}$ For air (20 °C): $k = 0.0257 \frac{\text{W}}{\text{m} \cdot \text{K}}$ For ethanol (20 °C): $k = 0.17 \frac{\text{W}}{\text{m} \cdot \text{K}}$
Throughput (continuous at 365 days/yr)	$1 \text{ yr} = 365 \text{ days} = 8760 \text{ hr} = 5.256 \times 10^5 \text{ min}$ $1 \frac{\text{kg}}{\text{hr}} = 16.67 \frac{\text{g}}{\text{min}} = 24 \frac{\text{kg}}{\text{day}} = 8760 \frac{\text{kg}}{\text{yr}} = 8.76 \frac{\text{MT}}{\text{yr}}$ $10 \frac{\text{MT}}{\text{yr}} = 10\,000 \frac{\text{kg}}{\text{yr}} = 27.4 \frac{\text{kg}}{\text{day}} = 1.14 \frac{\text{kg}}{\text{hr}} = 19.03 \frac{\text{g}}{\text{min}}$ $1 \text{ Billion} \frac{\text{tablets}}{\text{year}} = 2.74 \times 10^6 \frac{\text{tablets}}{\text{day}}$ $= 114,155 \frac{\text{tablets}}{\text{hr}} = 31.7 \frac{\text{tablets}}{\text{sec}}$ $10 \frac{\text{MT API}}{\text{yr}} = 10\,000 \frac{\text{kg} \cdot \text{API}}{\text{yr}}$ of API formulated as a 10 mg dose $\frac{\text{API}}{\text{tablet}}$ $= 1.0 \text{ Billion} \frac{\text{tablets}}{\text{yr}}$
Thermal diffusivity	$\alpha = \frac{k}{\rho C_p} = \left[\frac{\text{m}^2}{\text{s}} \right]$ $1 \frac{\text{m}^2}{\text{sec}} = 10.76 \frac{\text{ft}^2}{\text{sec}} = 387.49 \frac{\text{ft}^2}{\text{hr}}$ $1 \frac{\text{ft}^2}{\text{sec}} = 929.03 \frac{\text{cm}^2}{\text{sec}} = 0.092903 \frac{\text{m}^2}{\text{sec}}$ For air (20 °C): $\alpha = 2.12 \times 10^{-5} \frac{\text{m}^2}{\text{s}}$ For water (20 °C): $\alpha = 1.43 \times 10^{-7} \frac{\text{m}^2}{\text{s}}$
Viscosity	Dynamic viscosity (μ) $(1 \text{ Pa} \cdot \text{sec}) = \frac{1 \text{ N} \cdot \text{sec}}{\text{m}^2} = \left[\frac{1 \text{ kg}}{\text{m} \cdot \text{sec}} \right] = 1000 \text{ cP (centipoise)}$ $1 \text{ cP} = 0.01 \text{ poise} = 0.01 \frac{\text{g}}{(\text{cm} \cdot \text{sec})} = 0.001 \text{ Pa} \cdot \text{sec} = 1 \text{ mPa} \cdot \text{sec (milliPascal} \cdot \text{sec)}$

Quantity	Equivalent Values
Viscosity (<i>continued</i>)	$1 \text{ cP} = 3.6 \frac{\text{kg}}{(\text{m} \cdot \text{hr})} = 0.001 \frac{\text{kg}}{(\text{m} \cdot \text{sec})} = 2.419 \frac{\text{lb}_m}{(\text{ft} \cdot \text{hr})}$ $1 \text{ poise} = 1 \frac{\text{g}}{(\text{cm} \cdot \text{sec})} = 100 \text{ cP}$ <p>For liquid water (20 °C): $\mu = 1.002 \times 10^{-3} \text{ Pa} \cdot \text{sec} = 1.002 \text{ cP}$</p> <p>For gases (20 °C): $\mu \cong 10^{-5} \frac{\text{kg}}{\text{m sec}} = 0.01 \text{ cP}$</p> <p>For air (20 °C): $\mu \cong 1.8 \times 10^{-5} \frac{\text{kg}}{\text{m sec}} = 0.018 \text{ cP}$</p> <p>Kinematic viscosity (ν)</p> $\nu = \frac{\mu}{\rho} = \frac{\text{kg}}{\text{m sec}} \frac{\text{m}^3}{\text{kg}} = \left[\frac{\text{m}^2}{\text{s}} \right]$ <p>Stoke = $1 \frac{\text{cm}^2}{\text{sec}} = 1 \text{ St}$</p> <p>Centistoke = $1 \times 10^{-6} \frac{\text{m}^2}{\text{sec}} = 0.01 \text{ stoke} = 0.01 \frac{\text{cm}^2}{\text{sec}} = 1 \text{ cSt}$</p> $= 0.0036 \frac{\text{m}^2}{\text{hr}} = 0.0388 \frac{\text{ft}^2}{\text{hr}}$ $1 \frac{\text{m}^2}{\text{s}} = 10^4 \frac{\text{cm}^2}{\text{s}} = 10^4 \text{ stoke} = 10^6 \text{ centistoke}$ <p>$\nu(\text{H}_2\text{O } 20^\circ\text{C}) = 1.004 \times 10^{-6} \frac{\text{m}^2}{\text{sec}} = 1.004 \text{ cSt}$</p>
Gravitational force	$g = 9.8066 \frac{\text{m}}{\text{sec}^2} = 32.174 \frac{\text{ft}}{\text{sec}^2}$
Ideal Gas Law	$PV = nRT \text{ and } R = \frac{PV}{nT}$ $R = 8.314 \frac{\text{J}}{\text{mol K}} = 8.314 \frac{\text{m}^3 \text{ Pa}}{\text{mol K}} = 82.06 \times 10^{-6} \frac{\text{m}^3 \text{ atm}}{\text{mol K}}$ $R = 0.08206 \frac{\text{L} \cdot \text{atm}}{\text{mol K}} = 1.987 \frac{\text{cal}}{\text{mol K}}$ $R = 1.987 \frac{\text{Btu}}{\text{lb mol}^\circ\text{R}} = 0.729 \frac{\text{ft}^3 \text{ atm}}{\text{lb mol}^\circ\text{R}}$ $R = 82.057 \frac{\text{atm cm}^3}{(\text{mol K})} = 10.73 \frac{\text{psi ft}^3}{(\text{lb mol}^\circ\text{R})} = 62.36 \frac{\text{L} \cdot \text{torr}}{\text{mol K}}$ <p>At STP (Standard Temperature and Pressure), temperature is equal to 0 °C and pressure is equal to 1 atm. At STP, 1 mol of an ideal gas occupies 22.415 L.</p>
Raoult's Law (<i>approximation and generally valid for concentrated solutions when x_A is close to 1</i>)	<p>Raoult's Law:</p> $p_A = y_A P_T = x_A p_A^*(T)$ <p>where x_A is the mol fraction of A, y_A is the mol fraction in the vapor phase, P_T is the total pressure, $p_A^*(T)$ is the vapor pressure of A, and p_A is the partial pressure of A</p> <p>As x_A approaches 1, the partial pressure $p_A \approx p_A(T)$ approaches the vapor pressure of liquid A.</p> <p>Example: Drying 2-propanol (IPA) from a wetcake using a nitrogen stream (single-pass through the cake). How long will it take to remove 100 g of IPA from a wet-cake using nitrogen blow-through (the cake) at 1 L/min at 30 °C? Assume the drying cake remains isothermal at 30 °C and the nitrogen remains saturated and the system total pressure is 1 atm = 760 mmHg absolute.</p> <p>Solution:</p> <ol style="list-style-type: none"> Vapor pressure of IPA (30 °C) = 58.3 mmHg Assume $x_A \approx 1$ for IPA Ignoring solids and any mass transfer limitations, simply calculate the saturation condition for IPA in N₂. Assume 30 °C isothermal wet-cake. Mol fraction of saturated 2-PrOH in nitrogen: $y_{\text{IPA}} = \frac{p_{\text{IPA}}^*(T)}{P_T} = \frac{58.3 \text{ mmHg}}{760 \text{ mmHg}} = 0.0767 \text{ mol frac IPA}$

Quantity	Equivalent Values
Raoult's Law (continued)	<p>Mol fraction of nitrogen:</p> $y_{N_2} = 1 - 0.0767 = 0.9233 \text{ mol frac } N_2$ <p>5. Calculate mass of nitrogen per gram of IPA required to become fully saturated with 2PrOH (30 °C):</p> $\frac{0.9233 \times 28 \text{ g/mol}}{0.0767 \times 58.3 \text{ g/mol}} = 5.78 \text{ g of } N_2 / (\text{g of } 2\text{PrOH})$ <p>6. Flow of 1 L/min (N_2) at 30 °C: $\frac{1 \text{ L (273K)}}{\text{min (303K)}} \frac{\text{mol } 28 \text{ g}}{22.4 \text{ L mol}} = 1.13 \text{ g } N_2 / \text{min}$</p> <p>7. To saturate and remove 100 g of IPA using nitrogen at 30 °C and 1 L/min:</p> $100 \text{ g of } 2\text{PrOH} \left(\frac{\text{min}}{1.13 \text{ g } N_2} \right) \frac{(5.78 \text{ g } N_2)}{(\text{g } 2\text{PrOH})} = 511.5 \text{ min} = 8.5 \text{ hr}$ <p>Note: This is only a rough estimate (order of magnitude). The assumptions are idealized since there is typically a significant drop in temperature due to evaporative cooling which will lower the vapor pressure and slow the rate of drying. For a comprehensive treatment of mass transfer during flow-through drying see Treybal [2].</p>
Henry's Law (generally valid for dilute solutions when x_A is close to 0; and commonly applied to solutions of noncondensable gases)	<p>Henry's Law: At a constant temperature, the amount of a gas dissolved in a specific type and volume of liquid is directly proportional to the partial pressure of that gas in equilibrium with that liquid.</p> $p_A = y_A P_T = x_A H'_A(T)$ <p>where H'_A is the Henry's Law constant with units of (pressure/mol fraction), x_A is the mol fraction of gas dissolved in liquid, y_A is the mol fraction in the vapor, P_T is the total pressure, p_A is the partial pressure of A.</p> <p>Commonly used variation for pure gas ($y_A \approx 1$) over liquid: (such as a hydrogenation)</p> $p_A = y_A P_T \approx (1.0) P_{\text{sat}} = C_{\text{sat}} \cdot H_A(T)$ <p>where H_A is the Henry's Law constant with units of (pressure/(mol/L)), C_{sat} is the gas solubility in mol/L at saturation pressure P_{sat}.</p> <p>Example: Calculate the solubility of H_2 in methanol at 25 °C and 3.5 bar (absolute pressure of pure H_2): Solution: The Henry's constant for hydrogen in methanol at 25 °C:</p> $H_{H_2} = 268 \frac{\text{bar L}_{\text{soln}}}{\text{mol}_{H_2}}$ $\text{Solubility} = C_{\text{sat}} = \frac{P_{\text{sat}}}{H_{H_2}} = \frac{3.5 \text{ bar mol}_{H_2}}{268 \text{ bar L}} = \frac{0.0131 \text{ mol}_{H_2}}{\text{L}} = 0.0131 \text{ M}_{H_2} \text{ at } 3.5 \text{ bar and } 25 \text{ °C}$ <p>Other commonly used forms of Henry's Law:</p> $p_A = y_A P_T = \frac{C_A}{k_H}$ <p>where k_H is the Henry's Law constant, with units of $\frac{\text{mol}}{\text{L atm}}$:</p> <p>Example:</p> <p>a. Calculate the solubility of pure oxygen in equilibrium with water at 25 °C at 1 atm (absolute pressure of oxygen): Henry's constant, k_H for O_2 in water at 25 °C: $1.3 \times 10^{-3} \frac{\text{mol}}{\text{L atm}}$ (see table below)</p>

Quantity	Equivalent Values																																																		
Henry's Law <i>(continued)</i>	$C_{\text{sat},\text{O}_2} = p_{\text{O}_2} k_{\text{H},\text{O}_2} = \frac{1.0 \text{ atm} \cdot 0.0013 \text{ mol}}{\text{L atm}} = \frac{0.0013 \text{ mol O}_2}{\text{L}} = 0.0013 \text{ M O}_2$ <p>b. Instead of pure oxygen, calculate the solubility of oxygen in water while in equilibrium with air at 25 °C and 1 atm.</p> <p>Recognize the mol fractions in the gas phase: $y_{\text{N}_2} = 0.79$ and $y_{\text{O}_2} = 0.21$ $P_{\text{T}} = 1 \text{ atm}$ so $P_{\text{O}_2} = y_{\text{O}_2} P_{\text{T}} = (0.21)(1 \text{ atm}) = 0.21 \text{ atm}$</p> $C_{\text{sat},\text{O}_2} = p_{\text{O}_2} k_{\text{H},\text{O}_2} = \frac{0.21 \text{ atm} \cdot 0.0013 \text{ mol}}{\text{L atm}} = \frac{2.73 \times 10^{-4} \text{ mol O}_2}{\text{L}}$ <p style="text-align: center;">Forms of Henry's Law and Constants (Gases in Water at 298 K)</p> <table border="1" data-bbox="412 655 1476 1192"> <thead> <tr> <th data-bbox="412 655 634 737">Equation</th> <th data-bbox="634 655 849 737">$k_{\text{H}} = \frac{p_{\text{gas}}}{C_{\text{aq}}}$</th> <th data-bbox="849 655 1068 737">$k_{\text{H}} = \frac{C_{\text{aq}}}{P_{\text{gas}}}$</th> <th data-bbox="1068 655 1310 737">$k_{\text{H}} = \frac{p_{\text{gas}}}{x_{\text{aq}}}$</th> <th data-bbox="1310 655 1476 737">$k_{\text{H}} = \frac{C_{\text{aq}}}{C_{\text{gas}}}$</th> </tr> </thead> <tbody> <tr> <td data-bbox="412 737 634 814">Units</td> <td data-bbox="634 737 849 814">$\frac{\text{L}_{\text{soln}} \text{ atm}}{\text{mol}_{\text{gas}}}$</td> <td data-bbox="849 737 1068 814">$\frac{\text{mol}_{\text{gas}}}{\text{L}_{\text{soln}} \text{ atm}}$</td> <td data-bbox="1068 737 1310 814">$\frac{\text{atm mol}_{\text{soln}}}{\text{mol}_{\text{gas}}}$</td> <td data-bbox="1310 737 1476 814">Dimensionless</td> </tr> <tr> <td data-bbox="412 814 634 863">O₂</td> <td data-bbox="634 814 849 863">769.23</td> <td data-bbox="849 814 1068 863">0.001 3</td> <td data-bbox="1068 814 1310 863">42 590</td> <td data-bbox="1310 814 1476 863">0.031 80</td> </tr> <tr> <td data-bbox="412 863 634 911">H₂</td> <td data-bbox="634 863 849 911">1282.05</td> <td data-bbox="849 863 1068 911">0.000 78</td> <td data-bbox="1068 863 1310 911">70 990</td> <td data-bbox="1310 863 1476 911">0.019 07</td> </tr> <tr> <td data-bbox="412 911 634 959">CO₂</td> <td data-bbox="634 911 849 959">29.41</td> <td data-bbox="849 911 1068 959">0.034 0</td> <td data-bbox="1068 911 1310 959">1 630</td> <td data-bbox="1310 911 1476 959">0.831 7</td> </tr> <tr> <td data-bbox="412 959 634 1008">N₂</td> <td data-bbox="634 959 849 1008">1639.34</td> <td data-bbox="849 959 1068 1008">0.000 61</td> <td data-bbox="1068 959 1310 1008">90 770</td> <td data-bbox="1310 959 1476 1008">0.014 92</td> </tr> <tr> <td data-bbox="412 1008 634 1056">He</td> <td data-bbox="634 1008 849 1056">2702.7</td> <td data-bbox="849 1008 1068 1056">0.000 37</td> <td data-bbox="1068 1008 1310 1056">149 700</td> <td data-bbox="1310 1008 1476 1056">0.009 051</td> </tr> <tr> <td data-bbox="412 1056 634 1104">Ne</td> <td data-bbox="634 1056 849 1104">2222.22</td> <td data-bbox="849 1056 1068 1104">0.000 45</td> <td data-bbox="1068 1056 1310 1104">123 000</td> <td data-bbox="1310 1056 1476 1104">0.011 01</td> </tr> <tr> <td data-bbox="412 1104 634 1152">Ar</td> <td data-bbox="634 1104 849 1152">714.28</td> <td data-bbox="849 1104 1068 1152">0.001 4</td> <td data-bbox="1068 1104 1310 1152">39 550</td> <td data-bbox="1310 1104 1476 1152">0.034 25</td> </tr> <tr> <td data-bbox="412 1152 634 1192">CO</td> <td data-bbox="634 1152 849 1192">1052.63</td> <td data-bbox="849 1152 1068 1192">0.000 95</td> <td data-bbox="1068 1152 1310 1192">58 280</td> <td data-bbox="1310 1152 1476 1192">0.023 24</td> </tr> </tbody> </table> <p>where: C_{aq} = moles of gas per Liter of solution p_{gas} = partial pressure of gas above the solution in atmospheres x_{aq} = mole fraction of gas in solution https://chemengineering.wikispaces.com/Henry%27s+Law</p>	Equation	$k_{\text{H}} = \frac{p_{\text{gas}}}{C_{\text{aq}}}$	$k_{\text{H}} = \frac{C_{\text{aq}}}{P_{\text{gas}}}$	$k_{\text{H}} = \frac{p_{\text{gas}}}{x_{\text{aq}}}$	$k_{\text{H}} = \frac{C_{\text{aq}}}{C_{\text{gas}}}$	Units	$\frac{\text{L}_{\text{soln}} \text{ atm}}{\text{mol}_{\text{gas}}}$	$\frac{\text{mol}_{\text{gas}}}{\text{L}_{\text{soln}} \text{ atm}}$	$\frac{\text{atm mol}_{\text{soln}}}{\text{mol}_{\text{gas}}}$	Dimensionless	O ₂	769.23	0.001 3	42 590	0.031 80	H ₂	1282.05	0.000 78	70 990	0.019 07	CO ₂	29.41	0.034 0	1 630	0.831 7	N ₂	1639.34	0.000 61	90 770	0.014 92	He	2702.7	0.000 37	149 700	0.009 051	Ne	2222.22	0.000 45	123 000	0.011 01	Ar	714.28	0.001 4	39 550	0.034 25	CO	1052.63	0.000 95	58 280	0.023 24
Equation	$k_{\text{H}} = \frac{p_{\text{gas}}}{C_{\text{aq}}}$	$k_{\text{H}} = \frac{C_{\text{aq}}}{P_{\text{gas}}}$	$k_{\text{H}} = \frac{p_{\text{gas}}}{x_{\text{aq}}}$	$k_{\text{H}} = \frac{C_{\text{aq}}}{C_{\text{gas}}}$																																															
Units	$\frac{\text{L}_{\text{soln}} \text{ atm}}{\text{mol}_{\text{gas}}}$	$\frac{\text{mol}_{\text{gas}}}{\text{L}_{\text{soln}} \text{ atm}}$	$\frac{\text{atm mol}_{\text{soln}}}{\text{mol}_{\text{gas}}}$	Dimensionless																																															
O ₂	769.23	0.001 3	42 590	0.031 80																																															
H ₂	1282.05	0.000 78	70 990	0.019 07																																															
CO ₂	29.41	0.034 0	1 630	0.831 7																																															
N ₂	1639.34	0.000 61	90 770	0.014 92																																															
He	2702.7	0.000 37	149 700	0.009 051																																															
Ne	2222.22	0.000 45	123 000	0.011 01																																															
Ar	714.28	0.001 4	39 550	0.034 25																																															
CO	1052.63	0.000 95	58 280	0.023 24																																															
Reynold's Number (Re)	Reynolds Number for stirred vessel: $\text{Re} = \frac{\rho N D^2}{\mu}$ where ρ = density, N = stir speed, D = impeller diameter, and μ = viscosity For pipe or tube: $\text{Re} = \frac{\rho u D_{\text{pipe}}}{\mu}$ where u is fluid velocity, ρ = density, D = pipe inside diameter, and μ = viscosity Example (Stirred Tank): Calculate Re for a lab reactor containing water using the following parameters: Impeller diameter: $D = 5 \text{ cm} = 0.05 \text{ m}$																																																		

Quantity	Equivalent Values
Reynold's Number (Re) <i>(continued)</i>	<p>Stirrer speed: $N = 600 \text{ rpm}/(60 \text{ sec}/\text{min}) = 10 \text{ rotations}/\text{sec}$</p> <p>$\rho = 1 \text{ g}/\text{cm}^3 = 1000 \text{ kg}/\text{m}^3$</p> <p>Viscosity of water: $\mu = 1 \text{ cp} = 0.001 \text{ kg}/(\text{m}\cdot\text{s})$</p> $\text{Re} = \frac{1000 \text{ kg}/\text{m}^3 \times 10 \text{ rps} \times (0.05 \text{ m})^2}{0.001 \text{ kg}/(\text{m}\cdot\text{s})} = 25\,000$ <p>Example (pipe flow): What flow rate of water inside a 1/4" OD tube would provide a Reynolds number of 25 000: Assume inside diameter of tube, ID = 0.23 in = 0.584 cm Tube: cross-sectional area: $(3.14 \times (0.584)^2)/4 = 0.268 \text{ cm}^2$ $\rho = 1 \text{ g}/\text{cm}^3 = 1000 \text{ kg}/\text{m}^3$ Viscosity of water: $\mu = 1 \text{ cp} = 0.001 \text{ kg}/(\text{m}\cdot\text{s})$</p> $\text{Re} = \frac{(1000 \text{ kg}/\text{m}^3) \cdot u \cdot (0.00584 \text{ m})}{0.001 \text{ kg}/(\text{m}\cdot\text{s})} = 25\,000$ $u = \frac{(25\,000) \left(0.001 \frac{\text{kg}}{\text{m}\cdot\text{s}}\right)}{\left(1000 \frac{\text{kg}}{\text{m}^3}\right) (0.00584 \text{ m})} = 4.28 \text{ m/s}$ <p>Volumetric flow rate = $u \cdot \text{area} = 428 \text{ cm}/\text{sec} \times 0.268 \text{ cm}^2 = 114.7 \text{ cm}^3/\text{sec} = 6.88 \text{ L}/\text{min}$</p>
Power Number (N _P)	$N_P = \frac{P}{\rho N^3 D^5}$ <p>where P = power, ρ = density, N = stir speed, D = impeller diameter For a given reactor + agitator configuration the mixing power of an impeller can be uniquely characterized by the power number:</p> <ul style="list-style-type: none"> Turbulent flow where $\text{Re} > 2000$: $N_P = N_{P, \text{turbulent}} = \text{constant} \geq \text{Power} \propto \rho N^3 D^5$ <p>Example: If the measured power required to agitate a tank of liquid at 600 rpm with a 5 cm ID impeller is 1.0 W then what is the power at 1000 rpm (16.7 rps) in water? Solution: Check Reynolds number ($\text{Re} = 25\,000$; see above for calculation); Power Number, N_P, will be a constant for this reactor + impeller when $\text{Re} > 2000$</p> $N_P = \frac{P}{\rho N^3 D^5}$ <p>Calculate N_P :</p> <p>Note that $1.0 \text{ W} = 1.0 \frac{\text{kg m}^2}{\text{s}^3}$</p> $\text{Power Number, } N_P = \frac{1.0 \text{ kg m}^2/\text{s}^3}{1000 \text{ kg}/\text{m}^3 \times (10 \text{ rps})^3 \times (0.05 \text{ m})^5} = 3.2 \text{ will be constant for } \text{Re} > 2000$ <p>Calculate the Power required for the higher stir speed of 1000 rpm $P = N_P \rho N^3 D^5 = 3.2$</p> $P = 3.2 \times 1000 \frac{\text{kg}}{\text{m}^3} \times (16.7 \text{ rps})^3 \times (0.05 \text{ m})^5$ $P = 4.7 \text{ W}$ <p>Note the ~5× increase in power draw to increase stirring from 600 to 1000 rpm.</p>
Surface tension (γ)	$\left[\frac{\text{dyne}}{\text{cm}} \right]$ <p>Water-air interface: $\gamma_{\text{H}_2\text{O-AIR}} (20^\circ\text{C}) = 72.75 \text{ dynes}/\text{cm}$ CRC HB 62nd edition</p> $1 \frac{\text{dyne}}{\text{cm}} = 0.001 \frac{\text{N}}{\text{m}} = 1 \frac{\text{erg}}{\text{cm}^2}$

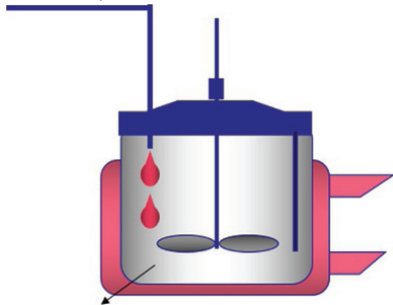
Quantity	Equivalent Values										
ppm	<table border="1" style="margin-left: auto; margin-right: auto;"> <thead> <tr> <th style="text-align: left;">Percent (%)</th> <th style="text-align: left;">ppm</th> </tr> </thead> <tbody> <tr> <td>0.0001</td> <td>1</td> </tr> <tr> <td>0.001</td> <td>10</td> </tr> <tr> <td>0.01</td> <td>100</td> </tr> <tr> <td>0.1</td> <td>1000</td> </tr> </tbody> </table>	Percent (%)	ppm	0.0001	1	0.001	10	0.01	100	0.1	1000
Percent (%)	ppm										
0.0001	1										
0.001	10										
0.01	100										
0.1	1000										
Moisture content	<p>Moisture content (wet basis) Range = 0–100%</p> $\%M_{\text{wet}} = 100 \frac{\text{mass}_{\text{solvent}}}{\text{mass}_{\text{solvent}} + \text{mass}_{\text{dry solids}}} = 100 \frac{\text{mass}_{\text{solvent}}}{\text{mass}_{\text{Total}}}$ <p>Note wet basis is the same as % Loss on drying (%LoD) where:</p> $\% \text{LoD} = 100 \frac{\text{mass}_{\text{loss on drying}}}{\text{mass}_{\text{initial wet cake}}}$ <p>Moisture content (dry basis) range = 0 to $\gg 100\%$</p> $\%M_{\text{dry}} = 100 \frac{\text{mass}_{\text{solvent}}}{\text{mass}_{\text{dry solids}}}$ <p>To convert wet basis to dry basis: $\text{wt}\%M_{\text{dry}} = 100 \frac{\%M_{\text{wet}}}{100\% - \%M_{\text{wet}}}$</p> <p>To convert dry basis to wet basis: $\text{wt}\%M_{\text{wet}} = 100 \frac{\%M_{\text{dry}}}{100\% + \%M_{\text{dry}}}$</p> <p>Example: 100 g of wetcake contains 40 g of water and 60 g of dry API, compare the moisture contents (wet vs dry basis):</p> $\%M_{\text{wet}} = \% \text{LoD} = 100 \frac{40 \text{ g}}{40 \text{ g} + 60 \text{ g}} = 40\%$ $\%M_{\text{dry}} = 100 \frac{40 \text{ g}}{60 \text{ g}} = 67\%$										
Humidity	<p>Absolute humidity for water in air: $h = \frac{kg_{\text{water}}}{kg_{\text{dry Air}}}$</p> <p>Example: Calculate the absolute humidity of air saturated with water at 30 °C at 1 atm pressure. Solution:</p> <ul style="list-style-type: none"> The vapor pressure of water at 30 °C, is 31.8 torr = 0.0418 atm Calculate the mole fraction water: $y_{\text{water}} = \frac{p_{\text{water}}^*(T)}{P_{\text{T}}} = \frac{0.0418 \text{ atm}}{1 \text{ atm}} = 0.0418 \text{ mol frac water}$ Mole fraction of air $y_{\text{air}} = 1 - 0.0418 = 0.9582 \text{ mol frac air}$ Calculate mass ratio of water to air to find the absolute humidity: $\frac{0.0418 \times 18.02 \text{ g/mol}}{0.9582 \times 29 \text{ g/mol}} = 0.0271 \frac{kg_{\text{water}}}{kg_{\text{dry air}}}$ <ul style="list-style-type: none"> This value can also be obtained be read from a psychrometric chart for water/air; Because the air is saturated, in this example, the relative humidity is 100%. 										

Quantity	Equivalent Values
Inertion of a tank via pressure cycles	$x_{\text{final}} = x_{\text{initial}} \left(\frac{P_{\text{Low}}}{P_{\text{High}}} \right)^k$ <p>Example: Determine the number of pressure purges required to reduce the oxygen concentration in a reactor from 21 to 0.1 vol % O₂ using nitrogen. A nitrogen source is used to pressurize the reactor to 50 psig and is vented down to 5 psig, in several cycles. Calculate the approximate number of cycles required.</p> $k = \frac{\ln \left(\frac{0.001}{0.21} \right)}{\ln \left(\frac{5 \text{ psig} + 14.7 \text{ psig}}{50 \text{ psig} + 14.7 \text{ psig}} \right)} = \frac{-5.347}{-1.189} = 4.5 \approx 5 \text{ cycles}$ <p>Source: Adapted from Kinsley [3].</p>

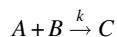
Polymath (6.10) program for Semi-Batch (i.e. Fed-Batch) with 1 hour Feed-Time. A Is Being Fed to B.

Assume Isothermal Kinetics

Feed stream A:
1 L fed over 60 minutes
(C_{A0} = 1 mol A/L)



at $t=0$ $V=1$ L
 $C_B(0)=1.0$ mol B/L

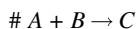


$$\text{rate} = -kC_A C_B \text{ where } k \equiv \left(\frac{\text{L}}{\text{mol min}} \right)$$

$$\Delta H = -30 \text{ kcal/mol}$$

Initial conditions at $t=0$

- Volume in the reactor, $V_0 = 1$ L
- Concentration of B in the reactor, $C_B(0) = 1$ M
- Concentration of A and C in the reactor = 0



A is fed to B

$$d(C_A)/dt = \text{if } (t > \text{dose}) \text{ then } r_a \text{ else } r_a + C_{A0} * v_0/V - C_A * V_0/V \text{ \# mols/(l \cdot min)}$$

$$d(C_B)/dt = \text{if } (t > \text{dose}) \text{ then } r_a \text{ else } r_a - v_0 * C_B/V \text{ \#}$$

$$d(C_C)/dt = \text{if } (t > \text{dose}) \text{ then } -r_a \text{ else } -r_a - v_0 * C_C/V \text{ \#}$$

Dose = 60 # minutes

$$V = \text{if } (t > \text{dose}) \text{ then } V_0 + v_0 * \text{dose} \text{ else } V_0 + v_0 * t \text{ \#}$$

$\text{molsAfed} = \text{if } (t > \text{dose}) \text{ then } C_{ao} * v_o * \text{dose} \text{ else } C_{ao} * v_o * t \#$
 $\text{molsB} = C_b * V$
 $V_o = 1 \# \text{ liter (initial volume of the reactor)}$
 $v_o = 1/60 \# \text{ L/min (volumetric flow rate of the feed)}$
 $k = 0.1 \# \text{ rate constant}$
 $C_{ao} = 1 \# \text{ mol/L (concentration of A in the feed stream)}$
 $C_{bo} = 1 \# \text{ mol/L (initial concentration of B in the reactor)}$
 $r_a = -k * C_a * C_b \# \text{ reaction rate expression}$
 $\text{rate} = -r_a \#$

#Heat of Reaction

$\Delta H = 30 \times 1000/0.23901 \# \text{ Exothermic heat of reaction, } (30 \text{ kcal/mol}) \times (1000 \text{ cal/kcal}) \times (J/0.23901 \text{ cal})$
 $Q = \Delta H \times \text{rate} \times V/60 \# (J/\text{mol}) \times (\text{mol}/(\text{L} \cdot \text{min})) \times (\text{L}) \times (\text{min}/60 \text{ sec}) = J/\text{sec} = W$
 $WL = Q/V \# W/L$

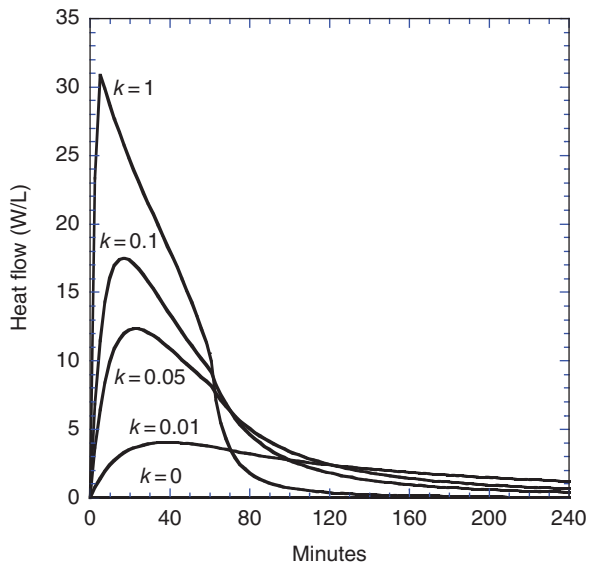
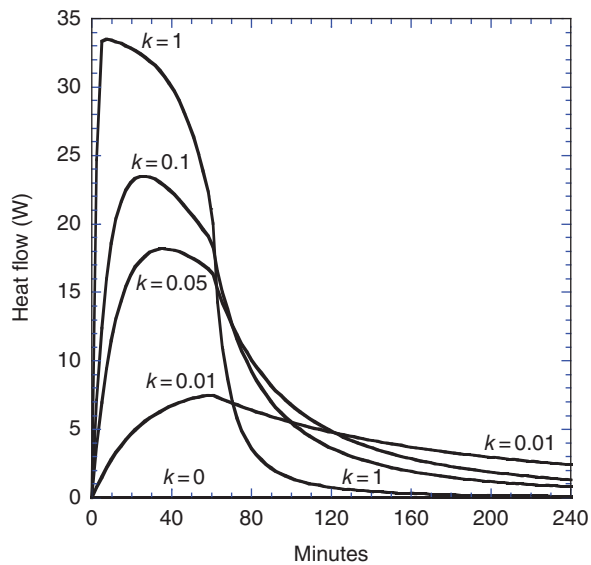
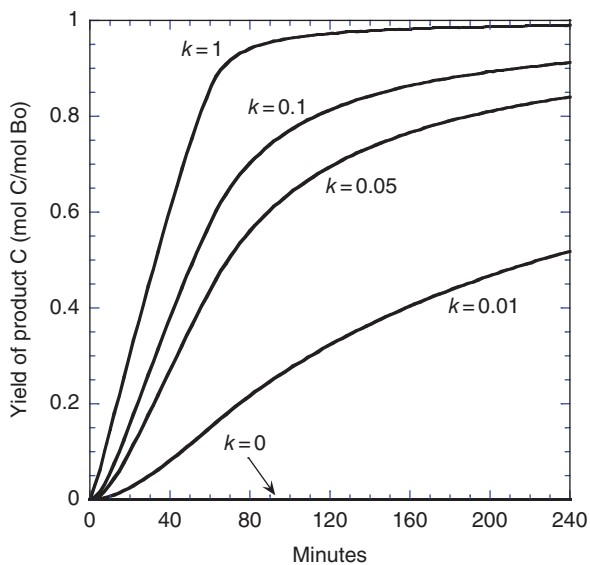
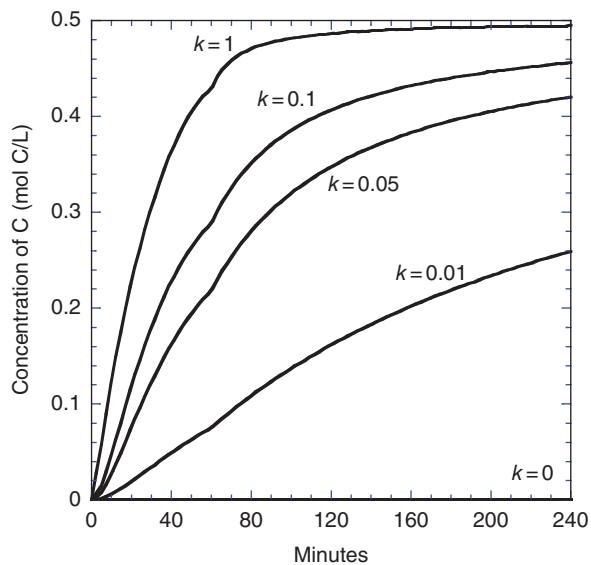
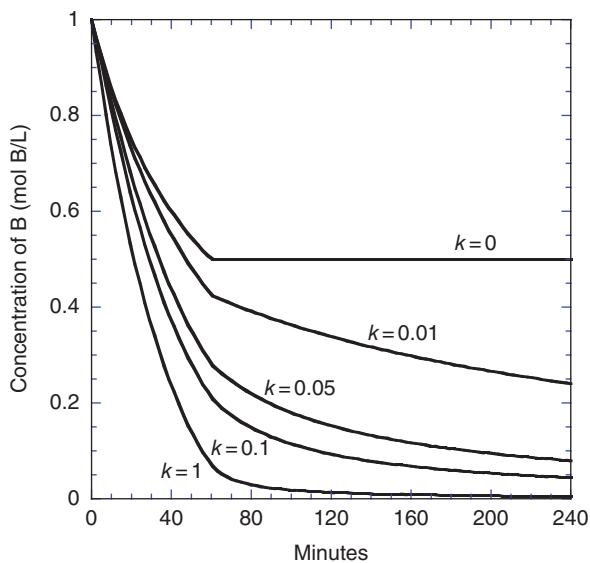
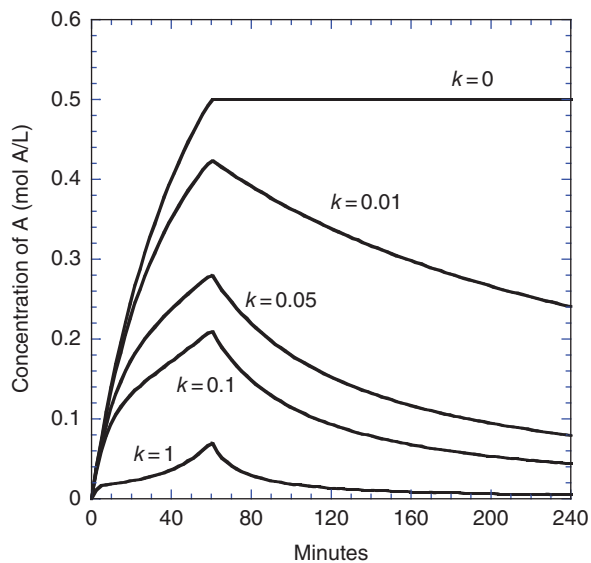
#Yields

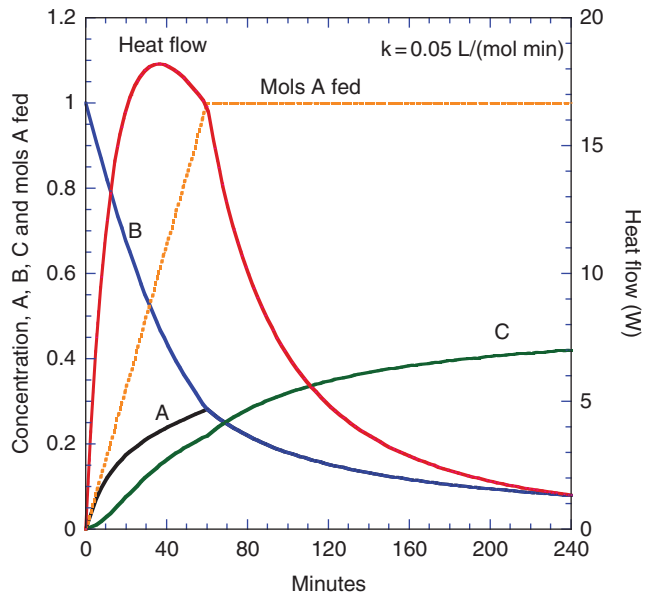
$YC = \text{if } (t > 0) \text{ then } (C_c * V)/(C_{bo} * V_o) \text{ else } 0 \# \text{Yield of C}$
 $XB = \text{if } (t > 0) \text{ then } (C_{bo} * V_o - C_b * V)/(C_{bo} * V_o) \text{ else } 0 \# \text{Conversion of B}$

#Initial Conditions

$t(0) = 0 \#$
 $C_a(0) = 0 \# \text{There is no A initially in the reactor}$
 $C_b(0) = 1 \# \text{ initial concentration of B (mol/L) initially in the reactor}$
 $C_c(0) = 0$
 $t(f) = 240 \# \text{ minutes}$

The plots below simulate concentration, heat, and yield profiles for rate constants of 0, 0.01, 0.05, 0.1, and 1 (L/(mol·min)) under Isothermal conditions.



**REFERENCES**

1. Pitts, D.R. and Sissom, L.E. (1997). *Schaum's Outline of Heat Transfer*, Schaum's Outline Series. New York: McGraw-Hill.
2. Treybal, R. (1980). *Mass Transfer Operations*, 3rd ed., 684–686. New York: McGraw-Hill.
3. Kinsley, G.R. (2001). Properly purge and inert storage vessels. *Chemical Engineering Progress* 97: 57–61.

PART I

INTRODUCTION

1

CHEMICAL ENGINEERING IN THE PHARMACEUTICAL INDUSTRY: AN INTRODUCTION

DAVID J. AM ENDE

Nalas Engineering Services, Inc., Centerbrook, CT, USA

MARY T. AM ENDE*

Process Modeling & Engineering Technology Group, Pfizer, Inc., Groton, CT, USA

Across the pharmaceutical industry chemical engineers are employed throughout research and development (R&D) to full-scale manufacturing and packaging in technical and managerial capacities. The chapters in these two volumes provide an emphasis on the application of chemical engineering science to process design, development, and scale-up for active pharmaceutical ingredients (APIs), drug products (DPs), and biologicals including sections on regulatory considerations such as design space, control strategies, process analytical technology (PAT), and quality by design (QbD). The focus of this introduction is to provide a high-level overview of bringing a drug to market and highlight industry trends, current challenges, and how chemical engineering skills are an exquisite match to address those challenges.

In general pharmaceuticals are drug delivery systems in which drug-containing products are designed and manufactured to deliver precise therapeutic responses [1]. The drug is considered the “active,” i.e. active pharmaceutical ingredient (API) or “drug substance,” and the formulated final dosage form is simply referred to as the drug product (DP).

This book focuses on API in volume 1 and DP in volume 2. The API and DP are designed and developed in R&D and then transferred to the commercial manufacturing sites by teams of organic chemists, analytical chemists, pharmaceutical scientists, and chemical engineers. Prior to the transition to the commercial site, co-development teams are formed with members from R&D and manufacturing working

together to define the computational and experimental studies to conduct based on risk and scientific considerations. The outcome of this multidisciplinary team effort forms the regulatory filing strategy for the API and drug products.

Once the commercial API and DP have been established, the co-development teams support three major regulatory submissions for a global product. A New Drug Application (NDA) is submitted to the US Food and Drug Administration (FDA), whereas in the Europe Union a Marketing Authorization Application (MAA) is submitted to the European Medicines Agency (EMA), and in Japan a Japan New Drug Application (JNDA) is submitted to the Pharmaceuticals and Medical Devices Agency (PMDA). Subsequently, the rest of world regulatory filings are led by the commercial division with no significant involvement by R&D since more commercial experience is available at the site by that time.

In the United States, federal and state laws exist to control the manufacture and distribution of pharmaceuticals. Specifically, the FDA exists by the mandate of the US Congress with the Food, Drug, and Cosmetics Act as the principal law to enforce and constitutes the basis of the drug approval process [1]. Specifically in the United States, “The FDA is responsible for protecting the public health by assuring the safety, efficacy, and security of human and veterinary drugs, biological products, medical devices, our nation’s food supply, cosmetics, and products that emit radiation. The FDA is also responsible for advancing the public health where possible by speeding innovations that make medicines and foods

*Current address: Lyndra Therapeutics, Watertown, MA, USA

more effective, safer, and more affordable. They also serve the public by ensuring accurate, science-based information on medicines and foods to maintain and improve their health.”¹ On 28 March 2018 the FDA announced organizational changes available on their website. Janet Woodcock remains the director of the small molecule division, referred to as Center for Drug Evaluation and Research (CDER).² Peter W. Marks is the director of the large molecule division, referred to as Center for Biologics Evaluation and Research (CBER).³ Further information can also be easily obtained from the FDA website, including the overall drug review process, current good manufacturing practices (cGMP), International Council on Harmonization (ICH), and mechanisms to comment on draft guidances, recalls, safety alerts, and warning letters that have been issued to companies.⁴

EMA is a decentralized body of the European Union with headquarters in London whose main responsibility is the protection and promotion of public and animal health, through the evaluation and supervision of medicines for human and veterinary use.⁵

The Japan Pharmaceutical Affairs Law (JPAL) is a law intended to control and regulate the manufacturing, importation, sale of drugs, and medical devices.⁶ It exists to assure the quality, efficacy and safety of drugs, cosmetics, and medical devices while improving public health and hygiene. The JPAL also provides guidance to pharmaceutical companies on how to translate their QbD control strategy, which was found to align well with the three levels of criticality initially used in early QbD filings for noncritical, key, and critical process parameters. Japan’s Ministry of Health, Labour and Welfare (MHLW) has issued clear guidance in English for three key ministerial ordinances to assure compliance requirements for manufacturers.

Japan, Europe, and United States collaborate as the International Council on Harmonization – Quality (ICH) to establish greater expectations for science and risk-based approaches to transform the pharmaceutical industry over the past decade. Critical to that transformation were the QbD guidances, Q8, Q9, and Q10 [2–4]. The final versions of the guidances are readily available on the CDER website, including the more recent QbD guidance for drug substance composed in Q11.⁷

¹<http://www.fda.gov/AboutFDA/WhatWeDo/default.htm>

²<https://www.fda.gov/AboutFDA/CentersOffices/OrganizationCharts/ucm350895.htm>

³<https://www.fda.gov/AboutFDA/CentersOffices/OrganizationCharts/ucm350556.htm>

⁴<https://www.fda.gov/default.htm>

⁵<http://www.ema.europa.eu/htms/aboutus/emeaoverview.htm>

⁶<http://www.jouhoukoukai.com/repositories/source/pal.htm>

⁷<https://www.fda.gov/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/ucm065005.htm>

1.1 GLOBAL IMPACT OF THE INDUSTRY

The value of the pharmaceutical industry to the American economy is substantial. In 2016, the industry employed over 854 000 people with each job indirectly supporting an additional 4 jobs. Thus as an aggregate, the industry supported 4.4 million jobs and generated nearly \$1.2 trillion in annual economic output when direct, indirect, and induced effects were considered for 2016.⁸

As an industry sector, the pharmaceutical industry is considered profitable, in spite of the high attrition rate for new chemical entities (NCEs).⁹ For example, *Forbes* estimated the profit margin for the health-care technology industry in 2015 to be approximately 21%, clearly placing near the top for profitable industries.¹⁰ The companies that are most profitable in this sector were major pharmaceutical and generics companies. As far as total revenues in pharmaceutical sales, the top 20 pharmaceutical companies are listed in Table 1.1.

Based on revenue, the pharmaceutical and biopharmaceutical companies are based in the following countries: 9 (United States), 2 (Switzerland), 2 (United Kingdom), 1 (France), 3 (Germany), 1 (Israel), 1 (Denmark), and 1 (Republic of Ireland). Only 1 company in the top 20 revenue producing is privately held.

Global prescription drug sales are on the order of \$800 billion in 2017. These drug sales are forecasted to grow at 6.3% compound annual growth rate (CAGR) between 2016 and 2022 to nearly \$1.2 trillion (as shown in Figure 1.1),¹¹ while generic drugs account for approximately 10% of those sales figures.

There is considerable value in being the first company to deliver a new medicine that treats a new indication (e.g. breakthrough therapy designation from regulators) or uses a new mechanism of action to benefit patients. Therefore, new developments in pharmaceutical R&D that speed quality drug candidates to the market are important investments for the future.

1.2 INVESTMENTS IN PHARMACEUTICAL R&D

R&D is the engine that drives innovation of new drugs and therapies. Significant investment is required to discover and advance potential NCEs and new molecular entities





















⁸<http://phrma.org/industryprofile>

⁹<http://money.cnn.com/magazines/fortune/global500/2009/performers/industries/profits>

¹⁰<https://www.fool.com/investing/2016/07/31/12-big-pharma-stats-that-will-blow-you-away.aspx>

¹¹<http://info.evaluategroup.com/rs/607-YGS-364/images/wp16.pdf>; some estimates are even higher, with 2017 global revenue of \$1.05 trillion (see, for example, <https://www.fool.com/investing/2016/07/31/12-big-pharma-stats-that-will-blow-you-away.aspx>).

TABLE 1.1 Top 20 Pharmaceutical Companies Based on 2017 Revenue as Listed in Wikipedia

#	HQ	Company	2017 (USD Billions)	2016 (USD Billions)	2015 (USD Billions)	2014 (USD Billions)	2013 (USD Billions)	2012 (USD Billions)	2011 (USD Billions)
1		Johnson & Johnson NYSE: JNJ	76.50	71.89	70.10	74.30	71.31	67.20	65.00
2		Roche OTCQX: RHHBY	57.37	50.11	47.70	49.86	48.53	47.80	45.21
3		Pfizer NYSE: PFE	52.54	52.82	48.85	49.61	51.58	58.99	65.26
4		Novartis NYSE: NVS	49.11	48.52	49.41	58.00	57.36	56.67	58.57
5		Sanofi NYSE: SNY	42.91	36.57	36.73	43.07	42.08	46.41	44.34
6		Glaxo SmithKline LSE: GSK	42.05	34.79	29.84	37.96	41.61	39.93	41.39
7		Merck & Co. NYSE: MRK	40.10	39.80	39.50	42.24	44.03	47.27	48.05
8		AbbVie NYSE: ABBV	28.22	25.56	22.82	19.96	18.79	–	–
9		Bayer FWB: BAYN	27.76	25.27	24.09	25.47	24.17	24.30	23.11
10		Abbott Laboratories NASDAQ: ABT	27.39	20.85	20.4	20.25	21.85	39.87	38.85
11		Gilead Sciences NASDAQ: GILD	25.70	30.39	32.15	24.47	10.80	9.70	8.39
12		Eli Lilly & Co NYSE: LLY	22.90	21.22	20.00	19.62	23.11	22.60	24.29
13		Amgen NASDAQ: AMGN	22.80	22.99	21.66	20.06	18.68	17.30	15.58
14		AstraZeneca LSE: AZN	22.47	23.00	24.71	26.10	25.71	27.97	33.59
15		Teva Pharmaceutical Industries NASDAQ: TEVA	22.40	21.90	20.00	20.27	20.31	18.31	16.12
16		Boehringer Ingelheim Private	21.67	17.54	16.41	17.70	18.68	18.89	18.34
17		Bristol-Myers Squibb NASDAQ: BMY	20.80	19.43	16.56	15.88	16.39	17.62	21.24
18		Novo Nordisk NYSE: NVO	18.77	16.61	16.06	15.83	14.88	13.48	11.56
19		Merck Group ETR: MRK	15.32	15.80	13.95	14.99	14.77	13.02	12.83
20		Shire NASDAQ: SHPG	14.40	11.40	6.42	6.00	4.76	4.68	3.50

Source: From https://en.wikipedia.org/wiki/List_of_largest_pharmaceutical_companies_by_revenue#cite_note-28. Licensed under CC BY 3.0.

(NMEs). For example, the pharmaceutical industry invested approximately \$150 billion into R&D in 2015. Worldwide pharmaceutical R&D spending is expected to grow by 2.8% (CAGR) to \$182 billion in 2022 (Figure 1.2).¹² The cost of advancing drug candidates and entire pharmaceutical portfolios in R&D is significant. In 2001 the average cost for

an approved medicine was estimated to be \$802 million, and by the end of 2014, the average cost escalated to \$2.6 billion as reported by Tufts Center for the Study of Drug Development.¹³ Although these figures clearly depend on the drug type, therapeutic area, and speed of development, the bottom

¹²<http://info.evaluategroup.com/rs/607-YGS-364/images/wp16.pdf>

¹³Based on estimated average pre-tax industry cost per new prescription drug approval (inclusive of failures and capital costs: source, DiMasi et al. [5]).

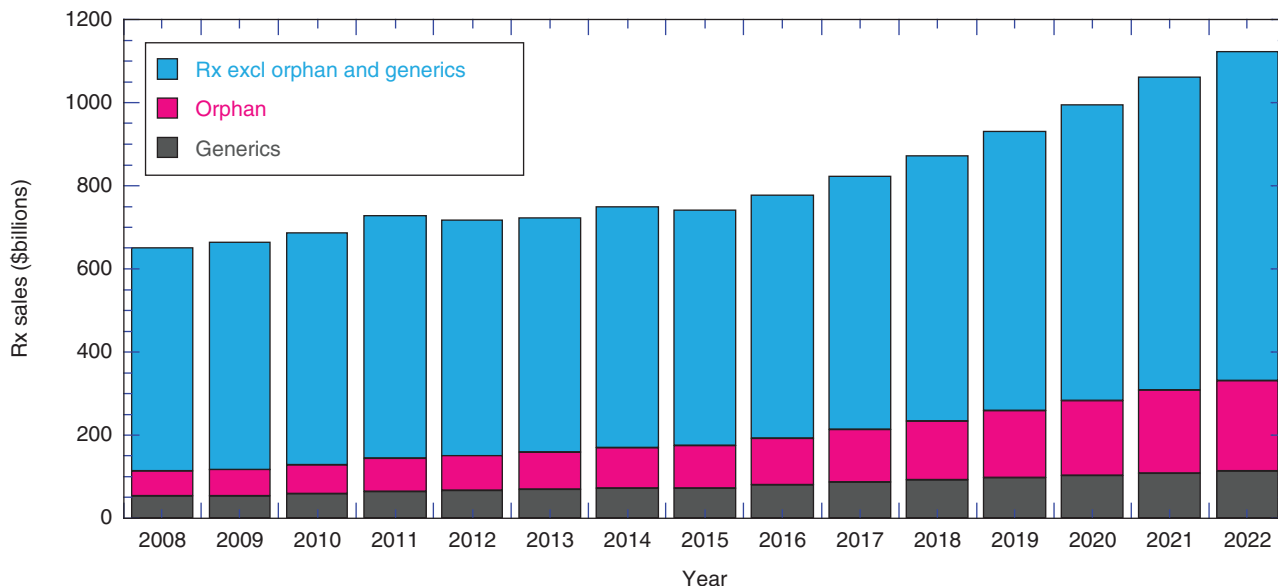


FIGURE 1.1 Global pharmaceutical prescription sales as a function of the type of drug. Global prescription drug sales were on the order of \$800 billion in 2017. These drug sales are forecasted to grow at 6.3% compound annual growth rate (CAGR) between 2016 through 2022 to nearly \$1.2 trillion while generic drugs account for approximately 10% of those sales figures. *Source:* From <http://info.evaluategroup.com/rs/607-YGS-364/images/wp16.pdf>.

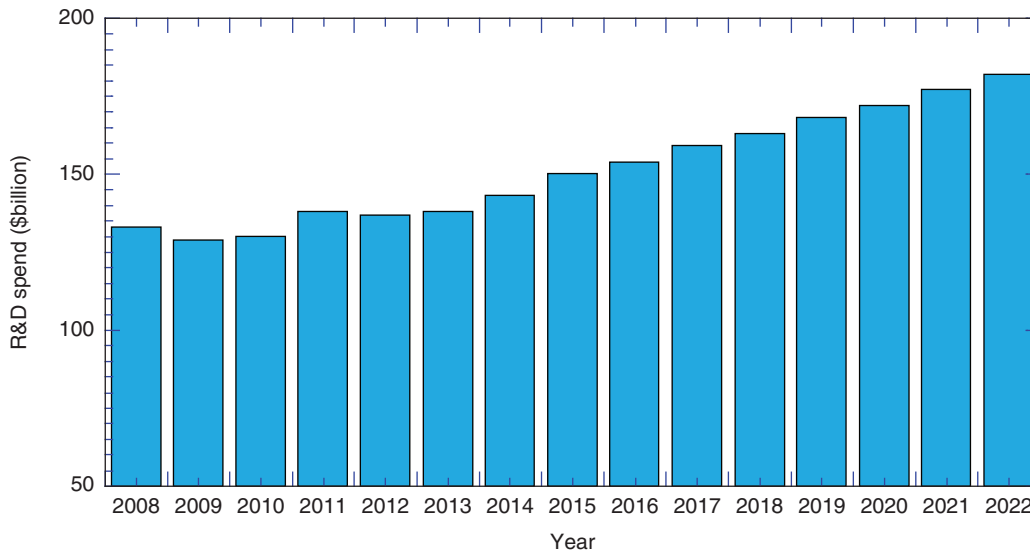


FIGURE 1.2 World-wide pharmaceutical R&D spend in 2015 was approximately \$150 billion. Growth rate in R&D spend is projected to grow at a rate of 2.8%. *Source:* EvaluatePharma®.

line is that the up-front investments required to reach the market are massive especially when considering the uncertainty whether the up-front investment will payback.

Given there might be 10 or more years of R&D costs without any revenue generated on a NCE or NME, the gross margins of a successful drug need to cover prior R&D investments and candidate attrition and to cover the

continuing marketing and production costs. Figure 1.3 shows the classic cash flow profile for a new drug developed and marketed. First there is a period of negative cash flow during the R&D phase. When the drug is approved and launched, only then are revenues generated, which have to be priced high enough to recoup the extensive R&D investment and provide a return on the investment.

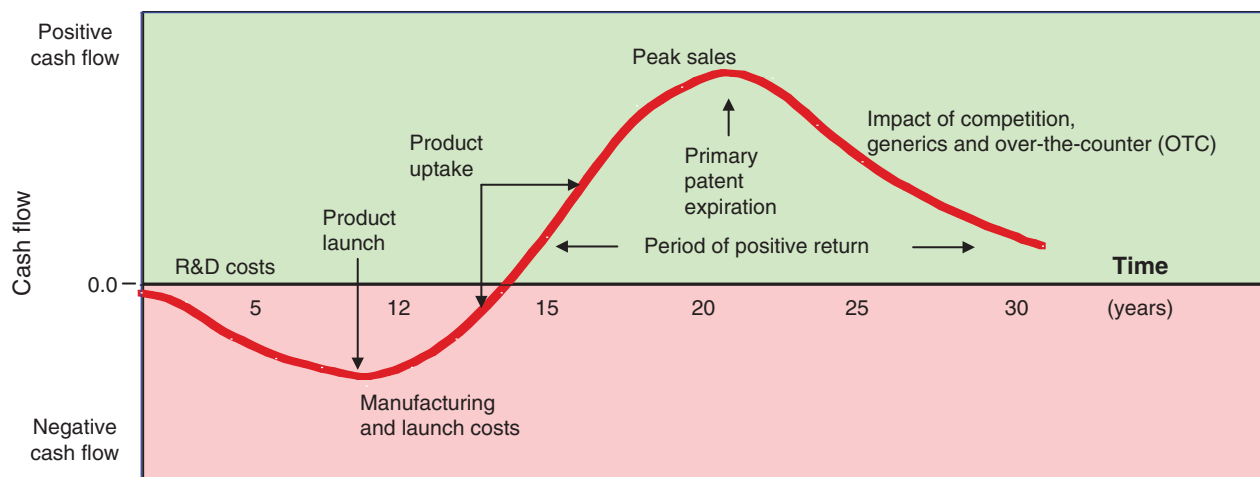


FIGURE 1.3 A hypothetical cash flow curve for a pharmaceutical product includes 10–15 years of negative cash flows of typically \$1–3 billion. Reasonably high margins are needed, once the drug is on the market, if it is to recuperate and provide a positive return on investment (ROI) over its lifecycle.

The net present value (NPV) calculation is one way to assess return on investment with a discount rate of 10–12% generally chosen in the pharmaceutical industry as the rate to value products or programs for investment decisions [6]. The highest revenues for a new drug are achieved during the period of market exclusivity (where no competitors can sell the same drug). So it is in the company's best interest to ensure the best patent protection strategy is in place to maximize the length of market exclusivity. Patents typically have a validity of 20 years from the earliest application grant date based on applications filed after 1995. In some cases the time of market exclusivity can be extended through new indications, new formulations, and devices, which may themselves be patent protected (see Table 1.2).

Once market exclusivity ends, generic competition is poised to immediately introduce an alternative cheaper option that will erode sales for the patent owner. A dramatic example of patent cliff can be seen in the sales of Lipitor (Figure 1.4). Peak sales occurred in 2006 with sales nearing \$13 billion in revenue, but at the end of patent exclusivity in 2011, sales dropped off precipitously to less than \$4 billion in 2012. The trend continued to drop off through 2017 to less than \$2 billion.

It now takes 10–15 years for a new medicine to go from the discovery laboratory to the pharmacy. Figure 1.5 shows the typical development activity timeline from discovery to launch. From thousands of compounds evaluated for potential therapeutic effect, very few will clear all the safety, efficacy, and clinical hurdles to make it to approval. Figure 1.5 also shows how a general range of volunteers, and clinical supplies, increases through phases I–III of clinical trials with clinical development typically lasting six years or more.

Before entering human clinical studies, the drug candidate is tested for safety and efficacy in preclinical studies. When

TABLE 1.2 Periods of Exclusivity Granted by the FDA

Specific FDA Applications	Period of Exclusivity
New Chemical Entity Exclusivity (NCE)	5 years
Orphan Drug Exclusivity (ODE)	7 years
Generating Antibiotic Incentives Now (GAIN)	5 years added to certain exclusivities
New Clinical Investigation Exclusivity	3 years
Pediatric Exclusivity (PED)	6 months added to existing patents/exclusivity

Source: Form https://www.fda.gov/Drugs/DevelopmentApprovalProcess/ucm079031.htm#What_is_the_difference_between_patents_a

the candidate looks promising for a targeted indication or potential therapeutic effect, the company files an Investigational New Drug Application (IND) for regulatory agency and clinical site approval. At this time, referred to as phase I, the drug candidate will be tested in a few healthy volunteers ($n \sim 10$'s) in single and multiple dose studies to test for safety and understand human pharmacokinetics. If the phase I evaluations are positive, then the candidate can progress to a larger population of healthy volunteers ($n \sim 100$'s) pending approval by the regulatory agency on study design, i.e. doses, route of administration, detection of efficacy, and side effects. If the candidate passes the phases I and II hurdles ensuring safety and efficacy, then the clinical teams will design incrementally larger, broader, and worldwide clinical studies in test patients (phase III, $n \sim 1000$'s).

The two common exceptions to conducting phase II studies in healthy volunteers are for oncology or biological candidates. These candidates proceed directly into the patient population, referred to as phase III, to test treatment of the

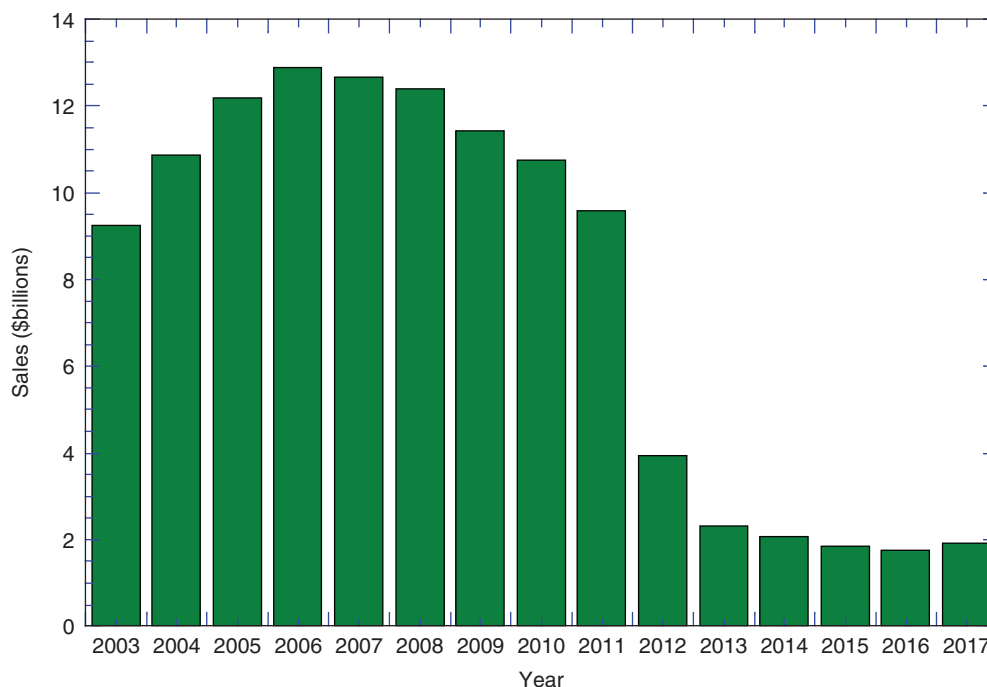


FIGURE 1.4 Sales of Pfizer's Lipitor (atorvastatin) between 2003 and 2017. In 2006 Lipitor generated nearly \$13 billion in revenue. Patent exclusivity ended in 2011 and its impact was significant as seen by the significant drop on revenue in subsequent years (known as the "patent cliff"). *Source:* Data from www.statista.com.

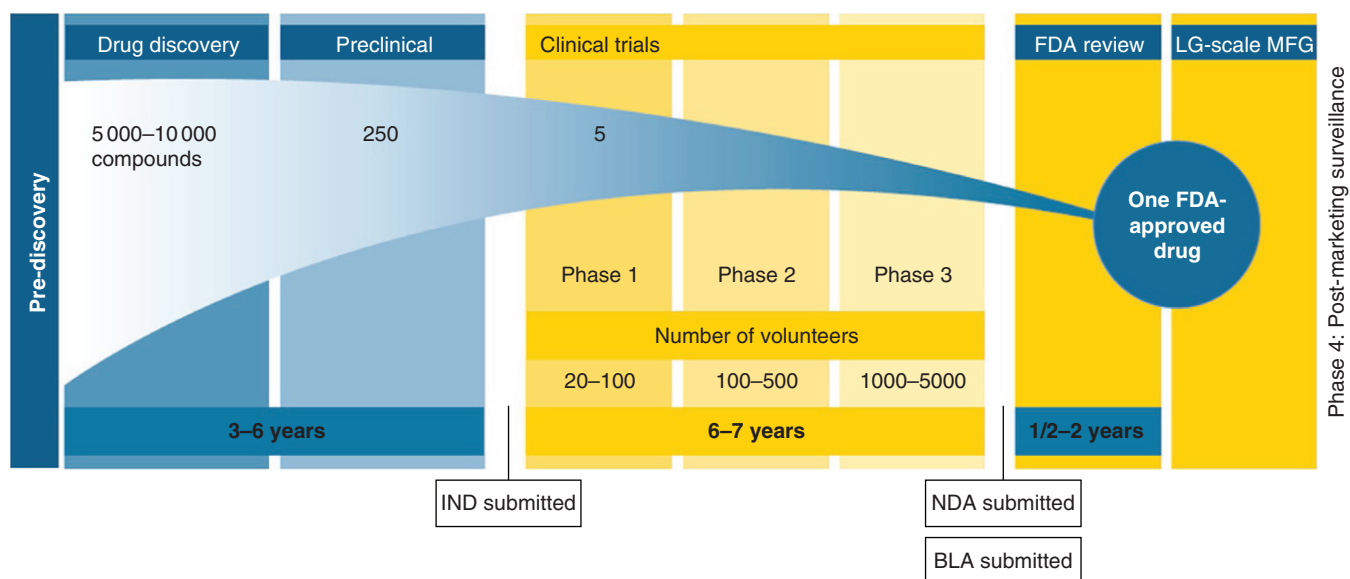


FIGURE 1.5 Drug research and development can take 10–15 years with one approval from 5 to 10 000 compounds in discovery. BLA, biologics license application; FDA, food and drug administration; IND, investigational new drug; NDA, new drug application. *Source:* Adapted from Pharmaceutical Research and Manufacturers of America (PhRMA), publication *Pharmaceutical Industry Profile 2009* (www.phrma.org).

indicated cancer or to progress the known safe and efficacious candidate derived from human antibodies or viruses, respectively.

After several years of careful study, the drug candidate may be submitted to the regulatory agency (e.g. FDA,

EMA, PMDA) for approval. Depending on the type of API, the regulatory submission may need to be filed differently. For example, in the United States, a small molecule is submitted as an NDA, while a biologic is submitted as a Biologics Licensing Application (BLA).

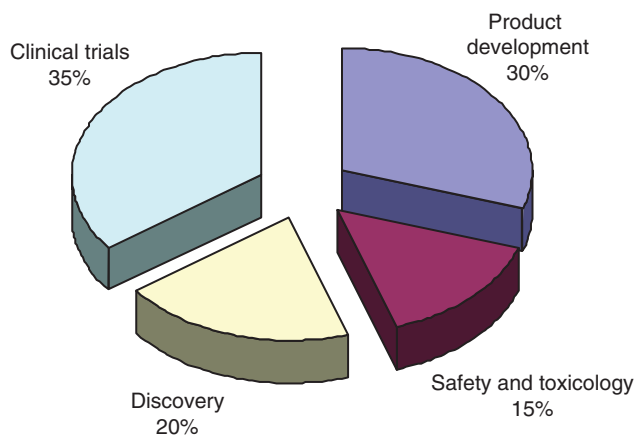


FIGURE 1.6 Estimated distribution of product development costs within R&D with the total cost to bring a new chemical entity NCE to market in the range of \$1–3.5 billion. *Source:* Adapted from Suresh and Basu [7].

As mentioned, the 2014 cost to advance a NCE or NME to market was estimated at \$2.6 billion. The cost of product development that includes the cost to manufacture clinical supplies is estimated to be in the range of 30–35% of the total cost of bringing a NCE/NME to market with the following other cost contributors: discovery 20–25%, safety and toxicology 15–20%, and clinical trials 35–40% [7]. The distribution is graphically displayed in Figure 1.6. Clearly the distribution will depend on the specific drug, its therapeutic area, dose, and specific company.

Chemical engineers, chemists, biologists, pharmaceutical scientists, and others make up the diverse scientific disciplines of product development that include API and formulation development including API and DP manufacture of clinical supplies.

1.3 BEST SELLERS

The top 20 drugs in sales are shown in Table 1.3 with Humira, topping the list with 2017 global sales of \$18.43 billion. Interestingly 11 of these top drugs are biologics, 1 is a vaccine, and the remaining 8 are small molecule drugs. The top 20 selling drugs in that year total nearly \$135 billion. This has changed significantly since the publication of the original version of this book in 2010 when the majority of top-selling drugs at that time were small molecules.

The majority of the 20 top sellers have remained in similar positions over the past 2 years; however a few have made significant moves in this short time. For instance, Harvoni was the second place with \$9.08 billion in sales in 2016 and dropped to seventeenth place in 2017 with \$4.37 billion sales. Another interesting move was Eylea from thirteenth to second place from 2016 to 2017 increasing sales from \$5.05

to 8.23 billion. It is also noteworthy that 9 of the top 20 products are partnerships, which further illustrates the significant cost to develop DPs are often sharing the risk.

In Table 1.4 the top-selling drugs of all time were analyzed by *Forbes*, utilizing the lifetime sales of branded drugs between 1996 and 2012 and company reported sales data between 2013 and 2016. It is noteworthy that the number one in sales, Lipitor, at \$148.7 billion is not even on the top 20 drug sales list for 2017 in Table 1.3. While there is a large gap between the top two selling drugs, amounting to \$53 billion for Lipitor above Humira, if Humira annual sales continue at \$18 billion, it will outperform Lipitor as the all-time best-selling drug in just under 3 years. However, the patent expiry for Humira was in 2016, and therefore sales may drop rapidly in the coming years if generics or biosimilars are able to penetrate the market.

1.4 PHARMACEUTICAL RESEARCH AND DEVELOPMENT EXPENDITURES

1.4.1 Pharmaceutical Development

In general, pharmaceutical product development is different than most other research intensive industries. Specifically in the pharmaceutical industry, there is the consistent need to ensure that clinical supplies are manufactured and delivered in a timely manner regardless of the current state of development or efficiency of the process. In other words, delivering clinical supplies when they are needed requires using technology that is good enough at the time even if it is not a fully optimized process. However, this is a regulated industry for clinical supplies as well as for commercial.

Further, process development, optimization, and scale-up historically tends to be an iterative approach [8] – clinical supply demands are met by scale-ups to kilo lab or pilot plant through phase I, phase II, and phase III, and it is through this period that R&D teams (including analysts/chemist/engineers, referred to as the ACE model) refine, optimize, and understand the API and DP processes to enable them to be eventually transferred to manufacturing. Manufacturing of clinical supplies in kilo lab, pilot plant, and solid dosage plants occurs under the constraints of cGMP conditions, which is discussed further in the chapter on kilo lab and pilot plant. The pilot plant and kilo lab are also sometimes used to “test” the scalability of a process. In this way, pilot plants serve a dual purpose, which make them unique as compared with non-pharmaceutical pilot plants. In terms of cost, however, large-scale experimentation in kilo lab or pilot plant can be significant – so there has been a shift toward greater predictability at lab scale to offset the need for pilot plant-scale “technology demonstration” experiments. Engineers through their training are well suited to scale-up and scale-down processes and can effectively model the chemical and physical

TABLE 1.3 Top 20 Global Pharmaceutical Products (2017 Sales)

Rank	Brand Name	API	Marketer	Indication	2017 Sales (\$ Billion)
1	Humira	Adalimumab	AbbVie	Autoimmune diseases and rheumatoid arthritis	18.43
2	Eylea	Aflibercept	Regeneron Pharmaceuticals and Bayer	Macular degeneration	8.23
3	Revlimid	Lenalidomide	Celgene	Multiple myeloma	8.19
4	Rituxan	Rituximab (MabThera)	Roche and Biogen	Treatment of cancer	8.11
5	Enbrel	Etanercept	Amgen and Pfizer	Autoimmune diseases including rheumatoid arthritis, psoriasis, and other inflammatory conditions	7.98
6	Herceptin	Trastuzumab	Roche and Biogen	Treatment of cancer, mainly breast and gastric	7.55
7	Eliquis	Apixaban	BMS and Pfizer	Anticoagulant, mainly used to treat atrial fibrillation and deep vein thrombosis	7.40
8	Avastin	Bevacizumab	Roche and Biogen	Advanced colorectal, breast, lung, kidney, cervical, and ovarian cancer and relapsed glioblastoma	7.21
9	Remicade	Infliximab	Johnson & Johnson and Merck	Autoimmune diseases	7.16
10	Xarelto	Rivaroxaban	Bayer and Johnson & Johnson	Anticoagulant	6.54
11	Januvia/ Janumet	Sitagliptin	Merck	Treatment of type 2 diabetes	5.90
12	Lantus	Insulin glargine	Sanofi	Long-acting human insulin analog for the treatment of diabetes	5.65
13	Prevnar 13/ Prevener	Pneumococcal 13-valent conjugate vaccine	Pfizer	Pneumococcal vaccine	5.60
14	Opdivo	Nivolumab	BMS	Melanoma	4.95
15	Neulasta/ Peglasta/ Neupogen	(Pegfilgrastim and Filgrastim)	Amgen and Kyowa Hakko Kirin	Neutropenia; decreases the incidence of infection during cancer treatment	4.56
16	Lyrica	Pregabalin	Pfizer	Anti-epileptic and neuropathic pain	4.51
17	Harvoni	Ledipasvir (sofosbuvir)	Gilead Sciences	HCV/HIV-1 infection	4.37
18	Advair	Fluticasone and Salmeterol	GlaxoSmithKline	Asthma	4.36
19	Tecfidera	Dimethyl fumarate	Biogen	Multiple sclerosis	4.21
20	Stelara	Ustekinumab	Johnson & Johnson	Plaque psoriasis	4.01

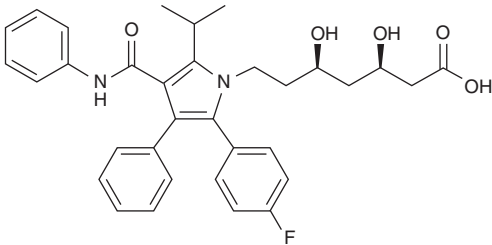
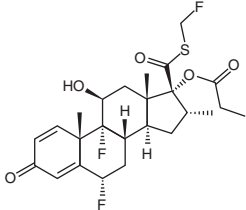
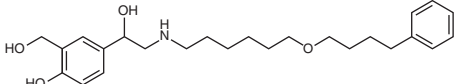
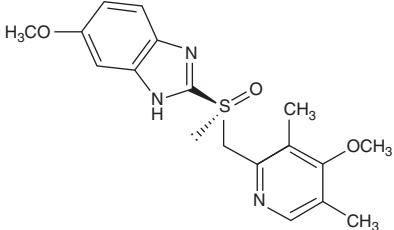
Source: From <https://igeahub.com/2018/04/07/20-best-selling-drugs-2018>

Shaded row indicates API is a new chemical entity; non-shaded row indicates API is a biologic.

behaviors in the lab to ensure success on scale. Many chapters in these two volumes discuss how scale-up/scale-down of various unit operations is performed. Chemical engineers are well trained in process modeling and optimization that support the reduction of experimentation and rehearsal

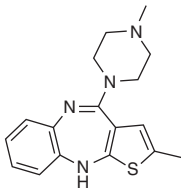
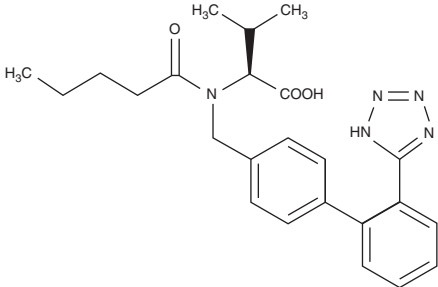
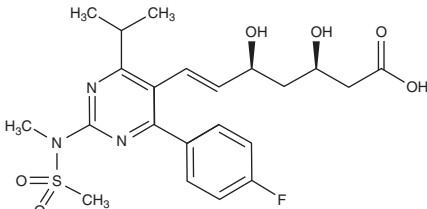
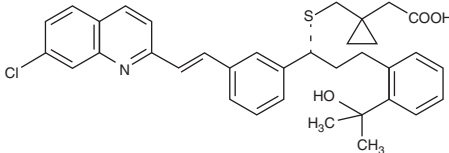
batches prior to commercialization. This helps to reduce the number of larger-scale “experiments,” thereby lowering costs during R&D. In this way, with the recent trend toward increasing efficiency and continuous improvement, the pilot plant and kilo labs are preferentially utilized to manufacture

TABLE 1.4 Fifteen Top-Selling Drugs (2013–2016) for Cumulative Sales Through 2016

Drug/Drug Product Type ^a	API	Marketer	Approval	\$ Billion
1 Lipitor Atorvastatin/film-coated tablet		Pfizer	1996	148.7
2 Humira Adalimumab/solution for injection	Antitumor necrosis factor (TNF) monoclonal antibody	AbbVie	2003	95.6
3 Advair (United States) Seretide(EU) fluticasone + salmeterol/dry powder inhaler		GlaxoSmithKline	2001	92.5
4 Remicade/lyophilized powder for constitution, solution injection	Anti-TNF chimeric monoclonal antibody	Janssen	1998	85.5
5 Plavix clopidogrel/film-coated tablet		Bristol-Myers Squibb	1997	82.3
6 Enbrel Etanercept/subcutaneous injection	Fusion protein produced by recombinant DNA	Amgen/Pfizer	1998	77.2
7 Rituxan Rituximab/solution injection	Chimeric monoclonal antibody	Roche Genentech	1997	75.9
8 Herceptin Trastuzumab/intravenous (IV) infusion	Monoclonal antibody	Roche Genentech	1999	65.2
9 Avastin Bevacizumab/IV infusion	Monoclonal antibody	Roche Genentech	2004	62.3
10 Nexium Esomeprazole/delayed release capsule; IV injection		AstraZeneca	2001	60.2

(continued)

TABLE 1.4 (Continued)

Drug/Drug Product Type ^a	API	Marketer	Approval	\$ Billion
11 Zyprexa Olanzapine/oral disintegrating tablets; injections		Eli Lilly	1996	60.2
12 Diovan Valsartan/film-coated tablet		Novartis	1997	60.1
13 Lantus Insulin glargine/Subcutaneous Injection	Long-acting basal analog	Sanofi-Aventis	2001	58.3
14 Crestor Rosuvastatin/film-coated tablet		AstraZeneca	2003	55.2
15 Singulair Montelukast/chewable tablet		Merck	1998	47.4

Source: Data from <https://www.fool.com/amp/investing/2017/03/13/the-19-best-selling-prescription-drugs-of-all-time.aspx>

^awww.drugs.com source of dosage form type for originator drug.

supplies for toxicological and clinical supplies rather than being used to “test” or verify that the chemistry or process will work on scale.

A primary focus of process development is to drive down the cost contribution of the API to the final formulated pharmaceutical product cost while at the same time optimizing to ensure quality and process robustness. The impact of API costs on overall manufacturing costs is approximated in Figure 1.7. The cost contribution of API is expected to increase with increasing complexity of molecular structures of APIs, e.g. biologics. It is interesting to note that API molecular complexity can often impact API cost more than formulation or packaging costs. As Federsel points out that, “Given the importance of ‘time to market’ which remains one

of the highest priorities of pharmaceutical companies, the need to meet increasingly stretched targets for speed to best route has come to the forefront in process R&D” [9]. In the not too distant past it was considered satisfactory to have a good-enough synthetic route that was fit for purpose (i.e. could support the quantities of material needed) but not one considered best or lowest cost (\$/kg of API). The prevailing view was that the market would bear higher product pricing as compensation for higher cost of goods (COGs). Further cost reduction through new routes could be and were pursued post-launch with savings realized later in the life cycle. According to Federsel, and evidenced frequently in contemporary R&D organizations, this approach is no longer viable, at least not as a default position. Instead the best

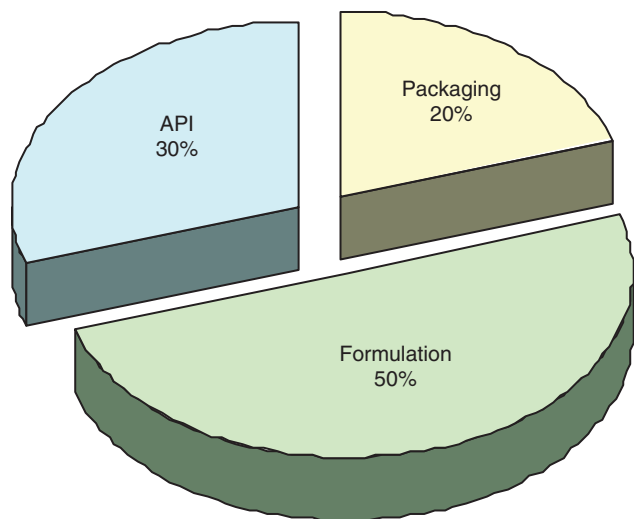


FIGURE 1.7 Average cost of goods (COG's) components in final dosage form across a large product portfolio – may vary widely for individual drugs (e.g. for API from 5 to 40%). *Source:* Reprinted with permission from Federsel [9]. Copyright (2006) Elsevier.

synthetic route to API (i.e. route with ultimate lowest cost materials) coupled with best process design and engineering (process with lowest processing costs) must be worked out as early as possible in API process development [9]. The optimal API process developed by the time of launch is necessary to extract additional revenues and respond to reduced COG margins. Achieving this requires continuous improvement in scientific and technical tools as well as multidisciplinary skill sets in the R&D labs, including chemical engineering science. The implementation of process design principles, drawing on the right skill sets, both from chemistry and engineering perspectives during clinical phase II, is considered such an important step toward leaner more cost-effective processes readied for launch that several portions of this book will expand on this concept.

1.5 RECENT TRENDS FOR PHARMACEUTICAL DRUG AND MANUFACTURING

During the past decade, the pharmaceutical field has evolved to a science and risk-based industry. It is now commonplace for the regulatory dossier to contain scientifically rigorous information and descriptions of the risk management approach used for decision making. Now, the industry is undergoing significant changes in the API (from small molecule to biologics), manufacturing (from batch toward continuous), medicinal approach (generalized to personalized), and complexity of manufacturing (from simple dosage forms toward additive manufacturing or 3D printing).

1.5.1 Drug Substances Trend Toward Biologics

Biologic medicines are revolutionizing the treatment of cancer, autoimmune disorders, and rare illnesses and are therefore critical to the future of the pharmaceutical industry. Cancer immunotherapy includes monoclonal antibodies, checkpoint inhibitors, antibody-drug conjugates (ADCs), and kinase inhibitors, to name a few.

From the 2017 top-selling drugs shown in Table 1.3, there is a strong trend toward drugs derived from biological origins dominating the market than small molecules. In fact, the majority of best sellers are biologics, often monoclonal antibodies, which treat new diseases such as Crohn's and ulcerative colitis previously unmet medical needs by small molecule APIs.¹⁴ It is also evident that the biologics retain their value even after patent exclusivity expires, e.g. Humira sales continue to grow post-patent expiry in 2016. The current generic industry is skilled in small molecule development but appears to be challenged to rapidly erode sales for biologics. In fact, in the coming years, it appears the first biologic medicine may take over as the all-time best seller from Lipitor.

Biological drug candidates include many different types of molecules including monoclonal antibodies, vaccines, therapeutic proteins, blood and blood components, and tissues.¹⁵ In contrast to chemically synthesized drugs, which have a well-defined structure and can be thoroughly verified, biologics are derived from living material (human, animal, microorganism, or plant) and are vastly larger and more complex in structure. Biosimilars are versions of biologic products that reference the originator product in applications submitted for marketing approval to a regulatory body and are not exactly generic equivalents. However, biosimilar DPs are far more complex to gain regulatory approval in developed markets than for chemical generics and may involve costly clinical trials. Those that succeed will also have to compete with the originator companies who are unlikely to exit the market considering their expertise and investments. The biosimilars market is expected to increase significantly with the first FDA approval for Sandoz ZAR-XIO subcutaneous IV injection product in 2015 that helped establish a clear pathway for gaining regulatory approval [10]. Recently, Hospira, a Pfizer company, received FDA approval of their epoetin alfa biosimilar, Retacrit, in May 2018 [11].

Biologic and biosimilar medicines are treating illnesses, with unmet needs while retaining value even after post-exclusivity period. These are clear advantages for the originator, biopharmaceutical company developing biologic medicines, and are expected to continue to increase in the coming

¹⁴www.blog.crohndology.com

¹⁵<https://www.selectusa.gov/pharmaceutical-and-biotech-industries-united-states>

years. While the major disciplines making advancements in this area are biologists and chemists, there is a role for chemical and biochemical engineers in the design and development of the processing and purification steps. Chemical engineers are skilled at developing predictive models, and scale-up/scale-down principles, which make them a key contributor to this growing field. In fact, for biologics, scale-down predictive models of process steps were established and helped pave the way for biological products to use them for validation [12].

Chemical engineers that include biochemical engineering are well trained to impact the biotech industry, which utilizes cellular and biomolecular processes for new medicines [13]. Chemical engineers can also support the design of protein recovery, purification, and scaling up from lab to commercial production of the therapeutic proteins.

1.5.2 Lean Manufacturing

Pharmaceutical production of APIs and DPs can be generally characterized as primarily batch-operated multipurpose manufacturing plants. At these facilities commercial supplies of API intermediates, APIs, and DPs are manufactured before being packaged, labeled, and distributed to customers. Pharmaceutical production plants were typically designed to be flexible to allow a number of different products to be run in separate equipment trains, depending on the demand. Further, these facilities have various degrees of automation, relatively high levels of documentation, and change control to manage reconfigurations, with relatively long downtimes for cleanup and turnover of the plant between product changes [14]. These considerations are in part to meet regulatory requirements for commercial manufacturing. Manufacturing costs or COGs often account for approximately one-third of the total costs with expenses exceeding that of R&D [15]. *For this reason COG's have received considerable focus as an area of opportunity for potential savings* [7, 16].

It has been claimed that through adopting QbD principles and principles of lean manufacturing, pharmaceutical companies, as an aggregate, could save in the range of \$20–50 billion/year by eliminating inefficiencies in current manufacturing [16]. This translates to 10–25% reduction in current COGs. An early QbD product approval of Chantix afforded an opportune chance to prove the benefits of these lean manufacturing and QbD principles. Chantix (varenicline tablet) was approved as an immediate release tablet commercially manufactured in the Pfizer Illertissen, Germany site. The OEB classification of this product required containment, available at small scale in this facility. Product demand increased dramatically in 1 year by 430%. The site employed lean manufacturing to eliminate process inefficiencies and wastes to increase production from 1 batch/day to 3 batches/day, delivering the desired 900 batches/year in the small-scale facility [17]. The lean manufacturing was indeed

proven when the manufacturing maintained an inventory of only approximately one week lead ahead of demand.

The principles of lean manufacturing are often cited as an approach to reduce COGs in pharmaceutical development and manufacturing. Lean manufacturing describes a management philosophy concerned with improving profitability through the systematic elimination of activities that contribute to waste – thus the central theme to lean manufacturing is the elimination of waste where waste is considered the opposite of value. Based on the work of Taiichi Ohno, creator of the Toyota Production System, wastes are considered based on the following [18]:

- Overproduction
- Waiting
- Transportation
- Unnecessary processing
- Unnecessary inventory
- Unnecessary motion
- Defects

All of these wastes have the effect of increasing the proportion of non-value-added activities. Lean thinking is obviously applicable to many industries including pharmaceutical manufacturing as well as pharmaceutical development. Continuous processing, for pharmaceutical APIs and DPs, is one application of lean thinking applied to pharmaceutical manufacturing. The challenge is that batch processing inherently leads to overproduction and specifically the buildup of excess inventory of intermediates and DPs to supply the market. This leads to longer cycle times and is addressed through the concepts of continuous manufacturing (CM).

According to Ohno, “The greatest waste of all is excess inventory” where in simplest terms, excess inventory incurs cost associated with managing, transporting, and storing inventories adding to the waste. Large inventories also tie up large amounts of capital. Excess inventory represents an opportunity cost where capital is held up in the form of work in process (WIP), API finished goods, and formulated finished goods versus what could be invested elsewhere or back into R&D. Implementation of lean manufacturing principles can be used to develop workflows and infrastructures to reduce inventories. One way to reduce inventories is through continuous processing. Several chapters discuss the technical benefits of CM. A reliable steady delivery of product API and DP through small product-specific continuous plants could potentially reduce the level of inventory required in a dramatic way if the workflows were designed to ensure consistent delivery of product to packaging and distribution. *The facilities of continuous production trains tend to be significantly smaller.*

The costs of inventory holdings are significant and include both the carrying cost and the cash value of the inventory.

The carrying costs of inventory include two main contributions – (i) weighted average cost of capital (WACC) and (ii) overhead [19].

Estimates for the combined carrying cost of WACC and overhead range from 14 to 25% that translates to approximately 20% return for every dollar of inventory eliminated [20]. Technology platforms and new workflows designed to minimize the need for stockpiling API and DP inventories across the industry therefore would seem to offer very rapid payback.

1.5.3 Continuous Manufacturing

For a large pharmaceutical company carrying \$5 billion in inventories, the holding cost based on the combined WACC and overhead of 20% is approximately \$1 billion/year. Considered another way, technologies that ensure a reliable and steady distribution of product with the result of eliminating the need to build and store massive inventories can return the company cost savings equivalent to a blockbuster drug (generating \$billion/year). Indeed one of the three factors having the largest impact on the profitability of a manufacturing organization is inventory with the other two being throughput and operating expense according to Goldratt and Cox [21]. Continuous processing if designed for reliable operations essentially year-round or in other cases simply “on demand” could potentially eliminate the need to accumulate significant inventories above and beyond two to four weeks of critical safety stocks of finished goods.

CM across API and DP integrated under one roof as a platform technology is one long-term approach to transforming the way the industry manages their commercial supply chain.

As one reference cites, “Even for very small processes, continuous processes will prove to be less expensive in terms of equipment and operating costs. Dedicated continuous processes often put batch processes out of business” [22]. The real point here is that continuous is one approach to lean manufacturing and to reducing inventories and costs but certainly not the only approach. Other lean systems can be devised that utilize the existing batch facilities as well. Since the publication of the first edition of this book in 2010, there has been a significant wave of interest in considering continuous processing for pharmaceutical API and DP.

In July 2015 FDA granted Vertex Pharmaceuticals approval of the first DP, Orkambi, a cystic fibrosis (CF) drug, to be produced using a CM process. Vertex’s second drug, Symdeko, for treating the underlying cause of CF occurred in February 2018. Janssen aims to manufacture 70% of their “highest volume” products using CM within eight years.¹⁶ In addition, they intend to increase yield by reducing waste by 33% and reduce manufacturing and testing cycle times by

80% through the use of CM. Their claim is that CM can reduce operating costs by as much as 50%, gain higher throughputs, and significantly reduce waste [23]. Janssen’s HIV drug Prezista is also manufactured via a continuous process after obtaining approval to convert from batch to continuous.¹⁷ Pfizer and Eli Lilly have made investments in CM and recently submitted an NDA or gained approval for a product, respectively. Merck states that CM will help achieve their goals of well-controlled processes with flexible sizes to handle small-to-large volume products localized closer to the customer.¹⁸

Merck Manufacturing Division targets a total lead time of 90 days formulation to the patient, reducing the current timing by one-quarter. Multiple companies are teaming together to leverage CM, e.g. Novartis-MIT Center for Continuous Manufacturing. A critical component of CM is that PAT is embedded into the overall plan for monitoring and control of the process. As stated by Kevin Nepveux, “one of the best ways to go about implementing CM processes is to develop the analytics in-line with the application.”¹⁹ In summary, CM requires significant focus by chemical engineers as there is more attention on cost savings and cost efficiencies.²⁰

1.5.4 Personalized Medicine

“One size doesn’t fit all is a tenet of personalized medicine, also called precision medicine,” states Lisa Esposito in a recent report [24]. In her article, she highlights the long-standing personalized medicine approach taken to treat cancer based on the individual patient’s disease state and conditions. There is a growing expectation that the pharmaceutical industry should deliver DPs targeted to the individual, tailoring the amount of drug based on their mass, metabolism, genetic factors, and disease state. In this section, we discuss two approaches for manufacturing personalized/precision medicine through a pharm-on-demand concept for military personnel and for complex dosage forms using additive manufacturing (referred to as 3D printing).

The Defense Advanced Research Projects Agency (DARPA) Battlefield Medicine program is keenly interested in miniaturized, flexible platforms for end-to-end manufacturing of pharmaceuticals to support the troops on location. As discussed in other chapters within this book, advances in continuous flow synthesis, chemistry, biological engineering, and downstream processing, coupled with online analytics, automation, and enhanced process control measures, are

¹⁷<https://www.lifescienceleader.com/doc/merck-s-path-to-continuous-manufacturing-for-solid-oral-dose-products-what-stands-in-the-way-0001>

¹⁸ibid.

¹⁹<https://www.pharmaceuticalonline.com/doc/pfizer-s-hybrid-approach-to-implementing-continuous-manufacturing-processes-0001>

²⁰http://www.contractpharma.com/contents/view_features/2018-01-30/pharmaceutical-industry-outlook

¹⁶<http://www.pharmtech.com/fda-approves-tablet-production-janssen-continuous-manufacturing-line>

proving that such capabilities are ready for implementation. The desire is to have a mobile, on-demand pharmacy located at the battlefield that could ensure readiness to treat threats of chemical, biological, radiological, and nuclear weapons [25].

1.5.5 Additive Manufacturing

Additive manufacturing, also referred to as three-dimensional (3D) printing, is an automated process of building layer by layer a complex dosage form personalized and manufactured on demand. FDA approved the first 3D printed DP in August 2015 for Aprexia Pharmaceuticals SPRITAM product as a disintegrating tablet [26].

The 3D printing in this case binds the powders while maintaining a porous structure (without the typical compression of a tablet press), providing a fast dissolving tablet. For example, 1000 mg of levetiracetam dissolves within seconds [27]. Extensions of 3D printing include printing extremely low dose APIs or highly potent APIs, but encapsulated with excipients, thus reducing potential exposure. Norman et al. [27] provide a thorough review of the different modalities of 3D printing for pharmaceutical manufacturing, which includes an analysis of the potential benefits of such products.

1.6 CHEMICAL ENGINEERS SKILLED TO IMPACT FUTURE OF PHARMACEUTICAL INDUSTRY

The fundamental principles taught in the chemical engineering curriculum ensure the chemical engineer is well poised to apply them to solve the coming challenging issues in the pharmaceutical industry. Chemical engineers are uniquely positioned to help address these needs in part derived from their ability to predict using mathematical models and their understanding of equipment and manufacturability. As Wu et al. highlighted, chemical engineers can help transform pharmaceuticals from an industry focusing on inventing and testing to a process and product design industry [28]. Significant pressure exists on what used to be a historically high-margin nature of the pharmaceutical industry to deliver safe, environmentally friendly, and economic processes in increasingly shorter timelines. This means fewer scale-ups at kilo and pilot plant scale, with expectation that a synthesis or formulation can be designed in the lab to perform as expected (and right the first time) at the desired manufacturing scale.

Chemical engineers are also uniquely positioned to influence regulators by incorporating advancements such as continuous processing coupled with PAT into a highly regulated industry. From R&D through manufacturing within the pharmaceutical industry, chemical engineering can be leveraged to bring competitive advantage to their respective organizations through process and predictive modeling that lead to process understanding, improving speed of development,

and developing new technology platforms and leaner manufacturing methods. The chapters in these two volumes are intended to provide examples of chemical engineering principles specifically applied toward relevant problems faced in the pharmaceutical sciences and manufacturing areas. Further the broader goal of this work is to promote the role of chemical engineering within our industry, to promote the breadth of skill sets therein, and to showcase the critical synergy between this discipline and the many scientific disciplines that combine to bring pharmaceutical drugs and therapies to patients in need around the world.

REFERENCES

1. Janusz Rzeszotarski, W. (2000). Encyclopedia of Chemical Technology.
2. FDA (2009). International conference on harmonisation: guidance on Q8(R1) pharmaceutical development. *Federal Register* 74 (109): 27325–27326.
3. FDA (2006). International conference on harmonisation: guidance on Q9 quality risk management. *Federal Register* 71 (106): 32105–32106.
4. FDA (2009). International conference on harmonisation: guidance on Q10 pharmaceutical quality system. *Federal Register* 74 (66): 15990–15991.
5. DiMasi, J.A., Grabowski, H.G., and Hansen, R.W. (2016). Innovation in the pharmaceutical industry: new estimates of R&D costs. *Journal of Health Economics* 47: 20–33.
6. Gregson, N., Sparrowhawk, K., Mauskopf, J., and Paul, J. (2005). Pricing medicines: theory and practice, challenges and opportunities. *Nature Reviews Drug Discovery* 4: 121–130.
7. Suresh, P. and Basu, P.K. (2008). Improving pharmaceutical product development and manufacturing: impact on cost of drug development and cost of goods sold of pharmaceuticals. *Journal of Pharmaceutical Innovation* 3: 175–187.
8. Dienemann, E. and Osifchin, R. (2000). The role of chemical engineering in process development and optimization. *Current Opinion in Drug Discovery and Development* 3 (6): 690–698.
9. Federsel, H.-J. (2006). In search of sustainability: process R&D in light of current pharmaceutical R&D challenges. *Drug Discovery Today* 11 (21/22): 966–974.
10. FDA (2015). FDA approves first biosimilar product Zarxio. *US Food and Drug Administration* (6 March 2015).
11. FDA (2018). FDA approves first epoetin alfa biosimilar for the treatment of anemia. *US Food and Drug Administration* (15 May 2018).
12. Li, F., Hashimura, Y., Pendleton, R. et al. (2006). A systematic approach for scale-down model development and characterization of commercial cell culture processes. *Biotechnology Progress* 22: 696–703.
13. Rovner, S.L. (2011). Process development shines in tough times. *Chemical and Engineering News* 89 (24): 41–44.
14. Behr, A., Brehme, V.A., Ewers, L.J. et al. (2004). New developments in chemical engineering for the production of drug substances. *Engineering in Life Sciences* 4 (1): 15–24.

15. Burns, L.R. (ed.) (2005). *The Business of Healthcare Innovation*. Cambridge: Cambridge University Press.
16. F-D-C Reports Inc. (2009). The Gold Sheet, Pharmaceutical & Biotechnology Quality Control, Attention turns to the business case of quality by design. F-D-C reports, January 2009.
17. am Ende, M.T., Bernhard, G., Lubczyk, V., Dressler, U. et al. (2008). Risk Management and Lifecycle Efficiency within Process Development. Leadership Forum 2008, Washington, DC, 29 May 2008.
18. Ohno, T. (1988). *Toyota Production System Beyond Large Scale Production*. New York: Productivity Press.
19. Cogdil, R.P., Knight, T.P., Anderson, C.A., and Drennan, J.K. (2007). The financial returns on investments in process analytical technology and lean manufacturing: benchmarks and case study. *Journal of Pharmaceutical Innovation* 2: 38–50.
20. Lewis, N.A. (2006). A tracking tool for lean solid-dose manufacturing. *Pharmaceutical Technology* 30 (10): 94–109.
21. Goldratt, E.M. and Cox, J. (2004). *The Goal: A Process for Ongoing Improvement*, 3e. London: Routledge.
22. Biegler, L.T., Grossmann, I.E., and Westerberg, A.W. (1997). *Systematic Methods of Chemical Process Design*. Upper Saddle River: Prentice Hall.
23. Kuehn, S. E. (2015). Janssen embraces continuous manufacturing for Prezista. *PharmaceuticalManufacturing.com* (8 October 2015), <https://www.pharmamanufacturing.com/articles/2015/janssen-embraces-continuous-manufacturing-for-prezista/?show=all>
24. Esposito, L. (2018). What does personalized medicine really mean? *US News & World Report* (26 January 2018).
25. Lewin, J.J., Choi, E.J., and Ling, G. (2016). Pharmacy on demand: new technologies to enable miniaturized and mobile drug manufacturing. *American Journal of Health-System Pharmacy* 73 (2): 45–54.
26. United States Food and Drug Administration (2015). Highlights of prescribing information: Spritam. https://www.accessdata.fda.gov/drugsatfda_docs/label/2015/207958s0001bl.pdf
27. Norman, J., Madurawe, R.D., Moore, C.M.V. et al. (2017). A new chapter in pharmaceutical manufacturing: 3D-printed drug products. *Advanced Drug Delivery Reviews* 108: 39–50.
28. Wu, H., Khan, M.A., and Hussain, A.S. (2007). Process control perspective for process analytical technology: integration of chemical engineering practice into semiconductor and pharmaceutical industries. *Chemical Engineering Communications* 194: 760–779.

PART II

DRUG PRODUCT DESIGN, DEVELOPMENT, AND MODELING

DESIGN OF SOLID DOSAGE FORMULATIONS

KEVIN J. BITTORF*, TAPAN SANGHVI, AND JEFFREY P. KATSTRA**
Formulation Development, Vertex Pharmaceuticals, Cambridge, MA, USA

2.1 INTRODUCTION

The oral route is the most common way of administering drugs. It not only represents a convenient (self-administered) and safe way of drug administration but is also more profitable to manufacture than the parenteral dosage forms that must be administered, in most cases, by trained personnel. This is reflected by the fact that well over 80% of the drugs in the United States that are formulated to produce systematic effects are marketed as oral dosage forms. Among the oral dosage forms (Table 2.1), tablets of various different types are the most common because of their low cost of manufacture (including packaging and shipping), increased stability, and virtual temper resistance.

Following oral administration of tablets, the delivery of the drug to the systemic circulation requires initial transport through the gastrointestinal (GI) membrane. The drug absorption from the GI tract requires that the drug is brought into solution in the GI fluids and that it is capable of crossing the intestinal membrane into the systemic circulation; therefore, the rate of dissolution of the drug in the GI lumen can be a rate-limiting step in the absorption of drugs given orally. Particles of drugs, e.g. insoluble crystalline forms or specific delivery systems such as liposomes, are generally found to be absorbed to a very small extent. The cascade of events from release of the drug from tablet, i.e. disintegration of tablet into granules or aggregates followed by dissolution of the drug in the gut lumen, interactions, and/or degradation within the lumen and the absorption of the drug across the intestinal membrane into the systemic circulation, is schematically shown in Figure 2.1. The slowest of these events (dissolution

and/or absorption) determines the rate of availability of the drug from tablet formulation. Many factors in each step influence the rate and extent of availability of the drug. Physical, chemical, and biopharmaceutical properties of the drug, as well as the design and production of the tablet, play a very important role in its bioavailability after oral administration. These considerations make the seemingly simple tablet formulation approach complex to formulate in reality. These realizations have resulted in a change in philosophy of tablet formulation design in last decade or more, wherein it is no longer considered an art but well-defined science.

The single greatest challenge to the tablet formulator is in the definition of the purpose of the formulation and the identification of the suitable materials to meet development objectives. A good formulation must not only be bioavailable but also be manufacturable, and be chemically and physically stable from manufacturing through the end of shelf life. In addition, many quality standards and special requirements must be met to ensure the efficacy and safety of the product.

All of these formulation goals can be described as the target product profile (TPP). A TPP is a summary of characteristics that, if achieved, will provide optimal efficacy, patient compliance, and marketability. A TPP (Table 2.2) often includes attributes like pharmacokinetic information (e.g. immediate release vs. extended release), dosage form (e.g. tablet vs. injectable), and shelf-life information (e.g. two years at 25 °C/60%Relative Humidity (RH)). There are also many other potential inputs for drug development that a formulator may or may not need, such as: Warnings and Precautions, Adverse Reactions, Drug Interactions, Use in Specific Populations, Drug Abuse and Dependence, Clinical Studies, and Patient Counseling Information.

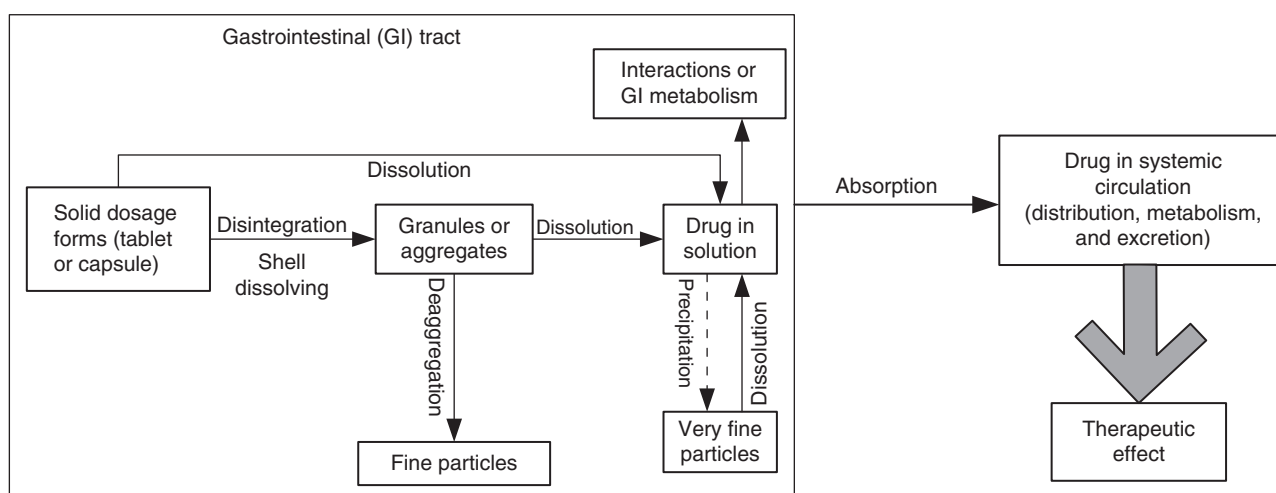
It is important to establish the TPP so that the formulation effort will be effective and focused. When the formulation

*Current affiliations: Simulation LLC, Boston, MA, USA

**Current affiliation: Agios Pharmaceuticals, Cambridge, MA, USA

TABLE 2.1 Types of Solid Oral Dosage Forms

Type of Oral Dosage Form	Characteristics
Immediate release tablets	Disintegrate in stomach after taken orally
Delayed release tablets	Enteric-coated tablets to keep tablets intact in stomach and disintegrate in intestine for absorption
Sustained/controlled release tablets	Release drug slowly over a period of time to decrease the frequency of administration
Chewable tablets	Tablets are broken by chewing before swallowing with water
Orally disintegrating tablets	Disintegrate in oral cavity without drinking water to form a suspension for ease of swallowing
Hard gelatin capsules	Two-piece capsule shells filled with granules, powders, pellets, sprinkles, semisolids, and oils
Soft gelatin capsules	One-piece capsule filled with oily liquid
Sachets	Single-dose unit bag containing granules

**FIGURE 2.1** Fate of solid dosage form following oral administration.**TABLE 2.2** Typical Target Product Profile (TPP) for an Immediate Release (IR) Tablet

TPP	How Used by a Formulator	Typical for IR Tablet
Indications and usage	Examine other products in the same class: examine improvements	Once a day (QD) Twice a day (BID) Three times a day (TID)
Dosage and administration	Good to know what is expected before one starts formulating	Oral tablet
Dosage forms and strengths	Multiple strengths may be needed depending on the population being targeted (adults vs. children)	Dependent on drug typically 10–500 mg
Overdosage	Useful if designing an extended release dosage, in which overdose (dose dumping) is a possibility	Dependent on drug
Description	This is up to the formulator and marketing: shape, size, and color of the tablet	A tablet with markings and color
Clinical pharmacology	Helps determine where the drug is absorbed and how fast the drug must get into solution	Dependent on drug
How supplied/stored/handled	Important as most people do not like refrigerated dosage forms	Two-year room temperature shelf life

requirements are defined by the TPP, a strategy must be established to facilitate effective formulation development. To establish a formulation strategy, one must consider the physical, chemical, and biopharmaceutical characteristics of the drug; optimal technologies to achieve formulation goals; and the manufacturing capabilities to support the product.

This chapter examines tablet-formulation design and development of an immediate release oral solid dosage form using a mix of Pharmaceutical Science, Statistical, and Engineering approaches. The chapter is aimed toward providing engineers an overview of the key physicochemical, mechanical, and biopharmaceutical properties of the drug and their

influence on the selection of formulation process platform. Subsequently, critical tablet characteristics that affect the stability and bioavailability of the drug product are discussed. Finally, strategy for tablet-process optimization and scale-up is defined to select proper equipment and to define operational design space (DS). A systematic scientific approach to tablet formulation and process development along with practical examples is discussed to expedite the drug product development.

2.2 UNDERSTANDING DRUG SUBSTANCE

Integrating the physicochemical, mechanical, and biopharmaceutical properties of a drug candidate is a prerequisite in developing a robust and bioavailable drug product that has optimal therapeutic efficacy. The measurement of physical, mechanical, and chemical properties helps guide not only the selection of dosage form but also provides an insight into their processability and storage to ensure

optimal drug product quality. Figure 2.2 lists the critical physicochemical, mechanical, and biopharmaceutical properties that need to be understood to aid in the design of tablet formulation.

2.2.1 Physicochemical Properties

Prior to the development of tablet dosage form, it is essential to understand to determine certain fundamental physical and chemical properties of the drug molecule along with other derived properties. This information dictates many of the subsequent approaches in tablet formulation development and is known as preformulation. It should be kept in mind that many of these properties are dependent on the solid form, and complete characterization of each of the most relevant solid forms is needed to provide a complete physicochemical picture.

2.2.1.1 Solubility and Drug Dissolution Solubility of a drug candidate may be the critical factor determining its

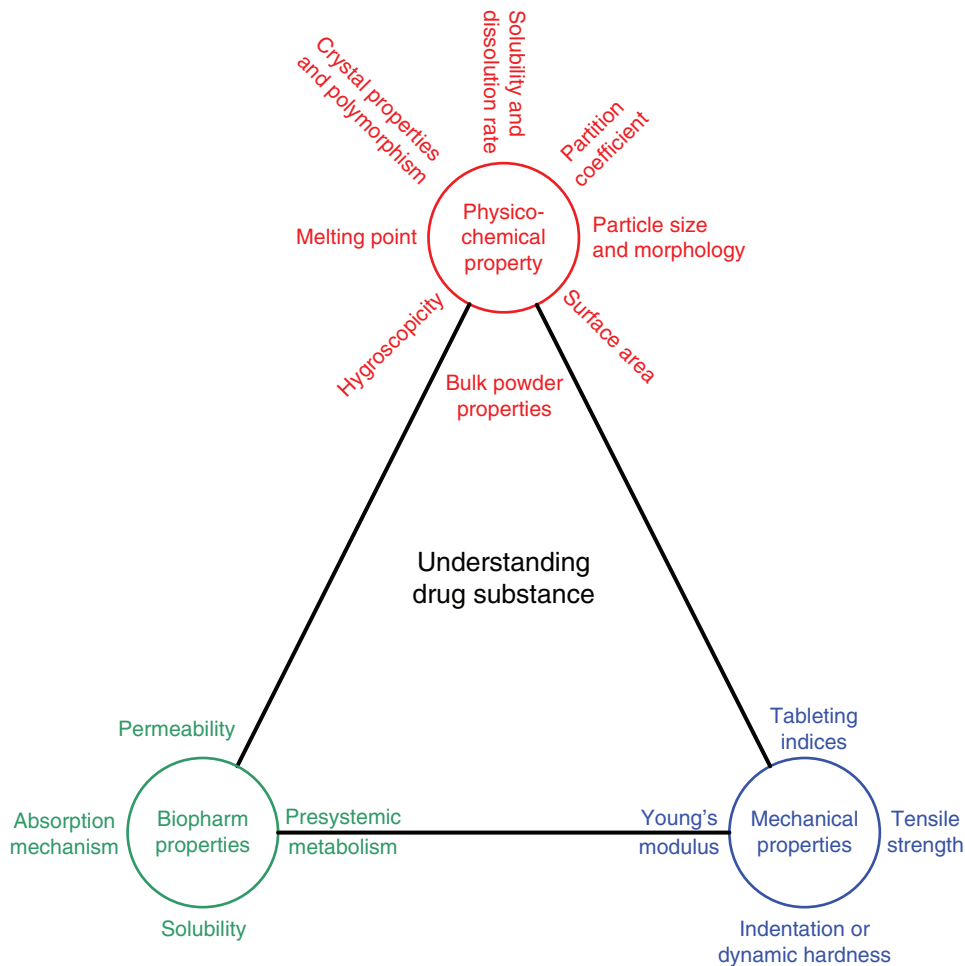


FIGURE 2.2 Understanding drug substance properties.

usefulness, since aqueous solubility dictates the amount of compound that dissolves; therefore, the amount available for absorption. A compound with low aqueous solubility could be subject to dissolution rate-limited absorption within the GI residence time.

Dissolution is the dynamic process by which a material is dissolved in a solvent that is characterized by a rate (amount dissolved per time unit), while *solubility* is the amount of material dissolved per unit volume of a certain solvent that is characterized as a concentration. Solubility is often used as a short form for “saturation solubility,” which is the maximum amount of drug dissolved at equilibrium conditions. Finally, *intrinsic solubility* is the solubility of the neutral form of an ionizable drug.

Dissolution rate is directly proportional to the aqueous solubility, C_s , and the surface area, A , of drug exposed to the dissolution medium. It is common, when developing an immediate release dosage form of poorly soluble drug, to increase the drug-dissolution rate by increasing the surface area of a drug through particle size reduction.

The dissolution rate of a solute from a solution is described by the Noyes–Whitney equation as follows [1]:

$$\frac{dC}{dt} = \left(\frac{D \times A}{h} \right) \times (C_s - C_t) \quad (2.1)$$

D is the diffusion coefficient of the drug substance (in a stagnant water layer around each drug particle with a thickness h , A is the drug particle surface area, C_s is the saturation solubility, and C_t is the drug concentration in the bulk solution at a given time.

The dissolution rate, rather than the saturation solubility, is most often the primary determinant in the absorption process of a sparingly low-soluble drug. Determining the dissolution rate is critical. The main area for dissolution-rate studies are evaluations of different solid forms of a drug (e.g. salts, solvates, polymorphs, amorphous, and stereoisomers) or different particle sizes of the drug. The dissolution rate can either be determined for a constant surface area of the drug in a rotating disc apparatus [2] or as a dispersed powder in a beaker with agitation (as detailed in pharmacopeias such as United States Pharmacopeia, etc.).

The impact of solubility and dissolution rate on formulation selection is discussed later in the chapter.

2.2.1.2 Partition Coefficient Partition coefficient is the relationship between chemical structure, lipophilicity, and its disposition *in vivo* and has been reviewed by a number of authors [3]. The lipophilicity of an organic compound is described in terms of a partition coefficient, $\log P$, which is defined as the ratio of the concentration of the unionized compound, at equilibrium, between organic and aqueous phases:

$$\log P = \frac{[A]_{\text{organic}}}{[A]_{\text{aqueous}}} \quad (2.2)$$

For ionizable drugs, the ionized species does not partition into the organic phase, the apparent partition coefficient, D , is calculated from the following:

$$\text{Acids: } \log D = \log P - \log \left[1 + 10^{(\text{pH} - \text{pK}_a)} \right] \quad (2.3)$$

$$\text{Bases: } \log D = \log P - \log \left[1 + 10^{(\text{pK}_a - \text{pH})} \right] \quad (2.4)$$

pK_a is the dissociation constant.

Since it is virtually impossible to determine $\log P$ in a realistic biological medium, the octanol/water system has been widely adopted as a model of the lipid phase [4]. There has been much debate about the suitability of this system [5], but it remains the most widely used in pharmaceutical studies.

Generally, compounds with $\log P$ values between 3 and 6 show good passive absorption, whereas those with $\log P$ s of less than 3 or greater than 6 often have poor passive transport characteristics. The role of $\log P$ in absorption processes occurring after oral administration has been discussed by Navia and Chaturvedi [6].

2.2.1.3 Crystal Properties and Polymorphism Most drug substances appear in more than one polymorphic form. Polymorphs differ in molecular packing (crystal structure), but share the same chemical composition [7]. Hydrates or solvates are often called “pseudo-polymorphs” because, in addition to containing the same given drug molecule, hydrates or solvates also contain molecules of solvents that are incorporated into the crystal lattice. Amorphous forms are characterized by the absence of long-range order.

Polymorphism has a profound implication on formulation development and biopharmaceutical properties, because polymorphs may exhibit significantly different solubility, dissolution rate, compactibility, hygroscopicity, physical stability, and chemical stability [7]. Figure 2.3 provides a detailed list (Ishikawa or Fishbone diagram) of physical properties that can differ among the polymorphs.

Higher solubility and faster dissolution rates of the metastable polymorph may lead to significantly better oral bioavailability. Chloramphenicol palmitate [9] (bacteriostatic antimicrobial) and ampicillin [10] (antibiotic example of this is, such as is the case with) are examples of the anhydrous form which gave higher blood serum levels than the less-soluble trihydrate form.

Although use of a faster-dissolving polymorph may have clinical benefit, it is important to keep in mind that a polymorph with a higher solubility or faster dissolution rate is also metastable (i.e. a higher energy form) and tends to convert to

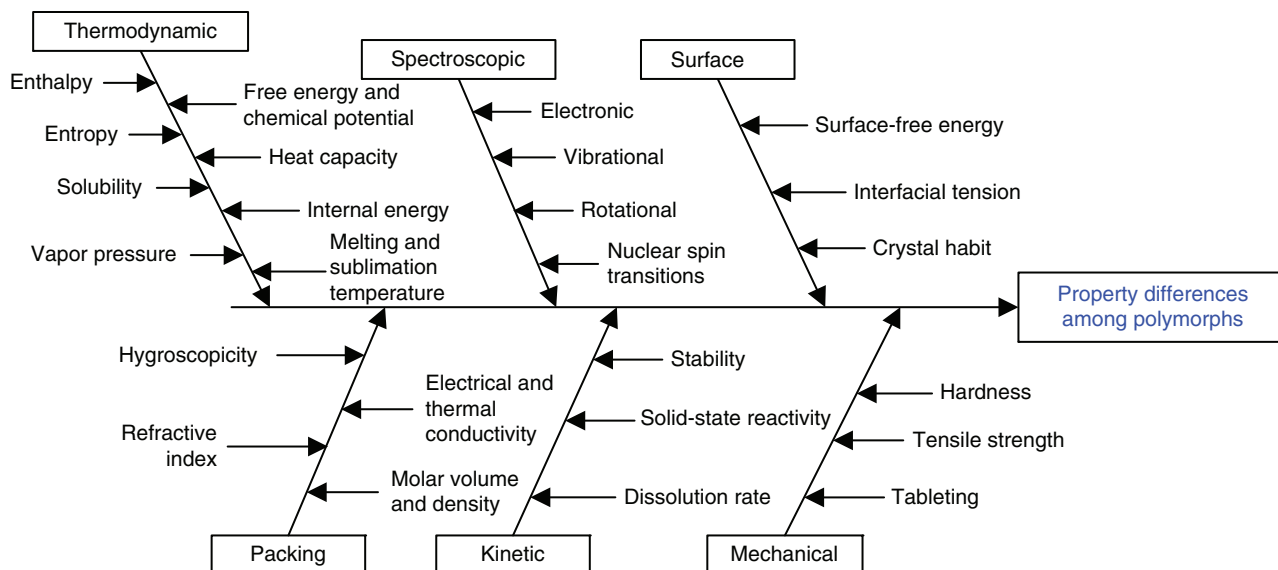


FIGURE 2.3 Fishbone schematic of physical property differences among polymorphs. *Source:* Adapted from Grant [8].

a thermodynamically more stable form over time. Conversion from a metastable form to a stable form could lower a drug's oral bioavailability, and lead to inconsistent product quality. From a formulating perspective, it is desirable to use the thermodynamically stable form of the API; however, biopharmaceutical and processability considerations may dictate the deliberate selections of a metastable form for processing.

It is important to keep in mind that polymorphic form conversion from the most stable form may still occur, even when a stable crystal form is chosen for development. Polymorphic transformations can take place during pharmaceutical processing, such as particle size reduction, wet granulation, drying, and even during the compaction process and compression process [11, 12] as each of these processes may add the energy required to move the drug to the unstable form.

2.2.1.4 Particle Size, Particle Morphology, and Surface Area Bulk flow, compactability, formulation homogeneity, and surface-area control dissolution and chemical reactivity, which are directly affected by size, shape, and surface morphology of the Drug/Active Pharmaceutical Ingredient (API).

Spherical particles have the least contact surface area and exhibit good flow, whereas acicular particles tend to have poor flow [13]. Milling of long acicular (or needle) crystals can enhance flow properties; however, excessively small particles tend to be cohesive and aggravate flow problems.

In addition to the flow properties, crystal shape and size have been demonstrated to impact mixing and tabletability. L-lysine monohydrate with plate-shaped crystals exhibited greater tabletability than the prism-shaped crystals [14]. Kaerger et al. [15] studied the effect of paracetamol particle size and shape on the compactibility of binary mixture with

microcrystalline cellulose, showing that compactibility increased with particle size and irregular crystals, whereas compactibility increased with a decrease in particle size.

Particle size affects drug content uniformity (CU). For low-dose direct compression (DC) formulations, where drug CU is of particular concern, the particle size of the drug substance has to be small enough to meet the US Pharmacopeia requirement on CU [16]. For example, Zhang and Johnson [17] showed that low-dose blends containing a larger drug particle size (18.5 μm) failed to meet the USP requirement, whereas a blend containing smaller particle sizes (6.5 μm) passed.

Surface areas of drug particles are important because dissolution is a function of this parameter (as predicted by the Noyes–Whitney equation). This is particularly true in those cases where the drug is poorly soluble. Such drugs are likely to exhibit dissolution rate-limited absorption. For such drugs, particle size reduction (e.g. micronization) is often utilized to increase the surface area which enhances the dissolution rate; e.g. micronization enhanced the bioavailability of felodipine when administered as an extended release tablet [18].

Methods to determine particle size and shape include light microscopy, scanning electron microscopy (SEM), sieve analysis, and various electronic sensing-zone particle counters. Methods available for surface area measurement include air permeability and various gas adsorption techniques.

2.2.1.5 Bulk Powder Properties Density and porosity are two important pharmaceutical properties that are derived from the information on particle size, particle shape, and surface area. A comparison of true particle density, apparent particle density, and bulk density can provide information on total porosity, interparticle porosity, and intraparticle

porosity. Generally, porous granules dissolve faster than dense granules, since pores allow water to penetrate more readily.

$$\text{Interparticle (interspace) porosity} = 1 - \frac{\text{bulk density}}{\text{apparent particle density}} \quad (2.5)$$

$$\text{Intraparticle porosity} = 1 - \frac{\text{apparent particle density}}{\text{true particle density}} \quad (2.6)$$

$$\text{Total porosity} = 1 - \frac{\text{bulk density}}{\text{true particle density}} \quad (2.7)$$

The increase in bulk density of a powder is related to the cohesivity of a powder. Bulk density and tapped density are used to calculate compressibility index and Hausner ratio, which are measures of the propensity of a powder to flow and be compressed. *A rule of thumb:* a compressibility index of higher than 30% indicates poor powder flow. The Hausner ratio varies from about 1.2 for a free-flowing powder to 1.6 for cohesive powders.

$$\text{Hausnerratio} = \frac{\text{tapped density}}{\text{bulk density}} \quad (2.8)$$

$$\begin{aligned} \text{Compressibility (Carr index)} \\ = \frac{100 \times (\text{tapped density} - \text{bulk density})}{\text{bulk density}} \end{aligned} \quad (2.9)$$

2.2.1.6 Melting Point and Hygroscopicity Low melting materials tend to be more difficult to manufacture and handle in conventional solid dosage forms. *A rule of thumb:* melting points below about 60 °C are considered to be problematic. Temperatures in conventional manufacturing equipment, such as fluid-bed dryers and tablet presses, can exceed 50 °C. During the milling process, hot spots in the milling chamber may have much higher temperatures.

Moisture uptake is a concern for pharmaceutical powders and is known to affect a wide range of properties, such as powder flow, compactibility, and stability [8, 17, 46]. On the other hand, moisture may improve powder flow and uniformity of the bulk density as well as an appropriate amount of moisture may act as a binder to aid compaction. Thus, knowledge of the type and level of moisture is critical for understanding its impact not only on deformation behavior but also on the attributes of the final product.

2.2.2 Biopharmaceutical Properties

Biopharmaceutics is defined as the study of the relationships between the physicochemical properties, dosage forms, and

routes of administration of drugs, and its effect on the rate and extent of absorption in the living body. Complete oral absorption occurs when the drug has a maximum permeability coefficient and maximum solubility at the site of absorption, which results in rapid and uniform pharmacological response. Based upon this premise, a key objective in designing a rational oral dosage form is having sound understanding of multitude of factors including physicochemical properties of the drug and dosage-form components, and physiological aspects of GI tract.

Generating formulations with relevant oral bioavailability depends on a number of factors including solubility, permeability,¹ and metabolic stability.² Absorbability is related to the first two factors whose importance has been recognized in the guise of the biopharmaceutics classification system (BCS) [19, 20]. This approach bins drugs and drug candidates into four categories based on their solubility and permeability properties. The Food and Drug Administration (FDA) has issued guidelines to define low and high solubility and permeability [21].

The primary objective of the BCS³ is to guide decisions with respect to *in vivo* and *in vitro* correlations and need for bioequivalence studies; it is also used to identify dosage-form strategies that are designed at overcoming absorption barriers presented by solubility and/or permeability-related challenges as depicted in Table 2.3.

The BCS nomenclature is centered on the premise that most orally administered drugs are absorbed via passive diffusion⁴ process through the small intestine and excludes other important factors such as the drug absorption mechanism (carrier-mediated,⁵ P-glycoprotein efflux,⁶ etc.) and pre-systemic degradation or complexation that may enhance or limit oral bioavailability.

¹Permeability determines the ability of drug to move across the lipophilic intestinal membrane in gastrointestinal tract (GIT). Permeability of a drug may be predicted using computational (*in silico*) models or measured using both physicochemical and biological methods (*in vitro*, *in situ*, or *in vivo*).

²Metabolic stability refers to ability of a drug to withstand metabolism or degradation in the gut wall and the liver.

³Biopharmaceutics classification system (BCS) is guidance for predicting the intestinal drug absorption using solubility and permeability provided by the U.S. Food and Drug Administration.

⁴Passive diffusion is a transport process, wherein drug molecules pass across the lipoidal intestinal membrane from a region of high concentration in the lumen (GIT) to a region of lower concentration in the blood (systemic circulation). Mathematically, it is described by Fick's first law of diffusion.

⁵Carrier-mediated transport may be subdivided into active transport and facilitated diffusion or transport. Active transport is a process whereby drug is bound to a carrier or membrane transporter and is transported against the concentration gradient across a cell membrane. Facilitated diffusion differs from active transport in that it cannot transport a substance against a concentration gradient of that substance.

⁶P-glycoprotein is one of the key counter-transport efflux proteins that expel specific drugs back into the lumen of the GIT after they have been absorbed.

TABLE 2.3 Dosage-form Options Based on Biopharmaceutical Classification System

Class I: High solubility, high permeability	Class II: Low solubility, high permeability
No major challenges for immediate-release dosage form	Formulation are designed to overcome solubility
Controlled-release dosage forms may be needed to limit rapid absorption	Salt formation
	Precipitation inhibitors
	Metastable forms
	Solid dispersions
	Lipid technologies
	Particle size reduction
Class III: High solubility, low permeability	Class IV: Low solubility, low permeability
Prodrugs	Formulation would have to use a combination of the approaches identified in Class II and III.
Permeation enhancers	
Ion pairing	Strategies for oral administration are not really viable. Often use alternative delivery methods such as intravenous administration.
Bioadhesives	

A drug is considered to be highly soluble where the highest dose is soluble in 250 ml or less of aqueous media over the pH range 1–8. A drug is considered to be highly permeable when the extent of absorption in humans is expected to be greater than 90% of the administered dose.

2.2.3 Mechanical Properties

Material mechanical properties play a role in manufacturing drug product. Particle properties influence the true areas of contact between particles and can affect unit operations, such as compression, milling, and granulation. Characterization of mechanical properties of drug substance is important in three areas: choosing a processing method, such as granulation or DC; selecting excipients with properties that mask the poor properties of the drug; and helping to document what went wrong, i.e. when a tableting process is being scaled-up or when a new bulk drug process is being tested. Since all of these can influence the quality of the final product, it is to the formulator's advantage to quantify and understand the importance of the mechanical properties of the active and inactive ingredients.

Pharmaceutical materials are elastic, plastic, viscoelastic, hard, tough, or brittle in the same sense that metals, plastics, or wood have similar properties. The same concepts that materials/mechanical engineers use to explain/characterize tensile, compressive, or shear strength are relevant to pharmaceutical materials. A number of characterization tools as outlined in Table 2.4 are available for understanding the mechanical properties of the material.

Based upon the analysis of the physicochemical, mechanical, and biopharmaceutical properties of the drug substance, selection of excipients and the formulation process are performed. The accompanying section discusses excipient, their

types, and the selection procedure based upon their effect on the drug substance properties.

2.3 EXCIPIENTS

Excipients facilitate formulation design to perform a wide range of functions to obtain the desired properties for the finished drug product. Historically, pharmaceutical excipients have been regarded as inert additives, but this is no longer the case. Each additive must have a clear justification for inclusion in the formulation and must perform a defined function in the presence of the active and any other excipients included in the formulation. Excipients may function, for example, as an antimicrobial preservative, a solubility enhancer, a stability enhancer, or a taste masker, to name a few.

Excipients are selected based on their chemical/physical compatibility with drugs, regulatory acceptance, and processability. First, excipients shall be chemically compatible with drug substances. Second, in the time of globalization, excipients are to meet the requirements of not only the FDA or EMEA but also the regulatory agencies of other potential marketing countries. Third, excipients impact the properties of a powder mixture, such as flowability, density, compactibility, and adhesiveness. For example, different fillers are selected carefully to balance the plasticity, elasticity, and brittleness of the pre-compaction powder mixture, in order to make large-scale production feasible.

For tablets, excipients are needed both for the facilitation of the tableting process (e.g. glidants) and for the formulation (e.g. disintegrants). Except for diluents, which may be present in large quantity, the level of excipient use is usually limited to only a few percent and some lubricants are required at less than 1%. Details of the types, uses, and mechanisms of action of various excipients for tablet production have been discussed at length in several articles and books. The types and functions of excipients for tablet production are summarized in Table 2.5.

It is worth noting that some of these tableting excipients may exert effects in opposition to each other. For example, binders and lubricants, because of their respective bonding and waterproofing properties, may hinder the disintegration action of the disintegrants. In addition, some of these tableting excipients may possess more than one function that may be similar (e.g. talc as lubricant and glidant) or opposite (e.g. starch as binder and disintegrant) to each other.

Furthermore, the sequence of adding the excipients during tablet production depends on the function of the excipient. Whereas the diluents and the binders are to be mixed with the active ingredient early on for making granules, disintegrants may be added before granulation (i.e. inside the granules), and/or during the lubrication step (i.e. outside the granules) before tablet compression.

TABLE 2.4 Characterization Tools for Understanding the Mechanical Properties of Materials

	Quasi-Static Testing	Dynamic Testing
API required	1–100 g	2–10 g
Advantages	“Independently” dissect out and investigate various mechanical properties	Understand the mechanics of materials at speeds representative of production tablet compaction
Limitations	Cannot determine properties at representative production scales	Difficult to factor out the individual mechanical property “component”
Characterization tests	<p>Tensile strength</p> <ul style="list-style-type: none"> describes the global strength of the material measured using traditional tablet hardness tester [22] or transverse compression in tensile tester [23] Typical desired value >1 MPa <p>Indentation/dynamic hardness</p> <ul style="list-style-type: none"> describes the “local” plasticity of the material measured using pendulum impact device or free-falling indenter [23, 24] <p>Young’s modulus</p> <ul style="list-style-type: none"> describes stiffness and toughness of the material measured using both four- and three-point beam bending, flexure testing [25] <p>Tableting indices</p> <ul style="list-style-type: none"> dimensionless numbers that integrates above described tests <p>Bonding index (BI)</p> <ul style="list-style-type: none"> defines the tendency of the material to remain intact after compression desired value >0.01 <p>Brittle fracture index (BFI)</p> <ul style="list-style-type: none"> measure of brittleness of a material BFI = 1 represents very brittle material and BFI < 0.3 is relatively non-brittle material <p>Strain index (SI)</p> <ul style="list-style-type: none"> indirect measure of elastic strain 	<p>Force–displacement profiles</p> <ul style="list-style-type: none"> indicator of tablet-forming ability of powder assessment of the elastic properties thermodynamic analysis of the process of compact formation <p>Tablet volume–applied pressure profiles</p> <ul style="list-style-type: none"> measured using hydraulic press, rotary press, compaction simulator, compaction emulator <p>Heckel equation</p> <ul style="list-style-type: none"> tablet porosity–applied pressure function

TABLE 2.5 Types and Functions of Tableting Excipients

Excipient	Function	Some Examples of Excipients
Diluents	Act as bulking/filling material	Sugars, lactose, mannitol, sorbitol, sucrose, calcium salts, microcrystalline celluloses
Binders and adhesives	Holds powder together	Sugars, glucose, polymers, starch, gelatin
Disintegrants	To facilitate the breakup of the tablet in the gastrointestinal tract	Croscarmellose sodium (CCS), sodium starch glycolate (SSG), crospovidone
Glidants	Improve the flow of granules, needed for compression	Silica, magnesium stearate, talc
Lubricants	Reduces friction between granules and the compression equipment	Magnesium stearate (MgSt), stearic acid, talc, sodium lauryl sulfate (SLS)
Antiadherents	To minimize the problems if sticking to the tablet punch head	Talc, cornstarch, SLS, MgSt
Colorants	For identification and marketing	Natural pigments and synthetic dyes
Flavors and sweeteners	To improve the taste of chewable tablets	Mannitol, aspartame

2.4 DRUG–EXCIPIENT COMPATIBILITY STUDY

Excipient compatibility testing provides a preliminary evaluation of the physical and chemical interactions that can occur. Testing is carried under stressed temperature and humidity conditions, between a drug and potential excipients. This helps excipient selection, particularly for tablet formulations in order to minimize unexpected formulation stability problems during product development.

Traditionally, a binary mixture of drug with the excipient being investigated is intimately mixed, and the ratio of drug to excipient is often 1 : 1; however, other mixtures may also

be investigated. These blends were stored at various temperatures and humidity, and analyzed for potential degradation products.

More recently, the use of a model formulation approach to excipient screening has become much more widespread across the industry. Model formulations include commonly used excipients in each functional category such as fillers, binders, disintegrants, and lubricants, and those with different chemical structures viz. celluloses, starches, and sugars. Both wet and dry model formulations may be prepared for stability testing. It is recommended that a Design of Experiment (DOE) be used to assist in the development and interpretation of results for these types of studies. Table 2.6 contains an example of the model formulation approach. It lists excipients and their approximate composition that would be found in a typical tablet formulation.

Powders are physically mixed and may be granulated or compacted to accelerate any possible interaction. Samples may be exposed in open pans or sealed in bottles/vials to mimic product packaging. The storage conditions used vary widely in terms of temperature and humidity, but a temperature of 40 °C for storage of compatibility samples is considered appropriate. Some compounds may require higher temperatures to make reactions proceed at a rate that is measured over a convenient time period. Methods of analysis also vary widely, ranging from thermal techniques (DSC) to chromatographic techniques (TLC and HPLC) to microcalorimetry.

An example of an excipient compatibility study utilizing partial factorial design is illustrated in Table 2.7. In this study, a model compound (BCS Class II) is blended with excipients (shown in Table 2.6) to make 16 formulations and stationed

TABLE 2.6 Typical Excipients Selected for a Model Formulation Study

Excipient Type	% Composition	Level 1	Level 2
API	10	—	—
Filler 1	38–40	MCC	Mannitol
Filler 2	38–40	Dicalcium phosphate	Spray-dried lactose
Surfactant	0–4	None	Sodium lauryl sulfate
Binder	4	PVP	HPC
Disintegrant	5	Sodium starch glycolate	Croscarmellose sodium
Lubricant	1	Magnesium stearate	Sodium stearyl fumarate
Wet granulation	20% (w/w) water	No	Yes

TABLE 2.7 Formulation Composition for Excipient Compatibility Study

Formulation Composition and Numbers	10%	38–40%	38–40%	0–4%	4%	1%	5%	20% (w/w) Water
1	API	MCC	ATab	None	PVP	MgSt	SSG	Dry (no)
2	API	MCC	ATab	None	HPC	MgSt	CCS	Wet
3	API	MCC	ATab	SLS	PVP	SSF	CCS	Wet
4	API	MCC	ATab	SLS	HPC	SSF	SSG	Dry (no)
5	API	MCC	Lactose	None	PVP	SSF	CCS	Dry (no)
6	API	MCC	Lactose	None	HPC	SSF	SSG	Wet
7	API	MCC	Lactose	SLS	PVP	MgSt	SSG	Wet
8	API	MCC	Lactose	SLS	HPC	MgSt	CCS	Dry (no)
9	API	Mannitol	ATab	None	PVP	SSF	SSG	Wet
10	API	Mannitol	ATab	None	HPC	SSF	CCS	Dry (no)
11	API	Mannitol	ATab	SLS	PVP	MgSt	CCS	Dry (no)
12	API	Mannitol	ATab	SLS	HPC	MgSt	SSG	Wet
13	API	Mannitol	Lactose	None	PVP	MgSt	CCS	Wet
14	API	Mannitol	Lactose	None	HPC	MgSt	SSG	Dry (no)
15	API	Mannitol	Lactose	SLS	PVP	SSF	SSG	Dry (no)
16	API	Mannitol	Lactose	SLS	HPC	SSF	CCS	Wet

on open dish stability at 25 °C/60%RH, 40 °C dry, and 40 °C/75%RH. The study duration is three months which is analyzed for physical and chemical stability.

Figure 2.4 shows a regression model that is defined for assessing the effect of formulation and time on degradation growth at a storage condition of 40 °C/75%RH. A regression analysis is completed for data at 40 °C/75% RH to determine which excipient affects the growth of degradation products. From the analysis (Table 2.8), Time, Filler 1, Disintegrant, and Granulation have effect on degradation as well there are some interactions between Time and Filler 1, Time and Disintegrant, and Time and Granulation (borderline as p -value ≈ 0.05).

The prediction profiler and the interaction profiles (Figure 2.5) provide information on the specific excipient within a significant class (from Table 2.7) and the sensitivity

of each of the variables on the degradation growth. As seen from the prediction profiler, within Filler 1 mannitol causes more degradation as compared to MCC. Similarly, SSG is better than CCS among disintegrant and dry blend is better than wet granulation as the latter causes more degradation.

These results suggest that both mannitol and CCS could be detrimental for the stability of the API and are not being assessed for formulation development. In addition, wet granulation is to be avoided to increase the shelf life.

2.5 PROCESSING OF FORMULATIONS

The properties of a drug substance dictate the design of formulation composition and the choice of formulation-processing platform technology. The most commonly used

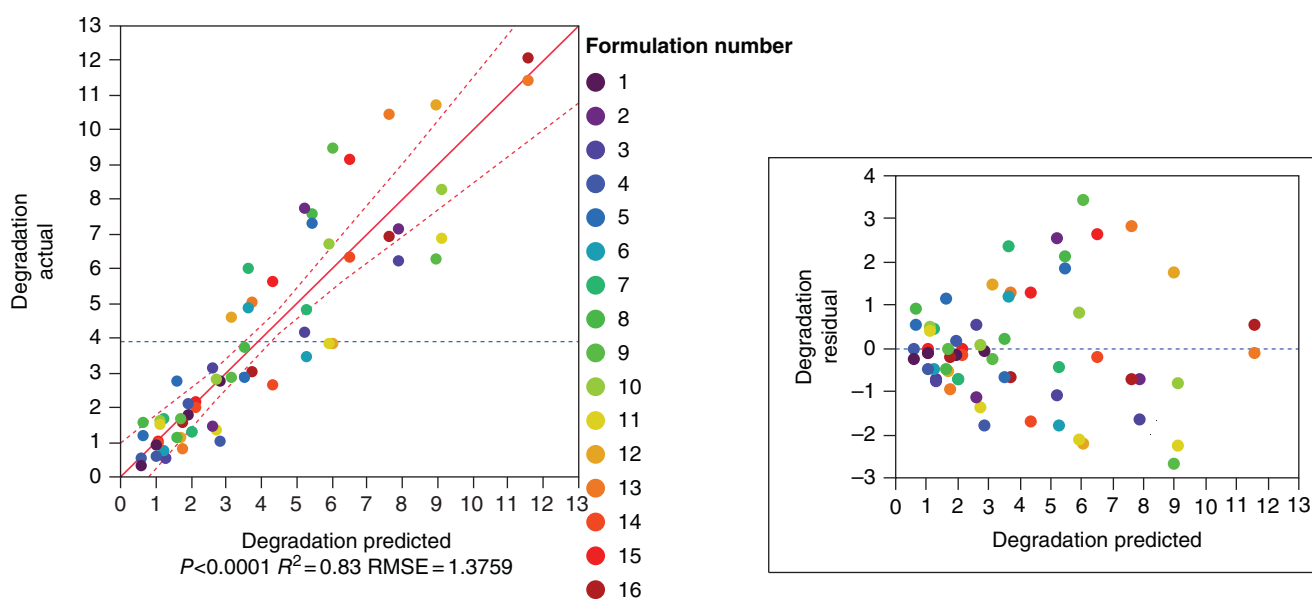


FIGURE 2.4 Degradation actual versus predicted plot and degradation residuals versus degradation predicted plot. The residuals are evenly distributed indicating there is no bias in the model. The formulation number refers to those listed in Table 2.7.

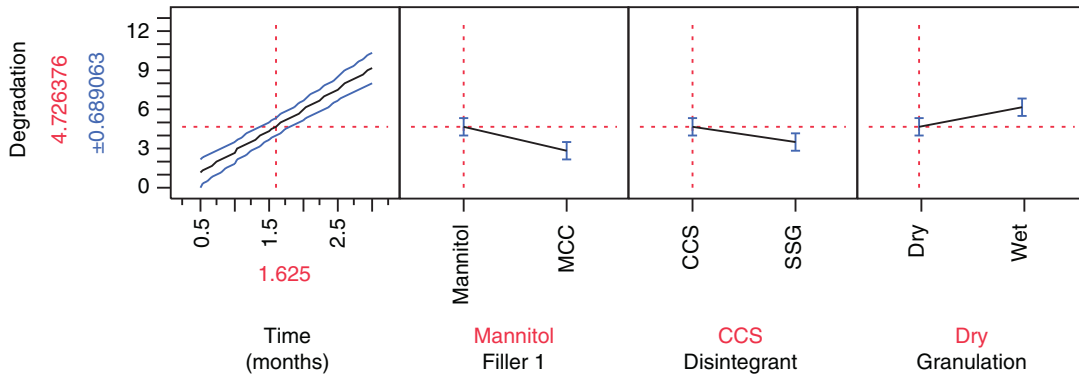
TABLE 2.8 Regression Results from the Excipient Compatibility Experiments

Term	Estimate	Std Error	t Ratio	Prob > $ t $
Intercept	-0.0292 67	0.338 094	-0.09	0.9313
Time (months)	2.406 077 3	0.179 126	13.43	<0.0001*
Filler 1 [mannitol]	0.959 366 9	0.171 987	5.58	<0.0001*
Disintegrant [CCS]	0.604 324 7	0.171 987	3.51	0.0009*
Granulation [dry]	-0.717 924	0.171 987	-4.17	0.0001*
(Time (months) - 1.625) \times filler 1 [mannitol]	0.640 884	0.179 126	3.58	0.0007*
(Time (months) - 1.625) \times disintegrant [CCS]	0.506 848 7	0.179 126	2.83	0.0065*
(Time (months) - 1.625) \times granulation [dry]	-0.358 106	0.179 126	-2.00	0.0505

All parameters shown are significant. The parameters analyzed that did not show significance were Filler 2, Surfactant, Binder, and Lubricant and were subsequently removed from the model during stepwise regression.

* Means that the term is statistically significant to a $p < 0.05$ value.

Prediction profiler



Interaction profiles

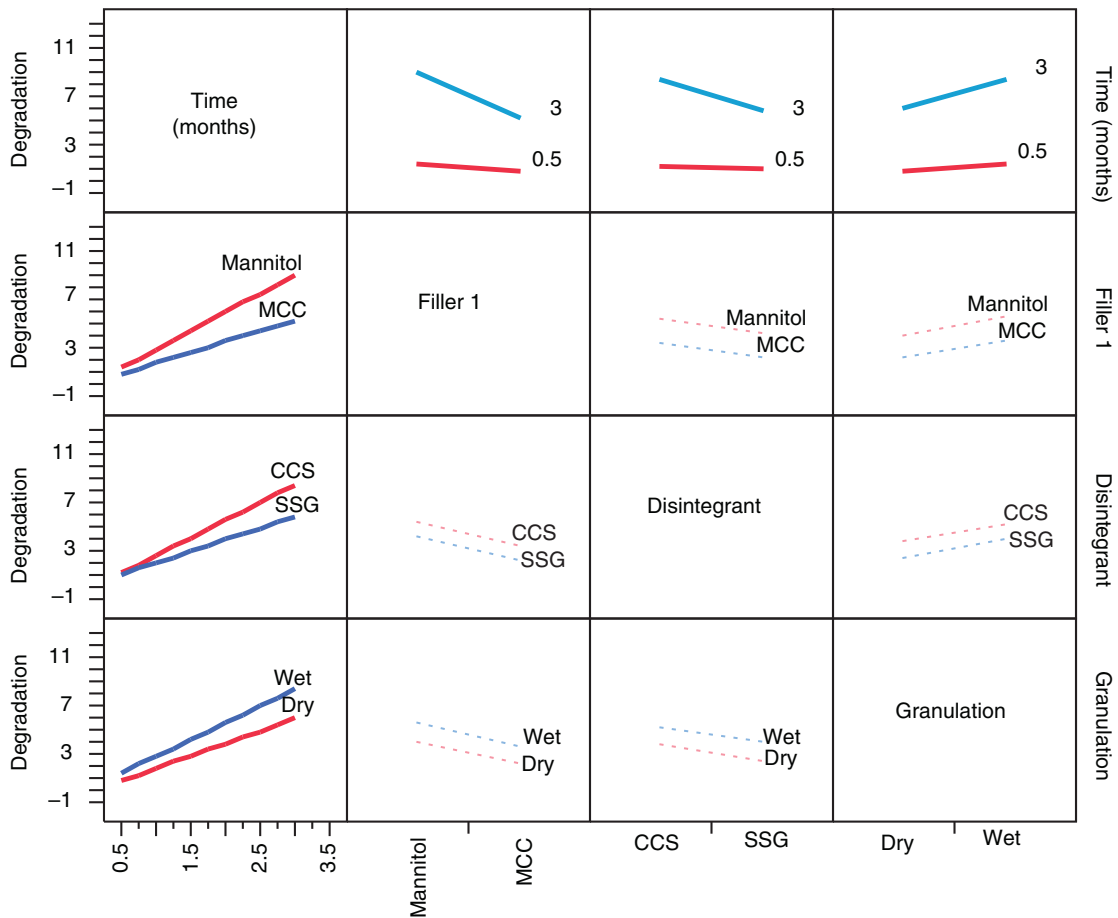


FIGURE 2.5 Prediction profiler and interaction profiles for drug–excipient compatibility studies.

processing platforms for solid oral dosage form include DC and granulation (wet and dry).

DC is the term used to define the process where powder blends of the drug substance and excipients are compressed

on a tablet machine. There is no mechanical treatment of the powder apart from a mixing process.

Granulation is a generic term for particle enlargement, whereby powders are formed into permanent aggregates.

The purpose of granulating tablet formulations is to improve the flow and compaction properties prior to compression. A number of methods are used to achieve the agglomeration; these are normally classified as either wet granulation, where a liquid is used to aid the agglomeration process, or dry granulation, where no liquid is used.

2.5.1 Dry Granulation

In the dry methods of granulation the primary powder particles are aggregated under high pressure. There are two main processes. Either a large tablet (known as a “slug”) is produced in a heavy-duty tableting press (a process known as “slugging”) or the powder is squeezed between two rollers to produce a sheet of material (“roller compaction”). In both cases, these intermediate products are broken using a suitable milling technique to produce granular material, which is usually sieved to separate the desired size fraction. The unused fine material may be reworked to avoid waste. This dry method may be used for drugs that do not compress well after wet granulation, or those which are sensitive to moisture.

2.5.2 Wet Granulation

Wet granulation involves the massing of a mix of dry **primary powder particles** using a **granulating fluid**. The fluid contains a solvent which must be volatile so that it is removed by drying, and be nontoxic. Typical liquids include water, ethanol, and isopropanol, either alone or in combination. The granulation liquid may be used alone or, more usually, as a solvent containing a dissolved **adhesive** (also referred

to as a **binder** or **binding agent**) which is used to ensure particle adhesion once the granule is dry.

The three main methods of producing pharmaceutical granulates are low-shear granulation, high-shear granulation, and fluid-bed granulation. Low-shear mixers encompass machines such as Z-blade mixers and planetary mixers which, as their name suggests, impart relatively low-shear stresses onto the granulate.

High-shear granulators are closed vessels that normally have two agitators; an impeller which normally covers the diameter of the mixing vessel and a small chopper positioned perpendicular to the impeller. The powders are dry-mixed using the impeller, and then the granulating fluid is added. Wet massing takes place using the impeller and the chopper, and granulation is usually completed in a number of minutes.

Fluid-bed granulation involves spraying the dry powder with a granulating fluid inside a fluid-bed drier. The powder is fluidized in heated air and then sprayed with the granulating fluid. When all the granulating liquid has been added, the fluidization of the powder continues until the granules are dry.

Seager et al. [26] produced a detailed analysis on the influence of the manufacturing method on the tableting performance of paracetamol granulated with hydrolyzed gelatin. The main difference in the granules produced by different methods is their final density, high-shear mixers producing denser granules than low-shear granulators, which in turn produced denser granules than fluid-bed granulations. Disintegration times were greater for tablets produced from the denser granulates. A detailed description of granulation process development and scale-up is found in the literature [27].

The advantages and disadvantages of each process are detailed in Table 2.9.

TABLE 2.9 Processing Platforms – Advantages and Disadvantages

Processing Platform	Advantages	Disadvantages
Direct compression	<ul style="list-style-type: none"> • Simple, cheap process • Suitable for heat and moisture-labile drugs • Prime particle dissolution 	<ul style="list-style-type: none"> • Generally limited to low-dose compounds • Potential to segregation • Expensive excipients
Dry granulation (slugging)	<ul style="list-style-type: none"> • Imparts flowability to formulation • Suitable for heat and moisture-labile drugs 	<ul style="list-style-type: none"> • Dusty process • Not suitable for all compounds • Slow process
Dry granulation (roller compaction)	<ul style="list-style-type: none"> • Imparts flowability • Suitable for heat and moisture-labile drugs • Limits segregation tendency 	<ul style="list-style-type: none"> • Slow process • Loss of compactibility for tableting • No hydrophilization of surfaces
Wet granulation (aqueous)	<ul style="list-style-type: none"> • Robust process • Improves flowability • Can reduce elasticity problems • Can improve wettability • Reduces segregation potential 	<ul style="list-style-type: none"> • Expensive • Specialized equipment • Stability concerns for moisture sensitive, thermolabile, and metastable drugs with aqueous granulation
Wet granulation (nonaqueous)	<ul style="list-style-type: none"> • Suitable for moisture-sensitive drugs • Vacuum drying techniques can reduce/remove need for heat 	<ul style="list-style-type: none"> • Expensive equipment • Explosion proof • Solvent recovery

Each processing platform has unique characteristics and complexity in terms of unit operations. Table 2.10 lists the unit operations required for manufacturing immediate release tablet using the processing platform discussed earlier.

Since more than one platform technology may be used to manufacture a drug product, selection of the most appropriate processing platform is affected by many factors as shown in Figure 2.6.

2.6 TABLET FORMULATION DESIGN

Having decided on a formulation design strategy, the process of preparing and screening initial formulation possibilities begins. It is important to appreciate that the goal is to develop a “robust” formulation and this objective facilitates identification of the factors that influence the selection of a design process as depicted in Figure 2.6. The first major design criterion is the nature of the API and in particular the possible

TABLE 2.10 Unit Operations Required for Various Processing Platforms

Unit Operation	Direct Compression	Dry Granulation	Wet Granulation
Raw materials (weighing and sieving)	✓	✓	✓
Blending	✓	✓	✓
Compaction		✓	
Wet granulation			✓
Wet screening			✓
Drying			✓
Milling		✓	✓
Tablet compression	✓	✓	✓

dosage level (described in Preformulation report and TPP). The knowledge of biopharmaceutical class to which the API belongs helps in deciding the formulation rationale. In particular, the implications of low permeability and low solubility must be carefully considered prior to the selection of the processing platform. For example, a poorly soluble drug often tends to be poorly wetttable, too. If the objective is to obtain a fast-dissolving and dispersing dosage form, inclusion of a wetting agent such as sodium lauryl sulfate or polysorbate 80 may be appropriate or even necessary.

Processing methods may also significantly impact dosage-form performance. For example, it may not be appropriate to wet-granulate amorphous drug because water may lower the glass-transition temperature and facilitate recrystallization during or after processing. In other situations, wet granulation can be used to avoid potential segregation and CU problems where there is a significant difference in particle size or bulk density between the drug and excipients.

Another major consideration must be the anticipated dosage level. It is worth emphasizing that in the case of a high dose active, a major proportion of the processing difficulties are traced to the physicochemical and mechanical properties of the API. Unfortunately, the key properties of the API may change during scale-up of the synthetic API process, or from lot to lot when outsourced. It follows that continuous monitoring of critical quality attributes (CQAs) of API that affect the process is an essential policy. Figure 2.7 depicts a decision-guiding flowchart for selection of the processing platform.

2.7 TABLET CHARACTERISTICS

There are two important classes of tablet characteristics. The first set examines the tablet immediately after manufacturing;

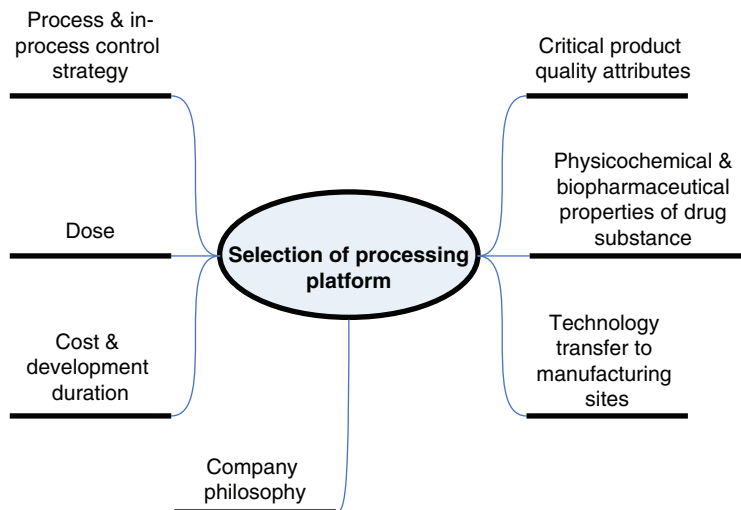


FIGURE 2.6 Factors affecting selection of processing platform.

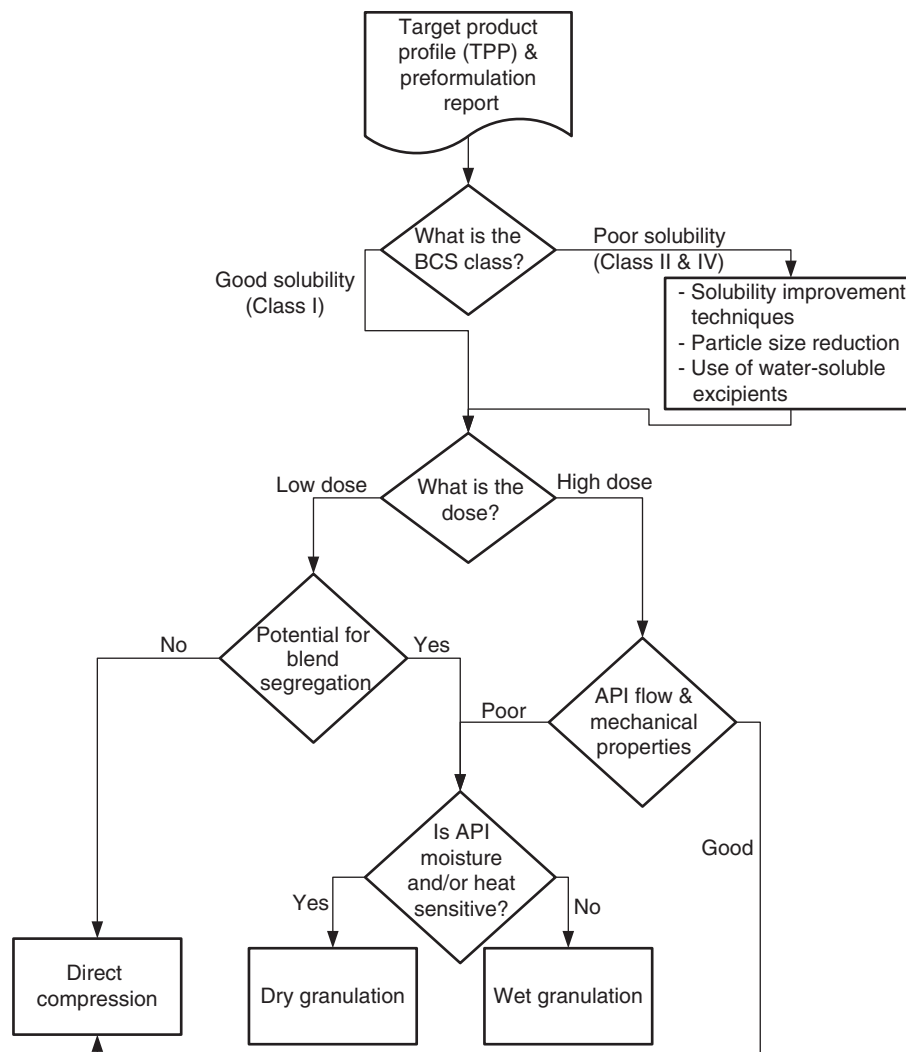


FIGURE 2.7 Flowchart for selection of adequate processing platform.

the second class examines what happens to the tablet over time.

Immediately after manufacturing and during the formulation process of a tablet, the release of the tablet is of utmost importance. If the tablet does not disintegrate or dissolve in the body, then the efficacious effect desired is likely not going to happen. There are many factors that can affect this from excipient choice to manufacturing.

After manufacturing, a tablet must maintain consistency over time. Similarly to drug release, excipients and processing can affect the shelf life of a tablet.

2.7.1 Release Profile: The Factors that Affect In-Vivo Performance

Release Profile of a tablet can affect in-vivo drug performance, as this is the case it is important to measure this

characteristic during development. The FDA guidance, Dissolution of Immediate Release Solid Oral Dosage Forms, states the dissolution requirements for an immediate-release drug. Dissolution testing is useful in development to determine how processing and formulations can potentially affect in-vivo performance. What is a dissolution test?

Dissolution is a test that provides some assurance of tablet performance by an indication of the mass transfer the drug into solution.

There are many stages in the development of a dissolution method. The final Quality Control (QC) form of the method is used in day-to-day production to ensure consistency of the tablets produced. In early development, dissolution testing is useful in screening formulations but this dissolution test may not be or even resemble the final QC test used when the drug has been approved. The development of a dissolution method

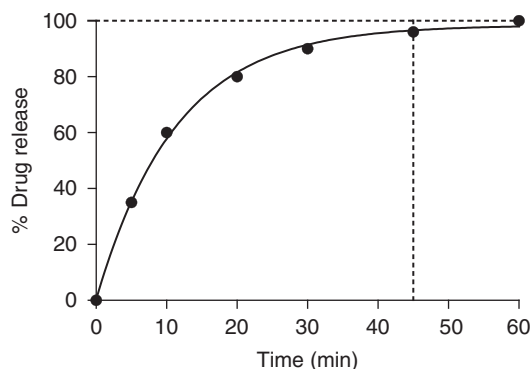


FIGURE 2.8 Typical drug-release profile; very fast initial release with a leveling off.

at each stage of development is the responsibility the analytical development (AD) group in a company. Figure 2.8 shows a “typical” immediate-release dissolution profile.

Even though it is generally the responsibility of the AD group to develop the dissolution method, it is critical for the drug developer to understand the final QC requirements from a regulation perspective as to aide in developing a final drug product. A final QC dissolution method is carried out according to the guidance is defined as:

Dissolution testing should be carried out under mild test conditions, basket method at 50/100 rpm or paddle method at 50/75 rpm, at 15-minute intervals, to generate a dissolution profile. For rapidly dissolving products, generation of an adequate profile sampling at 5- or 10-minute intervals may be necessary. For highly soluble and rapidly dissolving drug products (BCS classes 1 and 3), a single-point dissolution test specification of NLT 85% ($Q = 80\%$) in 60 minutes or less is sufficient as a routine quality control test for batch-to-batch uniformity. For slowly dissolving or poorly water-soluble drugs (BCS class 2), a two-point dissolution specification, one at 15 minutes to include a dissolution range (a dissolution window) and the other at a later point (30, 45, or 60 minutes) to ensure 85% dissolution, is recommended to characterize the quality of the product. The product is expected to comply with dissolution specifications throughout its shelf life. If the dissolution characteristics of the drug product change with time, whether or not the specifications should be altered will depend on demonstrating bioequivalence of the changed product to the original biobatch or pivotal batch. To ensure continuous batch-to-batch equivalence of the product after scale-up and post-approval changes in the marketplace, dissolution profiles should remain comparable to those of the approved biobatch or pivotal clinical trial batch(es) [28].

This is important knowledge to ensure compliance when developing and changing formulations. The QC method described above is not always the best method to use during development to assess potential impact on bioavailability; alternate media or methods may provide additional insight.

2.7.2 What Affects Dissolution: Problems and Trouble Shooting with Dissolution Testing

Beyond compliance dissolution is used to determine the performance of the tablet. Assuming a well-developed dissolution method, there are many things that can affect the dissolution of the tablet:

- Processing conditions: compressing the tablet too hard, overblending the lubricant
- Excipients: amount and choice
- API physical properties
- Storage: over time the tablet dissolution may slow down due to excipient interactions with the drug and excipient reaction with each other

A discriminating dissolution method is useful in developing a tablet formula and manufacturing process; however, a proper method may take time for the AD group to develop, just as it takes a while to develop a reliable process.

2.7.2.1 Problems with Dissolution: Non-Engineered Mixing Vessels and Trouble Shooting

Assuming a good dissolution method may not be the best assumption. Dissolution is a QC test required for regulatory compliance; however, there are many problems with the dissolution test.

Dissolution Apparatus 1 is a paddle mixer in a cylindrical vessel; from an engineering standpoint this does not provide good mixing. If an engineer is designing this, he/she would have put a baffle or two in there to promote top to bottom mixing. As is imagined, there may be problems with bottom settling and coning with tablets that disintegrate into large particles, which have a high density. In this case the dissolution results have significant variation as how the drug settles and the percentage of the drug setting has an effect on the results.

Apparatus 2 is a basket mixer in a cylindrical vessel; again from an engineering standpoint, this does not provide good mixing. There is little mixing power associated with the method, if the powder flows out of the basket the powder either settles, floats, or suspends depending on the powder’s buoyancy. If the powder stays in the basket, the method has a high probability to be reliable (Figure 2.9).

When examining dissolution-results method there are five considerations to determine if results are method-biased.

1. What is the media used in the dissolution bath? What is the solubility of the drug in the media? This determines the mass transfer-driving force for the drug to go into solution.
2. Does my drug change forms in the dissolution media? If it does the form it changes into may not have the



FIGURE 2.9 Typical dissolution apparatuses. http://www.lasany.org/pharma_lab equipments/dissolutiontestapparatus.html, <http://www.cecilinstruments.com/autodis-accs.html> www.bestlab.com.au/disso_access.htm

same solubility. Form conversion is a stochastic event and affects the consistency of the results.

3. Are the particles suspended and flowing? This also affects the mass transfer of the drug into the media.
4. Is the tablet submerged in the media? Often a floating tablet provides many problems and inconsistent results.
5. What is the dissolution media comprised of? The media may react with the API or excipients used in the tablet.

When analyzing a change in dissolution profiles ensures that the changes are due to changes made to the process and formula versus problems with the method. It is always a good idea to observe the dissolution testing so to see what is actually occurring.

2.8 USING DISSOLUTION TO DETERMINE CQAS

Assuming an acceptable dissolution method has been developed, dissolution is a useful tool to determine CQAs for the tablet. Dissolution can help determine the maximum tablet hardness, the optimal drug substance particle size and/or density, and the proper ratio or the amount of excipients.

2.8.1 Using Dissolution to Determine the Ratio of Excipients

A tablet formulation can affect the dissolution profile. A tablet often contains a mixture of water-soluble and -insoluble fillers/binders, and disintegrants that all have the potential to affect the dissolution profile. Determining the optimal loading of excipients is a difficult task even after the compatible excipients have been chosen.

Examining excipient optimization of a BCS class II tablet based on dissolution performance. For example, compressing a tablet consisting of 20% API with a particle size of 29 μm at a hardness of approximately 10 kP with remaining 80% of the tablet has different ratios of filler, binder, and disintegrant. Two commonly used fillers are Micro Crystalline Cellulose (MCC) and Calcium Di-Basic Phosphate (A-Tab), and a commonly used disintegrant Sodium Starch Glycolate (SSG), which are used based upon excipient-compatibility example. These are compressed into five different ratios; dissolution results are shown in Figure 2.10.

As is seen in Figure 2.10, different excipient ratios can affect tablet performance. For this example it looks like 71/25/4 MCC/A-Tab/SSG has the most optimal performance without putting an excess amount of disintegrant in the tablet (Table 2.11).

2.8.2 Using Dissolution to Determine the Optimal API Particle Size and Tablet Hardness

The next properties that can affect dissolution are API particle size and tablet hardness. API particle size has the potential to affect dissolution based on different surface area or particle morphology and the tablet hardness can affect how fast the tablet disintegrates into primary particles enabling the API to dissolve. As a rule of thumb about particle size:

There is never an instance where bigger particles will improve the immediate release performance but there are many instances where it will not change the performance.

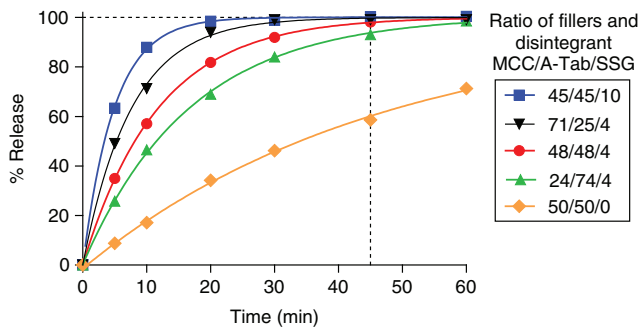


FIGURE 2.10 Dissolution comparison between different excipient ratios and a constant drug load of 20%.

TABLE 2.11 Percent Release Data from Figure 2.10

Time (min)	Release Data of Tablets with Different MCC/A-Tab/MCC Ratios				
	48/48/4	45/45/10	25/71/4	71/25/4	50/50/0
0	0	0	0	0	0
5	35.0	63.4	25.8	49.1	8.8
10	57.1	87.9	46.5	71.3	17.1
20	81.8	98.5	69.0	94.0	34.2
30	91.9	98.8	84.0	99.1	46.2
45	98.1	100.3	93.0	99.1	58.6
60	99.6	100.4	98.6	99.1	71.3

Due to method variability it is common to see tablet performance slightly above or below 100%.

In optimizing the release of the drug, first a target CQA must be defined which is determined from IVIVC⁷ or good scientific reasoning. A hypothetical CQA could be NLT (not less than) 70% release at 30 minutes, to ensure proper absorbance in the body; 30 minutes is chosen as it is the approximate gastric emptying time of an empty stomach [29].

Continuing with the example, determining the optimal hardness and API particle size-range dissolution is chosen at the CQA at $t = 30$ minutes. Starting with the “optimal” formulation from the example (71/25/4 MCC/A-Tab/SSG), the material is compressed at five hardness, ranging from approximately 10 to 30 kP, and four different API average particles sizes (d_{50} s), ranging from 29 to 73 μm . Table 2.12 indicates the results attained and from observation there is an effect of both hardness and API particle size.

In examining Figure 2.11, the data have a linear relationship between % release and hardness; as well, there is a relationship between release and particle sizes. It is noted that 8 of 20 experiments met the CQA requirement of NLT 70%

TABLE 2.12 Effect of API Particle Size and Tablet Hardness on the 30 Minute-Dissolution Time Point

T30 min	d_{50}	Hardness	T30 min	d_{50}	Hardness (kP)
99.1	29	9.7	87.18	50	10.3
85.17	29	15.6	75.91	50	14.4
77.45	29	20.8	65.03	50	20.4
64.3	29	25.6	52.42	50	24.7
54.15	29	30.2	40	50	30.2
89.04	42	10.7	72.65	73	9.6
84.35	42	14.7	60.13	73	15.3
69.17	42	20.9	50.84	73	20.9
61.31	42	24.1	43.62	73	25.4
46.61	42	29.4	26.97	73	29.6

⁷IVIVC: In-Vitro In-Vivo Correlation, but which benchtop data accurately correlates with human bioavailability.

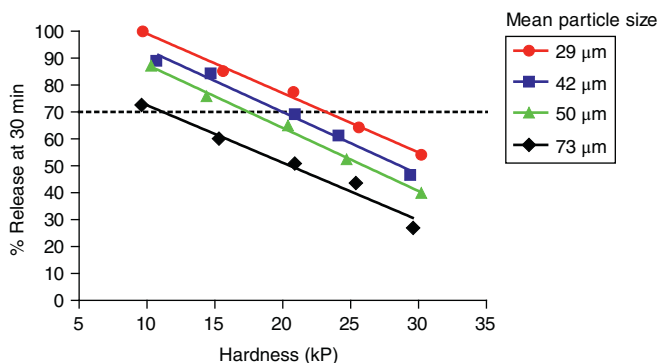


FIGURE 2.11 Comparison of hardness and API particle size to dissolution release at the 30 minute time point.

release at 30 minutes. From this point a design equation is developed to mathematically describe the DS.

Regression is completed providing an expression for the relationship of acceptable hardness and API particle-size combinations. The expression is used to describe the DS (Tables 2.13 and 2.14).

Based on this information the relationship between hardness, particle size, and % Release at 30 is:

$$\%R_{at30min} = 139.2 - 0.59 \cdot D_{50} - 2.25 \cdot \text{hardness} \quad (2.10)$$

This is not an ideal form of the equation as greater than 100% is predicted; however, it is used to determine the maximum range of hardness and particle size to attain release greater than 80%. The model is further developed to attain the curvature but more data above 30 kP and smaller particle sizes are required. For determining tablet and API properties, there is sufficient information for control.

To determine the acceptable combinations of hardness and particle size to maintain the CQA of NLT 70% Release at 30 minutes, Eq. (2.10) is rearranged.

TABLE 2.13 Results from the Linear Regression (Completed Using JMP8.0)

Term	Estimate	Std Error	t Ratio	Prob > t
Intercept	139.2	1.99	69.9	<0.0001 ^a
d_{50}	-0.592	0.0289	-20.5	<0.0001 ^a
Hardness	-2.247	0.0659	-34.0	<0.0001 ^a

^aIndicates the variable is significant.

TABLE 2.14 Summary of Fit of the Regression

R^2	0.989
R^2 Adj	0.988
Root mean square error	2.070
Mean of response	65.31
Observations	20

$$69.3 \geq 0.59 \cdot D50 + 2.25 \cdot \text{hardness} \quad (2.11)$$

As long as this equation is satisfied, there is confidence the CQA is maintained. The DS is described in Figure 2.12.

The last check is examining the residuals to ensure there is not a systematic error. Shown in Figure 2.13 is the data are randomly distributed, indicating the regression does not have a systematic error. Another way is to confirm that the model residuals are normally distributed by using a goodness of fit.

2.8.3 Physical Tablet Characteristics

The physical attributes of the tablet are important for processing and to ensure a consistent quality drug product is delivered to the customer. Physical attributes include tablet hardness, thickness, friability, disintegration, and weight. When determining tablet characteristics, consider how the material is to be handled after compression (Table 2.15).

2.9 DRUG PRODUCT STABILITY

Stability is critical in all drug product/formulation design; without stability there is no commercial product. Stability

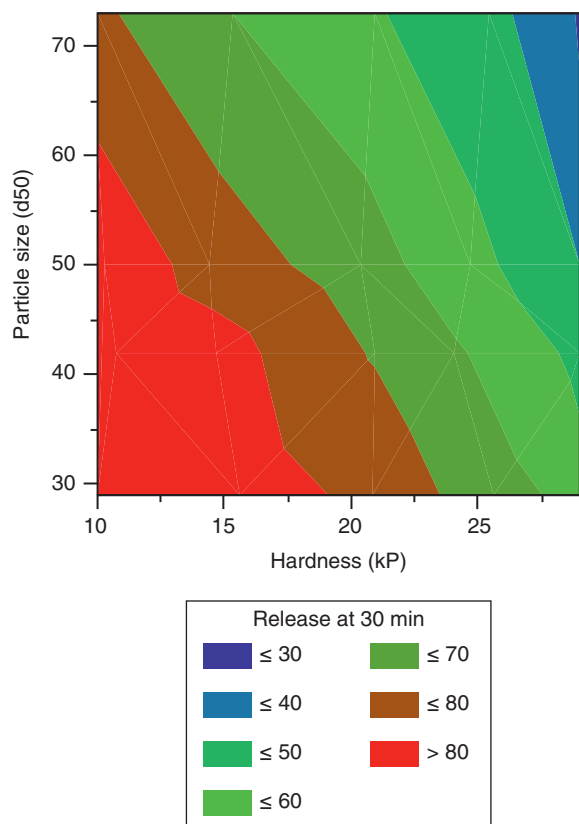


FIGURE 2.12 Contour plot showing dissolution as a function of particle size and hardness.

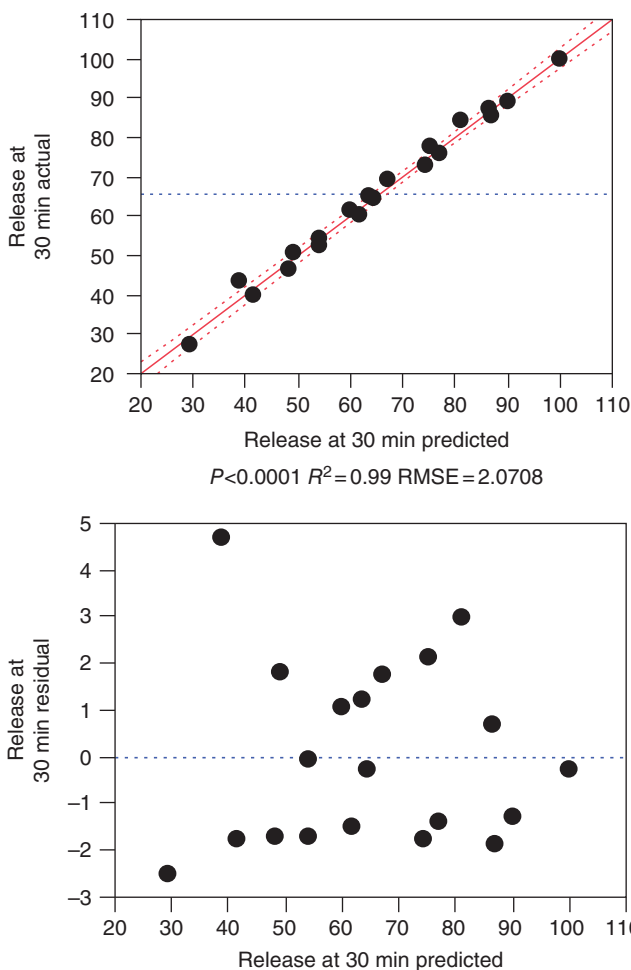


FIGURE 2.13 Residual plot from the regression. The data are randomly distributed and non-systematic.

is examined in two different manners: first meeting the minimum regulatory requirement needed to launch a drug and/or examining the root causes of degradation. From a scientific/engineering perspective, it is important to determine what affects the stability of the drug in order to design a drug product process around these factors.

2.9.1 Regulatory Requirements for Drug Product Stability

Regulatory requirements for stability need to be done in the intended primary commercial package. There are ICH guidance that governs the expectations of pivotal stability studies, see “Stability Testing Of New Drug Substances And Products Q1A(R2).” The minimum requirement for submission is shown in Table 2.16; real-time data are needed for shelf-life dating over two years.

This is an excellent guide for the regulatory requirements once a package(s) has been chosen for clinical trial,

TABLE 2.15 Tablet Attributes and Their Effect on Final Dosage Form

Attribute	Effect	Measurement
Hardness	Too soft: the tablet can break in storage shipping, coating, packaging. Too hard: the tablet cannot dissolve and may not have the required clinical effect	Hardness tester Typical units: kP, N
Friability	Too friable: Tablet cannot withstand further testing Not friable: nothing wrong	Friabilator: 100 revolutions; % weight loss; if capping/lamination occurs
Thickness	Too thick: May not fit into packaging equipment/package. May not be able to swallow (poor marketing compliance) Too thin: May clog packaging equipment	Caliper
Weight	Too heavy: The drug may be overpotent Too light: there may not be enough drug (poor clinical efficacy) Too much variability: may fail content uniformity. Too much yield loss during manufacturing.	Scale
Disintegration	Too slow: may not be efficacious Too fast: may have issues in humid environments and coating	Disintegration, dissolution bath
Elegance	Non-elegant: Shows inconsistent production and may turn off customers	Acceptable quality limit (AQL)

TABLE 2.16 Minimum Guideline from the ICH Q1A (R2) for Room Temperature Product

Study	Storage Condition	Minimum Time Period Covered by Data at Submission
Long term ^a	25 °C ± 2 °C/60%RH ± 5%RH or 30 °C ± 2 °C/65%RH ± 5%RH	12 months
Intermediate ^b	30 °C ± 2 °C/65%RH ± 5%RH	6 months
Accelerated	40 °C ± 2 °C/75%RH ± 5%RH	6 months

^aIt is up to the applicant to decide whether long-term stability studies are performed at 25 ± 2 °C/60%RH ± 5%RH or 30 °C ± 2 °C/65%RH ± 5%RH.

^bIf 30 °C ± 2 °C/65%RH ± 5%RH is the long-term condition, there is no intermediate condition.

registration, and commercial distribution. Registration batch minimum is three lots of at least 100 000 tablets and at least 1/10 of the expected commercial batch size, which is packaged into the intended commercial package, but there is much work required before selecting the primary package.

2.9.2 What Affects Stability and How to Predict Shelf Life?

In a QbD world the minimum is generally not sufficient to launch a product; the more a scientist determines what affects stability, the better engineered the product.

Packaging is usually not known in early development and it can range from blister packaging, bottles, and pouches. Each packaging type can vary significantly in the materials used. Different materials can protect from light, moisture, oxygen, and other environmental factors. Early research in the effect of heat, moisture, oxygen, and light enables

primary and secondary packaging selection. Though these studies are comprehensive, studies still need to be completed with the primary package to ensure no reaction between the packaging material and the drug product. The primary packaging may have leachables, extractables, or antistatic properties that may react with the drug product.

A structured approach helps determine the conditions a drug converts into a degradation product. This occurs given enough time in certain temperature, RH, or light conditions. Packaging is often used to prevent such occurrence, the proper choice and storage conditions are critical depending on the stability of the product.

Determining package type is as easy as answering a few questions: what is the drug product sensitive to, temperature, moisture, light? Is the drug sensitive to impurities/components in the packaging, impurities/components in the excipients, or starting impurities in the drug substance?

This section provides a framework to determine what affects stability and how. The simplest experiments are placing the product at open dish conditions, an example of this is provided in the excipient compatibility section, where a more integrated set of experiments is used to create a predictive model on how the tablet can degrade. From the initial readout of stability a more extensive experimentation is completed to model the stability of the drug.

2.9.3 Open Dish Experiments

These are the easiest experiments to get a quick read on how the drug product can degrade and what changes in formulation affect degradation. Using relatively harsh conditions of 40 °C dry and 40/75%RH provides immediate (i.e. one to four weeks) information of how the drug reacts with both temperature and humidity and the degradants or form change

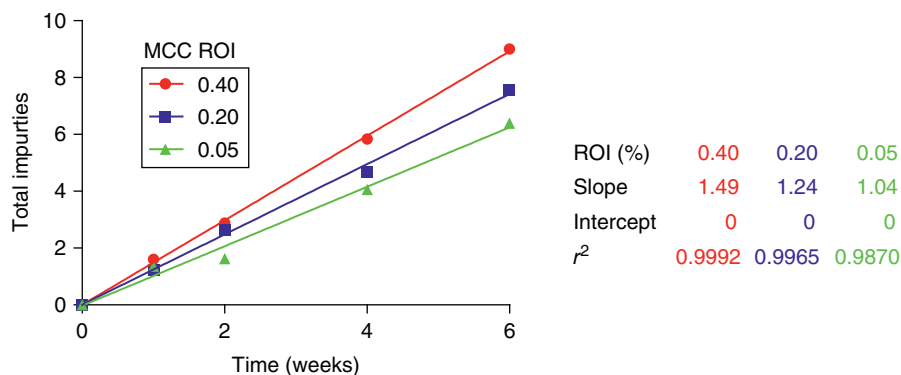


FIGURE 2.14 A hypothetical effect of MCC ROI and the growth of tablet impurities at 40 °C/75%RH open dish conditions.

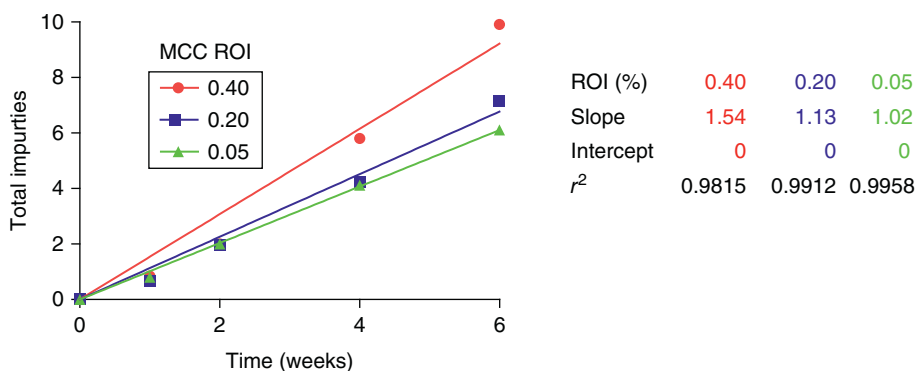


FIGURE 2.15 A hypothetical effect of MCC ROI and the growth of tablet impurities at 40 °C dry open dish conditions.

to expect upon stability. These types of experiments can provide an indication of how changes in the formulation can affect stability.

Open dish experiments are used to test specification of excipients. Possible effects on stability can occur from changing an excipient vendors or lot-to-lot variation within a vendor. For example, MCC, a commonly used binder, often has a residue on ignition (ROI)⁸ specification of not more than (NMT) 0.050%, so it is possible to receive material with ROI of 0.040, 0.005, and 0.020%. A tablet is compressed with these different lots of MCC and placed on open dish stability and depending on the drug the results could affect the stability. Figures 2.14 and 2.15 show the results from this example.

Both Figures 2.14 and 2.15 show a relationship between ROI and impurity growth. A regression analysis is completed and shown in Table 2.17. The regression show both Time and RH affect stability but RH does not affect stability.

Another way to examine the data is plotting the slopes from Figures 2.14 and 2.15 (degradation rate) against ROI of the MCC. Figure 2.16 shows how the ROI affects the

TABLE 2.17 Regression Results for the Material Stored at 40 °C Dry and 40 °C/75%RH

Term	Estimate	Std Error	<i>t</i> Ratio	Prob > <i>t</i>
Intercept	-0.902 499	0.136 978	-6.59	<0.0001 ^a
Time (weeks)	1.284 123 7	0.029 206	43.97	<0.0001 ^a
ROI (%)	3.293 04	0.438 798	7.50	<0.0001 ^a
RH conditions	0.1093 065	0.062 911	1.74	0.0946
Time (weeks) - 2.6) ^a (ROI - 0.21667)	1.583 377 6	0.203 707	7.77	<0.0001 ^a

^aIndicates the term is statistically significant (Prob < 0.05).

growth rate of impurities; this could be important and may provide justification setting excipient specifications.

Open dish studies are useful in determining what can degrade the drug but these are harsh conditions and do not simulate what would happen upon shelf life, but do give an indication of what to look for on stability.

2.9.4 How to Model and Predict Packaged Shelf Life

Using information gained from open dish studies, a more elegant study is then conducted in determining the drug shelf

⁸The residue on ignition (ROI) test measures the amount of residual substance not volatilized from a sample when ignited in the presence of sulfuric acid. The test determines the content of inorganic impurities. USP <281>.

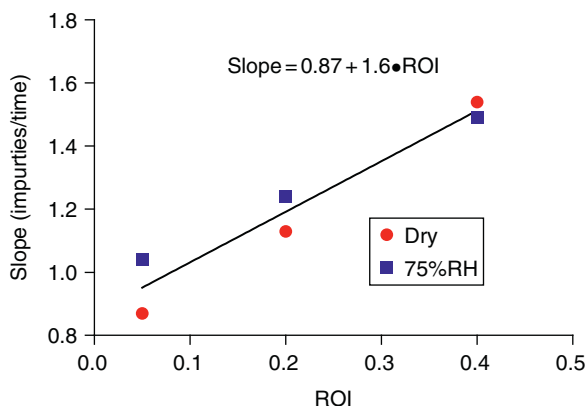


FIGURE 2.16 Comparison between impurity growth rate and ROI of excipient.

life. Experimenting on the effect of temperature and tablet moisture on impurity growth is used to develop a model to predict shelf life. This type of study is called the TRH study, which models the effect temperature (T) and RH on tablet shelf life.

Setting-up this study requires tablets, RH equilibration chambers, foil pouches, and a heat sealer for the pouches. Tablets are equilibrated at different RH conditions and then packaged in foil pouches to ensure the moisture content of the tablet remains constant throughout the time material is stable. In addition, every time point and condition shall be individually packaged to maintain the tablet moisture, as opening and closing packages could adulterate the tablets. The idea is to equilibrate separate tablets to a minimum of three different groups to equilibrate them at three different conditions (i.e. 15, 25, and 45%); equilibration may take up to two to seven days. The last three steps are:

1. Measure the tablet moisture content (Karl Fisher (KF) is one of the more effective measurements); separate the tablets into three moisture categories (i.e. 1, 2, and 5.5%)
2. Determine the amount of time pulls required (i.e. 1, 3, 6, 9, 12, 18, and 24 months)
3. Determine the storage temperatures (T) to place the tablets at; a minimum of three is recommended (i.e. 25, 30, and 40 °C), these are typical ICH temperatures.

The study described requires 63 foil pouches to cover each time point and condition. This is an extensive study but does not account for different lots of API or excipients. Much is learned from this study about packaging protection requirements. To expand the study, excipient ROI is examined as a factor, which increases the study 3X.

The experimental results are shown in Figure 2.17, the data set is extensive, but it is important to analyze interim data and

guide the packaging decisions. At the end of the study a complete predictive model for temperature, moisture, time, and excipient ROI is attained, to guide decisions made around storage temperature, shelf life, and excipient specifications.

A regression analysis is completed to provide a prediction equation to determine what and how much each of the variables affects impurity growth. The analysis indicates time, Temp, and ROI have effect on degradation; in addition, there are some interactions between Time and Temp, Time and ROI, and Temp and ROI. As expected, tablet moisture content did not affect stability as is seen from Figure 2.17 (Table 2.18).

Figure 2.18 shows that the regression is not biased and the residuals are evenly distributed.

Equation (2.12) is the expression for expected impurities at any given time, storage temperature, and MCC ROI. This is used to test different scenarios like: what would the ROI need to be to attain Room Temperature/25 °C (RT) storage condition with acceptable amount of impurities or what would the shelf life be at in warmer climates of 30 °C. It must be noted that impurity levels are set by a combination of toxicity and process capability.

$$\begin{aligned} \text{Impurities} = & 0.633 - 0.041 \cdot T - 0.020 \cdot t + 1.17 \cdot \text{ROI} \\ & + 0.00035 \cdot (t \cdot T) + 0.360 \cdot (t \cdot \text{ROI}) + 0.080 \cdot (T \cdot \text{ROI}) \end{aligned} \quad (2.12)$$

Contour plots (Figure 2.19) are useful as the sensitivity of each variable is more easily visualized.

Stability is important and knowing what can predict stability is valuable in determining packaging and excipient grade selection.

2.10 PROCESS OPERATIONS AND SCALABILITY OF DOSAGE FORM

There are many considerations in scaling-up unit operations that manufacture solid dosage forms. Scaling-up through pre-clinical → early clinical (phase I and phase II) → late clinical (phase IIb and phase III) → Registration → Engineering/Validation batches has many challenges. Scale-up usually takes the course of laboratory experiments, pilot-scale tests, and then commercial-scale operation and continuous improvement [30, 31] (Table 2.19).

Beyond development, scale-up or scale-down also occurs after approval, in which changes are governed by Post-Approval Changes (SUPAC⁹) guidelines as specified by the Center for Drug Evaluation and Research (CDER). Finally, Tech Transfer (TT) is needed if multiple plants or CMOs are required.

⁹<http://www.fda.gov/Drugs/default.htm>

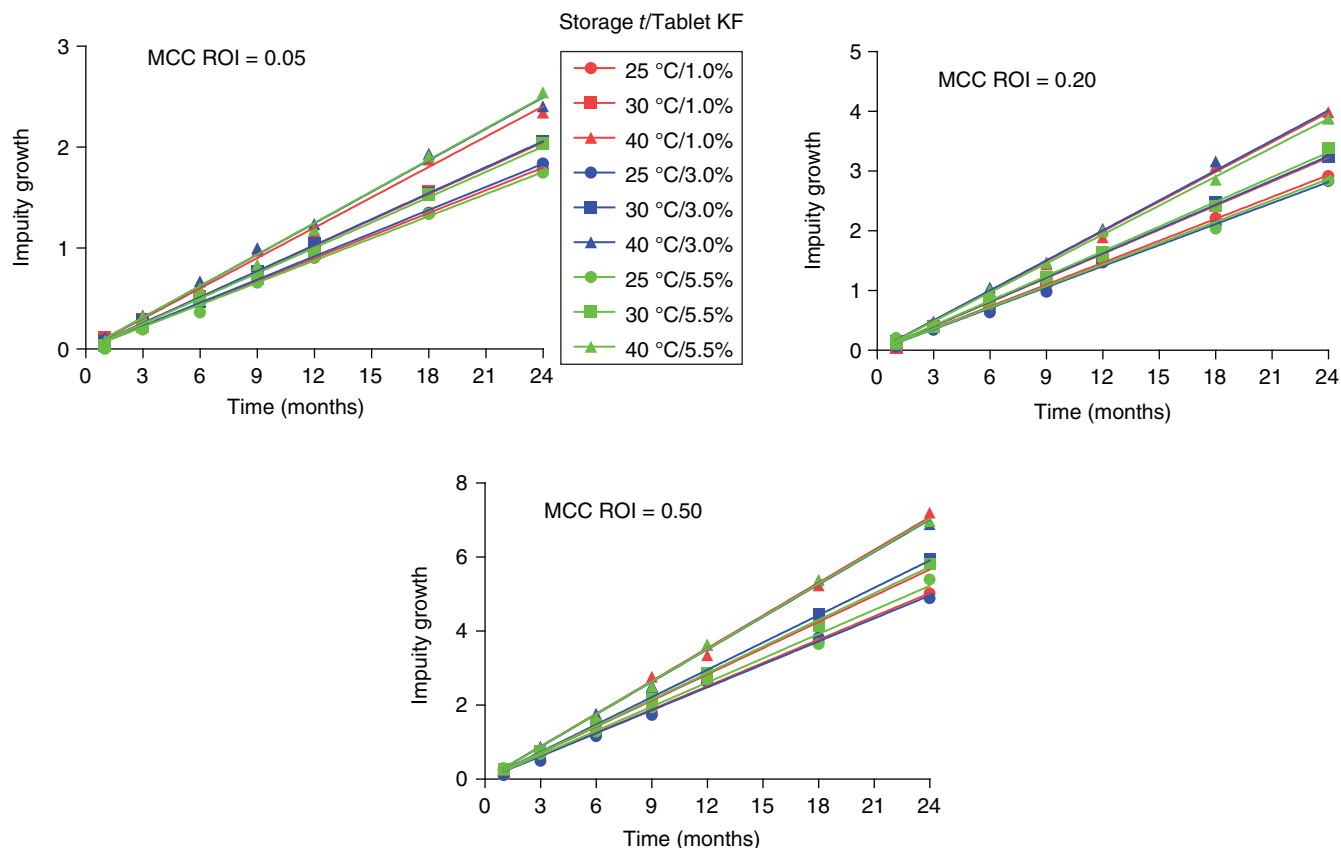


FIGURE 2.17 Result from the temperature moisture study completed by examining the effect of MCC ROI at three different levels 0.05, 0.20, and 0.50%. The figure seems to indicate that temperature time and ROI have the largest effect on impurities.

TABLE 2.18 Regression Results from the TRH Experiments

Term	Estimate	Std Error	t Ratio	Prob > $ t $
Intercept	-2.11	0.043	-49.12	<0.0001*
Time- T (months)	0.160	0.001 02	156.82	<0.0001*
Temp- t °C	0.037	0.001 25	29.60	<0.0001*
MCC ROI	3.71	0.042	89.00	<0.0001*
(Time (months) - 10.43) × (Temp °C - 31.67)	0.0035	0.000 16	21.60	<0.0001*
(Time (months) - 10.43) × (MCC ROI - 0.25)	0.360	0.0055	65.99	<0.0001*
(Temp °C - 31.67) × (MCC ROI - 0.25)	0.080	0.0067	11.91	<0.0001*

All parameters shown are significant. The parameters analyzed that did not show significance were tablet moisture and all tablet moisture interactions.

Limited and costly API or Drug Substance (DS) and resources may hinder the experimental understanding that could be gained; therefore, know-how prior to manufacturing is extremely valuable. Understanding and using engineering first principles, dimensional analysis, and DOE improves the likelihood that the process(es) and Drug Product (DP) are to succeed.

Pharmaceutical process scale-up shall consider formulation, process development, and marketing needs. A risk-based

approach is used to examine how the TPP of the drug is affected by CQAs of the final dosage form and the DS of the process. Quality by Design (QbD) principles are used to ensure a safe and efficacious product. DS, controls, and specifications are continuously improved through continuous learning.

Definitions referring to pharmaceutical manufacturing are given as follows; adapted from PQRI [32]:

Critical Quality Attribute (CQA). A quantifiable property of an intermediate or final product that is considered critical for establishing the intended purity, efficacy, and safety of the product. That is, the property must be within a pre-determined range to ensure final product quality.

Target Product Profile (TPP). A summary of characteristics that, if achieved, provides optimal efficacy, patient compliance, and marketability. A TPP often includes attributes like pharmacokinetic information (e.g. immediate release vs. extended release), dosage form (e.g. tablet vs. injectable), and shelf-life information (e.g. two years at 25 °C/60%RH).

Design Space (DS). The DS is the established range of process parameters that has been demonstrated to provide assurance of quality. In some cases DS can also be applicable to formulation attributes.

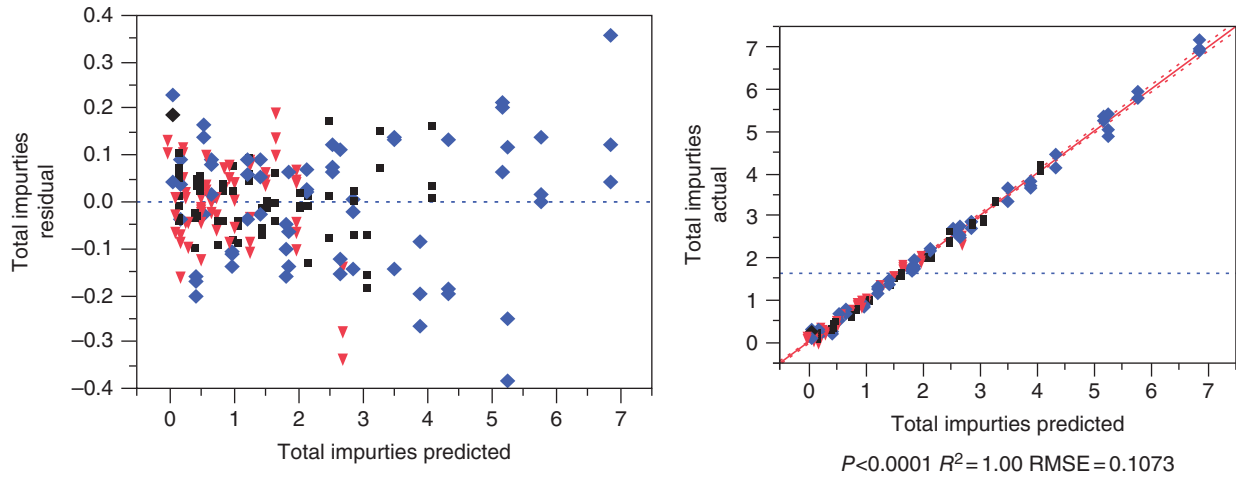


FIGURE 2.18 The residuals plot of predicted versus actual. The residuals are evenly distributed indicating there is no bias in the model. ∇ is 0.05 ROI, \square is 0.20 ROI, and \diamond is 0.50 ROI.

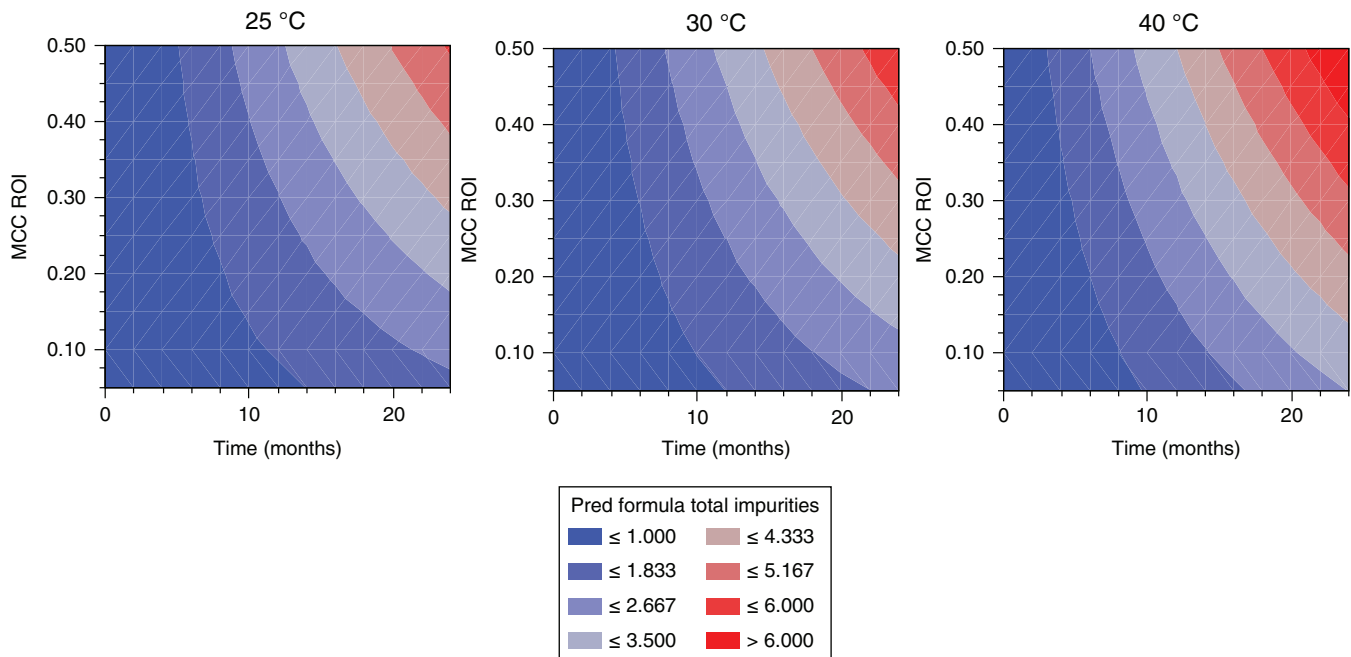


FIGURE 2.19 Contour plots examining total impurities versus time and MCC ROI.

Critical Process Parameter (CPP). A CPP is a process input that, when varied beyond a limited range, has a direct and significant influence on a CQA.

Critical Material Attribute (CMA). A CMA is an attribute that has direct impact on the processability or CQA of the drug product. CMAs could include impurities from the excipient or raw materials.

Normal Operating Range (NOR). A defined range, within the Proven Acceptable Range (PAR), specified in the manufacturing instructions as the target and range at which a process parameter is to be controlled, while producing unit operation

material or final product meeting release criteria and CQAs.

Proven Acceptable Range (PAR). A characterized range at which a process parameter may be operated within, while producing unit operation material or final product meeting release criteria and CQAs.

Another useful tool during scale-up is process Failure Modes Effects Analysis (pFMEA), which is used to understand the failure modes of the CQAs to help mitigate risks in unit operations [33, 34]. This is done by understanding the severity, occurrence, and detection of any or all potential

TABLE 2.19 Typical Scale-Up

Stage	Typical Material Required	REASON
Preclinical	0.05–1 kg	Early toxicology testing
Phase I and II	0.2–50 kg	Healthy volunteers and early proof of concept
Phase IIb and III	10–1000 kg	Proof of concept and verification trials
Registration	>100 kg and >100 000 unit dosages and minimum 1/10th commercial batch size	From FDA guidance
Engineering/validation	Based on registration and expected product demand	Final process testing and process confirmation runs

failure modes. pFMEA is different from Root Cause Analysis (RCA) as a RCA is performed after deviations have already occurred. Risk Prioritization Numbers (RPNs; scores from 1 to 100) are calculated using pFMEA ($RPN = \text{severity} [1-10] \times \text{occurrence} [1-10] \times \text{detection} [1-10]$) and point the engineer toward corrections that are implemented to reduce risk to the drug product. Usually, an engineer starts to look to ameliorate problems with high RPN numbers.

In addition to CQAs, Critical Business Attributes (CBAs) are also considered. Business decisions that involve scale-up can relate to choice of CMO, batch size, operators needed, equipment purchases, use of PAT tools, etc.

It is imperative to learn the CPPs of the unit operation at hand. This is done through an evolution of understanding the engineering principals and the processing knobs at the engineer's disposal. These CPPs affect any or all of the CQAs of the DP [32].

Unit Operation	CQAs	CPPs	Potential Failure Mode
Roller compaction	Ribbon density, degradants, downstream dissolution	Roll speed Feed screw speeds Roll force/pressure Roll separation/gap Room temperature/humidity	Ribbon density variation High degradation
Slugging	Hardness Dissolution	Slugging force	Too little or too much
Wet granulation	Particle size, powder density, degradants, downstream dissolution	Granulation fluid mixing time Granulation fluid mixing speed Granulating fluid amount Granulating fluid addition rate Granulating fluid temperature Spray nozzle air volume Dry mixing time Wet mixing time Impeller speed Chopper speed Power consumption	Too little or too much
Fluid bed granulation	Particle size Powder density Powder wetness Degradants Downstream dissolution	Granulation fluid mixing time Granulation fluid mixing speed Granulating fluid amount Granulating fluid addition rate Granulating fluid temperature Spray nozzle air volume Bed mixing time Supply air flow rate, temperature, dew point Product bed temperature Exhaust air temperature, dew point Filter shaking intervals	Loss of yield, powder degradation
Milling	Particle size Degradants	Impeller speed Feed rate Room temperature, humidity	Undesired particle size, degradation

(Continued)

Unit Operation	CQAs	CPPs	Potential Failure Mode
Lyophilization	Degradants Physical form Product wetness	Pretreatment, freezing, drying Temperature Cycle times Chamber pressure	Degradation, loss of stability, yield loss
Blending	Blend uniformity Content uniformity	Blend time (pre- and post-lube) Rotation rate Agitator speed Room temperature, humidity	Under-blending may lead to bad CU. Overblending may lead to poor compressibility
Encapsulation	Powder density, downstream dissolution Weight	Speed, dosing	Improper weight, broken capsules, too dense powder in capsule
Tableting	Hardness Thickness Weight Dissolution Degradants Content uniformity	Tablet weight Press (turret) speed Main compression force Pre-compression force Feeder speed Upper punch entry Room temperature, humidity	Capping if dwell time is too low Low weights or high weight variability if powder flow is bad
Tablet coating	Appearance Dissolution	Coating suspension mixing time Coating suspension mixing speed Coating suspension solids' load Atomization pressure Preheat time Jog time #, number of guns, type of guns Gun to bed distance	Twinning if tablet shape is not round Spray drying of coating suspension if temperature is too high Nonuniform coating if pan speed is too slow Tablet defects if pan speed is too fast
Tablet printing	Appearance Degradants	Ink dosage amount, force, location	Ink degrades product

There are many unit operations that are used for drug product manufacture, the easiest and most economical is DC. In oral solid-dosage manufacturing, DC process technology is the most effective and efficient way to make powder materials suitable for tableting or encapsulation without a step to increase the particle size [35]. In the example, the TPP is a DC Tablet, which focuses on the unit operations of (i) blending, (ii) compression, and (iii) coating.

2.10.1 Blending Scale-Up

Blending is a critical operation that determines how well the product is to perform in the next phases. Achieving and maintaining homogenous mixing of powders is critical, especially in formulations involving small amounts of high-potency components. Lack of blend uniformity at the blending stage may result in lack of CU in the finished product dosage forms.

Tumbling blenders are typically used. The most common types of blenders are In-Bin and V-shell blenders. In-Bin Blenders are typically used for high drug-load blends and are good for storage of said blend. V-shell blenders are used in intermediate drug-load blends. The main difference in these blenders is the geometry (Figure 2.20).

There are three mechanisms of particle mixing: convection, dispersion, and shearing [36]. In tumbling blenders, convective and dispersive mixing are dominant, unless intensifier bars or chopper blades are added to cause shear mixing. For example, within a V-shell blender, convective blending occurs within each shell side during tumbling, and dispersive mixing happens between shells.

Blending in a DC case consists of a pre-lubricant and a post-lubricant blend ahead of compression. Lubricants such as Sodium Starch Fumarate (SSF) and Magnesium Stearate (MgSt) are normally used.

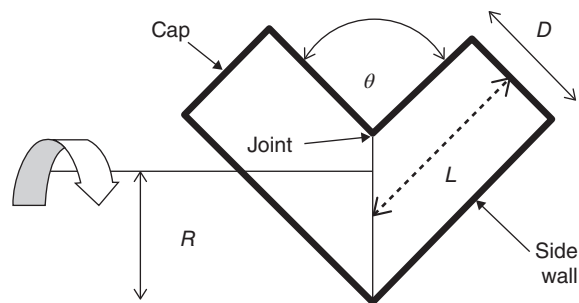


FIGURE 2.20 V-shell blender schematic.

Important parameters are:

CMA/ CPP	CPP/CQA/CBA It Can Affect
Particle shape and size	Blend uniformity, content uniformity, compressibility
Powder density	Blend uniformity, content uniformity, compressibility
Fill level	Blend uniformity, content uniformity
Loading procedure	Blend uniformity, content uniformity
Number of rotations (pre- and post-lube addition)	Blend uniformity, content uniformity
Rotation speed	Blend uniformity, content uniformity
Blender size	Throughput
Room humidity	Degradants, compressibility

There are many ways to determine if a blend is well mixed. Three simple ways are to

1. use online Process Analytical Technology (PAT) such as Near Infrared (NIR) technology
2. sampling blend using a thief over time and testing potency
3. simply compress the blend material and access compressibility, CU

The NIR region spans the wavelength range 780–2526 nm, in which absorption bands correspond mainly to overtones and combinations of fundamental vibrations. NIR spectroscopy is a fast and nondestructive technique that provides multi-constituent analysis of virtually any matrix. As NIR absorption bands are typically broad and overlapping, chemometric data processing is used to relate spectral information to sample properties.

The left-side graph in Figure 2.21 shows the second derivative of spectral data gathered of the API and the other components/excipients in the blend. Initial measurement is the purple curve and over time the spectral data line up over each other. The right-side graph shows when the API spectra to reach a less than 1%RSD distribution within the blend, which

is the blend endpoint. Commonly, 1 minute after less than 1% RSD for the API is accomplished is called as the blend endpoint but the engineer can see the asymptote of the line over time, sample number. Other determinations of blend endpoint are used as well, and method development is to be used for particular blends. This tool is useful as the engineer receives online data without sampling bias.

If PAT tools cannot be used, a more traditional sample method is used. Samples are commonly pulled from many locations (Figure 2.22) within the V-shell blender in order to understand if there is any location bias versus blend uniformity.

Usually blenders are scaled from V-shell (laboratory, pilot scale, and commercial scales); however, depending on the product and manufacturing needs the blending operation may be transferred to an In-Bin Blender (pilot, commercial scales). Change in geometric characteristics of tumbling blenders may lead to different mixing behaviors; therefore, a straightforward transition cannot be accomplished unless engineering principles are used. Some scale-up approaches are matching Froude (Fr) number, matching tangential/wall speed, or scaling particle surface velocities (Figure 2.23) [37–39].

2.10.2 Compression Scale-Up

Compression is important to make robust tablets. Tablets that are too soft cannot withstand the downstream coating or packaging processes without chipping or breaking and losing tablet weight/active component. Tablets that are too hard cannot dissolve effectively and therefore also cannot be efficacious when considering the TPP.

Compression is typically used to make solid oral dosage forms of core tablets. Many types of equipment are manufactured; some include single station, rotary presses. Typical manufacturers are Korsch, Elizabeth Hata, SMI, GEA Courtoy, and Manesty. Tablet presses are capable of using various-sized tooling such as A, B, and D.

Parameters which may be critical in tablet production are:

CPP	CQA It Affects
Incoming blend	Tablet weight (flowability), compressibility in general
Feeder speed	Tablet weight
Fill depth	Tablet weight
Press speed (Dwell time)	Appearance (defects via capping or lamination)
Pre-compression	Tablet hardness
Main compression	Tablet hardness
Upper punch entry	Tablet hardness
Room humidity	Compressibility in general, degradants, tablet water content
Press temperature over time	Tablet hardness, degradants, possible change in physical form

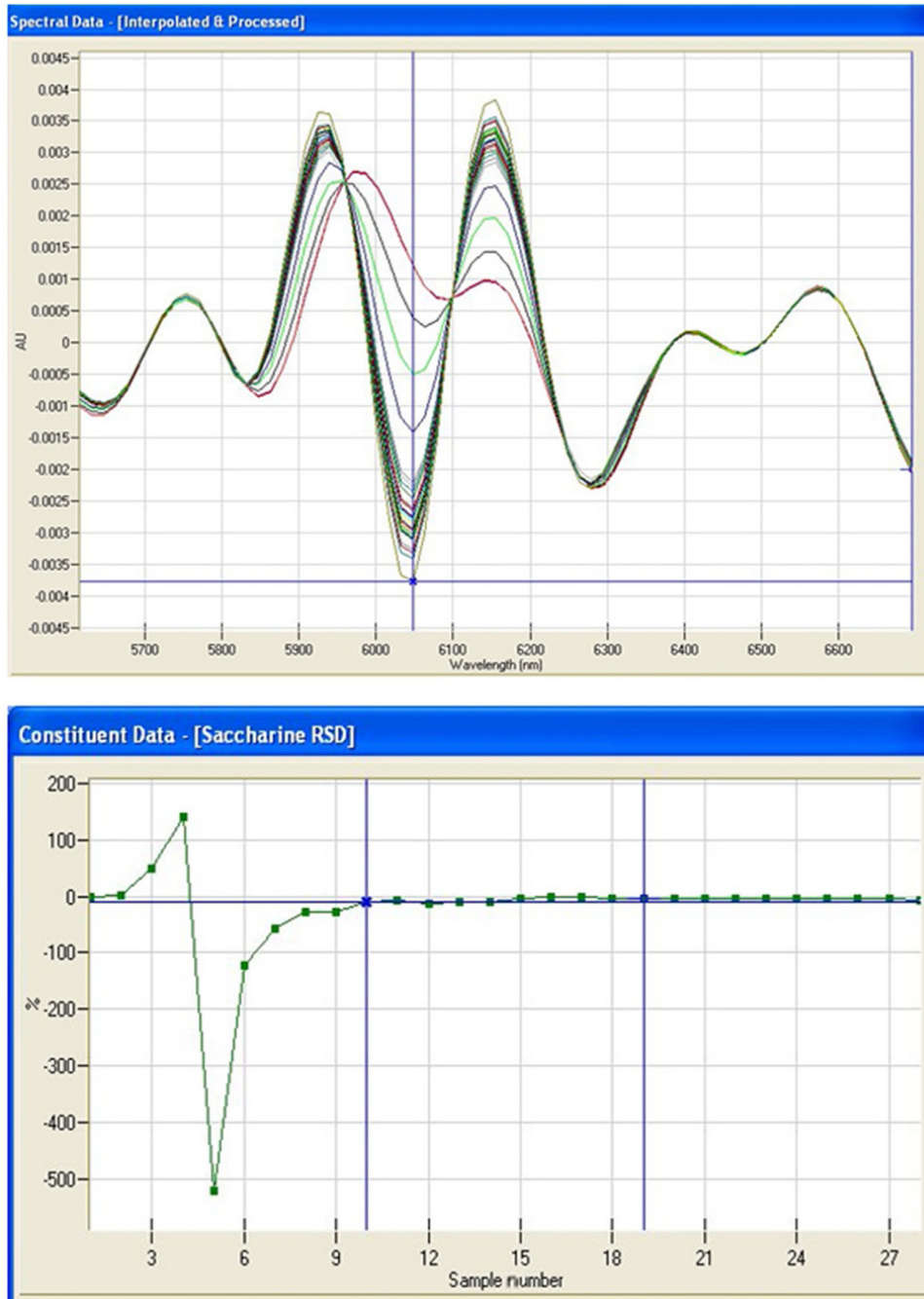


FIGURE 2.21 NIR spectral and constituent data for a blend containing saccharine as a model API. NIR technology from CDI pharma.

DOEs to access compressibility of the drug product are performed: Precompression Force, Main Compression Force, and Press Speed. Tests such as tablet weight, thickness, hardness, friability, and dissolution are performed to understand the processing effects on the CQAs. These data are used to determine processing targets, NORs and PARs (Figure 2.24).

Tablet dies and tooling may be the same from laboratory to pilot to commercial scale; the change is in tooling dwell time.

$$\text{Dwell time} = \frac{60\,000 \times \text{punch head flat diameter}}{\pi \times \text{pitch circle diameter} \times \text{press speed}} \quad (2.13)$$

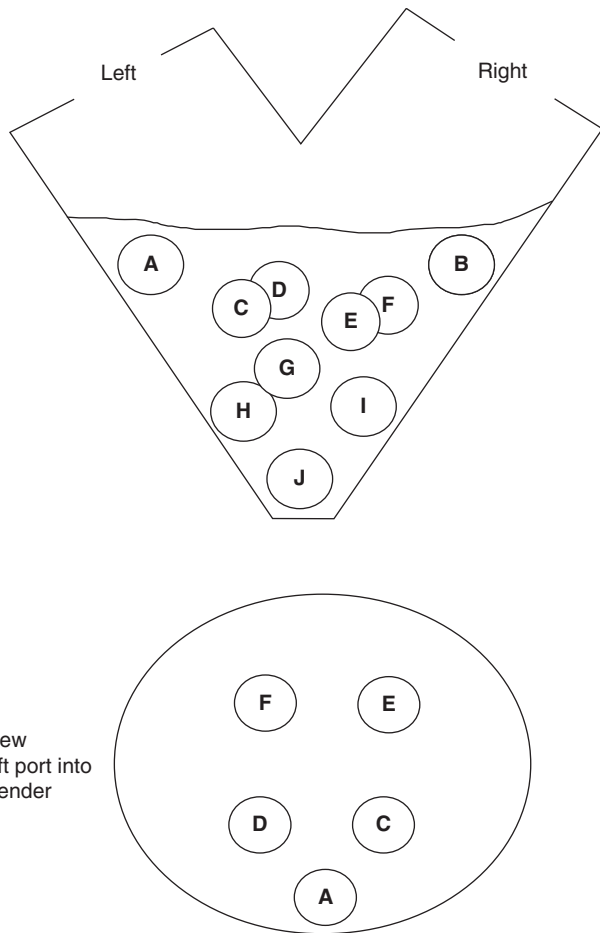


FIGURE 2.22 Example of sampling points within a V-shell blender.

Different dwell times can cause problems such as tablet capping or lamination. Tools such as compaction simulators could be used early on to save both time and money. Scaling up based upon mechanical similarity and quality attributes of

1. Matching of froude number (Fr), $Fr = [\Omega^2 R]/g$
2. Matching of tangential speed (wall speed) of blender, $2\pi\Omega R$
3. Scaling of particle surface velocities,

$$V = kR \Omega^{2/3} \left(\frac{g}{d} \right)^{1/6} \quad \text{for } \Omega \leq 30 \text{ rpm}$$

$$V = kR \Omega^{1/2} \left(\frac{g}{d} \right)^{1/4} \quad \text{for } \Omega > 30 \text{ rpm}$$

FIGURE 2.23 Common scale-up techniques for the process of blending.

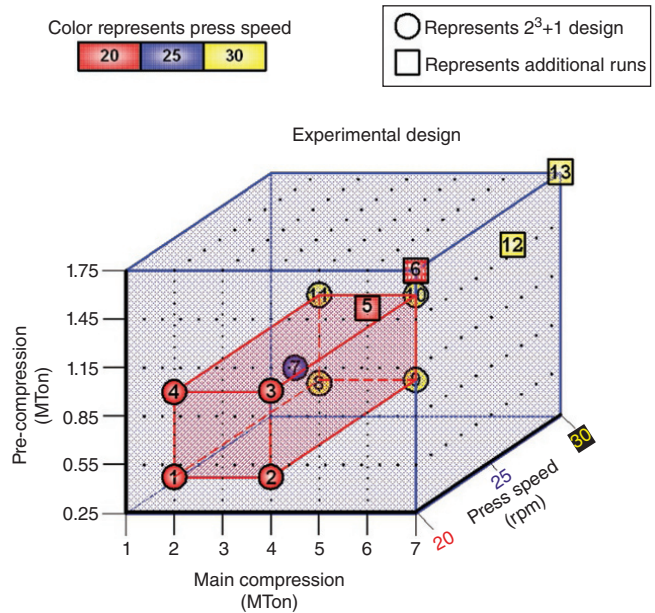
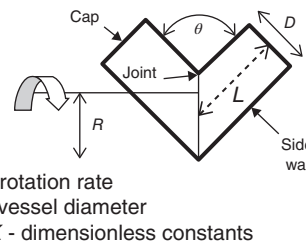


FIGURE 2.24 Fractional factorial experimental design for tableting.

the product is important, but sometime also is scaling down. Analytical techniques such as shear cells to understand powder flow and compaction simulators to understand compressibility behavior have been developed with the mindset of scaling down.

2.10.3 Coating Scale-Up

Tablet coating is the unit operation consisting of spray coating functional or nonfunctional/aesthetic coating onto the surface of the already compressed tablets. There are various sizes of tablet coaters, ranging up to 60" + coating pans.



Coating pans are either perforated or nonperforated. PAT tools may be implemented, for example, NIR for water content.

Important parameters are:

CPP	CQA It Affects	Potential Problems
Pan load	Appearance, tablet water content	Improper pan loading for the scale being used
Spray gun to bed distance	Appearance	Improper spray to tablet bed
Number of spray guns	Appearance	
Exhaust temperature	Degradants, tablet water content	Spray drying of coating suspension
Atomization air flow rate	Appearance, tablet water content	Improper spray
Pattern air flow rate	Appearance, tablet water content	Improper spray
Spray rate	Appearance, tablet water content	Improper spray
Spray formulation	Appearance, dissolution	May impede tablet dissolution
Weight gain	Appearance, dissolution	Too high may impede tablet dissolution, too low may not cover tablets/appearance
Pan speed	Appearance	Too high of pan speed
Jogging	Appearance	Too much or too little jogging of the tablet bed
Incoming tablets	Appearance, dissolution	Too soft tablets Too much disintegrant, especially on the surface of the tablets

Post-coated tablets are examined by Acceptable Quality Limit (AQL) for the number and type of defects (minor, major, or critical) [ANSI/ASQ Z1.4-2008]. Common reasons for defects stemming from film coating include:

- Improper EEF, thermodynamic conditions
- Incoming raw material including tablets
- Operating conditions (non-thermodynamic)

Thomas Engineering Inc. provides a Thermodynamic Analysis of Aqueous Coating (TAAC) model for coating scale-up which uses thermodynamic heat and mass transfer equations to characterize the environmental conditions inside a coating pan during a steady-state film coating process [40, 41].

The Environmental Equivalency Factor (EEF) is the most important piece of data output by the TAAC program. It is a dimensionless number proportional to the ratio of the dry area of the tablet bed to the wetted area and as such, is indicative of the drying rate of the film being applied. The dimensionless EEF lumps together all thermodynamic terms for ease of modeling or scaling up.

If there is a concern with water content increase, changing parameter values to increase the EEF helps; however, too high an EEF may cause unwanted spray drying of the coating suspension, leading to undesired tablet defects. A balance is usually found empirically. *A good rule of thumb*: use an EEF of 2–5, with 3.3 being a typical production value.

A spray-coating half-factorial design around the parameters of Spray Rate, Exhaust Temperature, Pan Speed, and Suspension Concentration is executed to better define the coating-processing DS with respect to tablet defects.

Table 2.20 shows the experiment and the number of defects seen in a sample size of 800.

Stepwise linear regression of the data yields the model shown in Figure 2.25, and Tables 2.21 and 2.22.

Spray Rate, Pan Speed, and Suspension Concentration were all seen to be significant parameters on the response of tablet defects. Exhaust Temperature is removed from the model as it is insignificant in the range studied. Two interaction terms were also found to be significant; Spray rate × pan speed and pan speed × suspension concentration. The Interaction Profiles are seen in Figure 2.26. The engineer optimizes the process by using the parameters of strongest leverage. For example, higher pan speeds were shown to have fewer defects.

TABLE 2.20 Spray-Coating Design and Number of Defects Observed

Experiment	Spray Rate (g/min)	Exhaust Temperature (C)	Pan Speed (rpm)	Suspension Concentration (wt%)	Defects
1	350	50	5	18	4
2	350	50	10	22	1
3	350	60	5	22	7
4	350	60	10	18	0
5	450	50	5	22	12
6	450	50	10	18	2
7	450	60	5	18	9
8	450	60	10	22	2
9	300	55	7.5	20	1
10	300	55	7.5	20	0

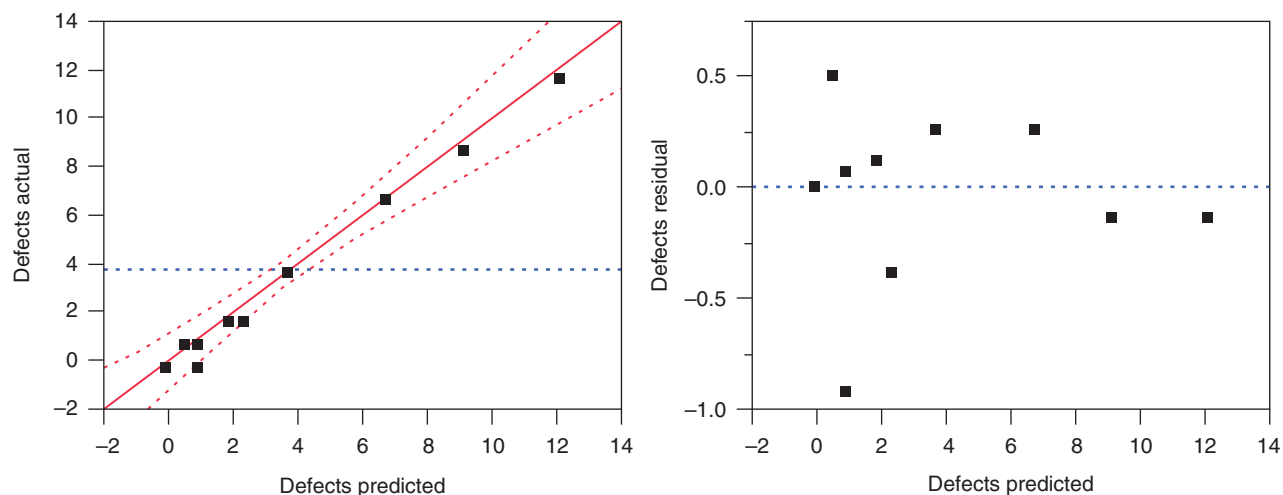


FIGURE 2.25 Defects actual versus predicted and defects residuals versus predicted.

TABLE 2.21 Summary of Fit for Defect Model

Summary of Fit	
R^2	0.990 806
R^2 Adj	0.979 314
Root mean square error	0.598 029
Mean of response	3.8
Observations (or sum wghts)	10

2.11 CONCLUDING REMARKS

This chapter demonstrated an approach to DP development. There are many different approaches; a good engineer examines the best approach for the situation. Defined in the chapter was a systematic manner in which to develop an immediate release tablet. First, account for the physical characteristics of the drug substance: particles characteristics, solubility, BCS classification, and stability. Next, the release and stability characteristics of the tablet become important. Finally, determine the processability and scalability of the tablet.

Overall, a wise selection of excipients and processes relies on a sound understanding of the physical, chemical, and mechanical properties of the drug and excipients. A formulation may be successfully scaled-up and consistently meet performance and manufacturing requirements only when one fully understands the complex relationship between the drug, excipients, processing, and the desired dosage-form performance criteria.

When formulating any pharmaceutical dosage form, it is important to remember that there is equilibrium between the bioavailability of the product, its chemical and physical stability, and the technical feasibility of producing it. Any changes made to a formulation in an attempt to optimize one of these properties are likely to have an effect on the other two parameters that must be considered. This is especially true of immediate-release solid dosage forms. Many of the properties require optimizing the bioavailability through rapid disintegration and dissolution of the active constituent, e.g. small particle size must be balanced with the manufacturability, where the fluidity and compactibility of a powder will often be enhanced by an increase in particle size.

TABLE 2.22 Parameter Estimates for Defects Model

Term	Parameter Estimates			
	Estimate	Std Error	<i>t</i> Ratio	Prob > <i>t</i>
Intercept	-9.702 778	2.531 304	-3.83	0.0186*
Spray rate (g/min)	0.036 388 9	0.003 152	11.55	0.0003*
Pan speed (rpm)	-1.21	0.091 089	-13.28	0.0002*
Suspension concentration (wt%)	0.4375	0.105 718	4.14	0.0144*
(Spray rate (g/min) - 380) × (pan speed (rpm) - 7.5)	-0.007	0.001 691	-4.14	0.0144*
(Pan speed (rpm) - 7.5) × (suspension concentration (wt%) - 20)	-0.125	0.042 287	-2.96	0.0417*

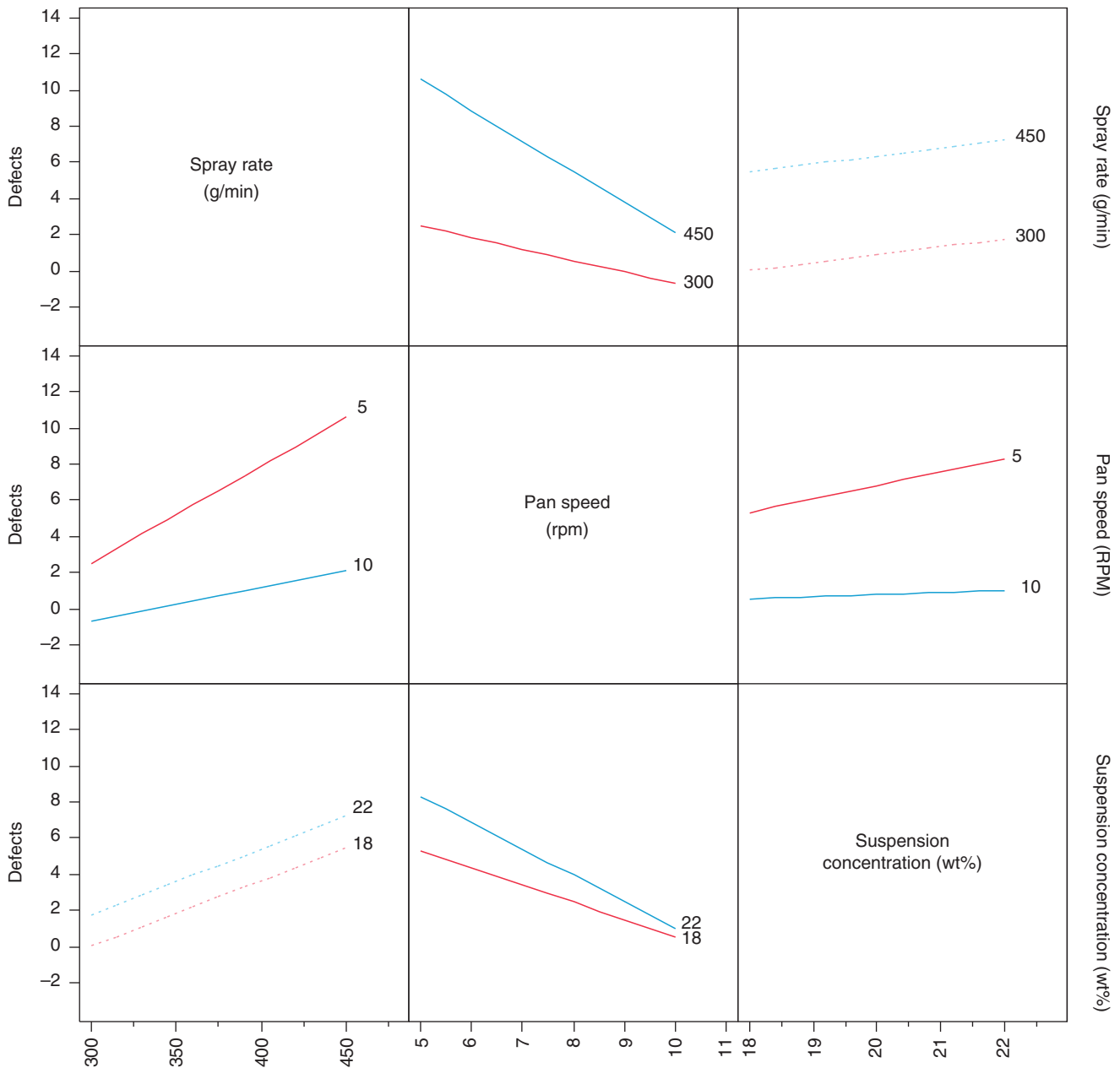


FIGURE 2.26 Interaction profiler for the tablet defects model.

REFERENCES

- Martin, A. (1993). *Physical Pharmacy*, 4e, 223. Philadelphia, PA: Lea & Febiger.
- Niklasson, M., Brodin, A., and Sundelöf, L.-O. (1985). Studies of some characteristics of molecular dissolution kinetics from rotating discs. *International Journal of Pharmaceutics* 23: 97–108.
- Koehler, M.G., Grigoras, S., and Dunn, W.J. III (1988). The relationship between chemical structure and the logarithm of the partition coefficient. *Quantitative Structure-Activity Relationships* 7: 150–159.
- Leo, A., Hansch, C., and Elkins, D. (1971). Partition coefficients and their uses. *Chemical Reviews* 71: 525–616.
- Dearden, J.C. and Bresnen, G.M. (1988). The measurement of partition coefficients. *Quantitative Structure-Activity Relationships* 7: 133–144.
- Navia, M.A. and Chaturvedi, P.R. (1996). Design principles for orally bioavailable drugs. *DDT* 1: 179–189.
- Byrn, S.R., Pfeiffer, R.R., and Stowell, J.G. (1999). *Solid State Chemistry of Drugs*, 2e. West Lafayette, IA: SSCI.
- Grant, D.J.W. (1999). Theory and origin of polymorphism. In: *Polymorphism in Pharmaceutical Solids* (ed. H.G. Britain), 1–34. New York: Marcel Dekker.

9. Aguiar, A.J., Krc, J., Kinkel, A.W., and Samyn, J.C. (1967). Effect of polymorphism on the absorption of chloramphenicol from chloramphenicol plamate. *Journal of Pharmaceutical Sciences* 56 (7): 847–883.
10. Poole, J.W., Owen, G., Silverio, J. et al. (1969). Physiochemical factors influencing the absorption of the anhydrous and trihydrate forms of ampicillin. *Current Therapeutic Research, Clinical and Experimental* 10: 292–303.
11. Matsumoto, T., Nobuyoshi, K., Iliguchi, S., and Otsuka, M.J. (1991). Effect of temperature and pressure during compression on polymorphic transformation and crushing strength of chlorpropamide tablets. *Pharmacy and Pharmacology* 43: 74.
12. Zhang, G.Z., Law, D., Schmitt, E.A., and Qiu, Y. (2004). Phase transformation considerations during process development and manufacture of solid oral dosage forms. *Advanced Drug Delivery Reviews* 56 (3): 371–390.
13. Venable, H.J. and Wells, J.J. (2001). Powder mixing. *Drug Development and Industrial Pharmacy* 27: 599–612.
14. Sun, C. and Grant, P.J.W. (2001). Influence of crystal shape on the tableting performance of L-lysine monohydrochloride dihydrate. *Journal of Pharmaceutical Sciences* 90: 569–579.
15. Kaerger, J.S., Edge, S., and Price, R. (2004). Influence of particle size and shape on flowability and compactibility of binary mixture of paracetamol and microcrystalline cellulose. *European Journal of Pharmaceutical Sciences* 22 (2–3): 113–119.
16. US Pharmacopeia USP29/NF24. (2006). US Pharmacopeial Convention, Rockville, MD, General Chapter <905>, pp. 2778–2784.
17. Zhang, Y. and Johnson, K.C. (1997). Effect of drug particle size on content uniformity of low-dose solid dosage forms. *International Journal of Pharmaceutics* 154 (2): 179–183.
18. Johansson, D. and Abrahamsson, B. (1997). In vivo evaluation of two different dissolution enhancement principles for a sparingly soluble drug administered as extended-release (ER) tablet. Proceedings of 24th International Symposium on Controlled Release of Bioactive Materials, pp. 363–364.
19. Amidon, G.L., Lennernäs, H., Shah, V.P., and Crison, J.R. (1995). A theoretical basis for a biopharmaceutical drug classification: the correlation of in vitro drug product dissolution and in vivo bioavailability. *Pharmaceutical Research* 12: 413–420.
20. Dressman, J.B., Amidon, G.L., Reppas, C., and Shah, V.P. (1998). Dissolving testing as a prognostic tool for oral drug absorption: immediate release dosage forms. *Pharmaceutical Research* 15: 11–22.
21. Food and Drug Administration guidance for industry (2000). *Waiver of in vivo Bioavailability and Bioequivalence Studies for Immediate-Release Solid Oral Dosage Forms Based on a Biopharmaceutical Classification System*. Rockville, MD: Food and Drug Administration, Center for Drug Evaluation and Research.
22. Fell, J.T. and Newton, J.M. (1970). Determination of tablet strength by the diametral-compression test. *Journal of Pharmaceutical Sciences* 59 (5): 688–691.
23. Hiestand, E., Wells, J.E., Peot, C.B., and Ochs, J.F. (1977). Physical processes of tableting. *Journal of Pharmaceutical Sciences* 66 (4): 510–519.
24. Hiestand, E.N., Bane, J.M., and Strzelinski, E.P. (1971). Impact test for hardness of compressed powder compacts. *Journal of Pharmaceutical Sciences* 60 (5): 758–763.
25. Ho, R., Bagster, D.F., and Crooks, M.J. (1977). Flow studies on directly compressible tablet vehicles. *Drug Development and Industrial Pharmacy* 3: 475.
26. Seager, H., Rue, P.J., Burt, I. et al. (1985). Choice of method for the manufacture of tablets suitable for film coating. *International Journal of Pharmaceutical Technology & Product Manufacture* 6: 1–20.
27. Parikh, D. (ed.) (2005). *Handbook of Pharmaceutical Granulation Technology*, 2e. New York: Marcel Dekker.
28. (August 1997). *Guidance for Industry Dissolution Testing of Immediate Release Solid Oral Dosage Forms*. U.S. Department of Health and Human Services Food and Drug Administration Center for Drug Evaluation and Research (CDER).
29. Ewe, K., Press, A.G., Bollen, S., and Schuhn, I. (1991). Gastric emptying of indigestible tablets in relation to composition and time of ingestion of meals studied by metal detector. *Digestive Diseases and Sciences* 36 (2): 146–152.
30. Monkhouse, D.C. and Rhodes, C.T. (eds.) (1998). *Drug Products for Clinical Trials; An International Guide to Formulation, Production, Quality Control*, Drugs and the Pharmaceutical Sciences, vol. 87.
31. Levin, M. (ed.) (2006). *Pharmaceutical Process Scale-Up*, Drugs and the Pharmaceutical Sciences, 2e, vol. 157.
32. Process Robustness – A PQRI White Paper. By PQRI Workgroup Members. November/December 2006. Pharmaceutical Engineering On-Line Exclusive
33. Marder, R. and Sheff, R.A. (2002). *The Step-by-Step Guide to Failure Modes and Effects Analysis*. Marblehead, MA: Opus Communications.
34. Stamatis, D.H. (1995). *Failure Mode and Effect Analysis; FMEA from Theory to Execution*. Milwaukee, WI: ASQC Quality Press.
35. Parikh, D.M. (ed.) (1997). *Handbook of Pharmaceutical Granulation Technology*, Drugs and the Pharmaceutical Sciences, vol. 81.
36. Lacey, P.M. (1954). Developments in the theory of particle mixing. *Journal of Applied Chemistry* 4: 257–268.
37. Swarbrick, J. (2006). *Encyclopedia of Pharmaceutical Technology*, 3e, vol. 5. Taylor & Francis.
38. Alexander, A., Shinbrot, T., and Muzzio, F.J. (2002). Scaling surface velocities in rotating cylinders as a function of vessel radius, rotation rate, and particle size. *Powder Technology* 126: 174–190.
39. Alexander, A.W. and Muzzio, F.J. (2001). *Batch Size Increase in Dry Blending and Mixing*. Pharmaceutical Process Scale-up, 2e. New York: Marcel Dekker.
40. Strong, J.C. (2009). Psychrometric analysis of the environmental equivalency factor for aqueous tablet coating. *AAPS PharmSciTech* 10 (1).
41. Ebey, G.C. (1987). A thermodynamic model for aqueous film-coating. *Pharmaceutical Technology* 11 (4): 40–50.

3

POWDER PROCESS CHALLENGES AND SOLUTIONS

THOMAS BAXTER AND JAMES PRESCOTT

Jenike & Johanson, Inc., Tyngsboro, MA, USA

3.1 INTRODUCTION

Many pharmaceutical processes include powder handling, such as blending, transfer, storage, feeding, compaction, and fluidization. Understanding powder-flow behavior is essential when developing, optimizing, or scaling-up a process, including designing new equipment or developing corrective actions for existing equipment for a batch or continuous manufacturing process. The robustness of a process is often adversely affected by flow or segregation problems that develop. Powder flow or segregation problems can have an adverse effect upon:

- Production costs, due to reduced production rates (e.g. tableting rate limitations and required operator intervention), restrictions on raw ingredient selection (e.g. percentage of lubrication used and excipients selected), method of manufacturing (wet granulation versus dry granulation versus direct compression), equipment selection (type of blender, bin, and press), and overall yield.
- Product quality, due to variation of tablet/capsule properties (weight, hardness, potency, and weight variability) or segregation and content uniformity concerns.
- Time to market, due to delays in product/process development and failed validation or commercial batches, since flow problems may not occur until the process has been scaled-up.

This chapter provides guidance in designing bulk solid (“powder”)-handling equipment to provide consistent,

reliable flow and the required product uniformity by minimizing segregation. The principles discussed can be applied to analyzing new or existing equipment designs, as well as comparing different powders using the various test methods discussed.

The chapter will focus on the equipment used from the final-blend step to the inlet of the press or encapsulation machine used to create the unit dose. Throughout this chapter we will focus on feed systems to tablet presses, but the design methods also apply to encapsulation feed systems, particularly the designs of intermediate bins, transfer chutes, and feed hoppers. This chapter is divided into the following primary topics:

- A review of introductory powder-handling concepts, such as flowability, blending, and segregation;
- Common powder-handling equipment and design parameters;
- Powder flow and segregation concerns, as well as the two primary flow patterns;
- Measurement of powder-flow properties and segregation potential properties;
- Trouble-shooting powder-handling problems and equipment design methods.

At the end of this chapter, you should have a working knowledge of what flow properties need to be measured and how to measure them, as well as applying the measured parameters to analyze and design handling equipment for reliable flow.

3.1.1 Introduction to Powder Flow

A bulk solid is defined as a collection of solid particles. The term “powder” is often used to describe a fine bulk solid, especially in the pharmaceutical industry. The concepts discussed in this chapter apply to many types of powders with different particle sizes, shapes, and distributions. The powders may include dust, granulations, and granules. The powder could either be a single substance, such as an excipient or active pharmaceutical ingredient (API) or a multicomponent blend of particles and/or granules (final blend). As the pharmaceutical industry begins to adopt more continuous manufacturing of oral solid dosages, reliable feed of the individual blend components, especially the API, will be more critical. The principles outlined in this chapter can be used to design for all these different types of powders.

A simple definition of “flowability” is the ability of a powder to flow through equipment reliably. By this definition, there is often a tendency to define flowability as a single parameter of a powder, ranked on a scale from “free-flowing” to “non-flowing.” Unfortunately, a single parameter is not sufficient to define a powder’s complete handling characteristics. In addition, a single parameter is not sufficient to fully address common handling concerns encountered by the formulator and equipment designer. In fact, several design parameters may need to be known for a successful design. The behavior of a powder will depend upon several different parameters or “flow properties.” Flow properties are the specific properties of a powder that affect flow that can be measured. Therefore, a full range of flow properties will need to be measured to fully characterize the powder. The measurement of these flow properties is discussed in Section 3.4.

Powder “flowability” is also a function of the design parameters of the handling equipment. For example, “poor-flowing” powders can be handled reliably in properly designed equipment. Conversely, “good-flowing” powders may develop flow problems in improperly designed equipment. Therefore, our definition of “flowability” is “the ability of powder to flow in the desired manner in a specific piece of equipment.”

The flow properties of the powder should be quantitative and scalable design parameters. The term “flow properties” often refers to the physical characteristics of the powder that were measured. In discussing or reporting flowability, both the powder-flow properties and the handling equipment must be included. Therefore, the measurement of the powder-flow properties can be used to predict behavior in specific equipment during scale-up.

3.1.2 Introduction to Blending

Powder-blending processes are used during the manufacture of products for a wide range of industries, including pharmaceuticals. In the pharmaceutical industry, a wide range of

ingredients may be blended together to create the final blend used to manufacture the solid dosage form. The range of materials that may be blended (excipients and API), presenting a number of variables that must be addressed to achieve products of acceptable blend uniformity. These variables may include the particle size distribution (PSD, including aggregates or lumps of material), particle shape (spheres, rods, cubes, plates, and irregular), presence of moisture (or other volatile compounds), particle surface properties (roughness and cohesivity), and many other variables.

The quality of the solid dosage form is dependent on the adequacy of the blend. Producing a uniform mixture of the drug and its excipients is paramount in being able to deliver the proper dose of the drug to the patient. Once an adequate blend is obtained, it is also critical to ensure that it does not segregate in the post-blending handling steps. Millions of dosage units may be created from a single batch, and each and every dose must be of acceptable composition, to ensure the safety and efficiency of the product. Therefore, the homogeneity of pharmaceutical blends and dosage units is highly scrutinized by both manufacturers and regulatory agencies throughout the world. Formulation components and process parameters involved with blending operations should be carefully selected, and validated, to ensure uniform blends and dosage units are produced. Blend and dosage unit uniformity data are provided in regulatory submissions, and often examined during preapproval inspections. This is to ensure that blending processes produce homogeneous blends that do not segregate upon further processing into dosage units. Finally, pharmacopeias require an assessment of content uniformity to be performed on every batch of solid dosage forms manufactured.

The scale of blending operations for the preparation of pharmaceutical dosage forms ranges from the extemporaneous compounding of a few capsules by pharmacists to large-scale production of batches containing millions of dosage units. The complexity of the blending process can vary substantially. Large-scale production batches often use equipment capable of blending hundreds of kilograms of material. Continuous blending operations require not only the mixing components within the blender itself, but also controlled feed of the incoming material, and monitoring, feedback, and control systems. Depending on the dose and characteristics of the drug substance, commercial-scale blending processes can be complex, and may require screening operations, the preparation of pre-blends, and/or the inclusion of milling operations to achieve acceptable content uniformity. Regardless of the scale of manufacture, the goal remains the same: to prepare a blend that is adequately blended, and can be further processed into dosage units that deliver the proper dose of the drug to the patient.

Blending should not be seen as an independent unit operation, but rather as an integral part of the overall manufacturing process. Blending includes producing an adequate blend,

maintaining that blend through additional handling steps, and verifying that both the blend and the finished product are sufficiently homogeneous. Therefore, a complete approach should be used to assess the uniformity of blends, and the subsequent dosage forms produced from them. A review of the common types of blenders is provided in Section 3.2, but the details of scaling-up common final blending-process steps (wet granulation, fluid-bed granulation, and roller compaction) are discussed in other chapters.

3.1.3 Introduction to Segregation

“Segregation” can be defined as particles of similar properties (size, composition, density, etc.) preferentially distributed into different zones within given equipment or processes. Segregation most notably affects the localized concentration of the drug substance. This can result in blend and content uniformity issues. In addition, the segregation of other components of the blend can be responsible for variations in properties such as dissolution, stability, lubrication, taste, appearance, and color. Even if the blend remains chemically homogeneous, variations in particle size can affect flowability, bulk density, weight uniformity, tablet hardness, appearance, and dissolution. Additionally, segregation can create concentrations of dust, which can lead to problems with agglomeration, yield, operator exposure, containment, cleanliness, and increased potential for a dust explosion. Segregation can occur in a batch or continuous manufacturing process.

Segregation can occur any time there is powder transfer, such as discharging the final blend from the blender into a bin. Segregation can also occur when force mechanisms acting on the particles in the blend, such as air flow or vibration, are sufficient to induce particle movement. This may include handling steps upstream of a blender, including segregation of raw materials at a supplier’s plant or during shipment, movement within the blender, during its discharge, or in downstream equipment. Of all of these potential instances where segregation can occur, the most common area for problems is post-blender discharge. Therefore, this chapter will focus on segregation of the final blend in the post-blending-handling steps.

Addressing segregation concerns can utilize empirical descriptors of segregation mechanisms (Section 3.4) and best practices to diagnose and address specific segregation behavior (Section 3.5). Unlike the flow property tests used to assess how a powder flows through equipment, there are no current “first principle” models that adequately describe the various segregation mechanisms that may occur in a process. Therefore, one cannot currently input the particle properties of the blend components (e.g. particle size and chemical composition of the excipients, and API) into a mathematical model, and obtain a prediction of segregation potential and the resulting content uniformity of the product. However,

computational models such as discrete element method (DEM) coupled with computational fluid dynamics (CFD) continue to evolve to simulate particle and fluid flow. These can be tuned to match specific segregation behaviors that are created in physical models to gain further insight into blending and segregation behavior. As these models evolve, they will become more powerful, and have fewer assumptions and limitations. Therefore, when assessing segregation concerns, it is critical to utilize as many resources as possible such as laboratory-scale tests to assess different segregation mechanisms (Section 3.4), as well as stratified blend and content uniformity data to assess potential segregation trends. The empirical tests that can be conducted to assess the segregation potential of a final blend (Section 3.4) and how to apply the results to a reliable equipment design (Section 3.5) are discussed later in this chapter.

3.2 COMMON POWDER-HANDLING EQUIPMENT

Numerous powder-process steps are commonly used in a batch process to produce the final blend used to manufacture the dosage units, including:

- processing steps prior to final blending, such as milling, screening, drying, and granulation;
- final blending and discharge from the final blender;
- intermediate bulk containers (“IBCs,” “totes,” and bins);
- transfer from the IBC to the press (or encapsulator);
- feed from the press feed hopper to the die cavity or encapsulation feed hopper to the dosing chamber.

As continuous manufacturing in the pharmaceutical industry continues to become more common, different processing equipment may become more crucial, such as reliable feed and metering of the API, but the equipment terms outlined in this section will still often apply to understanding how powders flow. The propensity for segregation due to long transfer chutes will likely be significantly reduced for continuous drug product processing.

For each of these different process steps, there are different key equipment parameters that affect the flowability and segregation of a powder. These typical handling steps serve as examples of the concerns with powder handling, but virtually any solid-handling application can be analyzed in the same way.

3.2.1 Processing Steps Prior to Final Blending

It is critical to understand the physical properties of the raw ingredients (API and excipients) and how they affect

flowability and segregation of the final blend, when selecting and designing the powder-handling equipment. A full understanding of the properties of the blend components is especially critical in a continuous manufacturing process. There are several common pre-blending process steps that may affect the final blend flow and segregation, including:

- Storage conditions of the raw ingredients, such as the temperature, relative humidity, container dimensions, and days stored at rest (inventory control), can all influence the flowability of the final blend. This is especially true if any of the raw ingredients are hygroscopic.
- Milling and screening steps that alter the raw ingredients', and thus the final blend's particle size, shape, and distribution. Therefore, milling and screening process parameters, such as the mill type, mill speed, screen size, and mill/screen feed method (controlled versus noncontrolled feed), may all have an influence on the flowability and segregation potential of the final blend, especially in a direct blending process.
- Pre-blending of selected raw materials, such as pre-blending a cohesive powder with a less cohesive powder to reduce the likelihood of flow problems during subsequent handling steps. Pre-blending may also be conducted to achieve a more uniform blend and reduce the segregation potential.
- Granulation (dry roller compaction, wet granulation, and fluid-bed granulation) of the API, together with select excipients, can often have a positive effect on the flowability of the final blend, especially for blends with highly active drug loadings. The granulation parameters, especially those that influence particle size/shape/distribution, will have a significant effect on flowability and segregation potential. For roller compaction, the process parameters that dictate the PSD and shape distribution of the final blend may include the roller compactor speed, roll compactor pressure, mill type, and the screen size. The wet granulation process parameters that affect PSD and shape, as well as the moisture content, are often critical to flowability. Therefore, wet granulation parameters, such as the blade and impeller design/speed, binder addition rate, and method and identification of granulation end point, are critical to flowability, as are the granule milling conditions. Similarly, the fluid-bed granulation parameters that affect the moisture content and particle size, such as the binder addition rate/method, inlet air flow rate and temperature, drying time, end-point determination (target moisture, powder temperature, and exhaust air temperature), and fluidization behavior for the powder bed, are all critical to flowability and segregation. The details of scaling-up different granulation processes are discussed in other chapters.

The measurement of flow properties and segregation potential, and the design parameters they provide are discussed in Section 3.4, but they can also be applied to trouble-shooting and developing corrective actions for flow problems in the pre-blending steps if needed.

3.2.2 Final Blending and Discharge from Final Blender

Final blending may be accomplished on a batch or continuous basis. In the pharmaceutical industry, batch blenders are most common, but continuous manufacturing of oral solid dosage forms has progressed significantly over the last decade. Therefore, this chapter focuses on different batch-blending processes, but continuous manufacturing is also used to produce dosage units and is discussed in another chapter. Batch-blending processes consist of three sequential steps: weighing and loading the components, blending, and discharging. Batch blenders come in many different designs and sizes, and make use of a wide range of blending mechanisms. All blenders use one or more of the following mechanisms to induce powder blending: convective blending, shear blending, and diffusive blending [1, 2].

Another classification system for blenders is based on their design. This system categorizes blenders into two categories: (i) blending within a moving vessel and (ii) fixed vessels that rely on internal paddles or blades to move the materials. There are many common blenders used in the pharmaceutical industry on a batch basis that are discussed in other chapters, including wet granulators, fluid-bed granulators, roller compactors, and direct blending via tumble blenders. Many of these blending processes, including high-shear wet granulation, fluid-bed granulation, roller compaction, and direct blending in mixer can also be conducted in a continuous manufacturing process.

As a result of the multiple classification systems, a number of terms have evolved throughout the industry to describe families of blenders. Regardless of the terminology used to classify the blender, the important thing is for the pharmaceutical scientist to understand the capabilities and limitations of the equipment when selecting an appropriate blender for a particular product. This is especially important during process scale-up, when equipment of different design and operating principle may need to be used. The optimization and scale-up of the wet granulation, fluid-bed granulation, and roller compaction processes are discussed in other chapters. The optimization and scale-up of tumble blenders are discussed elsewhere in the literature [3].

Note that some types of tumble blenders, such as bin blenders, serve a dual purpose. In addition to providing the container in which blending is accomplished, bin blenders can also be used to transfer the powder blend to the next unit operation in a batch process. This is of particular value when manufacturing blends that have the tendency to segregate

when discharging the blend onto the compression or filling equipment. This also makes the use of bin blenders desirable during the manufacture of potent drug products that must be processed in high containment facilities. Additionally, by decoupling the blending bin from the drive mechanism, the bin filling, discharge, and cleaning take place at a separate time and location, which increases the efficiency and utilization of equipment.

Powder that has been blended must be discharged from the blender for further processing, including the creation of the unit dose at the press or encapsulator. In many dry blending processes, such as tumble blending, the discharge is driven by gravity alone. As an example, the final blend step may be conducted in a V-blender or a double-cone blender. In these cases, the blender geometry often consists of a converging cross-section to the outlet, through which the powder must be discharged reliably. In these cases, the blender is essentially acting as a “bin,” so the equipment parameters of interest are those that are crucial to a bin design. These crucial bin design parameters are discussed in the following section on “Intermediate Bulk Containers,” which may also be referred to as “IBC.”

For fluid-bed granulation processes, it is not uncommon for a conical “hopper” to be attached to the “bowl” of a fluid-bed granulator, inverted, and discharged to a downstream process step via gravity. For wet granulators, the final

blend may be discharged using mechanical agitation by continuing to operate the plow blade (typically at a lower speed) to discharge the final blend through a central or side outlet. Although the plow blade typically ensures that the blend is discharged from the granulator “bowl” reliably, the design of the transition chute from the blender to the downstream process is also critical. This is especially critical if the equipment below the granulator has a converging cross-section that is full of material, as discussed further in Section 3.5.

The transfer of the material from the final blender to the downstream equipment, whether it is into an IBC or directly to the press, is critical to the segregation potential. Therefore, parameters such as the transfer method (e.g. manual, gravity, and pneumatic), transfer rate, transfer chute height and geometry, venting, and other items will be critical in assessing and minimizing the segregation potential during the blender-to-press transfer steps. Design techniques to minimize segregation during these transfer steps are further discussed in Section 3.5.

3.2.3 Intermediate Bulk Containers

The flowability and segregation potential of the final blend are especially critical during storage and discharge from an IBC (Figure 3.1). The IBC may be a bin (“tote”) or a drum

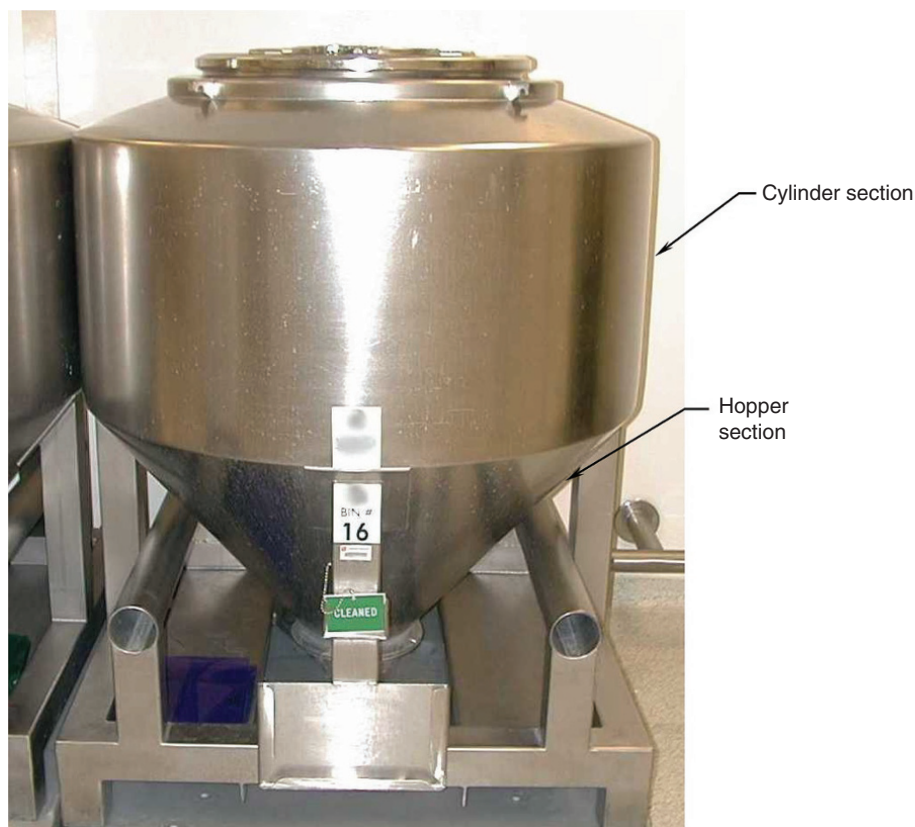


FIGURE 3.1 Example of a stainless steel intermediate bulk container (i.e. “IBC,” “bin,” and “tote”) commonly used in the pharmaceutical industry.

that is used to store and transfer the final blend from the blender to the press. When a drum is used, an attachment such as a conical hopper may be attached to the cone to mate the drum with downstream equipment with a smaller inlet (e.g. press hopper). This section focuses on the use of IBCs to store the blend in a batch process, but the concepts can also be applied and are critical to the storage of blend components in a continuous process.

The IBC consists of two primary sections:

- A cylinder or straight-sided section with a constant cross-sectional area that is often rectangular (with or without radiused corners) or circular;
- A hopper section with a changing cross-sectional area that is often a converging conical or pyramidal hopper.

IBCs may be used to store the blend for extended periods of time. During this time, the flowability may worsen as the blend is subjected to consolidation pressures, due to its own weight during storage at rest. In addition, IBCs may be used to move the blend from one process step to another, during which time the blend may be subjected to vibration that may adversely affect flowability. Therefore, it is important to determine what consolidation pressures will act on the powder as it is stored and transferred in an IBC.

Key IBC equipment parameters with respect to flowability and segregation potential include:

- Cylinder cross-sectional area and height, which affect the consolidation pressure on the blend.
- Hopper geometry (planar versus circular) and angles, which affect the flow pattern that develops during discharge. The flow pattern, discussed in Section 3.3, will also affect the segregation potential.
- Interior surface finish of the hopper section, which will affect the flow pattern that develops during discharge.
- IBC outlet size and shape (slotted versus circular), which will affect whether the blend will discharge reliably without arching or ratholing.
- General flow impediments, such as upward facing ledges or partially opened valves, that may act as flow obstructions.

Measurement of the flow properties and segregation potential that are used to obtain the key bin design parameters are discussed in Section 3.4. The application of these design parameters to provide reliable flow and minimize segregation from the IBC to the press or encapsulator is discussed in Section 3.5.

3.2.4 Transfer from Intermediate Bulk Containers to the Press/Encapsulator

The flowability and segregation potential of the final blend is also critical during transfer from the IBC to the press/

encapsulation machine. This transfer step may be a manual transfer (hand-scooping), in which case flowability may not be a primary concern. The transfer step may also be conducted via pneumatic conveying, in which case the flowability of the blend may not be a primary concern, but equipment and material parameters affecting conveying (conveying gas pressure and flow rate, conveying line diameter and layout, etc.) need to be considered. Pneumatically conveying the final blend may also raise segregation concerns, as further discussed in Section 3.3.

The transfer step may also be conducted via gravity transfer via a single or bifurcated chute, depending on the press configuration. Transfer chutes are often operated in a “flood-loaded” manner (i.e. full cross-section). Transfer chutes may consist of converging sections where the cross-sectional area of the chute is reduced, in which case the chute will need to be designed for reliable flow in a similar manner to the IBCs.

Key transfer chute parameters with respect to flowability and segregation potential include:

- Chute cross-sectional area and height, which will affect the consolidation pressure on the blend, and how it may segregate.
- Valving and venting of the transfer chute may affect how readily a blend segregates.
- For converging and non-converging sections of the chute, the chute geometry, angles, and interior surface finish will affect the flow pattern that develops during discharge through the chute.
- For converging sections of the chute, the outlet shape and size will affect whether the blend will discharge reliably without arching or ratholing.
- General flow impediments, such as upward-facing ledges (mismatched flanges), sight glasses, level probes, or partially opened valves, may act as flow obstructions.

The measurement of the flow properties and segregation potential that are used to obtain the key chute design parameters is discussed in Section 3.4. The application of these design parameters for equipment from the IBC to the press is discussed in Section 3.5.

3.2.5 Feed from the Press/Encapsulator Hopper to the Die Cavity

The press and encapsulation feed hopper must be designed to provide reliable flow and minimize segregation. Most modern presses/encapsulators consist of a small press hopper that is, in essence, a miniature IBC designed to provide a small amount of surge capacity. The press/encapsulation feed hopper often consists of a cylinder section and a hopper

section similar to a larger IBC. The hopper section may be asymmetric, as opposed to the symmetric hopper designs commonly used for larger IBCs. The press/encapsulation feed hopper is typically flood-loaded from the IBC/chute above, via gravity feed. However, in some instances, the material level in the press hopper may be controlled via a feeder at the IBC outlet (e.g. rotary valve or screw feeder) and/or use agitation to promote flow. Many encapsulation feed hoppers use internal agitators and/or feeders to facilitate feed to the dosating chamber. Some modern presses do not have press hoppers with a converging hopper, but instead consist of vertical, non-converging chutes from the press inlet to the feed frame inlet.

The key equipment parameters with respect to flowability, which are outlined in the preceding section for IBCs/bins, are also applicable to the press hopper. Since the press hopper outlets are often much smaller than an IBC outlet, flow problems such as arching or ratholing (discussed further in Section 3.3) may be more pronounced at this location. The same design parameters used for a reliable IBC design (see Section 3.4) can also be used to design a press hopper (see Section 3.5).

3.3 POWDER FLOW AND SEGREGATION CONCERNS

This section will focus on typical flow and segregation concerns that occur during transfer operations from the final blender to the press. This section will also provide a summary of common flow problems, segregation mechanisms, and the flow patterns that can occur during gravity discharge.

3.3.1 Common Flow Problems

A number of problems can develop as powder flows through equipment such as bins, chutes, and press hoppers. If the powder is cohesive, an arch or rathole may form, resulting in “no flow” or erratic flow. In addition, flooding or uncontrolled discharge may occur if a rathole spontaneously collapses. A deaerated bed of fine powder may experience flow-rate limitations or no-flow conditions due to the two-phase flow effects between the powder and the interstitial air. These flow problems are primarily reviewed with respect to batch processes, but also commonly apply to continuous processes. Each of these flow problems is discussed in more detail below.

“No flow”: no flow from a bin/hopper is a common and significant solid-handling problem, especially when handling fine/milled APIs and highly active loading blends. In production, it can result in problems, such as starving downstream equipment, production delays, and the requirement for frequent operator intervention to reinitiate flow. No flow can be due to either arching (sometimes referred to as “bridging”

or “plugging”) or ratholing (also referred to as “piping”), as described below:

- Arching: occurs when an obstruction in the shape of an arch or bridge forms above the bin outlet and prevents any further material discharge. It can be an interlocking arch, where the particles mechanically lock to form the obstruction, although this is less common with fine pharmaceutical powders. An interlocking arch occurs when the particles are large compared to the outlet size of the hopper. The arch could also be a cohesive arch where the particles pack together to form an obstruction (Figure 3.2a). Both of these problems are strongly influenced by the outlet size of the hopper the material is being fed through. Powder-flow properties, discussed in Section 3.4, can be used to determine if these problems will occur and used to address them during scale-up. In particular, the cohesive strength of a

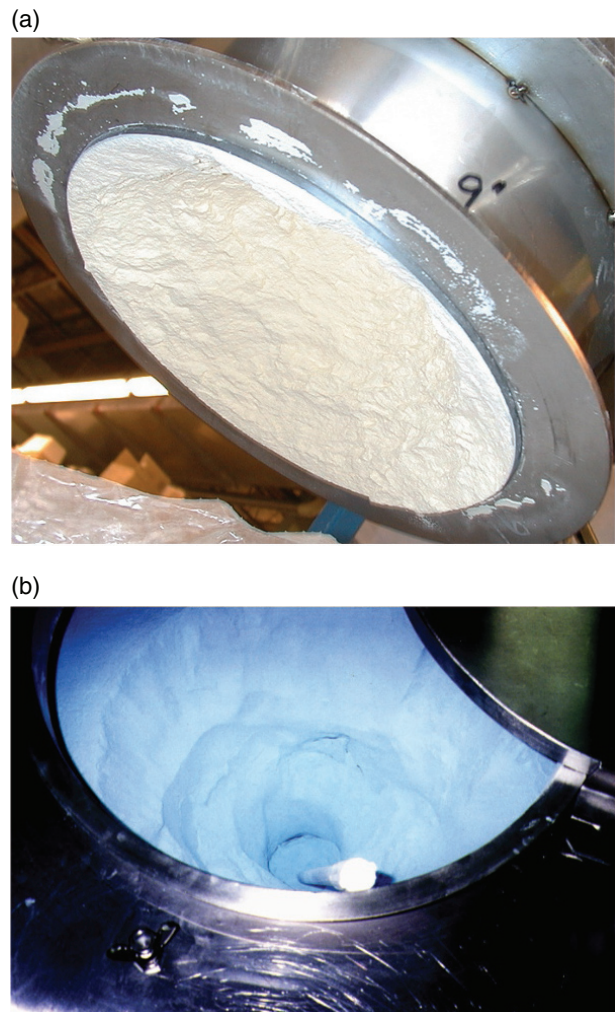


FIGURE 3.2 (a) Cohesive arch forming over the outlet of a bin. (b) Rathole (“piping”) formed for a cohesive powder stored in a bin.

powder will dictate what size outlet it can arch over (the greater the cohesive strength, the higher the likelihood of arching).

- Ratholing: can occur in a bin when the powder empties through a central flow channel, but the material at the bin walls remains stagnant and leaves an empty hole (“rathole”) through the material, starting at the bin outlet (Figure 3.2b). Ratholing is influenced by the bin/hopper geometry and outlet size the material is being fed through. Similar to the problem of arching, this problem will arise if the material has sufficient cohesive strength. In this case, the material discharge will stop once the flow channel empties.

“Erratic flow”: erratic flow is the result of obstructions alternating between an arch and a rathole. A rathole may collapse due to an external force, such as vibrations created by surrounding equipment or a flow-aid device, such as an external vibrator. While some material may discharge, falling material may impact over the outlet and form an arch. This arch may then break, due to a reoccurrence of the external force, and material flow may not resume until the flow channel is emptied and a rathole is formed again. This not only results in erratic feed to the downstream equipment (press) but can also result in a nonuniform feed density.

“Fine powder” flow concerns (two-phase flow effects): additional flow concerns can arise when handling fine powders, generally in the range below 100 μm in average particle size. These concerns are due to the interaction of the material with entrained air or gas, which becomes significant in describing the behavior of the material. This interaction can result in two-phase (powder/interstitial gas) flow effects. There are three modes that can occur when handling fine powders that are susceptible to two-phase flow effects: steady flow, flooding (or flushing), and a flow-rate limitation [4]. These three flow modes are discussed in more detail below:

- “Steady flow”: will occur with fine powders if the target flow rate (feed rate through the system) is below the “critical flow rate” that occurs when the solids’ stress is balanced by the air pressure at the outlet. The target flow rate is often controlled by a feeder, such as at the inlet to a compression machine (press feed frame). The critical flow rate and the flow property tests used to determine it are described in more detail in Section 3.4. At target flow rates exceeding the critical flow rate, unsteady flow can occur by two different modes, as described below.
- “Flooding”: (or “flushing”) is an unsteady two-phase flow mode that can occur as falling particles entrain air and become fluidized. Since handling equipment often cannot contain a fluid-like powder, powder can flood through the equipment (feeders and seals)

uncontrollably. Flooding can also occur when handling fine powders in small hoppers with high fill and discharge rates. In such situations, the powder does not have sufficient residence time to deaerate, resulting in flooding through the feeder. One adverse effect of flooding or flushing may be high variation in the tablet weight and strength.

- “Flow-rate limitation”: is another unsteady two-phase flow mode that can occur with fine powders. Fine powders have very low permeabilities and are affected by any movement of the interstitial air (air between the particles). This air movement will occur due to the natural consolidation and dilation of the powder bed that takes place as it flows through the cylindrical and hopper geometries. As the material is consolidated in the cylinder, the air is squeezed out. As the powder flows through the hopper and outlet, it dilates and additional air must be drawn in. The air pressure gradients caused as a result of this air movement can retard discharge from a hopper, significantly limiting the maximum achievable rates. This may be observed when the speed of a high-speed press is increased, and the tablet weight variation increases.

During unsteady two-phase flow modes, the material’s bulk density can undergo dramatic variations. This can negatively impact downstream packaging or processing operations. Problems can result, such as excessive tablet or capsule weight variations, a required reduction in filling speeds, and even segregation. Equipment and process parameters will govern whether such problems occur, and are further discussed in Section 3.5. These parameters include hopper geometry and outlet size, applied vacuum, and other sources of air pressure differences (such as dust collection systems), material level, time since filling, and of course the target feed rate. Material properties, such as permeability and compressibility (discussed in Section 3.4) will also play important roles, as will variations in the material’s state of aeration that can occur based on its residence time or degree of compaction from external forces and handling.

One of the most important factors in determining whether a powder will discharge reliably from a hopper is establishing what flow pattern will develop, which is discussed in the section below.

3.3.2 Flow Patterns

Two primary flow patterns can develop in a bin or hopper: funnel flow and mass flow. In funnel flow (Figure 3.3), an active flow channel forms above the outlet, which is surrounded by stagnant material. This results in a first-in–last-out flow sequence. It generally occurs in equipment with relatively shallow hoppers. Common examples of funnel flow

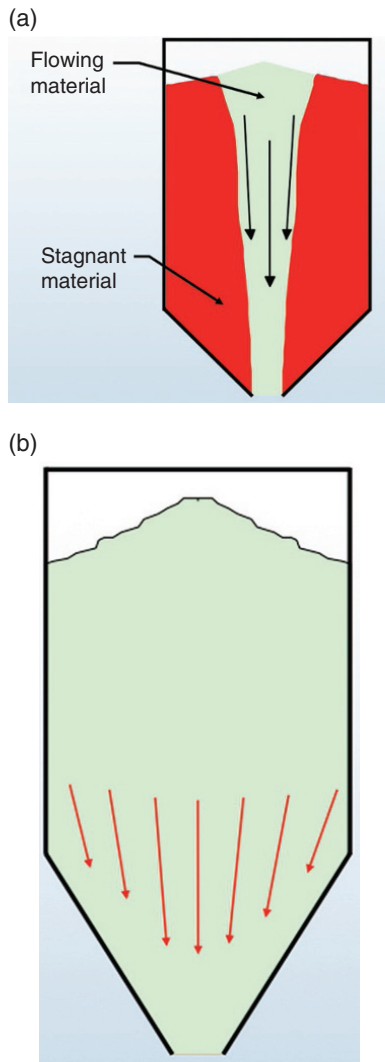


FIGURE 3.3 (a) Funnel-flow discharge pattern forming in a bin with flowing material in the center (green) and stagnant material on the outer area (red). (b) Mass-flow discharge pattern from a bin with all material in green in motion and velocity gradient shown (arrows).

bins include hopper geometries such as asymmetric cones and rectangular-to-round transitions. As the level of powder decreases in funnel flow, stagnant powder may fall into the flow channel if the material is sufficiently free-flowing. If the powder is cohesive, a stable rathole may develop. Funnel flow occurs if the powder is unable to flow along the hopper walls, due to the combination of friction against the walls and hopper angle. Funnel flow's first-in–last-out flow sequence can also have an adverse effect upon segregation and results in a nonuniform feed density to the downstream equipment (tablet or capsule weight variation).

In mass flow (Figure 3.3b), all of the powder is in motion whenever any is withdrawn. Powder flow occurs throughout the bin, including at the hopper walls. Mass flow provides a

first-in–first-out flow sequence, eliminates stagnant powder, provides a steady discharge with a consistent bulk density, and provides a flow that is uniform and well-controlled. Ratholing will not occur in mass flow, as all of the material is in motion. A mass-flow bin design may also be beneficial with respect to reducing segregation (further discussed in Section 3.5). The requirements for achieving mass flow are:

1. size the outlet large enough to prevent arch formation;
2. ensure the hopper walls are steep and smooth enough to allow the powder to flow along them.

These mass-flow design parameters are discussed further in Section 3.4 and their implementation is discussed in Section 3.5.

3.3.3 Common Segregation Mechanisms

The term “segregation mechanism” refers to the mode or motive force by which the components of the blend separate. There are many different segregation mechanisms that can adversely affect uniformity [5], but three common segregation mechanisms in typical pharmaceutical processes are:

1. sifting segregation (sometimes referred to as “percolation segregation”);
2. fluidization segregation (sometimes referred to as “air entrainment”);
3. dusting segregation (sometimes referred to as “particle entrainment in an air stream”).

The three segregation mechanisms are described in more detail below. These terms are not universally defined, so one must use caution when using them. Segregation may occur as a result of just one of these mechanisms or a combination of several mechanisms.

3.3.3.1 Material Properties that Affect Segregation

Whether segregation occurs, to what degree, and which mechanism or mechanisms are involved depends on a combination of the properties of the blend and the process conditions encountered. Primary material properties that influence segregation tendencies include:

- Mean particle size and PSD of the API, excipients, and final blend: segregation can occur with blends of any mean size, but different mechanisms become more pronounced at different particle sizes, as further discussed below.
- Particle density: the particle density will affect how the blend components fluidize.

- Particle shape: rounded particles may have greater mobility than irregularly shaped particles, which can allow more segregation.
- Particle resilience: this property influences collisions between particles and surfaces, which can lead to differences in where components accumulate during the filling of a bin, press, or an encapsulator.
- Cohesive strength of the blend: as a general rule, more cohesive blends are less likely to segregate. However, if enough energy is added to dilate the blend and/or separate particles from one another, even a very cohesive material can segregate.
- Electrostatic effects: the ability of components to develop and hold an electrostatic charge, and their affinity for other ingredients or processing surfaces, can also contribute to segregation tendencies.

Of all of these, segregation based on particle size is by far the most common [5, 6]. Particle size is the most important factor in the three primary segregation mechanisms considered here, as further described in the following sections.

3.3.3.2 Sifting Segregation Sifting segregation is a process by which fine particles tend to sift or percolate through coarse particles. For segregation to occur by this mechanism, four conditions must exist:

1. There must be a range of particle sizes. A minimum difference in mean particle diameters between components of 1.3 : 1 is often more than sufficient [6].
2. The mean particle size of the mixture must be sufficiently large, typically, greater than about 100 μm [7].
3. The mixture must be relatively free-flowing to allow particle mobility.

4. There must be relative motion between particles (inter-particle motion). This last requirement is very important, since without it even highly segregating blends of ingredients that meet the first three tests will not segregate. Relative motion can be induced in a variety of ways, such as when a pile is formed during bin filling, vibration from surrounding equipment (such as a tablet press), or as particles tumble and slide down a chute.

If any one of these conditions does not exist, the mix will not segregate by this mechanism. The result of sifting segregation in a bin is usually a side-to-side variation in the PSD. The smaller particles will generally concentrate under the fill point, with the coarse particles concentrating at the perimeter of the pile (Figure 3.4).

3.3.3.3 Fluidization Segregation Variations in particle size or density often result in vertically segregated material when handling powders that can be fluidized. Finer or lighter particles often will be concentrated above larger or denser particles. This can occur during filling of a bin or other vessel or within a blending vessel once the blending action has ceased. Fluidization segregation often results in horizontal gradation of fines and coarse material. A fine powder can remain fluidized for an extended period of time after filling or blending. In this fluidized state, larger and/or denser particles tend to settle to the bottom. Fine particles may be carried to the surface with escaping air as the bed of material deaerates. For example, when a bin is being filled quickly, the coarse particles move downward through the aerated bed, while the fine particles remain fluidized near the surface. This can also occur after blending, if the material is fluidized during blending.

Fluidization is common in materials that contain a significant percentage of particles smaller than 100 μm [8].



FIGURE 3.4 Example of sifting segregation in a 2-D pile. Note the gray and orange particles are approximately 1200 μm , while the light yellow particles are approximately 350 μm . *Source:* Courtesy of Jenike & Johanson, Inc.

Fluidization segregation is most likely to occur when fine materials are pneumatically conveyed, filled, or discharged at high rates, or if gas counter-flow occurs. As with most segregation mechanisms, the more cohesive the material, the less likely it will be to segregate by this mechanism.

Fluidization via gas counter-flow can occur as a result of insufficient venting during material transfer. As an example, consider a tumble blender discharging material to a bin with an airtight seal between the two (Figure 3.5a). As the blend transfers from the blender to the bin, air in the bin is displaced, and a slight vacuum is created in the blender. If both are properly vented, air moves out of the bin and, separately, into the blender. If the bin and blender are not vented, the air must move from the bin to the blender through the blender discharge. In doing so, the fines may be stripped off of the blend, and carried to the surface of the material still within the blender.

3.3.3.4 Dusting Segregation Like fluidization segregation, dusting is most likely to be a problem when handling fine, free-flowing powders with particles smaller than about $50\ \mu\text{m}$ [8], as well as a range of other particle sizes. If dust is created upon filling a bin, air currents created by the falling stream will carry particles away from the fill point (Figure 3.5b). The rate at which the dust settles is governed by the particle's settling velocity. The particle diameter is much more significant than particle density in determining settling velocity.

As an example of this mechanism, consider a mix of fine and large particles that is allowed to fall into the center of a bin. When the stream hits the pile of material in the bin, the column of air moving with it is deflected. The air then sweeps off the pile toward the perimeter of the bin, where the air becomes highly disturbed. After this, the air generally moves back up the bin walls in a swirling pattern. At this point, the gas velocity is much lower, allowing many particles to fall out of suspension. Because settling velocity is a strong function of particle diameter, the finest particles (with low settling velocities) will be carried to the perimeter of the bin. The larger particles will concentrate closer to the fill point, where the air currents are strong enough to prevent the fine particles from settling. Dusting segregation can also result in less-predictable segregation patterns, depending on how the bin is loaded, vented, and the use and location of dust collection.

The empirical tests used to assess these segregation mechanisms are reviewed in Section 3.4 and the application of their results to design equipment to minimize segregation is presented in Section 3.5.

3.4 MEASUREMENT OF FLOW PROPERTIES

This section summarizes different quantitative flow properties that should be measured, as well as the empirical segregation test methods used to assess segregation potential by different mechanisms. This section also reviews the

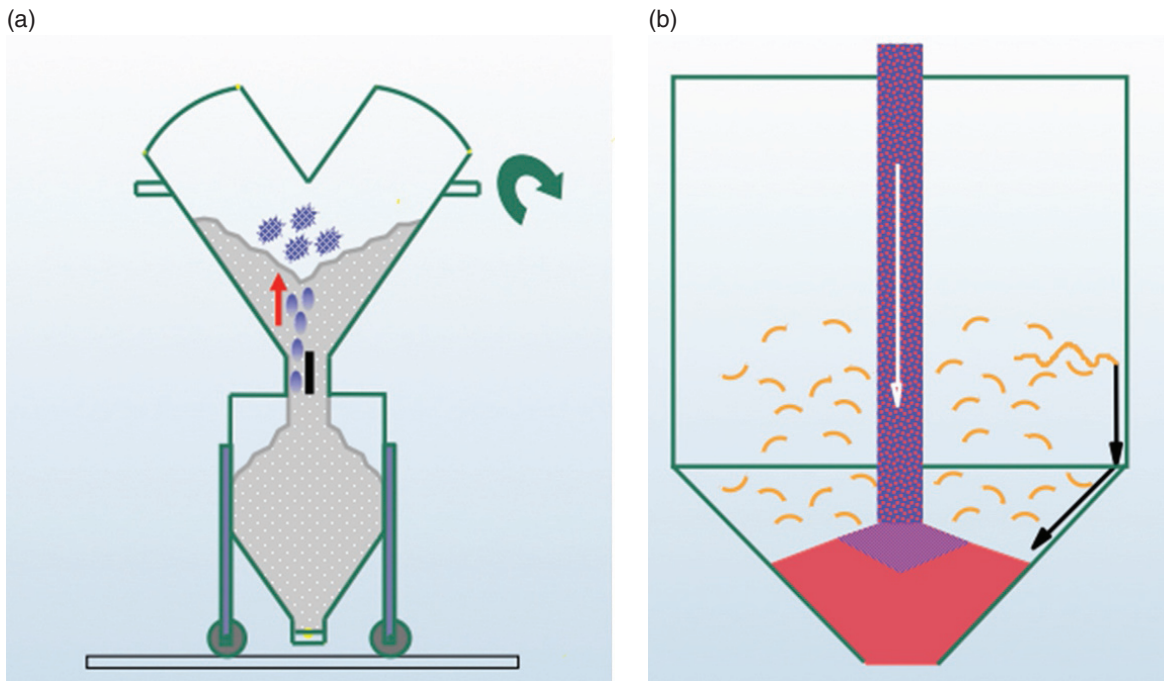


FIGURE 3.5 (a) Example of fluidization segregation from blender to bin. (b) Example of dusting segregation during filling of bin.

calculation of equipment parameters required for consistent, reliable flow in powder-handling equipment for gravity discharge.

3.4.1 Cohesive Strength Tests: Preventing Arching and Ratholing

Dr. Andrew Jenike developed his mathematical model of the flow of powders by modeling the powder as a rigid-plastic (not a visco-elastic) continuum of solid particles [9]. This approach included the postulation of a “flow–no flow” criterion that states the powder would flow (e.g. from a bin) when the stresses applied to the powder exceed the strength of the powder. The strength of a material will vary, depending on how consolidated it is. For example, the strength of wet sand increases as the consolidation pressure is increased (e.g. greater packing conditions). Therefore, it is critical to be able to measure the cohesive strength of a powder when scaling-up a process and designing equipment.

3.4.1.1 Cohesive Strength Test Method One of the primary flow problems that can develop in powder-handling equipment is a no-flow obstruction due to the formation of a cohesive arch or rathole. The required outlet size to prevent a stable cohesive arch or rathole from forming is determined by applying the flow–no flow criterion and using the results of a cohesive strength test. In order to apply the flow–no flow criterion, we need to determine:

1. The cohesive strength of the material as a function of the major consolidation pressure acting on the material: the consolidation pressure acting on the powder changes throughout the bin height, due to the weight of material above it. Therefore, the cohesive strength must be measured over a range of consolidation pressures. The cohesive strength can be measured as a function of major consolidating pressure using the test methods described in this section.

2. The stresses acting on the material to induce flow: gravity pulls downwards on a potential arch that may form. The stresses acting on the powder can be determined using mathematical models [9].

To further illustrate the concepts of strength and consolidation pressure, consider an “idealized” strength test, as shown in Figure 3.6. In this idealized test, the cohesive strength of the powder is measured in the following distinct steps:

1. Consolidation of the powder: the powder is consolidated using a prescribed consolidation pressure (P). In the idealized test shown, the sample is contained in a cylinder and the consolidation pressure is applied from the top (Step 1 in Figure 3.6).
2. Fracture of the powder: once the consolidation pressure is applied, the cylinder containing the powder would be removed without disturbing the powder sample (Step 2a in Figure 3.6). After this, the strength of the powder can be measured by applying pressure to the column of powder until it fails or fractures (Step 2b in Figure 3.6). The applied pressure at which the powder failed is referred to as the yield strength (F) (cohesive strength). This idealized test could be repeated several times to develop a flow function (FF), which is a curve illustrating the relationship between the yield strength (F) and the major consolidation pressure (P). An example of a flow function is shown in Figure 3.6.

This idealized strength test is not possible for the broad range of powders that might be tested. Therefore, several different cohesive strength test methods have been developed. The respective strengths and weaknesses of these different cohesive strength tests are reviewed in the literature [10, 11]. Although many different test methods can be used to

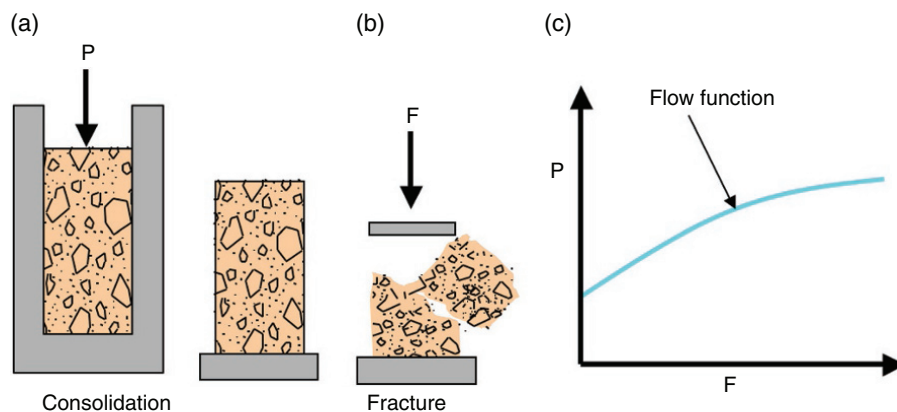


FIGURE 3.6 Schematic of “idealized” strength test: (a) Step 1; (b) Step 2a; and (c) Step 2b.

measure cohesive strength, this chapter focuses specifically upon the Jenike Direct Shear Test method. The Jenike Direct Shear Test method is the most universally accepted method, and is described in ASTM standard D6128 [12]. It is important that these tests be conducted at representative handling conditions, such as temperature, relative humidity, and storage at rest, since all these factors can affect the cohesive strength. An arrangement of a cell used for the Jenike Direct Shear Test is shown in Figure 3.7. The details of the Jenike Direct Shear Test method are provided in the work of Jenike [9], including the generation of:

- Mohr's circle to plot the shear stress (τ) versus the consolidation pressure (σ);
- Effective yield locus;
- Flow function.

The data generated experimentally from the Jenike Direct Shear Test can be used to determine the following derived parameters:

- The flow function (FF) that describes the cohesive strength (unconfined yield strength, F_c) of the powder as a function of the major consolidating pressure (σ_1). The flow function is one of the primary parameters used to calculate the minimum outlet diameter/width for bins, press hoppers, blender outlets, etc., to prevent arching and ratholing. The calculation of the minimum outlet diameter/width is discussed in more detail below.
- The effective angle of internal friction (δ) that is used to calculate the minimum outlet to prevent arching and the required hopper angles for mass flow (described below).

- The angle of internal friction for continuous flow or after storage at rest (ϕ and ϕ_r). The static angle of internal friction (ϕ_r) is used to calculate the minimum outlet to prevent ratholing (described in the following section).

Other testing methods exist that utilize the same principles of consolidation and shearing to determine the cohesive strength of a bulk powder. Annular (ring) shear testers produce rotational displacement between cell halves containing material, rather than a lateral displacement. The loading and shearing operations are more readily adapted to automation, since unlimited travel can be achieved with this type of test cell. The successful use of an annular ring shear tester to measure cohesive strength has been discussed in the industry [13–16].

3.4.1.2 Minimum Required Outlet Dimensions to Prevent Arching (Mass Flow Bin)

The flow of powders through bins and hoppers can be predicted by a complete mathematical relationship. This is very beneficial when scaling-up a process and designing powder-handling equipment. If gravity discharge is used, the outlet size required to prevent a cohesive arch or rathole from forming over a bin outlet can be calculated. The term B_c refers to the minimum outlet diameter for a circular outlet to prevent a cohesive arch from forming in a mass flow bin. The term B_p refers to the minimum outlet width for a slotted outlet, in which the length:width ratio exceeds 3 : 1, to prevent arching in a mass flow bin.

The majority of bins used in pharmaceutical processes utilize hoppers with circular outlets. Therefore, we will focus our discussion on the calculation of the B_c parameter. It is

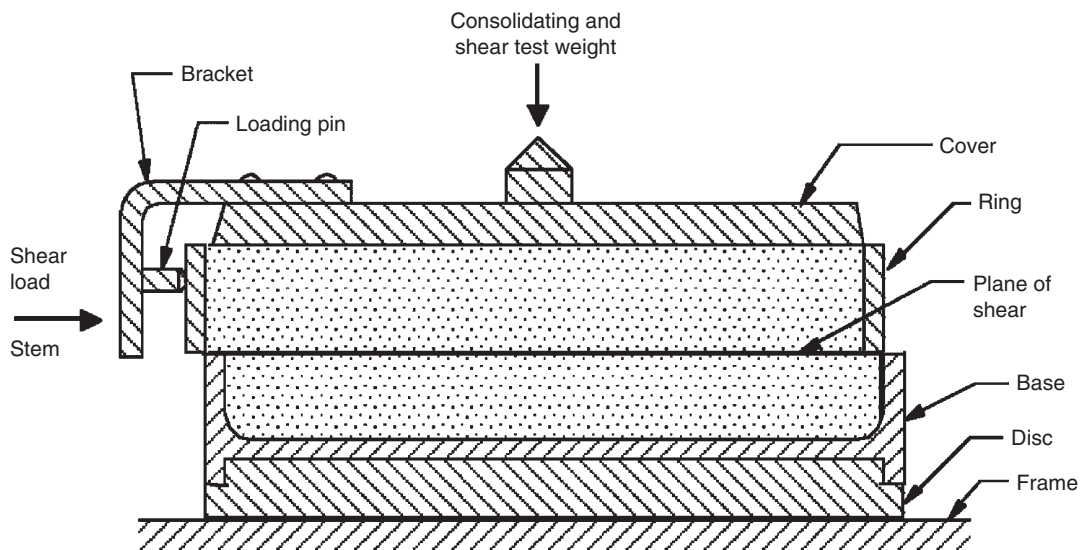


FIGURE 3.7 Jenike Direct Shear Test, cohesive strength test set-up.

worth noting that the outlet diameter required to prevent arching over a circular outlet (B_c) will typically be about two times greater than the required outlet width of a slotted outlet (B_p). The calculation of B_p is provided in the work of Jenike [9].

For mass flow, the required minimum outlet diameter to prevent arching is calculated in Eq. (3.1):

$$B_c = H(\theta') f_{crit} / \gamma \tag{3.1}$$

where:

$H(\theta')$ is a dimensionless function derived from first principles [9] (i.e. Jenike’s mathematical models of radial stress fields) and is given by Figure 3.8. The complete derivation of $H(\theta')$ is beyond the scope of this chapter, but is provided in the work of Jenike [9].

f_{crit} (units of force/area) is the unconfined yield strength at the intersection of the hopper flow factor (ff) and the experimentally derived flow function (FF), as shown in Figure 3.9. The flow factor (ff) is a mathematically

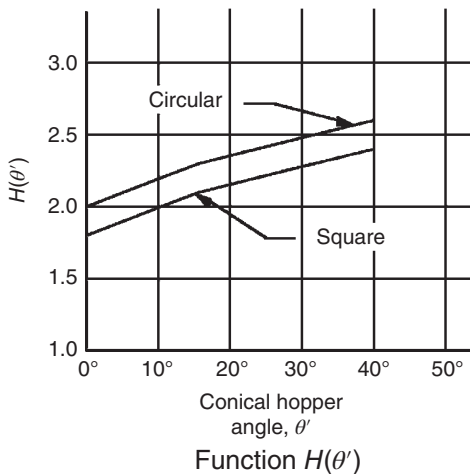


FIGURE 3.8 Plot of derived function $H(\theta')$ used to calculate arching dimensions for mass flow bins.

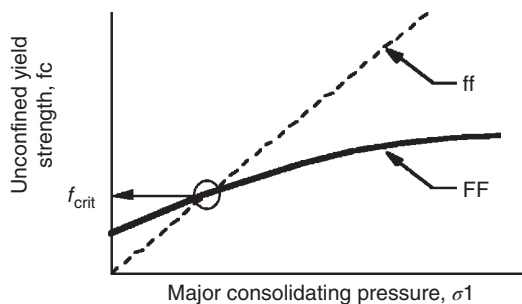


FIGURE 3.9 Example of flow function (FF) and flow factor (ff) intersection showing f_{crit} at their intersection.

determined value that represents the minimum available stress available to break an arch. The calculation of the flow factor (ff) is also beyond the scope of this chapter, but is provided in the work of Jenike [9], and is a function of the flow properties and the hopper angle (θ).

γ (units of weight/volume) is the bulk density of the powder at the outlet.

This calculation yields a dimensional value of B_c in units of length, which is scale-independent. Therefore, for a mass flow bin, the opening size required to prevent arching is not a function of the diameter of the bin, height of the bin, or the height-to-diameter ratio.

The determination of B_c is especially valuable in making decisions during process scale-up. As a formulation is developed, a cohesive strength test can be conducted early in the development process to determine the cohesive strength (flow function). This material-dependent flow function, in conjunction with Eq. (3.1), will yield a minimum opening (outlet) size to prevent arching in a mass flow bin. For example, this opening size may be calculated to be 200 mm. This 200 mm diameter will be required whether the bin holds 10 kg or 1000 kg of powder and is scale-independent. In this example, since a 200 mm diameter opening is required, feeding this material through a press hopper or similarly small openings would pose problems with an arch developing over the outlet. This information could then be used early in the development process, to consider reformulating the product to reduce the cohesive strength, and improve flowability.

3.4.1.3 Minimum Required Outlet Dimensions to Prevent Ratholing (Funnel Flow Bin) If the bin discharges in funnel flow, the bin outlet diameter should be sized to be larger than the critical rathole diameter (D_f) to prevent a stable rathole from forming over the outlet. For a funnel flow bin with a circular outlet, sizing the outlet diameter to exceed the D_f will also ensure that a stable arch will not form (since a rathole is inherently stronger than an arch). The D_f value is calculated in Eq. (3.2), and additional details of the calculation are provided in the work of Jenike [9].

$$D_f = G(\phi_t) f_c(\sigma_1) / \gamma \tag{3.2}$$

where:

$G(\phi_t)$ is a mathematically derived function from first principles models developed by Jenike [9] and is given by Figure 3.10.

$f_c(\sigma_1)$ is the unconfined yield strength of the material. This value is determined by the flow function (FF) at the actual consolidating pressure σ_1 (discussed below).

The consolidation pressure σ_1 is a function of the “head pressure” or height of powder above the outlet of the bin, and takes into account the load taken up by friction along the walls, as derived by Janssen [17] in Eq. ((3.3):

$$\sigma_1 = (\gamma R / \mu k) \left(1 - e^{(-\mu k h / R)} \right) \quad (3.3)$$

where:

γ is the average bulk density of the powder in the bin cylinder section

R is the hydraulic radius (area/perimeter) of the bin cylinder

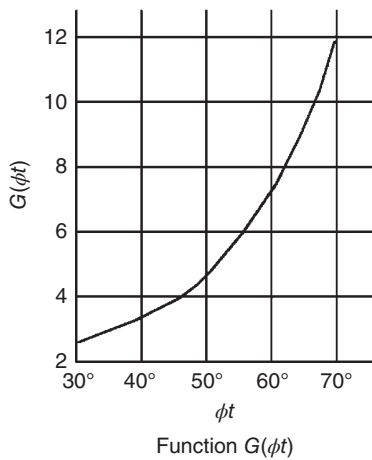


FIGURE 3.10 Plot of derived function $G(\phi_t)$ used to calculate critical rathole diameter for funnel flow bins.

μ is the coefficient of friction ($\mu = \text{tangent } \phi'$). The value of ϕ' is determined from the wall friction test (discussed in the next section)

k is the ratio of horizontal to vertical pressures. A k value of 0.4 is typically used for a straight-sided section

h is the depth of the bed of powder within the bin.

This relationship in Eq. (3.2) cannot be reduced further (e.g. to a dimensionless ratio), as the function $fc(\sigma_1)$ is highly material-dependent. The application of these parameters (Bc and Df) to design new equipment or develop corrective-equipment modifications is further discussed in Section 3.5.

3.4.2 Wall Friction: Determining Hopper Angles for Mass Flow

The wall friction test is crucial in determining whether a given bin will discharge in mass flow or funnel flow. Wall friction is caused by the powder particles sliding along the wall surface. In Jenike’s continuum model [9], wall friction is expressed as the wall friction angle (ϕ'). The lower the wall friction angle, the shallower the hopper or chute walls need to be for powder to flow along them.

3.4.2.1 Wall Friction Test Method This coefficient of friction can be measured by shearing a sample of powder in a test cell across a stationary wall surface using a Jenike Direct Shear tester [9, 12]. One arrangement of a cell used for the wall friction test is shown in Figure 3.11. In this case, a coupon of the wall material being evaluated is held in place on the frame of the tester and a cell of powder is placed above.

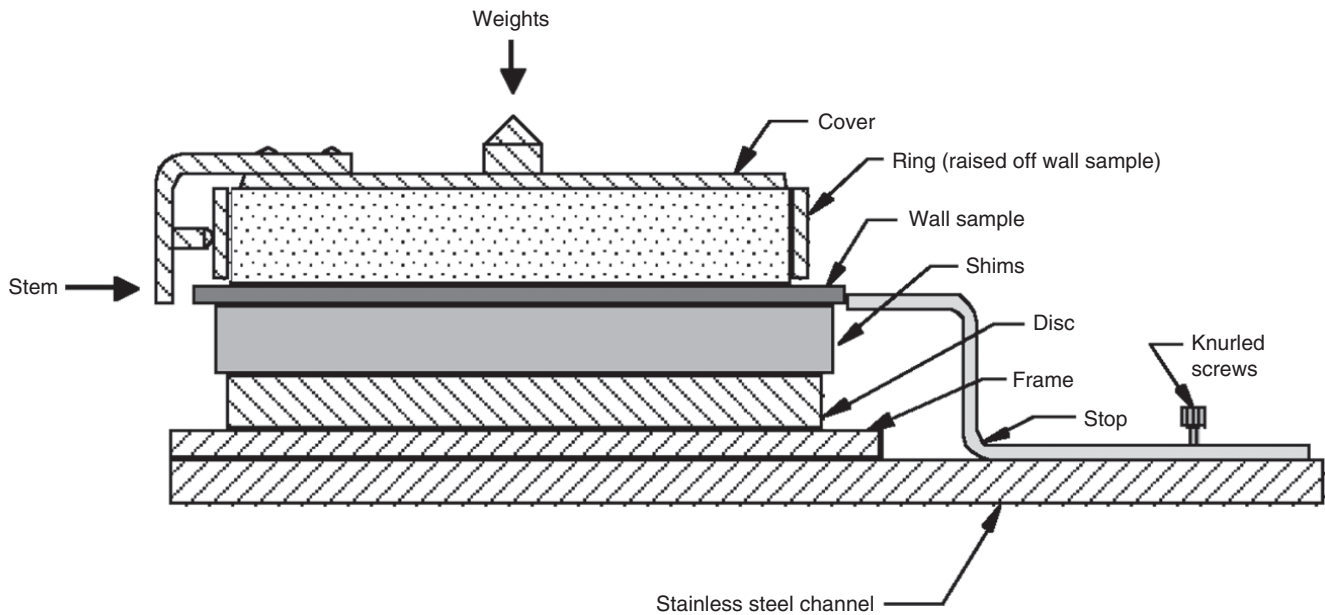


FIGURE 3.11 Jenike Direct Shear Test, wall friction test set-up.

The coefficient of sliding friction (μ , $\mu = \text{tangent } \phi'$) is the ratio of the shear force required for sliding (τ) to the normal force applied perpendicular to the wall material coupon (σn). A plot of the measured shear force (τ) as a function of the applied normal pressure (σn) generates a relationship known as the wall yield locus (see Figure 3.12).

The wall friction measured is a function of the powder handled and the wall surface (type, finish, and orientation) in contact with it. Variations in the powder, handling conditions (e.g. temperature and relative humidity), and/or the wall surface finish can have a dramatic effect on the resulting wall friction coefficient [18]. The results of the wall friction test are used to determine the hopper angles required to achieve mass flow, as discussed in the following section.

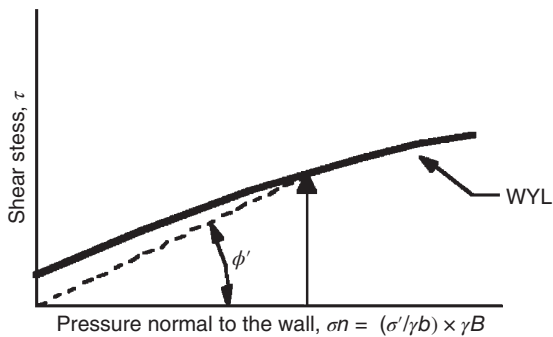


FIGURE 3.12 Example of wall yield locus generated from wall friction test data.

3.4.2.2 Calculation of Recommended Mass Flow Hopper Angles Based upon mathematical models, design charts [9] have been developed to determine which flow pattern is to be expected during gravity discharge from a bin. The design charts use the following inputs for the powder being handled and the bin design being considered:

1. hopper angle (“ θ_c ” for a conical hopper or “ θ_p ” for a planar hopper), as measured from vertical;
2. wall friction angle (ϕ'), as measured from the wall friction tests;
3. internal friction angle (δ), as measured from the cohesive strength tests.

This chapter will focus on the calculation of the recommended mass hopper angles for a conical hopper (θ_c), since the majority of pharmaceutical processes utilize bins with a conical hopper. The methods to calculate the recommended mass hopper angles for a planar hopper (θ_p) with a slotted outlet are similar in approach, and are outlined in the work of Jenike [9]. It is worth noting that the recommended mass flow angles for planar hopper walls (θ_p) can often be 8° – 12° shallower than those for a conical hopper (θ_c), for the same-sized opening.

An example of such a design chart for a conical hopper is shown in Figure 3.13. The design chart shown is specifically for a powder with an effective angle of internal friction (δ) of 40° ; the design charts will be different for different values of δ [9]. Hopper angles required for mass flow are a function of δ , since flow along converging hopper walls involves

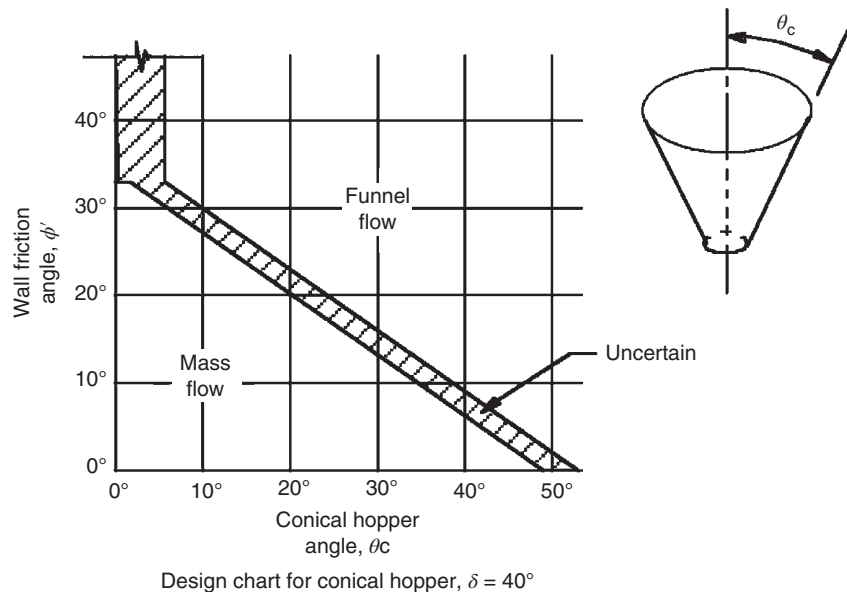


FIGURE 3.13 Mass-flow/funnel-flow design chart for conical hopper handling a bulk solid with an effective angle of internal friction (δ) of 40° .

interparticle motion of the powder. For any combination of ϕ' and θ_c that lies in the mass-flow region, mass flow is expected to occur. If the combination lies in the funnel-flow region, funnel flow is expected. The “uncertain” region is an area where mass flow is expected to occur based on theory, but represents a 4° margin of safety on the design, to account for inevitable variations in test results and surface finish.

As an example of using the design chart, a bin with a conical hopper angle (θ_c) of 20° from vertical is being used. Wall friction tests are conducted on the hopper wall surface and a wall friction angle of 35° is measured for the normal pressure calculated at the outlet. Based upon the design chart, this bin would be expected to discharge in funnel flow. In that case, the designer would need to find another wall surface, with a wall friction angle that is less than 20°, to ensure mass-flow discharge.

The wall friction angle (ϕ') is determined by the wall friction tests described above. The value of ϕ' to use for the hopper design charts will be selected for the expected normal pressure (σ_n), against the surface at the location of interest in the bin (e.g. the hopper outlet). For many combinations of wall surfaces and powders, the wall friction angle changes, depending on the normal pressure. When mass flow develops, the solids' pressure normal to the wall surface is given by the following relationship:

$$\sigma_n = \frac{\sigma'}{\gamma B} \times \gamma B \quad (3.4)$$

where:

$(\sigma'/\gamma B)$ is a dimensionless parameter that can be found in the work of Jenike [9];

B (units of length) is the span of the outlet: the diameter of a circular outlet or the width of a slotted outlet;

γ is the bulk density at the outlet.

Generally, ϕ' increases with decreasing normal pressure (σ_n). The corresponding normal pressure to the wall (σ_n) is the lowest at the outlet where the span (B) is the smallest. Therefore, it is at the outlet where the wall friction angle (ϕ') is the highest for a given design, provided the hopper interior surface-finish and angle remain constant above the outlet. As a result, if the walls of the hopper are steep enough at the outlet to provide mass flow, mass flow is to be expected at the walls above the outlet (regardless of total bin size).

The hopper angle required for mass flow is principally dependent on the outlet size selected for the hopper under consideration. The hopper angle required for mass flow is not a function of the flow rate, the level of powder within the bin, or the diameter or height of the bin. Since the wall friction angle generally increases with lower normal pressures, a steeper hopper is often required to achieve mass flow for a bin with a smaller outlet.

3.4.3 Bulk Density

The bulk density of a given powder is not a single or even a dual value, but varies as a function of the consolidating pressure applied to it. There are various methods used in industry to measure bulk density. One prominent method is utilizing different-sized containers that are measured for volume after being loosely filled with a known mass of material (“loose” density), and after vibration or tapping (“tapped density”), such as the USP Chapter <616> method [19]. These methods can offer some repeatability with respect to the conditions under which measurements are taken. However, they do not necessarily represent the actual compaction behavior of a powder being handled in a bin, chute, or press hopper. Therefore, it is necessary to measure the bulk density over a range of consolidation pressures, via a compressibility test [9, 20], for design purposes. The results of the compressibility test can often be plotted as a straight line on a log-log plot (see Figure 3.14). In powder-handling literature, the slope of this line is typically called the “compressibility” of the powder.

The resulting data can be used to determine capacities for storage and transfer equipment, and evaluate wall friction and feeder operation requirements. As an example, when estimating the capacity of a bin, the bulk density based upon the average major consolidation pressure in the bin can be used. For the calculation of the arching dimensions (B_c) and recommended mass flow hopper angle (θ_c), the bulk density based on the major consolidation pressure at the bin outlet can be used.

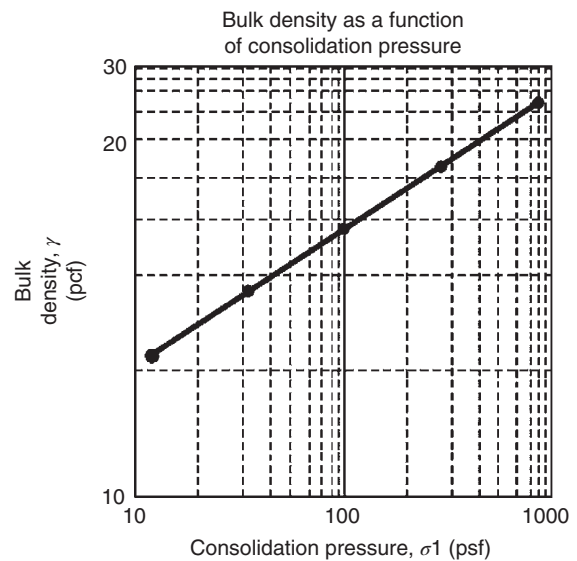


FIGURE 3.14 Example of bulk density versus consolidation pressure plot from compressibility test data.

3.4.4 Permeability

The flow problems that can occur due to adverse two-phase (powder and interstitial gas) flow effects were reviewed in Section 3.3. These problems are more likely to occur when the target feed rate (e.g. tableting or encapsulating rate) exceeds the “critical flow rate,” based on the powder’s physical properties. The results of the permeability test are one of the primary flow properties used to determine the critical flow rate. The permeability of a powder is a measurement of how readily gas can pass through it. The permeability will have a controlling effect on the discharge rate that can be achieved from a bin/hopper with a given outlet size. Sizing the outlet of a piece of equipment or choosing the diameter of a transfer chute should take into consideration the target feed rate.

Permeability is measured as a function of bulk density [20]. A schematic of the permeability tests is provided in Figure 3.15.

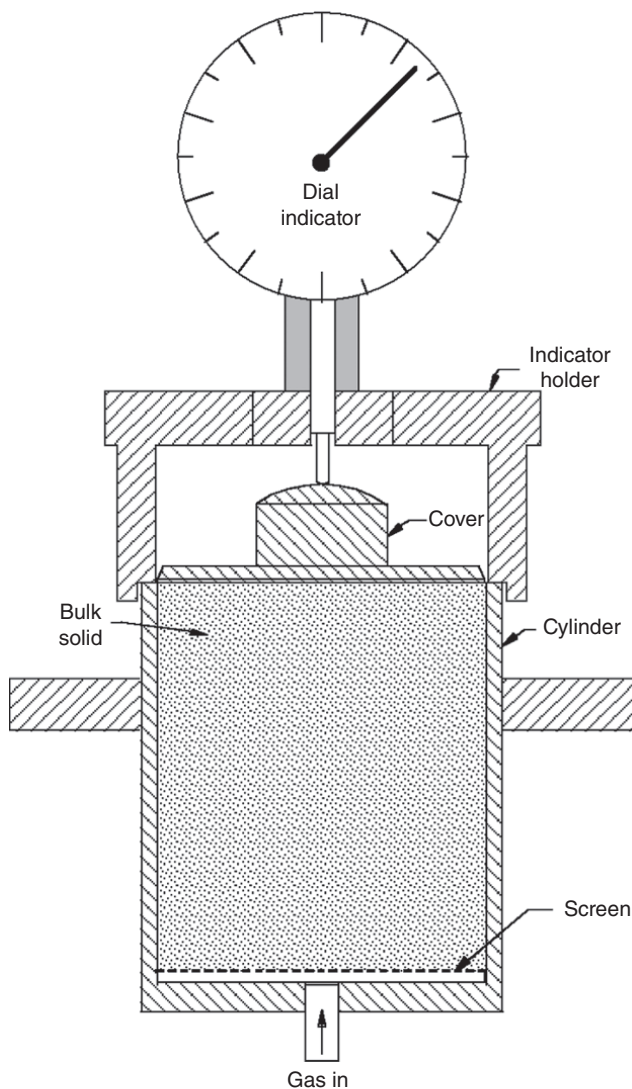


FIGURE 3.15 Schematic of permeability test set-up.

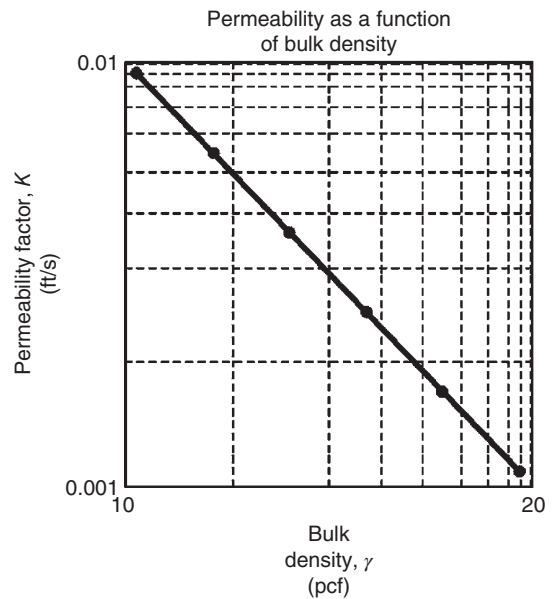


FIGURE 3.16 Example of permeability versus bulk density plot from permeability test data.

In this test set-up, gas is injected at the bottom of the test cell through a permeable membrane. During the test, the pressure drop and flow rate across the powder are measured. The method involves measuring the flow rate of air at a predetermined pressure drop through a sample of known density and height. The permeability is then calculated using Darcy’s law. The permeability of a powder typically decreases as the bulk density increases, so the test is conducted over a range of bulk densities.

Once the permeability–bulk density relationship is determined (see Figure 3.16), it can be used to calculate the critical flow rates that will be achieved for steady-flow conditions through various outlet sizes. The critical flow rate is dependent upon the permeability, bin geometry, outlet size, and consolidation pressure. The details of calculating critical flow rates are outside the scope of this chapter, but mathematical models have been developed for these calculations. Higher flow rates than the calculated critical flow rate may occur, but may result in nonsteady or erratic feed and adverse effects, as discussed in Section 3.3. Permeability values can also be used to calculate the time required for fine powders to settle or deaerate in equipment.

3.4.5 Segregation Tests

When developing a solid drug product or designing a process, it is beneficial to know whether the blend will be prone to segregation. If the blend is prone to segregation, it is beneficial to know what segregation mechanism(s) will occur, since this information can be used to modify the material properties (e.g. excipient selection, component PSD) to minimize the

potential for segregation. An understanding of the potential for segregation can alert the equipment or process designer to potential risks that may then be avoided during the scale-up process. In some cases, significant process steps, such as granulation, may be required to avoid potential segregation problems.

There are ASTM test methods for analyzing sifting and fluidization segregation [21, 22]. These tests are designed to isolate specific segregation mechanisms, and test a material's tendency to segregate by that mechanism. Several other researchers and companies have developed various segregation testers, including methods that induce vibration [23, 24], shearing in a cell [25], and methods that capture material from a pile after it has formed [25–27].

Improvements and variations to the ASTM test methods have also been made. A new fluidization segregation tester that utilizes a different mechanism to fluidize the bed has been developed [28]. It uses a smaller test sample, and provides unit-dose samples for analysis. An alternate way to run the sifting segregation test involves cycling the blend multiple times to strengthen the segregation “signal” [29].

Segregation tests are useful for identifying the:

1. segregation mechanism(s) that might be active for a given blend;
2. general segregation trend that may be observed in the process;
3. optimizing formulation design properties to minimize segregation potential.

However, the test results have limitations. Most notably, the segregation results are not scalable from the tester to the manufacturing process, and therefore cannot be tied quantitatively to the process. The segregation test results do not necessarily mean that a highly segregating material cannot be handled in a manner that prevents content uniformity problems. Therefore, the segregation test methods are primarily used as a stress test to identify the dominant segregation mechanism(s) expected to occur. This information enables the equipment designer to take the appropriate precautionary measures during scale-up or make corrective actions to existing equipment or add engineering controls, such as decelerators. Design techniques to minimize the potential for segregation are outlined in Section 3.5.

3.5 TROUBLE-SHOOTING POWDER-HANDLING PROBLEMS AND EQUIPMENT DESIGN METHODS

This section reviews basic trouble-shooting and design techniques for the bin-to-press/encapsulator feed system to provide consistent, reliable flow for gravity feed, and

minimize segregation. In particular, we will review the following basic design techniques:

- reliable funnel-flow bin design;
- reliable mass flow designs for bins, transfer chutes, and press/encapsulator hoppers;
- minimizing adverse two-phase flow effects (e.g. feed/tableting rate limitations and flooding);
- minimizing segregation during blender to press transfers.

For each of these different design concerns, we will review the key equipment design parameters. Regardless of whether the equipment being designed is a bin, transfer chute, or press hopper, a crucial first step in designing a reliable feed system is determining the flow pattern and designing the equipment accordingly. These design techniques will focus on equipment commonly used in batch processes, but can also be applied to continuous manufacturing equipment. The wall friction tests and design charts used to determine if a hopper will discharge in mass flow or funnel flow were discussed in Section 3.4.

3.5.1 Reliable Funnel Flow Design (Preventing a Rathole)

Funnel flow occurs when the hopper walls are not smooth and/or steep enough to promote flow at the walls. Funnel flow bins can be prone to ratholing if the material is cohesive enough. A funnel-flow bin design can be considered if all of the following design criteria are met:

1. Segregation of the final blend is not a concern. Since a funnel flow bin will discharge in a first-in–last-out flow sequence, any side-to-side segregation that occurred as the bin was filled will often be exacerbated in funnel flow discharge.
2. The final blend has low cohesive strength, so the formation of a stable rathole is not a concern. This can be checked by comparing the bin outlet diameter/diagonal length to the critical rathole diameter (D_f) for the estimated major consolidation pressure (σ_1) for the bin design in question as per Eqs. (3.2) and (3.3). If the outlet diameter is less than D_f , ratholing is a concern.
3. Flooding due to a collapsing rathole is not a concern. Flooding can result in highly aerated (low density) powder being fed from the bin to the press, which may have an adverse effect on the tablet properties (weight, hardness, and dissolution variation), and can result in segregation.
4. A nonuniform feed density is not a concern. Since tablet presses operate as volumetric feeders, variation of

the feed density into the press feed frame can result in tablet weight variation.

If all of these design criteria are met, a funnel-flow bin design can be used in the manufacturing drug product process. If a funnel-flow bin design is acceptable, the first concern is checking that the outlet diameter is greater than the critical rathole diameter (D_f), to ensure that a stable rathole will not form. If the diameter of the funnel flow bin is not greater than D_f , the following steps can be considered to reduce the likelihood of ratholing:

1. Enlarge the bin opening: this may require using a slotted outlet, which would require a feeder capable of feeding uniformly across the entire outlet (e.g. mass-flow screw feeder) or a valve capable of shutting off such an outlet. Using a larger outlet diameter may not be a practical modification, since the opening may need to be increased to be larger than standard valve or feeder sizes.
2. Reduce the material level in the bin: the critical rathole diameter typically decreases with a reduction in the major consolidation pressure (σ_1), which is a function of the fill height.
3. Use an internal, mechanical agitator: an internal, mechanical agitator, such as an agitator with “arms” that rotate about a central vertical shaft, may be a practical modification on a small scale for a hopper. A bin with an internal discharge valve could also be considered as a means of failing a stable rathole, but may need to be assessed via full-scale trials to determine the operating parameters required (valve “stroke” setting, etc.).
4. Use external vibrators: the effectiveness of external vibrators to collapse a stable rathole would need to be assessed via full-scale trials prior to installation, since vibration may actually increase the consolidation pressure acting on the powder and potentially the cohesive strength on the powder in the bin and increase the likelihood of ratholing. Trials would be required to assess the optimum vibrator type (high-frequency/low-amplitude versus low-frequency/high-amplitude), number of vibrators required, location, frequency settings, and other design considerations.

There are several adverse effects of using a bin that discharges in funnel flow (first-in–last-out flow sequence, non-uniform feed density, and exacerbation of segregation). In addition, the potential options to prevent a rathole are often limited or impractical. Therefore, a common design technique for preventing ratholing is to redesign the bin for mass flow. The design techniques for mass flow are discussed in the following section.

3.5.2 Reliable Mass Flow Designs for the Bin, Chute, and Press Hopper

Mass flow discharge from a bin occurs when the following two design criteria are met:

1. The bin walls are smooth and/or steep enough to promote flow at the walls;
2. The bin outlet is large enough to prevent an arch.

The wall friction tests and design charts used to determine if a bin will discharge in mass flow or funnel flow were discussed in Section 3.4. This section focuses on design techniques for mass flow bins, but these techniques may also be extended to obtaining mass flow in a transfer chute and press/encapsulator hopper. These techniques may be applied to designing new equipment or modifying existing equipment to provide mass flow.

When designing the bin to provide mass flow, the following general steps should be taken:

1. Size the outlet to prevent a cohesive arch: the bin designer should ensure that an arch will not form by making the outlet diameter equal to or larger than the minimum required outlet diameter (B_c , see Figure 3.17). If a slotted outlet is used (maintaining a 3 : 1 length : width ratio for the outlet), the outlet width should be sized to be equal to or larger than the minimum required outlet width (B_p , see Figure 3.17). The outlet may also need to be sized based upon the feed rate and two-phase flow considerations, as discussed below. If the outlet cannot be sized to prevent an arch (e.g. press hopper outlet that must mate with a fixed feed frame inlet), an internal mechanical agitator or external vibrator could be considered, as discussed in the preceding section.
2. Make the hopper walls steep enough for mass flow: once the outlet is sized, the hopper wall slope should be designed to be equal to, or steeper than, the recommended hopper angle for the given outlet size, and selected wall surface. For a conical hopper, the walls should be equal to, or steeper than, the recommended mass flow angle for a conical hopper (θ_c , see Figure 3.17). If the bin has a rectangular-to-round hopper, the valley angles should be sloped to be equal to or steeper than θ_c . For planar walls, the walls should be equal to, or steeper than, the recommended mass flow angle for a planar hopper (θ_p , see Figure 3.17).
3. Specify the interior wall surface finish: when conducting the wall friction tests, it is beneficial to conduct tests on several different finishes (e.g. 316 stainless steel with a #320 grit finish, #2B cold rolled finish, #2B electro-polished finish). This will provide the bin designer with a range of design options for the

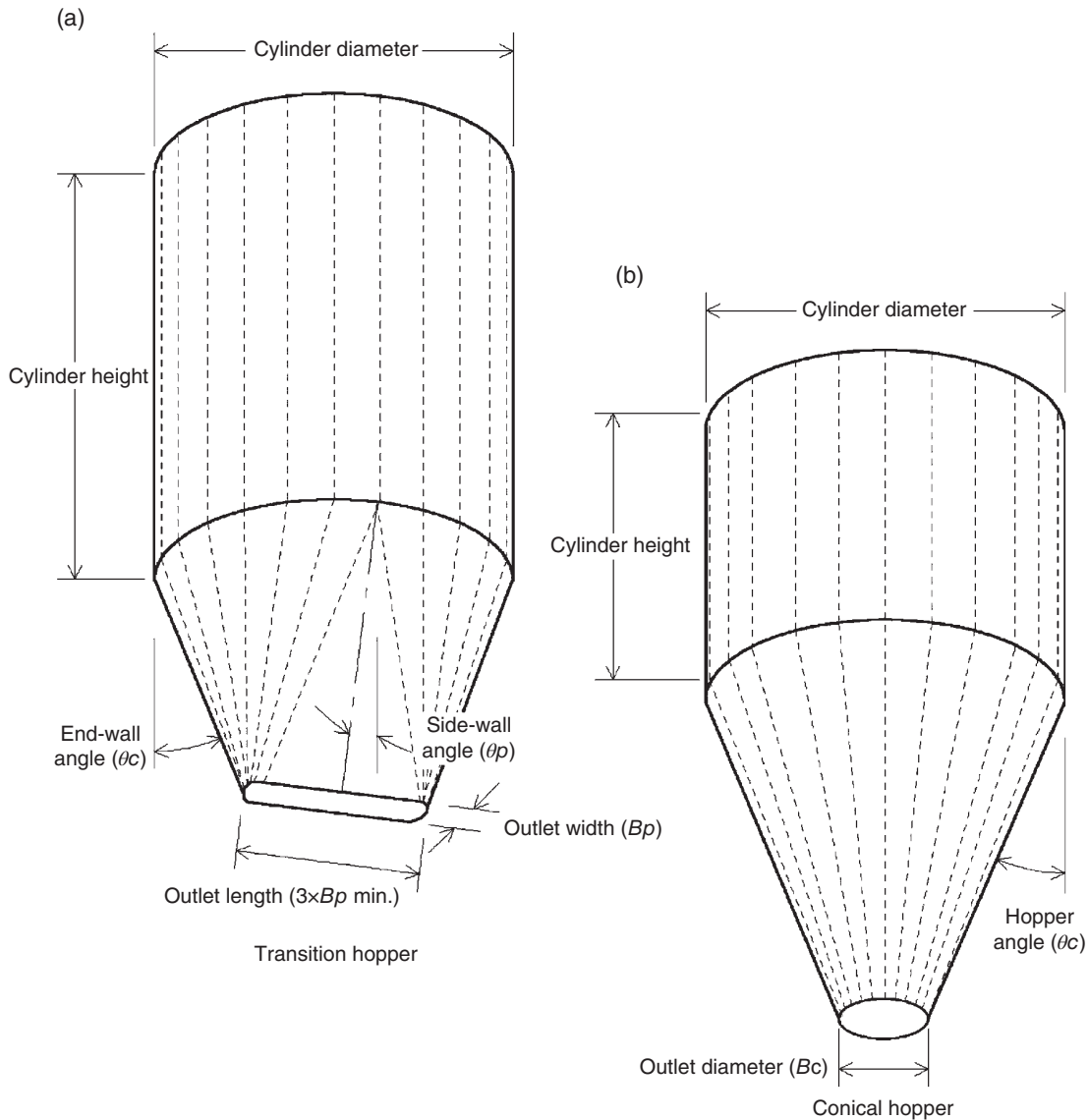


FIGURE 3.17 Mass-flow design parameters (B_c , B_p , θ_c , and θ_p), where (a) is the transition hopper and (b) is the conical hopper.

bin. Testing multiple-wall surfaces will also enable the designer to assess the sensitivity of the wall friction results to different finishes. It is not sufficient to simply test a 304 or 316 stainless steel with no regard to the interior finish. The wall friction of the blend may vary significantly from finish to finish. The orientation of directional finishes, such as a mechanical polish, is also critical to assess and control during fabrication. In addition, it cannot be assumed that an interior surface finish with a lower average roughness (R_a) will provide the best wall friction properties.

4. Consider velocity gradients: even when a bin is designed for mass flow, there still may be a velocity gradient between the material discharging at the hopper walls (moving slower) versus the center of the

hopper (moving faster), assuming a symmetric bin with a single outlet in the center. Depending upon the application, the bin designer may want to increase the velocity gradient to enhance blending between vertical layers of material in the bin, or may want to reduce the velocity gradient to enhance blending on a side-to-side basis. The decision to increase or decrease the velocity gradient will depend upon the segregation that occurs upon filling the bin and its effect upon content uniformity. The velocity gradient is reduced by making the hopper slope steeper, with respect to the recommended mass flow hopper angle (θ_c). The velocity gradient is increased by making the hopper slope closer to (but still steeper than) the recommended mass flow hopper angle. Changing the interior surface to

reduce friction or using an insert (discussed more below) are other methods used to control the velocity gradient. Asymmetric hoppers, which are common for press hoppers, are especially prone to velocity gradients since the material will move faster at the steeper hopper wall. In addition, a velocity gradient cannot be completely eliminated, especially as the material level in the hopper empties. Velocity profiles, and their effect on blending material, can be calculated a priori, given the geometry of the bin (θ_c) and measured flow properties that were discussed in Section 3.4 (i.e. ϕ' , δ , and ϕ).

5. Avoid upward-facing lips/ledges due to mismatched flanges, level probes, view ports, partially opened valves, etc., especially in the hopper section. Ideally, interior protruding devices should be located in the straight-sided, (non-converging) section of a bin/press hopper if possible, where they will be less detrimental in upsetting a mass flow pattern.

If the bin designer is modifying an existing funnel flow bin to provide mass flow, several different options can be considered, including:

1. Using a different interior surface finish with better wall friction properties (i.e. lower wall friction). The bin designer should conduct wall friction tests on alternative wall surfaces, to assess if changing the surface finish (e.g. electro-polishing an existing #2B finish) will convert the bin from funnel flow to mass flow. This is often one of the most cost-effective modifications to obtain mass flow.
2. Using a flow-controlling insert to obtain mass flow within the same bin. A properly designed insert can change the stresses that develop in the bin during discharge, so that mass flow can be obtained at a wall where the material was previously stagnant.
3. Modifying the hopper geometry. Use a different geometry that is more likely to provide mass flow (e.g. conical instead of a rectangular-to-round hopper with shallower valley angles). If the hopper is modified to have a slotted outlet, it is crucial that the feeder the hopper mates with withdraws material across the entire outlet.

In addition to these design techniques for bins, there are several additional techniques for designing transfer chutes for reliable mass flow including:

1. For converging sections that have a full cross-section (i.e. hoppers), use the same design criteria discussed above for a mass flow bin.
2. For non-converging sections of the chute, the chute should be sloped to exceed the wall friction angle

(ϕ') by a least a 10° margin of safety. As an example, if the measured wall friction angle for the given wall surface (from the wall friction test results) is 40° from horizontal, the recommended chute angle for the non-converging portion of the chute would be at least 50° from horizontal.

3. If a bifurcated chute is used, the sloping chute legs should be symmetric, to prevent velocity gradients and the possibility of stagnant material in the shallower leg.
4. Use mitered joints between sloping and vertical sections.

A scientist can use the design principles outlined in this chapter to determine if a funnel flow bin can be designed for reliable discharge and, if not, design a mass flow bin, as outlined in the example problem.

EXAMPLE 5.1

Design Criteria

A scientist has been asked to review if an existing bin will provide reliable discharge of a product, including ensuring that a stable rathole will not form in the bin. The existing bin is shown in Figure 3.18 and includes the following critical design parameters:

- Rectangular cylinder section: 1125 mm \times 1325 mm \times 1080 mm tall.
- Rectangular-to-round hopper section: Hopper is 920 mm tall and converges to a 250 mm internal diameter (ID) with side walls sloped at 25° and 30° (from vertical), resulting in valley angles of 39° , and is fabricated from 316 stainless steel with a #320 grit finish.
- The maximum storage time at rest in the bin is twenty-four (24) hours.
- The bin has a total capacity (i.e. "liquid" volume) of 2200 l and a minimum 1500 l total capacity is required if an alternative bin is used.
- From the wall friction test results, the recommended mass flow angles for a conical hopper (θ_c) with a 250 mm ID outlet were calculated to be 20° for the existing #320 grit mechanically polished finish and 30° for an alternative #2B cold-rolled finish, for 24 hour storage at rest conditions. In addition, the wall friction angle (ϕ') for the normal stress in the cylinder is 10° for the existing #320 grit finish.
- From the cohesive strength test results, the critical rathole diameter (D_f) was calculated to be 2100 mm after 24 hours at rest for a 1600 mm Effective Head (EH, see Eq. 3.5). The recommended outlet diameter required to prevent arching over a circular outlet (B_c) was calculated to be 200 mm after 24 hours at rest.

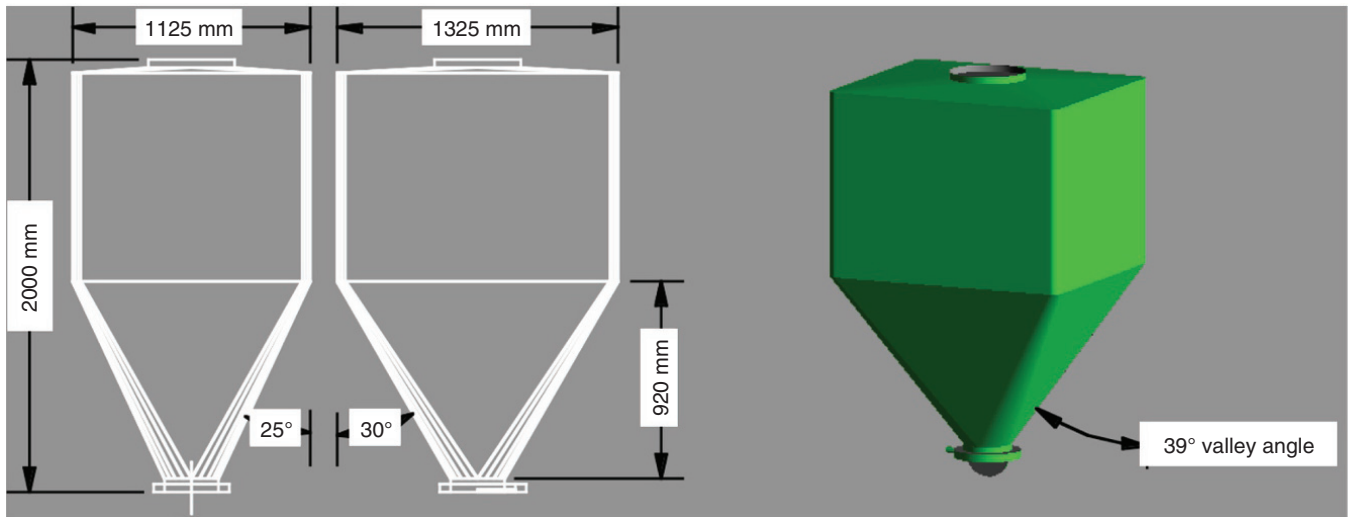


FIGURE 3.18 Existing bin design (funnel flow discharge).

Solution

- Step 1: Determine if the existing bin will discharge in mass flow or funnel flow. Based upon the wall friction results listed above, the recommended mass flow angle is 20° for a conical hopper with a 250 mm ID outlet for a #320 finish. The valley angles of the existing bin are 39°. Therefore, funnel flow discharge is predicted from the existing bin since the valley angles are shallower (larger) than the recommended mass flow angle. In addition, even if the existing bin was fabricated from 316 stainless steel with a #2B finish that provided lower wall friction (i.e. a shallower recommended mass flow angle of 30°), the existing bin is still expected to discharge in funnel flow.
- Step 2: Calculate the approximate EH and determine if a rathole will form in the bin. The EH in the cylinder section of a bin can be calculated using Eq. (3.5):

$$EH = \frac{R}{\mu k} \left[1 - e^{-\left(\frac{\mu k H}{R}\right)} \right] \quad (3.5)$$

where:

- R = Hydraulic radius (area/perimeter) = $(1.1 \text{ m} \times 1.3 \text{ m}) / (2(1.1 \text{ m} + 1.3 \text{ m})) = 0.3 \text{ m}$;
- H = Height of the cylinder section = 2.0 m, since for simplicity and conservatism we will use the total height of the column of material in the bin (i.e. overall bin height) versus just the cylinder section;
- k = Ratio of horizontal to vertical solids' stresses = 0.4, which is a typical value in the cylinder section of a bin and will be used for the entire bin height in this example for simplicity and conservatism;

μ = Coefficient of friction = $\tan \phi' = \tan (10^\circ) = 0.18$.

Based upon these inputs, the calculated EH = 1.6 m (1600 mm). For the calculated EH = 1600 mm, the critical rathole diameter (D_f) is 2100 mm. Since the critical rathole diameter of 2100 mm is much larger than the existing outlet of 250 mm, a stable rathole is expected to form in the bin. Therefore, a mass flow bin will need to be designed, as outlined in Step 3.

- Step 3: Design a bin for mass flow discharge. We will design a mass flow bin using the same outlet size (250 mm) as the existing bin, since an outlet exceeding the B_c dimension of 200 mm will suffice. The new mass-flow bin design has a conical hopper and is designed for the minimum required total capacity of 1500 l. Based upon the wall friction test results, we will use the #2B surface finish that provided better results and the recommended conical hopper angle (θ_c) of 30°. Based upon these design criteria, a mass flow bin with a 30° conical hopper (760 mm tall) and 1125 mm ID × 1240 mm tall cylinder was designed to provide the required capacity and mass flow discharge, as shown in Figure 3.19.

3.5.3 Minimizing Adverse Two-phase Flow Effects

The primary focus in preventing adverse two-phase flow affects is to ensure that the powder-handling equipment is designed so that the critical flow rate through a given outlet is greater than the target feed rate. The critical flow rate is determined using mathematical models, with permeability and compressibility test results as primary inputs. The critical

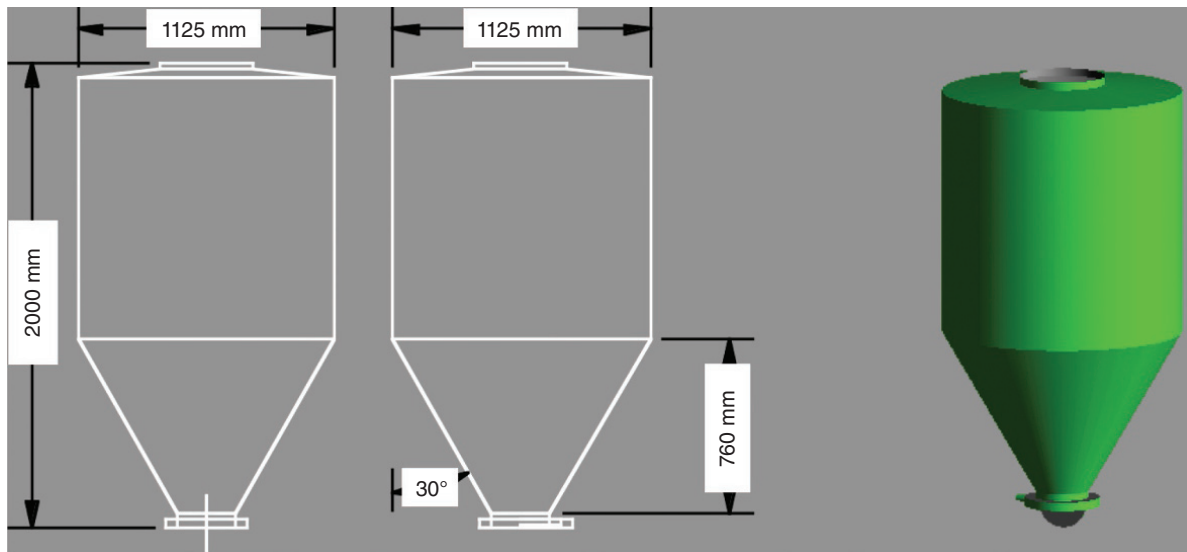


FIGURE 3.19 New bin design (mass flow discharge).

flow rate is also a strong function of the outlet size, and increases as the outlet size increases. The target feed rate is often set by the required maximum tableting/encapsulating rate for the process. Adverse two-phase flow effects are typically most pronounced at the press feed hopper, since it often has the smallest outlet in the entire press feed system. Therefore, the press hopper will typically have the lowest critical flow rate.

When designing the powder-handling equipment to minimize adverse two-phase flow effects, the following general design techniques are beneficial:

1. Design the equipment for mass flow: mass flow will provide consistent feed and a more uniform consolidation pressure acting on the powders. In addition, having a first-in–first-out flow sequence will allow the material more time to deaerate before being discharged through the outlet. This will reduce the likelihood of flooding. Mass flow will also prevent collapsing ratholes that can result in the powder aerating and flooding as it falls into the central flow channel. It is worth noting that mass flow can result in a lower critical flow rate than funnel flow, but will be more stable. Therefore, simply using a mass-flow bin design may not be the only corrective action required if a flow-rate limitation occurs. However, designing the equipment for mass flow is often the first step in addressing adverse two-phase flow effects.
2. Use larger outlets for the handling equipment: the critical flow rate is a strong function of the cross-sectional area of the outlet. Therefore, increasing the outlet can often be highly beneficial in reducing two-phase flow effects. The goal would be to increase the outlet size until the critical flow rate for the selected outlet size exceeds the target flow rate. Since this may not be feasible for a press feeder hopper, in which the outlet size is fixed to mate with the press feed frame inlet, additional design techniques are discussed below. Computer software can be used to model the two-phase flow behavior to assess the effect of changing the outlet diameter, e.g. DEM coupled with CFD.
3. Reduce the fill height in the handling equipment: the critical flow rate through a given outlet increases as the major consolidation pressure (σ) decreases. Therefore, reducing the fill height will be beneficial, but will be much less effective than increasing the outlet size.
4. Reduce the target feed rate: if possible, reducing the target feed rate (tableting rate) to be less than the critical flow rate will be beneficial, but is often impractical since it will result in decreased production rate.
5. Consider gas pressure differentials: a gas pressure differential can have a beneficial or adverse effect upon two-phase flow effects. A positive gas pressure differential at the outlet (i.e. bin at a higher gas pressure than the equipment downstream) may be beneficial in overcoming a feed-rate limitation. In this case, the air pressure is forcing the material in the direction of flow. Conversely, a negative gas pressure differential at the outlet can further reduce the critical flow rate, since the negative gas pressure acts to further retard the flow rate.
6. Add air permeation: air permeation may be added to the system actively via an air injection system or passively through a vent. In particular, adding judicious (often very small) amounts of air at the location in the press feed system where the interstitial gas pressure

is lowest can often be beneficial. However, this can be very unstable for small systems and/or very low-permeability materials.

7. Changing the PSD of the powder: the permeability of a powder is a strong function of its PSD. Powders with a finer PSD are often less permeable and, therefore, more prone to adverse two-phase flow effects. Even a reduction in the percentage of fines can often be beneficial in increasing the permeability of a powder, and decreasing the likelihood of adverse two-phase flow effects.

The key to implementing any corrective actions designed to reduce adverse flow effects will be using a mathematical two-phase flow analysis to assess the effects on the bulk solids' stresses and interstitial gas pressure [30]. This analysis utilizes inputs such as key flow properties (permeability and compressibility), equipment parameters, and process parameters (tableting rate and gas pressure gradients) to assess the effect of the potential corrective actions outlined above.

3.5.4 Minimizing Segregation in the Blender-to-Press Transfer Steps

It may be a challenging process for a designer to determine which segregation mechanism(s) is dominant and develop appropriate corrective actions. This requires knowledge of the material's physical and chemical characteristics, as well as an understanding of the segregation mechanisms that can be active. One must identify the process conditions that can serve as a driving force to cause segregation. Flow property measurements (wall friction, cohesive strength, compressibility, and permeability) can help to provide understanding of the behavior of the material in storage and transfer equipment. Consideration should be given to the fill/discharge sequence, including flow pattern and inventory management, which gives rise to the observed segregation. Testing for segregation potential (see Section 3.4) can provide additional insight about the mechanisms that may be causing segregation. Sufficient sampling is required to support the hypothesis of segregation (e.g. blend samples and final product samples, and samples from the center versus periphery of the bin). Finally, one must consider the impact of analytical and sampling errors specific to the blend under consideration, as well as the statistical significance of the results, when drawing conclusions from the data.

From the previous discussion about segregation mechanisms, it can be concluded that certain material properties, as well as process conditions, must exist for segregation to occur. Elimination of one of these will prevent segregation. It stands to reason, then, that if segregation is a problem in a process, one should look for opportunities to either: (i) change the material or (ii) modify the process equipment or conditions. It is realized that at certain milestones during

drug product development, changes in formulation of components and/or processing cannot be easily implemented. This chapter focuses on the equipment and process-design techniques to minimize segregation in batch processes.

Some generalizations can be made when designing equipment to minimize segregation. Complete details on how to implement these changes correctly are beyond the scope of this chapter. However, all equipment must be designed based on the flow properties and segregation potential of the blends being handled. Primary equipment and process-design techniques to minimize segregation during the final blender-to-press transfer include:

1. Minimize the number of material transfer steps. The tendency for segregation increases with each transfer step and movement of the bin. Ideally, the material would discharge directly from the blender into the tablet press feed frame, with no additional handling. In-bin blending is as close to this as most firms can practically obtain. This assumes that a uniform blend can be obtained within the bin blender in the first place.
2. Storage bins, press hoppers, and chutes should be designed for mass flow. In mass flow, the entire contents of the bin are in motion during discharge. In funnel flow, stagnant regions exist.
3. Minimize transfer chute volumes to reduce the volume of displaced air, and the volume of potentially segregated material. However, the chute must remain large enough to provide the required throughput rates.
4. Use a storage bin with a larger height: diameter aspect ratio. A mass flow bin with a tall, narrow cylinder will minimize the potential for sifting segregation, compared to that of a short, wide bin. A downside is that taller drop height may exacerbate other segregation mechanisms.
5. Bins and blenders should be vented to avoid gas counter-flow. Air in an otherwise "empty" bin must be displaced out of the bin as powder fills it. If this air is forced through material in the blender, it can induce fluidization segregation within the blender. A separate pathway or vent line to allow the air to escape without moving through the bed of material can reduce segregation.
6. Velocity gradients within bins should also be minimized. The hopper must be significantly steeper than the mass flow limit to achieve this. A steeper hopper section may result in an impractically tall bin. Alternate approaches include the use of inserts (discussed previously). If an insert is used, it must be properly designed and positioned to be effective. Asymmetric bins and hoppers should be avoided if possible, and symmetrical ones should be used whenever possible.
7. Dust generation and fluidization of the material should be minimized during material movement. Dust can be

controlled by way of socks, sleeves, or commercially available decelerators (e.g. contain the material as it drops from the blender to the bin).

8. Drop heights should be minimized where possible. Drop heights may aerate the material, induce dust, and increase the momentum of the material as it hits the pile. This will increase the tendency for each of the three segregation mechanisms to occur.
9. Valves should be operated correctly. Butterfly valves should be operated in the full open position, not throttled to restrict flow. Restricting flow will virtually assure a funnel flow pattern, which is usually detrimental to uniformity.
10. Use a symmetrical split whenever a process stream is divided. A symmetrical split, such as a bifurcated chute to feed two sides of a press (a Y-branch), will eliminate potential differences in the flow between the two streams. Consideration must be given to any potential for segregation upstream of the split. Even seemingly minor details, such as the orientation of a butterfly valve prior to a split, can affect segregation. Proper designs should be utilized for Y-branches to avoid stagnant material and air counter-flow.

Other specific solutions may be apparent once the segregation mechanism has been identified. As an example, mass flow is usually beneficial when handling segregation-prone materials, especially materials that exhibit a side-to-side (or center-to-periphery) segregation pattern. Sifting and dusting segregation mechanisms fit this description. It is important to remember that mass flow is not a universal solution, since it will not address a top-to-bottom segregation pattern. As an example, consider the situation in a portable bin where fluidization upon filling the bin has caused the fine fraction of a blend to be driven to the top surface. Mass flow discharge of this bin would effectively transfer this segregated material to the downstream process, delivering the coarser blend first, followed by the fines. In summary, when addressing segregation concerns it is crucial to know your process and how the blend will segregate before implementing equipment designs or corrective actions.

REFERENCES

1. Rippie, E.G. (1980). Powders. In: *Remington's Pharmaceutical Sciences*, A. Osol, G.D. Chase, A.R. Gennaro, M.R. Gibson & C. B. Granberg, (eds), 16th edn. Mack Publishing Company, Easton, PA. pp. 1535–1552.
2. Venables, H.J. and Wells, J.I. (2001). Powder mixing. *Drug Development and Industrial Pharmacy* 27 (7): 599–612.
3. Alexander, A.W. & Muzzio, F.J. Batch size increase in dry blending and mixing. In Michael Levin *Pharmaceutical Process Scale-Up*, 2nd edn. Taylor & Francis, (2005). pp. 161–180.
4. Royal, T.A. and Carson, J.W. (2000). Fine powder flow phenomena in bins, hoppers and processing vessels. Presented at Bulk 2000, London. Barham Press.
5. Bates, L. (1997). User Guide to Segregation. British Materials Handling Board.
6. Williams, J.C. (1976). The segregation of particulate materials: a review. *Powder Technology* 15: 245–251.
7. Williams, J.C. and Khan, M.I. (1973). The mixing and segregation of particulate solids of different particle size. *Chemical Engineer* January: 19–25.
8. Pittenger, B.H., Purutyan, H., and Barnum, R.A. (2000). Reducing/eliminating segregation problems in powdered metal processing. Part I: segregation mechanisms. *P/M Science Technology Briefs* 2 (1): 5–9.
9. Jenike, A.W. (1964, Revised 1980). Storage and flow of solids. Bulletin 123 of the Utah Engineering Experimental Station 53(26).
10. Schulze, D. (1996). Measuring powder flowability: a comparison of test methods Part I. *Powder and Bulk Engineering* 10 (4): 45–61.
11. Schulze, D. (1996). Measuring powder flowability: a comparison of test methods part II. *Powder and Bulk Engineering* 10 (6): 17–28.
12. Standard Shear Testing Method for Powders Using the Jenike Shear Cell. (2016). ASTM Standard D6128-16. American Society for Testing and Materials.
13. Bausch, A., Hausmann, R., Bongartz, C., and Zinn, T. (1998). Measurement of flowability with a ring shear cell, evaluation and adaptation of the method for use in pharmaceutical technology. Proceedings of the 2nd World Meeting APGI/APV, Paris, pp. 135–136.
14. Hausmann, R., Bausch, A., Bongartz, C., and Zinn, T. (1998). Pharmaceutical applications of a new ring shear tester for flowability measurement of granules and powders. Proceedings of the 2nd World Meeting APGI/APV, Paris, pp. 137–138.
15. Nyquist, H. (1984). Measurement of flow properties in large scale tablet production. *International Journal of Pharmaceutical Technology and Product Manufacturing* 5 (3): 21–24.
16. Ramachandruni, H. and Hoag, S. (1998). Application of a modified annular shear cell measuring lubrication of pharmaceutical powders. Poster presentation at the AAPS Annual Meeting, University of Maryland, San Francisco, CA.
17. Janssen, H.A. (1895). Versuche uber Getreidedruck in Silozellen. *Verein Deutscher Ingenieure, Zeitschrift* 39: 1045–1049.
18. Prescott, J.K., Ploof, D.A., and Carson, J.W. (1999). Developing a better understanding of wall friction. *Powder Handling and Processing* 11 (1): 27–35.
19. The United States Pharmacopeia. (2015). Chapter, 616. Bulk and Tapped Density. US Pharmacopeial Forum 28(3).
20. Carson, J.W. and Marinelli, J. (1994). Characterize powders to ensure smooth flow. *Chemical Engineering* 101 (4): 78–90.
21. Anonymous. (2010). Standard Practice/Guide for Measuring Sifting Segregation Tendencies of Bulk Solids, D6940-10. ASTM International.

22. Anonymous. (2012). Standard Practice for Measuring Fluidization Segregation Tendencies of Powders, D6941-12. ASTM International.
23. Ahmed, H. and Shah, N. (2000). Formulation of low dose medicines – theory and practice. *American Pharmaceutical Review* 3 (3): 9–15.
24. Globepharm's Powdertest™. <http://www.globepharm.com/html/powertest.html> (accessed September 2006).
25. Johanson, K., Eckert, C., Ghose, D. et al. (2005). Quantitative measurement of particle segregation mechanisms. *Powder Technology* 159 (1): 1–12.
26. Massol-Chaudeur, S., Berthiaux, H., and Doggs, J. (2003). The development and use of a static segregation test to evaluate the robustness of various types of powder mixtures. *Trans IchemE* 81 (Part C): 106–118.
27. De Silva, S.R., Dyroy, A., and Enstad, G.G. (1999). Segregation mechanisms and their quantification using segregation testers In: *Solids Mechanics and Its Applications*, Volume 81, IUTAM Symposium on Segregation in Granular Flows, pp. 11–29.
28. Hedden, D.A., Brone, D., Clement, S.A. et al. (2006). Development of an improved fluidization segregation tester for use with pharmaceutical powders. *Pharmaceutical Technology* 30 (12): 54–64.
29. Alexander, B., Roddy, M., Brone, D. et al. (2000). A method to quantitatively describe powder segregation during discharge from vessels. *Pharmaceutical Technology Yearbook*: 6–21.
30. Baxter, T.J. (2009). When powders flow like water: addressing two-phase flow effects in tablet press feed system. *Tablets & Capsules* 7 (2): 26–32.

4

DESIGN AND SCALE-UP OF DRY GRANULATION PROCESSES

HOWARD J. STAMATO

Global Regulatory, Safety and Biometrics, Research and Development, Bristol-Myers Squibb, Hopewell, NJ, USA

OMAR L. SPROCKEL

Engineering Technologies, Product Development, Bristol-Myers Squibb, New Brunswick, NJ, USA

The pharmaceutical industry continues to have a large proportion of solid drug products composed of powders. The composition, morphology, and flow properties of the powders are important to the manufacturability and patient use of the product.

Dry granulation, most often accomplished by roller compaction, is a way to modify powder properties quickly, without solvent, in a way amenable to rapid development and continuous manufacturing. This chapter provides an in-depth review of the relevant literature as well as a practical guide to completing the product design and commercialization via dry granulation. It is organized with an overview of the process components, material considerations and testing, and a stepwise look at the equipment, processes, and scale-up techniques.

4.1 OVERVIEW OF THE DRY GRANULATION PROCESS

This section gives the rationale for powder modification, the evolution of dry granulation as a process to modify powders, and an overview of roller compaction – the dominant technique in the pharmaceutical industry for dry granulation.

Granulation offers the opportunity to alter the properties of solid particulate material. An increase of particle size reduces potential hazards or nuisance from dust. The granule structure can improve the flow and compression properties of

the powders and helps assure a more uniform composition throughout the powder mass if a mixture of materials is used.

Nishii and Horio [1] have described how dry granulation can be achieved during mixing or alternatively as a powder is passed through a screen. As reported by Horio [2], particles can be dry granulated by a fluidization technique where the particle size of the starting materials is small, for example, lactose at a particle size of approximately three microns. However, granulation without a solvent to help with the binding is usually accomplished by subjecting powders to pressure.

Slugging is a form of dry granulation achieved in a way similar to tablet compression where pressure is applied to powders in a cavity formed by a set of tools and a die. Pietsch [3] described dry granulation by slugging among the methods and referenced a patent for a tableting machine from 1843. Because slugging and tableting use similar technology, the referenced patent shows how long slugging may have been available. When comparing slugging with tableting, it is apparent that: tablets are typically much smaller than slugging compacts and can be the final product, whereas slugging compacts are milled and reduced to granules feeding a subsequent tableting or capsule-filling operation. As the purpose of the slugging operation is to form granules with larger particle size, better flow, and improved compression properties, the feed to a slugging press often consists of very fine powder with poor flow properties when compared to a tableting operation.

Johnson [4] described dry granulation by roller compaction noting that passing powders through the nip of two rollers to produce larger-sized material was originally used in coal processing. One of the earliest patents for a roller compaction system, "A Method for Converting Fine Coal into Lumps," was issued in 1848. The largest volume applications of roller compaction are in the coal, soda ash, potash, calcined lime, and magnesium oxide industries. Bakele [5] gives an example of soda ash production of 160 000 tons/year. Discussion of roller compaction from other industries may focus on materials with different properties than those typical to pharmaceuticals and different parameters such as energy consumption [6, 7].

In pharmaceutical applications, Kleinebudde [8] has reviewed the advantages of granulation and the application of roller compaction to form granules used in tableting and filling of hard gelatin capsules. Mouro et al. [9] and Majerova et al. [10] showed how granulation via roller compaction improved the processing of low bulk-density materials in a capsule-filling operation. Bejugam et al. [11] showed how granulation by roller compaction eliminated punch filming in the downstream process of tablet compression compared to directly compressing the powder.

The roller compaction unit operation is comprised of several subprocesses. Once the components and composition have been selected, the powder to be compacted must be mixed so that the feed to the compactor is relatively homogeneous. The container with the prepared powder is placed in position to feed the roller compaction machine and the powder must be metered to the area in front of the compaction rollers. A variety of configurations including valves, stirring devices, and feed screws are used to convey the powder from the holding container to the roller compaction unit, and into the area in front of the rollers. The resulting powder bed is dragged by the rollers through the nip area where it is compacted. During the compaction the powders are deformed to a degree that causes the formation of a continuous sheet. The compacted output is commonly referred to as a ribbon but has also been referred to as a compact, flake, briquette, or tape.

The last step of the roller compaction unit operation is the reduction of the ribbon to granules. This can happen in one or more steps where mechanical elements impact the ribbon, break it into pieces, and force the broken pieces through a screen. The size-reduction subprocess of roller compaction may be integral to the roller compaction machine or performed as a separate operation. In either case the size reduction typically yields particles with a distribution of sizes. Because one of the objectives of the compaction process is to increase particle size, a large fraction of granules below a minimum size may be undesirable. In some cases fines are separated from the granulation and returned for another pass through the compaction step.

4.2 GENERAL CONSIDERATIONS FOR ROLLER COMPACTION OPERATIONS AND EQUIPMENT

In this section the component parts of a typical roller compactor and the selection of various change parts or equipment configurations based on some leading manufacturers' offerings are presented.

Selection of the process and composition for pharmaceutical products often includes matching the active pharmaceutical ingredients (API) and formula properties to the type of processing. Tests and batches at small scale to explore the feasibility of various techniques are recommended to assure efficient development and commercialization. Dalziel et al. [12] have described a screening design for an API with poor physical properties. Although the roller compaction process is used to improve the flow of material, the incoming feed needs to have sufficient flow properties to be delivered with as consistent a composition, flow, and density as possible. The requirements may include limitations on the micromeritic properties of the active ingredient.

The API must also be screened for the effect of significant pressures experienced during roller compaction on its downstream processability or polymorphic form. Hadzovic et al. [13] studied the effect of roller compaction on anhydrous theophylline monohydrate. Studies showed theophylline retained its polymorphic form and hydration as determined by differential scanning calorimetry (DSC). Furthermore, the theophylline behavior in tablet compression was essentially independent of the roller compaction. However, there was a difference noted between the hydrate and the anhydrous polymorphic forms of theophylline which could influence formulation choices.

Excipients are usually selected to dilute the API to an appropriate concentration and enhance various aspects of the powder properties. Materials referred to as binders often improve flow properties, enhancing powder feed to the roller compactor, and compression properties, to assist in ribbon formation (e.g. microcrystalline cellulose (MCC), lactose, etc.). Including these materials is often necessary to achieve the desired granulation properties. As with API the moisture content (Gupta et al. [14]) or crystallinity (Omar et al. [15]) may alter the performance of the material.

Materials used in small amounts can have a large effect on performance measures for the powders such as flow, compression, and particle strength. Flow aids (e.g. silicon dioxide) may be needed to achieve the desired mixture flow properties. A lubricant (e.g. magnesium stearate), typically added to modify interaction between the formulation powder/granulation and the equipment surfaces, may also be necessary to achieve the desired performance of the roller compactor. Other components used in small amounts are necessary for performance of the final product, rather than processing characteristics, such as disintegrants (e.g.

croscarmellose sodium, cross-linked polyvinyl pyrrolidone, or sodium starch glycolate) may also be used.

The next selection for the overall roller compaction operation is the unit operation that will prepare the powder blend for roller compaction. A diffusion mixer is one of the more popular choices for the mixing subprocess of roller compaction used to prepare the roller compactor feed. The batch size, sequence of loading, number and speed of revolutions, and the shape of the blender may be considered when designing the process. Multiple steps involving a geometric dilution strategy or intermediate milling between blending steps may be necessary to distribute the ingredients and achieve the appropriate level of uniformity and handling properties. Although a homogeneous powder feed will reduce variability in the roller compaction operation, content uniformity of the powder blend is not required at the same limits that would normally be applied to pharmaceutical dosage forms. Issues with blocking, bridging, segregation, or adhesion to equipment surfaces may result from poor selection of materials or insufficient blending prior to roller compaction.

The arrangement of the blender, bin, or other container relative to the roller compactor is also to be considered. Powder needs to flow evenly from the container to the equipment without bridging, flooding, or segregating between the bin and the roller compactor feed hopper. A general arrangement of the elements of a roller compaction process is shown in Figure 4.1.

Once the powder reaches the integrated roller compaction machinery, there will be various means of conveying the powder toward the rollers depending on the manufacturer, model, and user requirements. Typically, there is a feed hopper to receive the material from the main powder container. The hopper may have a device to break any powder bridges and help move the material for further processing. The hopper may feed directly to the rollers by gravity, but the majority of roller compactor designs also use one or more feed screws to deliver the powder and consolidate it in front of the rollers. In all cases the feed screw is oriented in-line with the roller nip. The orientation of the screw will depend on the orientation of the rollers.

The powder bed must be contained to assure it travels to the rollers and through the gap. The most typical system involves plates mounted to cover the powder bed before the gap which extend over the gap itself preventing powder migration outside the compaction area. It is important to assure proper mounting of the plates and sufficient maintenance to prevent the leakage of powder from the compaction area. Unprocessed fines could join the product stream if powder leaks from the seals. A different type of sealing system also uses ridges on the roller edges such that one roller fits into a channel in the surface of the other to form a tighter seal.

Kleinebudde [8] surveyed various manufacturers regarding their roller and screw configurations, offering a list of suppliers correlated to references focused on the various

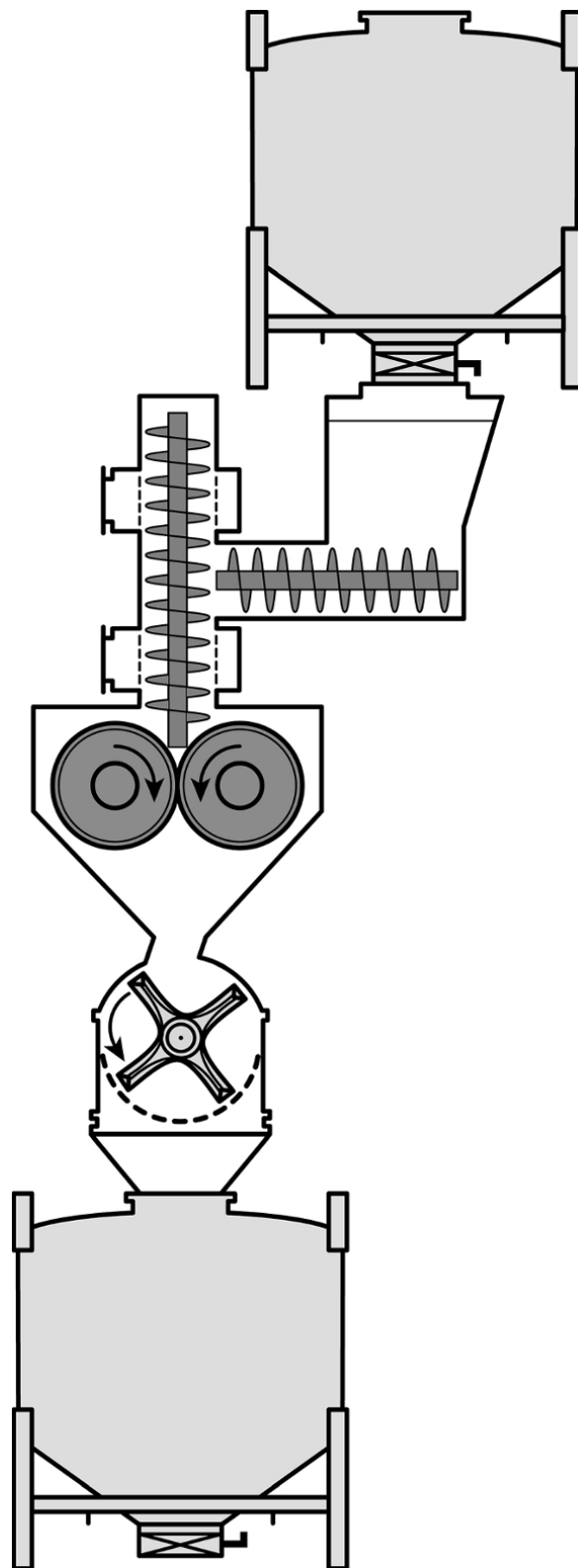


FIGURE 4.1 General arrangement of the roller compaction process. Courtesy: The Fitzpatrick Co.

designs. Some machines have the powder conveyed to the nip vertically such that the open space between the rollers - the gap - is vertical and a line drawn between the roller axes is horizontal. A system with this configuration may lose a small amount of powder through the nip as conditions are stabilized in the beginning of the run.

Other roller compaction machines have this configuration turned on their side so that the powder feed and the nip opening are oriented horizontally and a line drawn between the roller axes is vertical. The horizontal powder feeding configuration reduces the possibility of powder flooding the machine at startup and depends entirely on the action of the feed screw to move powder to the rollers.

A third configuration, used in other types of roller compactors, orients the powder flow and nip at an angle to the vertical giving some of the advantages of both the horizontal and vertical configurations. An illustration of the various configurations is shown in Figure 4.2. Typical roller compactors from laboratory to production scale from different manufacturers are shown in Figure 4.3. A unit shown in an expanded view so that all of the parts are visible is shown in Figure 4.4.

During roller compaction the porosity of the powder is reduced in the compaction step to form the ribbon. The air filling the interstitial space in the powder must escape through the forming compact and adjacent powder bed. In some types of roller compaction machines, powder is deaerated before reaching the nip by passing over a porous plate with vacuum applied underneath. The deaeration is intended to reduce any disruption from air moving through the powder as the compact is formed. Miller [16, 17] described the effects

of air entrainment and improvements in the roller compaction operation due to deaeration of the powder feed.

In the simplest case the roller faces in contact with the powder are smooth. However, other types of roller finishes, such as grooved or other inscribed patterns, have been used in an attempt to change the interaction between the powder and the surface of the rollers. Rollers with different surface finishes and the resulting ribbons are shown in Figures 4.5 and 4.6. Daugherty and Chu [18] studied the effect of roller surfaces on the compacted ribbon properties. Rambali et al. [19] also noted differences attributed to the use of smooth versus grooved rollers. Further study by Tan et al. [20] showed that the engagement of material in the rollers as measured by the nip angle is altered by the roller surface treatment. Pietsch [3, 21] described a briquetting application in which the faces of the rollers are indexed so that shaped indentations in each roller face form the desired final product shape during the compaction step. Rollers with internal channels for cooling liquid may be used for temperature-sensitive products which might melt from the energy of compaction as mentioned by Pietsch [3, 21].

Even on a relatively small compaction machine, the roller pressure may be set to a nominal value of 50–70 bar. This pressure is typically applied by use of a hydraulic cylinder pushing on the roller axle. The pressures involved are in a range where deflection of the axle can affect the operation of the machine. Most manufacturers of roller compaction machinery have carefully considered the design of the hydraulic system and axle deflection so that the width of the gap is consistent across the roller width. It should be noted that the pressure on the powder may be different than the

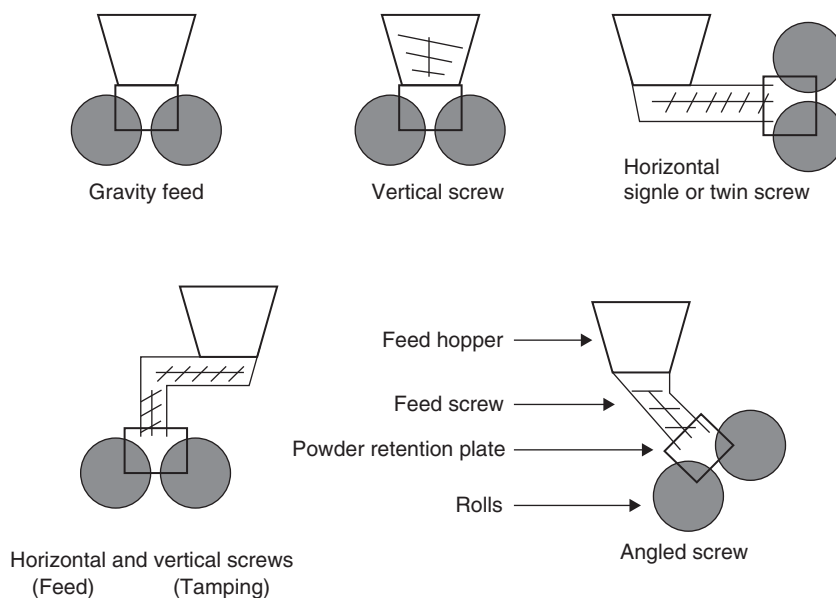


FIGURE 4.2 Common feed, feed screw, and roller configurations.

(a)



(b)

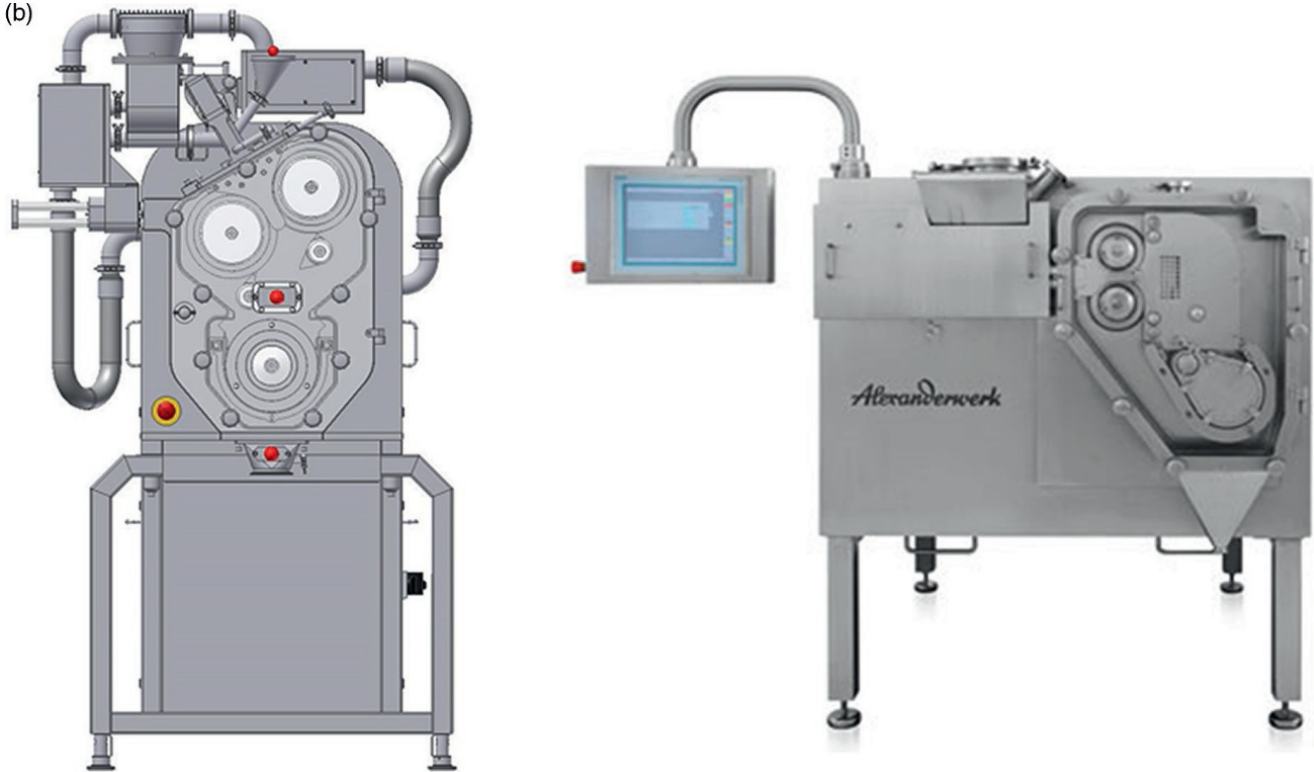


Figure 4.3 (a–c) Examples of roller compactor units from laboratory to commercial scale. Courtesy: Freund Vector Corporation, The Fitzpatrick Co., Gerteis Maschinen + Processengineering AG, Alexanderwerk AG. (a) Laboratory-scale roller compactors (images not to scale). (b) Pilot-scale roller compactors. (c) Production-scale roller compactors (images not to scale).

(c)



Figure 4.3 (Continued)



FIGURE 4.4 Expanded view of a roller compactor showing the components. Courtesy: The Fitzpatrick Co.

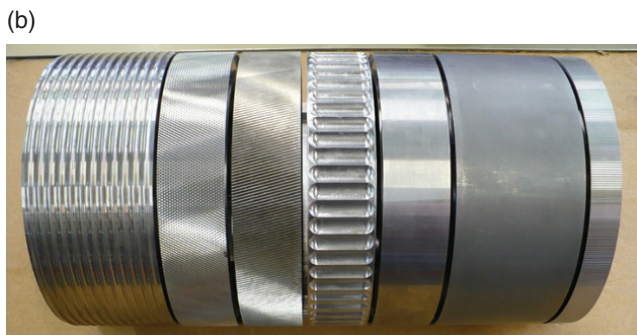
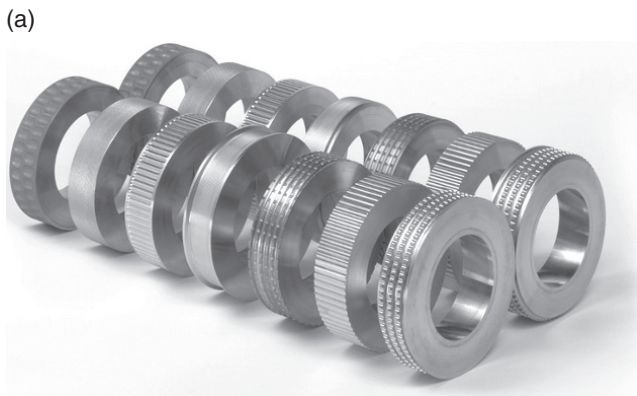


FIGURE 4.5 (a and b) Rollers with different surface treatments. Courtesy: The Fitzpatrick Co. and Gerteis Maschinen + Process-engineering AG.



FIGURE 4.6 Examples of product produced from rollers with different surfaces. Courtesy: The Fitzpatrick Co.

nominal pressure set on the hydraulic system although the two values are related.

Guangsheng et al. [22] described an application for roller compaction of magnesium alloy powder to form sheets. In this case, after some post treatment, the roller-compacted ribbon is the final product. However, in most roller compaction operations the ribbon itself is not the useful product and is broken into granules that have the desired properties for further processing. Although used in some high-volume applications, briquetting rollers, as described by Pietsch [21], are not used to form the final product in pharmaceuticals due to the need for tight uniformity of the finished product.

Different types of output from roller compaction operations including broken ribbons with different surfaces, granules, and long briquettes are shown in Figure 4.7.

The size-reduction step for the ribbon typically consists of an impeller without a screen to start the process followed by one or two stages of milling, this time with an impeller mounted in proximity to the screen. The mill configuration may take the form of an oscillating screener, hammer mill, or conical mill. The screening mill may or may not be integrated with the roller compaction unit.

Fines may be present in the resulting granulation from several sources. The sealing mechanism at the edges of the rollers may not contain all of the powder or powder may not be



FIGURE 4.7 Examples of roller compactor output. Courtesy: The Fitzpatrick Co.

completely incorporated into the ribbon and is carried over into the final product. The milling operation itself can create fines. In some cases the amount of fines are considered unacceptable for the final granulated material.

For operations where more fines are present, the equipment may be fitted with mechanisms to separate the fine fraction. Some fines reduction can be accomplished by collecting powder by use of a chute into which the fines fall while the ribbon is collected separately. However, this method will not completely assure a low fines fraction as some fines are generated while milling the ribbons. In other cases the entire output after the milling operation is size separated to remove fines from the granulation caused by the milling operation along with any carryover of uncompact fines. In both cases the undersized material may be returned to the roller compactor to improve yield of granules with the desired size. Some roller compaction systems are specifically designed to segregate fines and return them for further processing (Sandler and Lammens [23] and Saarinen et al. [24]).

It should be noted that repeating the compaction step can reduce the performance of the granules in subsequent processing. If reprocessing is necessary, the number of passes through the roller compactor should be limited and the combination of reprocessed material with the main granulation should be monitored. Material behavior and the characterization of raw material, finished granules, and compacted ribbon are discussed in the next section.

4.3 MATERIAL BEHAVIOR, QUALITY ATTRIBUTES, AND TESTING

4.3.1 Material Behavior During the Roller Compaction Process

The performance of the roller compaction unit process and the resulting products (e.g. ribbon, granules, tablet, or capsules) depend on the properties of the powder being processed. Some considerations of the choices for the roller compaction unit operation related to material behavior are described in this section.

4.3.1.1 Powder Handling of Feed and Granulated Product Powder handling is important in the feed of the roller compaction unit and also for the granules produced by roller compaction when they are processed downstream. The particle size distributions, densities, and flow properties of the powders going into the roller compactor must be matched, manipulated, or modified by additives (e.g. lubricants and flow aids) and/or preprocessing (e.g. milling and blending) to feed material evenly, and with the appropriate levels of the components to the roller compactor.

The powder mix will have to maintain these properties throughout the compaction run. Blend consistency may need

to be held for an extended time, because roller compaction is typically performed as a semicontinuous operation. The powder to be compacted will be subjected to various conveying and consolidation operations which may induce segregation due to material or air motion, and vibration. Bacher et al. [25] reported that the shape and size distribution of calcium carbonate and sorbitol used to prepare roller-compacted granules affected the granule content as a function of granule size.

The output granules must have appropriate flow, resist segregation, and have sufficient strength to withstand handling when being conveyed to various downstream process or packaging into the final product presentation. Sophisticated control schemes with combinations of control methods (e.g. feedback, feed forward, multivariate, and model based) have been developed for continuous operation of an entire process train, which could be applied to roller compaction, as reported by Singh et al. [26, 27].

4.3.1.2 Compaction In addition to feeding performance, raw material properties can also have an effect on the compaction of the powder into ribbons. Material properties were considered as an input for the earliest mathematical models, such as Johanson [28], who suggested measuring the internal and surface friction of the powder, to more recent studies including process analytical technology (PAT), such as Soh et al. [29, 30], who studied 20 material parameters and suggested particle size, span (a measure of size distribution), and angle of fall as the most interesting measurements to characterize the input material.

Many studies have focused specifically on the behavior of an excipient or class of excipients in the roller compaction operation and have given insight into excipient selection. In other studies, a formulation including a model-active ingredient was used to simulate the response of the excipient under typical-use conditions (Gupte et al. [31], Allesso et al. [32]). Other researchers are attempting to aggregate data across formulations and processing techniques such as the work of Leane et al. [33]. Shi and Hilden [34] aggregated data across 136 formulations to predict roller compaction parameters.

Because the material in a roller compactor undergoes some of the same physical processes as in a tablet press, many of the same considerations of material behavior apply. For example, Sheskey and Dasbach [35] noted that slower roller speeds, giving a longer dwell time under pressure, allowed plastically deforming materials to perform better as binders. Falzone et al. [36] noted differences in behavior between plastically deforming MCC and lactose which exhibited behavior influenced by brittle fracture.

The number of studies of specific materials, and combinations of materials, in roller compaction continues to grow. A summary is presented in Table 4.1. Of note are publications related to roller compaction of APIs so that the entire formulation is considered.

The addition of a small amount of components to influence the powder properties is common practice, but may have an effect on the compaction operation. He et al. [53] studied the effect of lubricating the roller compactor feed with magnesium stearate on the strength of tablets compressed from

TABLE 4.1 Selected References Reporting Material Behavior in Roller Compaction

Material	References
Microcrystalline cellulose	Inghelbrecht and Remon [37] and Mitra et al. [38]
Lactose	Riepma et al. [39], Inghelbrecht and Remon [40], and Omar et al. [15]
Hydroxypropylmethylcellulose	Sheskey et al. [41]
Hydroxypropylmethylcellulose, mannitol	Heiman et al. [42]
Hydroxypropylcellulose, methylcellulose, polyvinylpyrrolidone, starch, and microcrystalline cellulose	Sheskey et al. [41]
Magnesium carbonate	Freitag et al. [43]
Magnesium carbonate in combination with powdered cellulose	Freitag et al. [44]
Hydroxypropylmethylcellulose, hydroxypropylcellulose, microcrystalline cellulose, and polyvinylpyrrolidone	Herting et al. [45]
Hydroxypropylmethylcellulose, microcrystalline cellulose	O'Mahony et al. [46]
Methacrylic acid copolymers	Dave et al. [47, 48]
Calcium carbonate and sorbitol	Bacher et al. [25]
Polyethylene oxide, hydroxypropylmethylcellulose	Gangurde et al. [49]
Microcrystalline cellulose co-processed with dicalcium phosphate, polyvinylpyrrolidone-vinyl acetate	Gupte et al. [31]
Theophylline	Hadzovic et al. [13]
Clopidogrel	Khomane and Bansal [50]
Ibuprofen	Matji et al. [51] and Lopes et al. [52]

the resulting granules. The authors noted that a lubricant is often added to roller compaction feed to prevent sticking to the rollers. One caution offered was that the lubricant can change the interaction of the powder and rollers resulting in a smaller nip angle and reduced process efficiency. It was found that tablets produced from roll-compacted granules of un-lubricated MCC did not show a significant decrease in strength at two of the three levels of ribbon densification studied. At the highest level of ribbon densification there was approximately a 30% drop in tablet strength. For MCC lubricated with 0.5% magnesium stearate, the reduction in tablet strength was 90% over a much broader range of ribbon densities.

More recently Dawes et al. [54] showed how the presence of magnesium stearate is coupled with roller surface type. An increased throughput for rough surfaces and a reduction in input for smooth surfaces was found for lubricated materials ostensibly due to more slip of the lubricated powder at the surface of the smooth roller. A reduction of feed-screw efficiency was also noted for powders with higher lubrication, attributed to slip at the machine interface. A counteracting phenomenon was noted when reduced sticking at the roller increased the effective gap width seen by the powder and allowing higher mass flow. Pre-lubricating the machine surface had the same effects as powder lubrication. A system to continuously lubricate the roll surfaces was also investigated [55] and had similar effects with the option of adding less magnesium stearate as it was only dosed to the equipment surface.

Many excipients can absorb water from the environment under typical processing conditions. The moisture can affect the material properties, perhaps causing unwanted effects in the roller compaction operation. The effect of moisture on the behavior of MCC in roller compaction with ambient relative humidity ranging from 15 to 75% was studied by Gupta et al. [14]. It was found that as the water content increased, the powder yield strength decreased, indicating better powder rearrangement; tensile strength of the resulting ribbons decreased, indicating poorer bonding. Inghelbrecht and Remon [56] intentionally added water to blends of lactose, MCC, hydroxypropylmethylcellulose, and other ingredients in order to reduce the amount of fines in the roller-compacted granules. They found that the fines produced during the compaction step were negligible and the fines fraction produced during the milling operation were reduced. The resulting granules and tablets had lower friability and the tablets had higher strength.

The effect of a second component to soften materials was shown for methacrylic copolymers plasticized by triethyl citrate (TEC) in the work of Dave et al. [48]. Roller-compacted granules were typically larger for higher TEC content. Tablets made from granules formed with higher roll pressures had a slower release profile for the active ingredient.

A more typical strategy to reduce fines is the recycle of the fine fraction to the inlet of the roller-compaction step. However, the change in material properties when subjected to roller compaction can be cumulative and has been reported by many investigators. Bultmann [57] studied this phenomenon as a function of the number of repeated roller compactions for MCC and found the repeated compactions not only decreased the amount of fines but also decreased the compressibility of the resulting granules. Up to 10 cycles of roller compaction were studied with most of the losses in material properties seen in the first and second processing by the roller compactor.

Sheskey et al. [35, 41] found a similar reduction in compressibility for hydroxypropylmethylcellulose which decreased with increasing roller pressure after a single pass through the machine. In contrast, Riepma et al. [39] found that for material displaying brittle fracture characteristics, such as lactose, dry granulation had minimal influence on the compatibility during subsequent tableting. However, lactose was found to increase in ribbon strength with increasing content of amorphous lactose by Omar et al. [15].

Insight into the behavior of materials in the milling process is gained by determining the breakage rate experimentally [58, 59] or a breakage function as part of an effort to model the milling step [60].

A review of specific material attributes and their effect on the products from a roller compaction process are discussed in the following sections.

4.3.2 Material Attributes and Testing

Powder properties continue to be a fascinating topic as they are difficult to predict. Powder behavior has a strong dependency on environmental conditions of storage and handling or imposed by the process. In this section the relative importance of material properties important when considering roller compaction are presented. In the language of Quality by Design used in the pharmaceutical industry, these material properties are considered Material Attributes.

The roller-compacted materials (ribbons, briquettes, or granules) depend on input material characteristics as well as on process parameters. For both the input powder performance in the hopper, feeder, slip, and nip region as well as the downstream performance of the post-milling processes, the flow of the material is paramount. Measures of size, size distribution, shape, response to shear, porosity, powder flow, compressibility/compactability, and friability can be used to select materials which mitigate risks and predict ranges for acceptable performance.

The ribbon and downstream tablets or capsules have the compact strength, porosity, breakage mode, breakage rate, and friability as useful quality attributes for determining performance, whether in the milling of ribbons or the packaging, shipping, and use of tablets. The material has a macroscopic

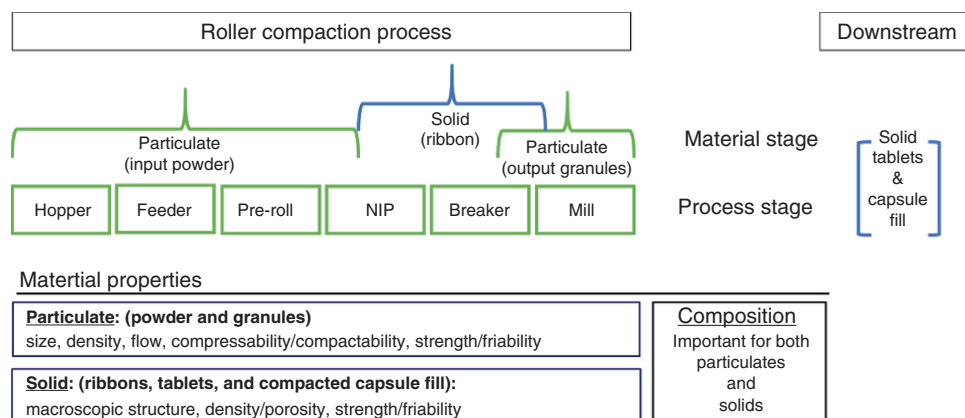


FIGURE 4.8 Visualization of the roller compaction process, the material structure at each stage, and a brief listing of relevant material properties.

structure at these stages of manufacture. Changes across the structure of the product are often key to understanding what happened to the material during the process. This information can be used to optimize the formulation and process parameters.

A visualization of the roller compaction process, the material structure at each stage, and a brief listing of relevant material properties are presented in Figure 4.8.

Tremendous growth in process analytical testing, which has increased in sophistication and accuracy, is evident in the recent literature. The interest in this area, developing techniques off-line and then moving them to at-line and eventually online, enables the drive toward continuous manufacturing of pharmaceuticals. An improvement in quality and consistency of the process performance is possible by having online sensors providing the data to adjust the parameters, thus reducing the effects of feed and process variability. Research regarding PAT is mentioned throughout the chapter and in this section according to its measurement. An overview of PAT techniques for continuous manufacturing of powders was published by Fonteyne et al. [61]. The market for new sensors changes more rapidly than updates in reviews of these techniques and so a current market and literature search is recommended when selecting online techniques.

Other sections of the chapter are organized, for simplicity, by their position in the roller compaction unit operation. This section describes material properties, and is therefore organized from a material standpoint starting with the small scale of structure, particles whether powder or resulting granules, and ending with the largest scale of structure, a solid form whether ribbons or tablets.

4.3.3 Particle Properties

Particles occur in the roller compaction process at the inlet, with the ribbon as a small amount of unincorporated fines,

and as granules resulting from the milled ribbon. In each case the size, size distribution, density, flow properties, friability, and compaction/compression characteristics have an impact on the roller compaction process.

Kona et al. [62] looked across the entire process to relate the input material properties to the characteristics of the resulting tablets. Near infrared spectroscopy (NIR), a common PAT technique, and standard measurements were used to characterize the relation between process parameters and powder properties with results seen in the finished tablets. The multivariate model generated was able to distinguish the effect of roll pressure on tablet compression force. Prediction of tablet crushing force and disintegration time were possible from the upstream measurements. Interestingly, the model and measurements detected distinct differences when running on a different roller compactor of the same make and model. This is a phenomenon experienced by the authors. While this situation is often difficult to resolve, some troubleshooting has shown that minor differences in construction and calibration from machine to machine could have caused the discrepancies (Authors own experience, and personal communication with Vishwas Nesarikar who was key to efforts using the Johanson model and instrumented rolls when all three worked at Bristol-Myers Squibb to resolve technology transfer challenges.).

4.3.3.1 Flow For processing both in and downstream from the roller compaction operation, particulate flow dominates process stability: the continuous running of the operation with little variation in the input and output measurements. There are several techniques used to gauge the ability of a powder to flow. A mass flow determination can be obtained by measuring the flow time of a certain mass of powder passing through a certain orifice. An alternative method would be to determine the minimum orifice opening that would support continuous flow (Flow index). Other

techniques rely on indirect means to assess flow, such as Carr Index, shear cell, and other measures of powder rheology. New testers and techniques are developed and should be reviewed for their applicability to roller compaction, for example, in the asphalt industry a shear box compactor test described was by Qiu et al. [63].

Miguel-Moran [64] showed that the size of the three flow regions in the roller compaction zone and the transition from one flow region to the other (such as the areas where powder slips along the surface of the roller to where the powder travels with the surface of the roller) depends on the properties of the input powder. Examples of powder properties studied that affect the powder flow behavior in the compaction zone are: internal friction, cohesion, and friction between the powder and the rollers and side shields.

Downstream from the roller compactor, flow of the powder blend into the compression operation controls the variability in tablet weight. Weyenberg et al. [65] reported that the fastest powder flow was obtained with a combination of low roller speed and high roller pressure. These conditions yielded ribbons with high strength that resisted attrition during the milling operation. The rate of powder feeding had a minor impact on granule flow properties. These conditions also produced the largest and strongest granules with the lowest friability. Granule size increased with higher roller pressure, lower roller speed, and higher powder feed in the order of importance.

Freeman et al. [66] used the FT4 apparatus, and correlated the rheological parameters measured to variance in die filling during subsequent tablet manufacture. Roll compaction force, roll gap, and initial powder bulk properties correlated well with the flow measurements, operational characteristics in tablet manufacture, and the resulting tablet quality attributes.

4.3.3.2 Particle Size Particle size and size distribution are common parameters to characterize both the input powders and resulting granules. Traditional techniques require taking a sample and testing it by size exclusion with a stack of sieves or an offline instrument of a variety of designs. The measurement may or may not be representative due to the difficulty of sampling.

There are several examples in the literature of technologies for determining the particle size of powders online. Zhang and Yan [67] made the point that development of cost-effective online particle size instruments is challenging based on their use of an electrostatic sensor combined with digital signal processing to determine particle size distribution. A slip stream of particles was carried on an air stream toward the sensor causing the particles to acquire a charge; the magnitude of the charge was size dependent.

NIR continues to be used for determining particle size and correlating the spectral signal to process performance with studies reported by Frake et al. [68], Rantanen and Yliruusi

[69], Gupta et al. [14], Kona et al. [62], and McAuliffe et al. [70]. Difficulty in developing accurate techniques was also shown in a study by Rantanen et al. [71] where the NIR spectrum was affected by physical properties such as surface conditions and density.

Two additional means of ascertaining the particle size distribution of the granules as they exit the milling chamber post-ribbon production are laser light scattering (LLS) and high-speed image analysis. For the LLS method, a slip stream sampling of the falling granules is needed. Traditional LLS techniques can be used on the sample of granules [72]. Image analysis would require high-speed image capture that is processed against a predictive model of pixels versus volume or diameter. Liao and Tarn [73] developed a high-speed optical inspection system to determine particle size. A Charge Coupled Device (CCD) or Complementary Metal-oxide Semiconductor (CMOS) camera was used to acquire the image that was processed and analyzed against a reference. Other researchers have used different imaging instruments such as the Eyecon or Camsizer to determine particulate particle size [70].

Several authors have reported on the importance of selecting the proper diluent particle size. Herting et al. [45, 74] reported on the effect of MCC particle size on granule and tablet properties. Under similar processing parameters, a reduction in the particle size of the input material (MCC and theophylline) resulted in larger mean granule size and higher compactability. Inghelbrecht and Remon [37] evaluated the effect of MCC particle size in ibuprofen/MCC drug mixtures. They found that smaller MCC particles produced stronger granules. The irregularity of the MCC particles was ascribed a secondary role for the differences seen in granule strength.

In evaluating the effect of particle size of sorbitol on granules' properties, Bacher et al. [25] found that smaller-sized sorbitol produced granules that had higher compactability due to the increase in surface area with smaller particles of sorbitol. Inghelbrecht and Remon [40] evaluated the influence of lactose particle size as well as the type on granule and tablet properties. As with MCC and sorbitol, they found that reducing the lactose particle size (regardless of type) improved the granule quality (less friable).

The work cited above shows that selecting a smaller particle for the input material (MCC, sorbitol, or lactose) results in granules that are larger, stronger, and less friable. The effects were ascribed to the increased surface area.

As with input materials, particle size of the resulting granules is often used as a metric to determine input or parameter effects. The granule size is controlled by the roller compaction process as well as downstream processing. The initial granule size is a result of the input material properties and the compaction and milling process parameters. Granule attrition during post-roller compaction processing (blending) will determine the final particle size distribution.

4.3.3.3 Strength/Friability Processing downstream from the roller compaction and milling operations can subject the granules to stresses causing attrition or breakage. This is assessed by determining the friability of these granules by measuring the change in particle size distribution under a specified challenge (e.g. due to additional mixing). There are two ways by which the friability of the granules can be determined. An indirect way is to compare the particle size distribution of the granules after milling with that of the final blend after additional mixing. After accounting for any extragranular materials added, the change in size distribution would be indicative of granule friability.

Inghelbrecht and Remon [37] quantified granule friability by tumbling granules of a particular particle size range with glass beads for certain duration, and then determined the change in particle size. The reduction in mean particle size was ascribed to particle attrition. Patel et al. [75] used stress-strain analysis on single particles using a 2 mm flat probe to determine the particle fracture potential. Inghelbrecht and Remon [40] ranked the process parameters for their effect on granule friability in order of decreasing influence: roller pressure, roller speed, and feed rate.

4.3.3.4 Particulate Material Density Soh et al. [29] and Freitag et al. [43] investigated the effect of raw material attributes on their performance in roller compaction. One attribute tracked was material density. Since, material density was confounded with other material properties, such as particle size or morphology, a clear relationship between input density and output (ribbon or granules) properties was difficult.

Traditional density measurement techniques typically used include bulk density, tap density, and true density by helium pycnometry. Derived parameters, such as porosity, Carr Index, Hausner ratio, also have been used [29].

4.3.3.5 Compressability One of the common objectives of a roller compaction process is to improve the performance of granules in a downstream capsule or tablet operation. Compressability, the ability to form tablets of a desired strength at an acceptable pressure, is one of the main attributes to consider for the roller-compacted granules. Compressability is typically studied by experiments on a tablet press or compaction simulator to assure that the desired tablet properties can be achieved.

Herting and Kleinebudde [74, 76] characterized the hardening of granules post-roller compaction by measuring the yield pressure. They showed an increasing relationship between the applied roller pressure and the apparent yield pressure. They ascribed the increased resistance to deformation as granule hardening.

Several other, earlier, authors report a reduction in powder compressability of the roller-compacted material [77–79]. Investigators have identified two possible causes for this reduction: work hardening and size growth. Malkowska

and Khan [78] describe the effect as a loss of bonding capacity between the particles. Plastically deforming materials were found to be more susceptible to work hardening.

Sun and Himmilspach [77] ascribed the reduction in compressability to an enlargement of granule size relative to the input particle size. The growth in particle size reduces the area available for bonding. Herting and Kleinebudde [74, 76] concluded from their investigation that the reduction in compressability is due to both size enlargement and hardening of the granules. They observed that work hardening could be countered by producing smaller granules.

Mitra et al. [38] attempted to settle the question of whether increased tablet strength is due to size, surface area effects, or material properties of the particle by preparing small compacts with uniform geometry as model granules with a single size. The compacts, simulating different particle size granules, were varied in porosity, thus the work on the material resulting in a particular porosity could be separated from surface area effects due to particle size. It was found that at low porosity, where the material sample had more energy imparted and was less likely to deform further, led to weaker tablets. When powders with multiple sizes and porosities were adjusted to account for the higher porosity of fine particles, the tablet strength was as predicted by the overall porosity.

4.3.4 Ribbon or Solid Compact Properties

Solid structures are created in the roller compaction process as ribbons or downstream from the roller compaction process as the final tablets produced from granules made by roller compaction. The dimensions, strength, density, and variations in properties as a function of position are all important to the performance of the materials in the process or for the patient.

4.3.4.1 Thickness Some Investigators use a micrometer to measure the thickness of the ribbons produced by roller compaction [64, 80]. This measurement has to be repeated at several places across the ribbon sample to account for variation in the ribbon thickness. This average thickness then is used for density calculations using the sectioning technique or for feedback control of the roller compaction process.

In lieu of measuring the ribbon thickness, some investigators use the roller gap as a surrogate. This approach ignores the relaxation that may occur post consolidation.

Alleso et al. [81] used a dual laser technique to determine ribbon thickness, and with a known ribbon width and sample weight, the resulting density.

Rambali et al. [19] showed that the mean granule size produced from thicker ribbons was smaller than that for thinner ribbons. However, the ribbon thickness effect on granule size was marginal. Another study by Sun and Sun [82] found that the thicker ribbons had a higher percentage of fines in the

granules, ostensibly due to lower material density at the center of the ribbon.

4.3.4.2 Strength Ribbon strength is an attribute that is indicative of performance during the milling operation post-ribbon production. Solid fraction can sometimes be used as a surrogate for ribbon strength within a single composition. However, when comparing ribbons made from different materials, solid fraction is insufficient as a sole descriptor of quality.

As a measure of strength Zinchuk et al. [83] argued that a ribbon's resistance to milling post-ribbon production is a better metric. They also determined the tensile strength of ribbons using a three-point beam bending analysis and related it to the granule output. Miguelez-Moran et al. [80] used a microindentation technique to determine the hardness of ribbons at a microscale. The size of the indentation made depends on the shape of the indenter, the force used to make the indentation, and the hardness of the ribbon.

Within a composition, the ribbon tensile strength or hardness varies directly with the solid fraction of the ribbons. Similar to tablets, the ribbon tensile strength or hardness is directly proportional to the pressure used to make the ribbons. Farber et al. [84] hypothesized that during roller compaction the particles deform under load, causing them to interlock. This is the reason for the increased strength of ribbons post-compaction.

4.3.4.3 Density The solid fraction of the ribbons produced is a result of powder properties, process parameters, and equipment geometry factors. A method of determining the solid fraction by oil absorption was reported by Khorasani et al. [85]. Other researchers [70] have used a helium gas pycnometer (AccuPyc). The increase in solid fraction is related to ribbon density and affects the mechanical properties of both the ribbons and granules.

Two physical density measurement techniques were reported by various authors for determining ribbon density: sectioning and enveloping. With the sectioning method, a portion of the ribbon is cut and carefully measured to calculate the volume. The mass is measured, taking care to keep the sample intact and density is then calculated. Variability in the data can be encountered by the difficulty of accurately measuring the dimensions or keeping the sample together for determining mass.

With the enveloping method, a volume displacement approach, the sample is immersed in a chamber of defined dimensions with material of known density. A popular instrument used by many researchers is the GeoPyc [70]. Using this type of method reduces variation as one less measurement is necessary to generate the results. In determining the ribbon solid fraction, Soh et al. [30] argued that envelope volume is required for higher precision rather than using sectioning.

The reason stated for this is the imprecise nature of the edges during sectioning.

Various techniques for measuring the surface with three-dimensional laser scans are also available with potential benefit to roller compaction [86]. Lasers have been used to determine the ribbon thickness and calculate the density of roller-compacted ribbons [81].

Two general methods were reported by various authors: physical (sectioning, enveloping) and associative (ultrasound, NIR). The physical methods were described previously and can be used only at-line. Several investigators have reported on the associative methods via ultrasound [86, 87] and NIR [14, 70, 85, 87, 95]. As discussed for other applications of NIR, there is variability in the measurements which has to be carefully considered and addressed.

Even though the fundamental principles of the associative methods may differ, most rely on a correlation or predictive model that relates density to the underlying measurement. The associative methods lend themselves to online or at-line measurements, whereas the physical methods can only be used at-line.

Other associative methods have also been explored. For example, Austin et al. [92] in addition to studying NIR also explored microwaves to simultaneously determine moisture content and density of ribbons.

Ghorab et al. [96] reported on their evaluation of the relationship between ribbon physical attributes (such as density and strength) and thermal effusivity. They found strong correlations between thermal effusivity and ribbon density or strength. The relationships were first- or second-order polynomials depending on the composition. Even though these relationships are composition dependent, the utility of this method is intriguing.

There are always new techniques to consider. Herting et al. [45] reported a means of calculating the in-gap porosity by calculating the volume of ribbon produced per unit time. The mass of granules corresponding to this time unit was used to calculate the ribbon porosity. Other techniques might be adopted from other industries such as the roller vibration frequency analysis to determine substrate density described by Commuri et al. [7].

In place of the techniques cited above that provide the average overall density of a ribbon specimen, Miguelez-Moran et al. [80] and Akseli et al. [88] used X-ray micro-computed tomography to obtain the distribution of the densities in a ribbon sample. This enabled the investigators to determine the effect of roller-compaction process parameters on average ribbon density and also the density distribution within the ribbon.

The quality of the ribbons produced (density, solid fraction, and strength) has a direct impact on the granule size distribution. Ribbons with higher density and higher solid fractions (lower porosity) yielded granules with larger mean sizes and better flow [45, 74]. Under similar milling

conditions (mill speed and screens) ribbons with higher strength produced larger granules compared to ribbons with lower strength [37, 40, 65, 97].

Farber et al. [83] hypothesized that during roller compaction the particles deform under load, causing them to interlock. Upon milling, the break in the ribbon occurs at the weakest interlocking junction, but the deformed particles remain intact in the resulting granules.

4.3.4.4 Macroscopic Variation The packing of particles in a ribbon after roller compaction can be described by three related terms: density, solid fraction (density relative to the true density), and porosity (measured or calculated from the solid fraction). Ribbon solid fraction is an attribute that indicates the degree to which the powder has been compacted [83]. The strength of ribbon varies across the width of the ribbon, due to the variation in ribbon solid fraction across the width of the ribbons. This variation is due to a nonuniform distribution of stress across the roller width.

Guigon and Simon [98] used a series of pressure transducers on the rollers to obtain the distribution of the pressure on the rollers during roller compaction. They observed that the pressure varied with a period that coincides with the screw feeder pushing the powder into the nip region. Guigon and Simon [99] showed this variation in applied force across the rollers visually by adding charcoal particles to lactose. The areas of high stress were identified by the large number of broken charcoal particles.

The density across the width of a ribbon can vary [80].

Density is often highest in the center and lowest at the edge. This density gradient is caused by the friction between the powder and the face plates covering the rollers [99] and can be reduced by the inclusion of lubricant. This result was reproduced in the work of Akseli et al. [88], who found that the effect decreases with increasing powder lubrication as well as Talwar et al. [95]. Funakoshi et al. [100] showed that density distributions are related to the force distribution across the ribbon.

Funakoshi et al. [100] and Parrott [101] evaluated an alternate roller design (concave–convex roller pair) to address the nonuniform distribution of pressure on the rollers during roller compaction. Several incline angles were evaluated and a 65° angle was shown to be optimum for a uniform pressure distribution over the roller surface [100]. Some manufacturers have adopted design features which account for the phenomena of flow observed in these results.

The force distribution is related to the flow patterns of the powder passing between the rollers [64, 85, 93]. This effect is accentuated at high roller speeds as described by Funakoshi et al. [100] and Miguelez-Moran et al. [64]. The mean ribbon density is higher at a narrower gap setting compared to a wider gap setting. Sun and Sun [82] had similar results. Guigon and Simon [99] also showed that if the screw feeding in a roller compactor is nonuniform in time or space, this

produces ribbons whose solid fraction varies over the width of the ribbon.

4.3.4.5 Composition and Uniformity for Particulates and Solids

Across the roller compaction and downstream processes it is important that uniformity of the material be maintained, so that the excipients can perform their function and the active is dosed at the right level. The distribution of components can also give clues to the conditions in the process and how to improve the process performance. The improved capability in spectroscopic methods has led to an increase in the ability to determine the distribution of components within granules, ribbons, and tablets. As discussed later in Section 4.4, NIR chemical imaging data has been used to enhance images capturing variation in ribbon density along the width and length of ribbons.

In an example relating uniformity to processing, Bacher et al. [25] investigated the cause for the nonhomogeneous distribution of calcium carbonate in roller-compacted granules containing sorbitol. They showed that the particle size of the sorbitol diluent was the main contributing factor. Using the smaller particle size sorbitol produced granule fractions with near-theoretical mean calcium carbonate content. When sorbitol with larger particle sizes was used, a higher content of calcium carbonate was seen in the fines.

They postulated that the weakest interparticulate bond in the granule was the calcium carbonate–calcium carbonate bond. It is at this juncture that the ribbons fractured during the milling process. This rupturing of the ribbons at the weakest point exposed the calcium carbonate to attrition.

The use of uniformity testing and chemical imaging extends to the tablets made from roller-compacted granules. Romero-Torres et al. [102] used chemical imaging of tablets to screen which manufacturing process would produce better tablets in a study that included roller compaction as an option. In a similar evaluation, Alleso et al. [103] used tablet imaging for the uniformity of roughness to select between processes to produce the tablets.

4.4 PRINCIPLES OF OPERATION

To design and operate a dry granulation process, an in-depth understanding and, where feasible, mathematical model of each place where the machinery acts on the powder are needed. This section reviews the different suboperations in the equipment. More details on components common to other processes (e.g. powder filling and milling) can be found by following the references to the literature in these fields. Detailed knowledge for those suboperations when specific to roller compaction is presented here along with an in-depth look at compaction between the rollers. Recent trends show more accuracy and sophistication in small-scale experiments

and tests which predict process performance as well as various modeling techniques.

4.4.1 Powder Modification and Feeding to the Roller Compactor

The preparation and handling of powder fed to the roller compactor has challenges common to similar operations in other equipment trains. Powder mixing and handling are specialized areas with adequate sources from which to assemble information for the design of the mixing and handling subprocesses of a roller compaction process. Further useful information, not specific to roller compaction, regarding powders can be helpful such as blending and blend batch size increase as discussed by Alexander and Muzzio [104]. In addition, Prescott [105] presented aspects of powder handling, and the metering and dispensing of powders was reviewed by Yang and Evans [106].

Flow is an important consideration in the selection of material for a roller compaction process. Some common themes from powder handling technology are important to the powder handling for roller compaction. For example, it is commonly held that powder flow can be affected by particle size and shape. Yang and Evans [106] mentioned how humidity and electrostatic charge can affect powder flow, and this knowledge can be applied for conveying powder to the roller compactor. Dawes et al. [54, 55] showed the changes in powder feeding behavior and roller compaction performance from the presence of magnesium stearate.

Pietsch [21] has described how air entrainment can influence powder handling, including the densification step, and the post-compaction recovery in the roller compaction process. A study by Miller et al. [16] showed that the leakage of uncompacted powder was reduced from 20–30% to <2% of the material by use of a vacuum deaeration system fitted to the roller compactor. The throughput of the roller compactor was also increased by 20–40% with the deaeration system activated.

Most roller compaction-equipment configurations move the powder toward the rollers via a screw feeder. Sander and Schonert [107] showed that the delivery of powder from an unconstricted feed screw follows a linear relation with the screw speed. It was found that the screw feeder needed to have a minimum speed in order to move sufficient material for the rollers to compact properly at a given roller speed. At higher screw speeds the screw throughput was less than that predicted by the unconstricted delivery rate indicating that the screw exerts a pressure on the powder before the rollers (Figure 4.9). The screw feed increased the throughput of the roller compactor by causing consolidation of the powders before the nip. Similar results, showing the behavior of the screw feeder and its effect on the roller compaction throughput, were also reported by Guigon and Simon [99, 108].

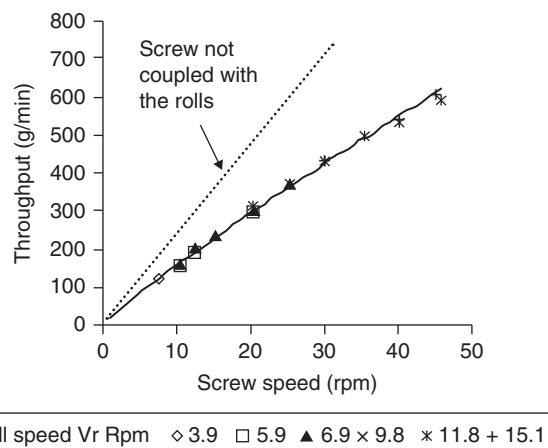


FIGURE 4.9 Screw throughput as a function of screw speed with and without the influence of the rollers. *Source:* Reprinted with permission from Guigon et al. [99].

The design and rotation of the screw can have unintended effects on the compacted ribbon. Patterns of color or differences in ribbon properties oriented along the main axis of the ribbon in a sinusoidal pattern were reported by Simon and Guigon [99, 108] for operations with a single screw (Figure 4.10). They used piezoelectric sensors in the rollers to study the pressure during compaction and image analysis of the ribbons to determine the density of the ribbon. Both sets of data showed a periodic variation that could be correlated with the frequency of the feed-screw rotation. Experiments with a piston device to feed the powder under constant pressure did not show periodic variations confirming that the screw feeder was causing the powder to consolidate differently depending on the screw rotation. It was postulated that the screw is preferentially applying pressure where the clearance between the flight and the nip are at the minimum, a position corresponding to the screw flight terminus.

Lecompte et al. [109] did not find screw-related variations in similar experiments with an instrumented roller compactor. It was proposed that because the screw terminated further from the rollers, variations due to screw rotation were not carried forward into the ribbon in the experimental setup used by Lecompte. The authors also examined how various combinations of parameters such as feed screw speed, roller speed, and gap could be adjusted to assure that the powder spread evenly across the rollers and was converted to a ribbon with consistent properties. Settings which increased the amount of powder entering the nip region, such as high screw speed, low roller speed, and a narrow gap setting, promoted the distribution of powder across the roller width. Miller [17] reported that a dual-screw feed design in combination with vacuum deaeration of the powder minimized any influence of the screw rotation on the ribbon.

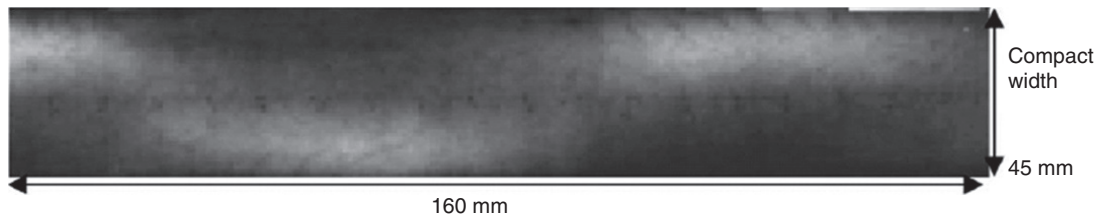


FIGURE 4.10 Variation in ribbon color due to variation in ribbon compaction induced by the motion of the feed screw. *Source:* Reprinted with permissions from Guigon et al. [99].

Studies with NIR imaging show similar patterns to that reported by Gigon and Simon. Both Lim et al. [93] and Khorasani et al. [85] report images showing sinusoidal variation along the axis of ribbon production when spectral data are processed to show ribbon density. This finding confirms the effect of the screw on the intake of powder at the roller nip.

From the end of the screw feed to the gap, the powder is typically contained by plates which may also be considered as part of the feeding system. The sealing plates can have an influence on the distribution of powder and compaction of the ribbon. This phenomenon was studied by Miguelez-Moran et al. [64, 80] using an instrumented laboratory roller compactor and several techniques to characterize ribbon density. A distribution of density was found across the ribbon (parallel to the roller axes, transverse to the ribbon motion) with lower density found at the ribbon edges. The effect was less pronounced for slower roller rates, smaller gaps, and powders which slipped along the seal surfaces more readily. The authors suggested that the lower density at the ribbon edges was attributed to drag induced in the powder feed by the sealing plates. This result has been confirmed by recent work using NIR [94, 95].

Proper installation and maintenance of the plates, which are typically constructed of Teflon or Nylon material, is important to product quality. Some leakage of powder from the seals typically occurs and increases the proportion of fines in the granulation. If not properly installed so that the leakage is at a minimum, and not sufficiently maintained so that the performance of the seal plates is consistent, the sealing plates may cause an isolated batch-to-batch variation or a trend in the granule properties over time.

To improve the powder sealing and process performance, some manufacturers offer rollers which interlock. One roller has a rim on the edge such that the edge of the second roller fits the channel and seals the powder into the nip. A system of this type was explored by Funakoshi et al. [100]. Several roller designs with different modifications to the rim geometry were explored. The amount of leaked powder was reduced from ~20% to ~5% by use of the interlocking rollers. When examining the resulting ribbons it was found that without the interlocking rollers, the ribbon experienced the highest pressure and achieved the highest density in the center.

With the interlocking roller system in place, the pressure experienced by the ribbon and resulting density was even across the ribbon width.

4.4.2 Consolidation and Compaction in Between the Rollers

Once the powder is delivered to the area before the nip, it begins its interaction with the rollers, moves forward, and becomes compacted into ribbon as it passes through the gap between the rollers. Johanson [28] developed one of the most-referenced descriptions and mathematical models of the roller compaction process. The roller compaction operation was described based on the machine geometry and assumptions of powder behavior.

Johanson took the pressure exerted on the powder bed from the feed screw as an input to the mathematical model but did not otherwise discuss the powder motion up to the rollers. From the feed area the powder moves forward and begins to be influenced by the roller motion. This region is typically referred to as the slip region because the rollers move faster than the powder with a boundary condition of slip between the powder and the roller surface. As the powder is dragged forward, the space in between the rollers narrows so that the powder bed consolidates and the pressure between the roller surfaces and the powder increases. A schematic diagram of the powder and roller interactions is presented in Figure 4.11.

To describe the process in the slip region, Johanson built upon earlier work describing steady-state powder flow originally developed by Jenike in 1961. The Jenike model and other powder flow concepts, developed more recently, have been summarized by Podczeczek [110].

It was proposed that for a cohesive, compressible, isotropic powder, a shear test could give information about the effective angle of friction and the surface friction angle. This information could then be used to plot yield loci describing the response of the powder to shear and normal stresses (Figure 4.12). The intersection of the wall yield locus and the stress circle gives the resolution of shear and normal stresses at the roller surface. The angle between the normal stress and the tangent to the roller is then described by Eq. (4.1)

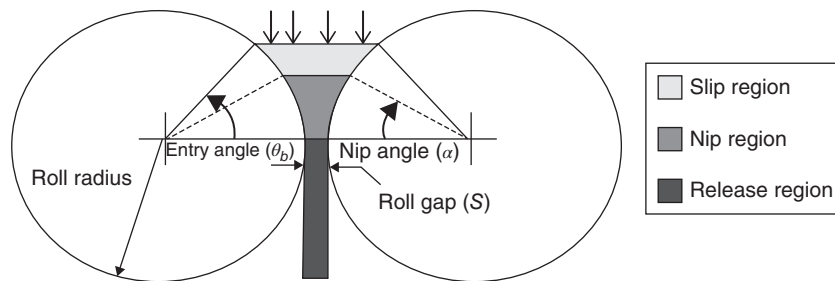


FIGURE 4.11 Schematic diagram of the roller and powder interactions. *Source:* Reprinted with permission from Bindhumadhavan et al. [111].

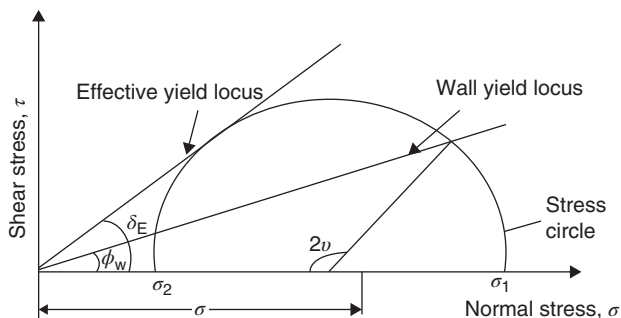


FIGURE 4.12 Jenike-Shield yield criterion for the slip region. *Source:* Reprinted with permission from Bindhumadhavan et al. [111].

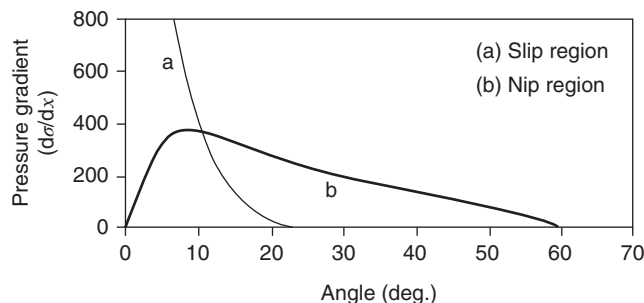


FIGURE 4.13 Pressure gradient versus angle from the nip. *Source:* Reprinted with permission from Bindhumadhavan et al. [111].

(Eqs. 4.1–4.4 were prepared with permission from Bindhumadhavan et al. [111]).

$$2\nu = \pi - \arcsin \frac{\sin \phi_w}{\sin \delta} - \phi_w \quad (4.1)$$

Johanson used the combination of the incoming pressure, roller geometry, and powder properties to predict a pressure distribution as a function of position as shown in Eq. (4.2).

$$\frac{d\sigma}{d\chi_{slip}} = \frac{4\sigma(\pi/2 - \vartheta - \nu)\tan \delta_E}{(D/2)[1 + (S/D) - \cos \vartheta][\cot(A - \mu) - \cot(A + \mu)]} \quad (4.2)$$

where A is given by:

$$A = \frac{\vartheta + \nu + (\pi/2)}{2}$$

The pressure at the roller surface is typically plotted as pressure versus angle from the closest approach of the rollers, the gap. For the slip region (see Figure 4.13, line a), the pressure decreases as the powder is taken up by the roller. Typical curves constructed for the roller surface with this approach show a low starting pressure followed by a rapid nonlinear

increase of pressure moving toward the gap (see Figure 4.13, line b).

At some point, the pressure on the rollers increases such that the powder no longer slips along the roller surface but moves with the roller surface until the powder exits the gap as compacted ribbon. This region close to the gap is referred to as the nip region. In order to predict the pressure in this region, Johanson considered the densification of a cross-section of powder moving between the rolls. This section of powder is compressed as if in uniaxial compression with density increasing as the distance between the rollers decreases. Powder property measurements from compression experiments with a die can determine the functional relationship between pressure and density. Johanson developed the expression shown in Eq. (4.3) to describe the pressure as a function of position in this region.

$$\left(\frac{d\sigma}{d\chi}\right)_{Nip} = \frac{K\sigma\vartheta(2\cos \vartheta - 1 - (S/D))\tan \vartheta}{(D/2)[(1 + (S/D) - \cos \vartheta)\cos \vartheta]} \quad (4.3)$$

This function can be used to construct a plot of pressure as a position from the gap. For this case, in the nip region where there is no-slip at the roller surface and powder moves with the roller, the pressure at the roll surface rises rapidly as the space between the rollers decreases. The powder compacts

and the pressure falls off as the material travels through to the other side of the gap (see Figure 4.13, line b).

The difficulty is predicted when the powder behavior will switch from slip to no-slip at the roller surface, thereby transitioning from the slip to the nip region. Since the pressure is most often plotted as an angle from the gap, the position of this transition was referred to by Johanson, and is commonly called the nip angle. The lowest pressure is predicted by the slip and no-slip assumptions for the roller surface boundary condition. At the nip angle the pressure predictions from the slip and no-slip condition are equal and the powder no longer slips at the roller surface from this point onward as no-slip is the lowest stress condition and is the pressure needed to induce powder motion. This relationship is represented in Eq. (4.4).

$$\left(\frac{d\sigma}{d\chi}\right)_{slip} = \left(\frac{d\sigma}{d\chi}\right)_{nip} \tag{4.4}$$

or

$$\frac{4((\pi/2) - \alpha - \nu)\tan\delta_E}{[\cot(A - \mu) - \cot(A + \mu)]} = \frac{K(2\cos\alpha - 1 - (S/D))\tan\alpha}{\cos\alpha}$$

A schematic illustration of the concepts from the Johanson model showing the two pressure predictions as a function of angle from the gap and the intersection of line a (slip condition) and line b (no-slip condition) defining the nip angle is shown in Figure 4.13. A graphic representation of the stress on the rollers is shown in Figure 4.14.

The Johanson equation is still the most established and convenient way to understand the process and material

interaction through the nip. Assumptions regarding the screw-feed pressure on the powder, the level of slip at the roller surface and within the powder, and the transition or nip angle are all areas of research and improvement. Once the Johanson equation is understood, more sophisticated mathematical or computational models can be built. Further discussion is included in the section on mathematical modeling.

The mechanistic understanding of roller compaction developed by Johanson and others is useful to understand the processing history of the powder as it is compacted into ribbon. The approach of these investigators grew from an interest in designing roller compaction equipment. For most applications, the machine geometry, speeds, and pressure ratings have been predetermined by an equipment manufacturer for a range of materials similar to the proposed process. However, the preselected machine capabilities still leave the choice of the various processing parameters to be made in designing the roller compaction operation.

Other investigators reported studies focused on the interaction of the process variables and particular materials or combinations of materials to determine the most effective processing parameters. Most of these studies used a statistical design of experiments to evaluate the effect of parameter changes on the ribbon, granule, and finished tablet properties for a fixed formula, but may also include raw material changes, composition, or other variables as part of the study.

Falzone et al. [36] studied the effect of the horizontal and vertical feed-screw speeds, and roller speed on granule size and compactability for MCC, lactose, and a model-active blend consisting of 60% acetaminophen. The results for MCC granules and tablets were successfully modeled by a

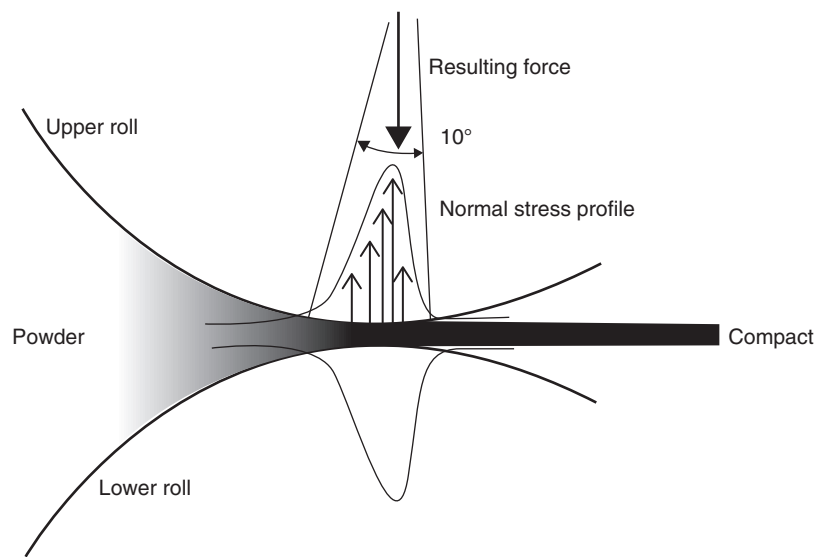


FIGURE 4.14 Graphic representation of the stress profile in the roller compactor nip. *Source:* Reprinted with permissions from Guigon et al. [99].

quadratic regression which included the horizontal feed speed and the roller speed with high feeder speed and low roller speed giving the highest values for both granule size and compactability. For lactose, the vertical feed speed also had to be used to successfully model the results. The acetaminophen blend results showed granule size equal to or less than the starting material due to fracture of the acetaminophen crystals. The results for granule size could not be modeled by a quadratic fit of the data, but the acetaminophen granulation compactability could be described by the quadratic regression and showed a dependence on both the feed-screw speeds and the roller speed.

Hervieu et al. [112] used a model powder to study the effect of feeder speed, roller speed, and compaction force on granule properties and the hardness and friability of the final tablet. The Box Wilson experimental design required 15 batches to complete. It was observed that at low feeder speeds compared to the roller speed, the powder could not be effectively compacted and had high friability. If the feed speed was too fast in comparison to the roller speed, the material temperature increased and the roller compactor jammed. Compaction force only had a secondary effect on the results. Similar results showing dependence on the feeder to roller speed ratio were reported by Guigon and Simon [99]. At higher feed speeds, overfeeding resulted in a poor-quality compact with loss of uncompacted powder, while at lower feeder speeds underfeeding resulted in no compact being formed. The effect of feeder and roller speed selections at constant pressure on the gap opening is shown in Figure 4.15 with lower feed resulting in a smaller gap at

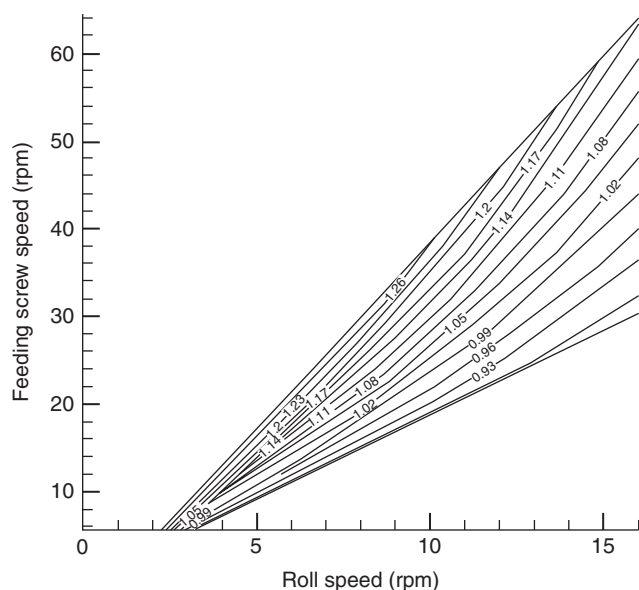


FIGURE 4.15 The effect of feed screw and roller speed on the gap opening at constant pressure. *Source:* Reprinted with permission from Guigon et al. [99].

constant roller speed and lower roller speed resulting in a larger gap at constant feed speed.

Four parameters (pressure, gap, roller type, and sieve aperture) were studied for their effects on a buccal tablet formulation by Rambali et al. [19]. A partial factorial design showed that smooth rollers promoted larger granule size and higher tablet strength but resulted in slower dissolution compared to ribbed rollers. It was noted that smaller granule size typically gave higher tablet strength. The lower than expected tablet strength for small granules produced by ribbed rollers was attributed to the higher frictional force and powder rearrangement induced by the ribbed surface. Smaller gaps and higher pressures produced larger granules. However, the tablet strength was found to be inversely proportional to gap and pressure.

A factorial design was also used by Weyenberg et al. [65] to study the effects of roller speed, pressure, and horizontal feed speed on granule flow, granule size, granule friability, and the resulting tablet strength for a bioadhesive tablet formulation. The roller speed and the compaction force had the largest effects on the granule properties followed by the horizontal screw feed speed. Higher roller speeds combined with low compaction force had reduced granule flow, increased granule friability, and smaller granule size. The horizontal screw speed did not have a large influence on granule properties. The highest tablet strength was obtained from smaller granules prepared at high roller speed and low compaction force. Tablets from the smaller granules best matched the dissolution profile of tablets which were previously prepared by direct compression, tested in vivo, and used for comparison with tablet from a roller compaction granulation.

In a study by am Ende et al. [113], a two-stage design was used to examine the effects of roller compaction on the content uniformity, granule properties, and tablet properties of a low-active concentration granulation. The first stage of the experiment used a full factorial design to study the effects of compaction pressure, and gap width, where the feed screw was automatically adjusted to maintain the gap. The study responses were the amount of uncompacted fines and the potency of the uncompacted fines. It was found that the amount of uncompacted fines was affected by the pressure and gap settings with a significant interaction between the pressure and gap variables. Higher levels of fines were present for low pressures and high gap widths. The potency of the fines, a measure of the uniformity of the granules, could not be predicted by these two variables alone.

The second stage of the experiment used a D-optimal design to examine the effects of roller speed, gap width, compaction pressure, and post-compaction mill sieve opening on granule and tablet properties. The granulation size increased with increasing pressure but had lesser effect as the mill sieve opening decreased showing the interaction between these two variables. Compaction pressure and sieve opening also

had an influence on the uniformity of both the granulation and the tablets. As screen size increased, the variability of the granulation potency increased. It was found that lower pressures and larger gaps during the roller compaction led to lower compaction forces to form a tablet of the desired strength. Tablet friability was unaffected by any of the roller compaction variables studied.

While the studies mentioned focused on the response of a model or actual formula to the roller compaction parameters, other studies have focused on a particular setting or a minor formula component introduced for a particular effect. By considering both the overall process response and risk around a specific setting range or functional excipient level assures a robust process.

An example of such studies for materials is the work of Dawes et al. [54] who studied the effect of magnesium stearate and found reduced throughput in cases where the rollers are smooth and increased throughput for rough-surfaced rollers. Zhao et al. [114] studied the influence of surfactant on the processability of the powder. Interestingly, it was found that the lubrication properties of the material counteracted the expected surfactant effect. At the levels and process conditions studied, the tablet compression required more force resulting in a corresponding increase in tablet disintegration negating the surfactant effect of the sodium laurel sulfate. The effect of roll pressure on tablet disintegration was confirmed in a study by Rajkumar et al. [115], where a novel disintegration apparatus was used to measure differences in the effects of formulation, roll compaction pressure, and tablet compression pressure.

Prediction of roller compaction settings for the four main variables: feed speed, roller speed, compaction pressure, and gap width is possible to get the desired system responses in terms of ribbons, granules, and downstream processing. As discussed in this section, the mathematical descriptions and mechanistic understanding combined with some knowledge of the materials being roller compacted and verified with experimental studies of the system response to the parameter settings give practical estimates of performance.

4.4.3 Ribbon Breaking and Size Reduction

The roller-compacted ribbon is typically broken and reduced in size to form the granulation. The interparticle bonds in the ribbon need to be strong enough such that fractures form and lead to granules rather than a loosely compacted ribbon returning to powder. Ribbon of a consistent strength and density should give consistent granule particle size distributions. Bacher et al. [25] suggested that the ribbon breaks at the weakest interparticle bonds and that ribbons with uniform interparticle bonds give the better compactability.

The milling operation can be integral to the machine or a separate step and is usually accomplished with a screening

mill. General considerations common to screening mills such as choosing the impeller type, screen type, speed (impeller or screen depending on type of mill), and spacing of the screen and impeller are steps to achieving the desired granulation particle size distribution. Proper selection, setup, and maintenance can eliminate the need for a metal detector at the mill exit. Effects of screen selection affecting the granule and tablet properties in a roller compaction operation have been reported by am Ende et al. [113].

Information on the operation of mills can be applied directly to the milling subunit operation of roller compaction. Rekhi and Sidwell [116] described how Kick's Law, Rittinger's Law, and Bond's Law have all been developed to relate the mill energy input to the size reduction process. It was reported that only a small part of the energy used by the mill is consumed in breaking the particles making the ability to predict performance limited in practice.

Advances in computational modeling have been applied to the milling step for roller compaction. Barrasso et al. [60] have used a population balance model and verified it with experimental data from a roller compactor. With this approach several material dependent parameters, such as the description of material breakage rate or breakage kernel, need to be tuned to the real-world performance. Eight of the 12 experiments performed were used for model tuning and the last two were used to check the model-predictive capability. Models such as these are extremely useful to understand the process, accelerate process development, and conserve resources. Further development of the science around particle breakage will expand their utility.

Typically, some experimentation is still necessary to select the mill change-parts (e.g. screen, impeller, and spacers) and operating parameters due to the limitations of the predictive methods. These experiments can also identify common milling problems such as screen blinding, heat generation (with possible melting), and interactions with moisture either from the environment or liberated during the milling process. Consideration of these operational factors are necessary to verify model results.

Mirtic and Reynolds [58] studied the breakage of two commonly used roller-compacted materials: MCC and mannitol. They used the change in particle size as a fraction of mass throughput to determine the breakage parameters for a population balance model. Their experiments showed differences in the milling behavior and interdependence of milling variables when based on difference in material properties between MCC and mannitol as well as the ribbon properties such as porosity. For example, one of the breakage parameters described whether particle fractured or abraded to produce smaller size materials. For mannitol, lower porosity ribbons showed parameter values consistent with fracture, while higher porosity ribbons had values consistent with abrasion.

4.4.4 Process Models for Roller Compaction

Predictive techniques include all types of logical, mathematical approaches to estimating outcomes. The means to collect the data from which to perform those calculations or test the accuracy of the estimate are also necessary to include in the definition. Even the collated observations from early experiments can provide good estimates of system performance if the potential causes of the outcome are analyzed and potential results estimated. Building on the analysis with more technology and data, statistical methods or simple mechanistic equations, such as Johanson's, can be built. Advances in hardware and software over the last three decades have allowed process simulations to become more powerful, more affordable, and more timely. What were approximations, lists of numbers, and crude graphics are now movie quality visualizations of an ocean of data regarding the material experience in the process at a fine level of geometric detail.

Not all questions need the most sophisticated model. Conversely, seemingly simple phenomenon often need a combination of different types of models considering different scales of action (microscopic, macroscopic, and system based) to achieve a useful result. Choosing the right questions, approach, level of sophistication, and accuracy to have the desired impact on quality at a reasonable commitment is where a scientist's skill and experience in modeling shine. Much of the value of modeling is in thinking through those choices and the effect that has in handling the challenges of the process. Even better is that today's modeling effort often can be recycled as the foundation for future advances. Rapid growth in capability can be achieved at reasonable time and resources if a long-term vision is applied to building the model in stages as each challenge is faced.

The area of the roller compaction operation that is unique to the unit operation and of most interest to the topic of this chapter is the area where the powder is consolidated, begins to move with the rollers, and is compacted into the ribbon. Powder flow, screw conveying, and milling are also important and can be found in the extensive literature on these topics.

Several approaches to process predictions for roller compaction have been explored: mathematical models, multivariate regression of designed experiments, predictions from compaction properties, computational modeling of the material in the process, and artificial intelligence/neural network construction to predict outcomes from a set of training or even "big" data.

The foundation of a mathematical description for the roller compaction process was developed by Johanson [28] and described earlier in the chapter. Even recent researchers have used it to understand process data. For example, nine data sets were used to determine the internal and wall friction angles in the Johanson equation, four data sets were used to validate the predictions which were then used in predicting three

conditions outside the training set as described by Sajjia et al. [117].

Other studies have explored the accuracy and updates to the equation such as the work of Bindhumadhavan et al. [111], Yusof et al. [118], and Sommer and Hauser [119]. Alternate approaches were presented by Yehia [120] and reviewed by Dec et al. [121]. Several studies have used multivariate regressions to characterize the roller compaction operation such as Falzone et al. [36], Hervieu et al. [112], Rambali et al. [19], Weyenberg et al. [65], and am Ende et al. [113].

Bindhumadhavan et al. [111] explored the Johanson model 40 years after its publication. They compared the results of calculations with experiments for MCC. In their work, the roller compaction was performed with the benefit of better instruments to measure powder properties, pressure sensors mounted in the roller faces to measure the compaction forces, and better computing power to calculate the model and evaluate the data. Good agreement was found between the predictions using the method of Johanson and the experimental values as shown in Figure 4.16 relating the pressure as a function of angle from the nip for different gap settings. Yusof et al. [118] also reviewed Johanson's model and conducted experiments with maize powder to compare the model and experimental values. Once again good agreement between the model and the experiments was found.

Sommer and Hauser [119] reviewed the Johanson model and found that predictions were useful given the limited number of inputs. However, they examined the assumptions which Johanson used and concluded that there could be limitations that might cause the model to deviate from the results found in practice. One example presented was the model's possible sensitivity due to the method of choosing the boundary condition for the point at which the rollers engage the powder feed. The boundary condition assumption could cause an unrealistic sensitivity to the feed pressure in the model results. Simple material models were used for the

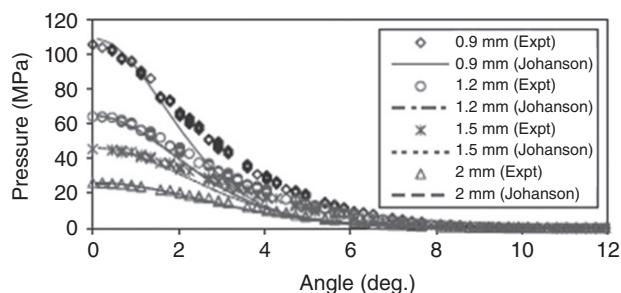


FIGURE 4.16 Predicted and experimental roller surface pressure as a function of gap. *Source:* Reprinted with permission from Bindhumadhavan et al. [111].

description of the nip region, and it was proposed that the limitations in these models could also lead to inaccuracies.

An additional mechanism of slip has been explored by Schonert and Sander [122]. They reviewed several theoretical models and concluded that there could be slip between the compacted ribbon from the point of maximum stress to the exit of the gap. Instrumentation embedded in the rollers capable of resolving the normal and tangential stresses at the roller surface confirmed that the powder does begin to move with the rollers at some level of consolidation. However, the maximum stress occurred slightly before the line of centers between the two rollers. From the point of maximum stress it was found that the compacted ribbon accelerates, moving faster than the rollers, toward the gap exit. A similar measurement showing the maximum pressure before the line of centers and an acceleration of the ribbon was obtained by Lecompte et al. [109].

Tan et al. [20] examined the effect of the roll surface as predicted by both the Johanson equation and a mass balance method where the mass output of the roller compactor is assumed to be a function of the mass in the nip area and the speed of the roller. The Johanson equation depends on the measurement of wall friction, which were deemed difficult to measure for serrated roll surfaces. By using the mass flow measurement, which depends on powder density measurements, a correlation between stress and density was easy to establish with simple compression experiments. In experiments where dicalcium phosphate, lactose, and MCC were mixed in different ratios to create a range of material properties from primarily plastically deforming to brittle fracture, these researchers found that the predicted nip angle was more accurate for serrated rolls when using the mass flow method.

Other approaches to predicting the process conditions have also been proposed. Yehia [120] described a simplified approach in which the change in bulk density going from the feed to the ribbon is considered when estimating the pressure during the roller compaction process. Rather than subjecting powder to a shear test, only measurement of the input and output bulk densities and the pressure response to uniaxial compression is needed. The author assumed that most of the powder densification occurs in the nip region and that the pressure at the beginning of the nip region could be estimated by a density measurement and information about the density-pressure relationship for the material studied. The speed and geometry of the rollers and gap width is used to estimate the change in volume, and therefore, density of the material as it is processed. It was proposed that the performance at scale could be predicted from the limited material testing required for the calculation and information obtained on a lab-scale roller compactor. The lab-scale machine would need to be adjusted to a variety of geometries and equipped with different roller finishes to determine an appropriate regime to predict the performance at scale.

Dec et al. [121] reviewed various models including a method of estimating roller compaction conditions. The “slab method,” originally developed for the compaction of metal powders, considers a section of material or “slab” passing as a single element through the nip region. The stepwise calculations assume a condition at the inlet of the nip and uses experimental data to determine the nip angle. Pressure is predicted by successive iterations of the calculation until the prediction matches the measured properties of the ribbon.

Although achieved by passing between rollers, the compaction of ribbon causes the same physical processes within the powders that are induced during compaction within a die. Borrowing or combining concepts from tablet compaction can be useful to understanding roller compaction. Tan et al. [20], Zavaliangos [121, 123], Hilden [124], and others all reference the utility of die compression experiments for understanding the relation of applied stress and density to predict roller compaction behavior and parameters. Bi et al. [125] used die compression data and measurements from laboratory roller compactor experiments with Johanson’s equation to define the roll pressure needed to achieve ribbon quality on another machine. Farber et al. [84] examined the loss in tensile strength when comparing tablets made via a roller compaction granulation to a direct compression prototype. The roller compaction process was considered as part of the overall compaction history of the materials. A “unified compaction curve” was constructed which described both the roller compaction and tablet-forming processes. It was found that, for materials which bond primarily from plastic deformation, compaction information generated with a compaction simulator or single-station press could be used with information about the pressure exerted by the rollers to estimate the tensile strength of the final tablets.

Hein et al. [126] used a three-dimensional model of compaction properties populated by data from a single-station press to predict the change in final tablet properties as a result of roller compaction. A reduction in final tablet strength was shown for materials with plastic deformation. A minimal reduction in tablet strength was shown for the material with primarily brittle fracture. The model was considered effective at screening materials for use in roller compaction.

A modified partial least squares (PLS) analysis was used to compare data sets with overlapping but different parameter maps from the different roller compaction installations to select values for the new equipment from the successful parameters on the existing equipment. The technique, described by Liu et al. [127], clearly identified differences in roller compaction equipment of the same design and different size. Using the PLS analysis with three principal components, a space was mapped for the second machine. The model was trained with data sets from seven existing products. Predictions for parameters on the new machine for the same formula and across different formulas were found to be accurate using the PLS predictions.

Computational methods can be used to understand and predict process behavior and can be applied to roller compaction. Dec et al. [122] reviewed several finite element method (FEM) applications as part of a review of modeling methods. However, most of the models reviewed were published from researchers in the metals industry. The simulations were dependent on estimates of the feed stress and friction to predict the process conditions and post-compacted material performance.

More recently, Michrafy et al. [123], in their review, found that researchers were using the FEM (most commonly used for flows), the discrete element method (DEM, most commonly used for particles), and combination methods where a mesh is combined with a multiparticulate model. DEM was most commonly applied to the feeding mechanism and feed zone of the roller compaction, while FEM was the most common technique for the powder consolidation and compaction. The challenge in the FEM models was distortion of the elements as the powder goes through the compaction process. The three choices for this type of model are moveable mesh, fixed mesh, and a combination approach. Michrafy et al. [123] emphasized the importance of model inputs via materials testing, and summarizing that information in a behavior model, to feed the FEM predictions.

Both two-dimensional and three-dimensional FEM simulations were performed by Muliadi et al. [128, 129]. The model construction included a modular design in that the material behavior was described by the Drucker–Prager Cap model. The material deformation during compaction was accounted for by using an adaptive mesh technique where the inlet and exit grid are defined and the elements from one condition to fit the other as the process progresses. For the two-dimensional model it was found that the nip angle had good agreement with the one-dimensional Johanson equation. However, the normal stress and resulting density predicted were higher for the Johanson model. As the compressibility of the material increased, the results converged.

For the three-dimensional version, comparison to experiments using MCC was used to verify the results. To reduce variability in the experimental data the inlet of the machine was modified, replacing the feed screw with a pneumatic piston, to assure a consistent powder feed. The model was able to capture fine details of ribbon density across the ribbon width with good agreement to the experimental data. Significantly, this model was constructed without adjustable parameters or fitting such that the correspondence between experimental and model results was even more impressive.

The use of neural networks and artificial intelligence approaches to modeling have been explored to correlate different types of inputs, process parameters, and granule or resulting tablet performance. Inghelbrecht et al. [130] studied a 60-experiment data set and then predicted the results for an additional 20 experiments. The speed of two feed screws

(horizontal and vertical), roller pressure, and roller speed was used to predict the granule friability and particle size. It was found that the neural network was more effective than a quadratic mathematical model approach in predicting the granule performance results.

The binder type, binder concentration, number of compaction passes, and addition of MCC extra-granularly were used as inputs to predict the performance of acetaminophen tablets by Turkoglu et al. [131]. The results were poorly predicted using a typical neural network learning algorithm of adjusting the weighting of parameters relating model inputs and outputs. A second calculation method using a “genetic” algorithm which progressively selects best-fit solutions in “generations” of calculations gave better predictions.

A variety of material inputs were used with projected process settings for roller speed and roller gap to predict the roller pressure, nip angle, ribbon density, and ribbon porosity by Mansa et al. [132]. The commercial software package employed used neural network, genetic algorithms, and fuzzy logic in order to predict the outputs. Good agreement was found inside the training range. However, some rules generated by the system did not seem to correlate with the physical system (e.g. roller gap not having an effect on ribbon density) and predictions outside the training range did not compare well with experimental values.

Commuri et al. [7] correlated roller vibration with other process parameters of the compaction of an asphalt mix under a pavement rolling compactor. Neural net techniques trained by density measurements for experimental operation were able to predict the density of the compacted asphalt in normal operation. Frequency analysis of the roller vibration combined with asphalt temperature and a few other parameters were all that was needed. While this work is from a different industry and for a different type of compaction, it may be a useful approach to modeling roller compaction in the future.

The term “big data” has come into favor with fanfare. However, the big in big data can come from the impact of looking at aggregated data sets which are mere hundreds rather than thousands of points and gigabytes rather than terabytes of data. Shi and Hilden [34] did just such an impactful analysis of data from 136 formulations subjected to tap density, roller compaction, and tablet compression tests. Much like the work of Farber et al. [84], Hilden and Earle [124] built a correlation between roller compaction and tablet experiments such that deformation can be related to the applied stress. Extending the research on powder densification from feed to tablet was considered and found to follow a relationship between the solid fraction (inverse of porosity) and a simple quadratic equation of the applied stress. By aggregating and fitting the data for 136 formulation, the constants of the quadratic equation were determined and could be used in reverse to calculate the compaction pressure necessary to achieve a desired ribbon density.

4.4.5 Control Strategies for Roller Compaction

The overall roller compaction operation includes the powder preparation and handling, compaction by the rollers, and the subsequent breaking and sizing of the resulting ribbon. The variables affecting powder preparation and handling as well as sizing operations are not specific to roller compaction and can be chosen for the roller compaction operation from a base of information devoted specifically to behavior of powders.

For the compaction portion of the operation, several authors such as Johanson [28], Yehia [121], and others have discussed how powder measurements with or without experiments can be used to design a roller compactor which will subject the material to a compaction history resulting in the desired output ribbon and granule properties. The typical situation in many areas of manufacture, including pharmaceutical applications, is the reverse. An equipment vendor has already spent the effort to design a piece of equipment of general applicability for the typical range of powder properties encountered by their clients. The decisions regarding screw feeder arrangement, design, and placement in proximity to the rollers, as well as the roller diameter, width, and a mechanism to maintain position of the rollers thus applying pressure on the powder, along with the milling equipment have already been chosen.

The challenge in most practical applications of roller compaction technology is to find the appropriate formulation and settings of adjustable parameters given a set piece of equipment and control system previously engineered by the manufacturer. The four common variables discussed in most of the literature are the feeder speed, roller speed, gap, and pressure. These variables depend on one another and need to be set in combinations which are appropriately balanced.

The experience of the vendor who configured the machinery and who has the benefit of the knowledge from working with many clients can help in designing the roller compaction process and selecting parameter values. Extreme settings of the controllers are usually not effective at making the best product unless there is something very unusual about the system under study. Several investigators, such as Rambali et al. [19], Weyenberg et al. [65], and am Ende et al. [113], used statistically designed studies to understand and even predict the granule and tablet properties as a function of the roller compactor variables and select the parameter values. More recently, Shi and Sprockel [133] or Liu et al. [127] have used similar approaches. A “big data” approach to aggregate data and predict parameters was taken by Shi and Hilden [34].

The simplest method is to choose a feed speed, roller speed, and pressure which delivers the desired gap or ribbon thickness. The challenge for this method is to deliver the powder very consistently. A balance of the powder feed speed and the roller settings must be achieved to avoid over or under feeding the roller compactor. As discussed by Lecompte et al. [109], it is necessary to feed enough material

to the nip region such that a uniform of powder across the roller width is achieved with consistent ribbon properties along its width and length as the result. Guigon and Simon [99] discussed how the feed speed and compactor throughput should be mediated, avoiding leakage of powder from the sealing mechanism or air entrainment which could disrupt the powder flow.

Given the difficulty of feeding powders without variation, control systems have been developed such that a roller speed, pressure, and gap distance can be fixed. The feeder speed is then adjusted in a feedback loop with the gap measurement to maintain a consistent ribbon thickness, density, and strength. More recent developments in control systems improve upon the feed control loop by monitoring the feed-screw output, or the actual mass throughput of the roller compactor.

More sophisticated methodology is now possible due the ability to integrate tests, small-scale experiments, online data, control loops, and models (either statistical or computational into a single web). This can be done off-line to set stable conditions using individual models, or can be combined in a single block using integrated software, such as the commercially available gPROMS from Process Systems Enterprise [134], or in an integrated control system such as the work of Singh et al. [26, 27].

Two different approaches to setting and scaling the operation of the roller compactor are discussed in the next section.

4.5 SCALE-UP OF ROLLER COMPACTION

Equipment parameters are adjusted to impose a physical environment that will transform material properties in a way beneficial to the intended product performance. As equipment size changes, the environmental factors acting on the material do not all scale at the same rate. Thus, maintaining the balance of factors impacting the material on a change in scale is difficult to do. The challenge is to understand and choose workable parameters as discussed in this section.

4.5.1 Scale-Up Strategies

Successful scale-up of the roller compaction process begins in development. Selecting and testing a blend to avoid segregation and have flow properties immune to electrostatic, bridging, funneling, or other flow anomalies makes roller compaction easier to control and to scale. Compression experiments to characterize the material density in response to stress serves as a basis to select roll pressure. Leane et al. [33] have developed a classification system to assist in determining if the selected materials are appropriate for use with a particular process such as roller compaction.

A common method for scaling up a roller compaction process from development to commercial equipment is to use a parametric strategy. The parametric strategy focuses on determining the commercial equipment parameter values by using equivalency factors to compute the settings from development data. Equivalency factors are based on aspects of the equipment. Some of the values used may be taken from the equipment manufacturer, who has assembled information from the design and testing of the unit as well as the collected experience of the client base in order to develop scale-up factors.

In selecting the roller pressure for the commercial equipment, the hydraulic pressure required to generate a needed force on the rolls can be estimated by considering the applied pressure divided by the roller width (pressure/length). Roll diameter plays a role where the rpm can be set to obtain a linear velocity of the roller surface to match conditions in development and commercial settings. An approximate flux of material at the gap can serve as a metric similar to “dwell time” in tablet compression to estimate the work on the material at a given pressure. A calculation of the mass throughput/linear roll speed \times roll width should yield values which can be compared across development and commercial machines. The selection of roll speed, roll gap, and an approximate density is needed to make selections which provide the desired throughput. This approach seeks primarily to choose parameters and then measure attributes such as ribbon density or granule size to assure similar performance.

The simple parameters above can be augmented by information from statistical models, as noted earlier in the chapter and again under Control Strategies. Other researchers have used a mechanistic model based on Johanson’s work to help in assuring that the material experience in development and commercial equipment is the same. Sajjia et al. [117] and Nesarikar et al. [135, 136] have taken this approach. Miguez-Moran et al. [64] and Nesarikar et al. [135, 136] have increased the accuracy by using instrumented roll technology to determine the pressure response and nip angle used in the Johanson model. A Johanson equation calculation tool has been posted to the web (www.pharmahub.org) by Kuriyan and Reklaitis [137].

An alternative strategy focuses on the attributes of ribbons and calls for adjusting the ribbon-production parameters to attain attribute values on the commercial equipment equivalent to that produced in development. The intent of controlling the quality of the ribbons produced is to control the downstream granule properties (size distribution, solid fraction, and compressability). The hypothesis underlying this approach is that under similar milling conditions (mill speed and screen opening) the output particle size distribution is determined by the ribbon input quality [59, 138].

The roller compaction unit operation is composed of sub-processes focused on achieving two sequential, independent, but linked manipulations of the material: ribbon production

and granule production by milling. The two quality attributes of interest for the ribbons are the thickness after recovery due to relaxation, and the density or solid fraction. As with the general consolidation theory, these two ribbon attributes control the breaking strength of the ribbon. Both of these attributes should, therefore, influence the behavior of the ribbons during milling. The indirect effects on ribbon recovery should be carefully studied, since extensive recovery could affect ribbon strength.

One crucial decision to be made when contemplating scaling up a roller compaction process is whether or not to use the automatic gap feedback control system or to proceed with a preset feed rate. This decision will determine what parameters need to be considered in scale up. If the roller compactor is run under gap control, the effect on ribbon thickness is muted, unless purposely varied (by changing the gap). To determine what process parameter to alter to obtain the desired intermediate attribute requires a detailed knowledge of the interplay between process parameters and ribbon/granule attributes.

CASE STUDY 4.1 PARAMETERIC-BASED SCALE-UP

Case Study 4.1 reports on the roller compaction of a model MCC/lactose blend with 5% active using the manual operation. In manual operation the three parameters for ribbon production (screw speed, roller speed, and roller pressure) were set at predetermined values that remained constant for the duration of the run. For reproducible ribbon production, this implied that the powder flow into the rollers had to be constant.

The behavior of the powder during roller compaction is dependent upon the region it is in [37, 40, 83, 97, 111, 118]. The powder in the slip region is densified slightly by rearrangement as it travels toward the rollers. In this region there is slippage between the powder and the rollers. In the slip region the velocity of the powder is slower than the linear velocity of the rollers. In the nip region the powder undergoes densification by particle rearrangement and by deformation. In the compaction region further densification by deformation occurs proceeded by bonding to form the ribbon.

In the parameter optimization study three factors were evaluated: screw speed, ratio of roller speed to screw speed (powder delivery rate), and roller pressure. Based on the data set, various parameter combinations were identified that produced granules of the desired flow and particle size distribution. The process parameters (screw speed, roller pressure, and roller speed) for commercial-scale equipment were determined using equivalency factors.

4.5.1.1 Adjusting Pressure Several researchers have reported on the loss of powder compactability during tablet compression for roller-compacted powders [43, 57, 139]. The reduced compactability at tablet has been ascribed to a reduction in the binding potential due to the consolidation that occurs during roller compaction [8, 36, 78]. To mitigate this reduction it is recommended that only sufficient pressure should be applied during roller compaction to form granules with improved powder flow.

More recently, investigators have looked at instrumented rollers to gather experimental data on the actual pressure curve on the rolls [80]. Farber et al. [84] argued that a powder undergoes similar volume reduction during roller compaction and tableting. Therefore, instrumented rolls can be used in setting roller pressure on commercial equipment to achieve equivalent pressure on the commercial equipment (and hence similar powder behavior). The underlying principle being that similar ribbons would be produced on scale-up if the pressure is maintained. Nesarikar et al. [135, 136] has taken this further and combined the data for use with the Johanson model.

4.5.1.2 Choosing Screw and Roller Speed The roller speed for the commercial equipment in Case Study 4.1 was set to obtain the desired process efficiency. In manual mode, the ratio of screw to roller speeds dictates, for a given material and roller pressure, the gap between the rollers. To attain a roller gap on the commercial equipment similar to that on the development equipment, the value of ratio of screw speed to roller speed obtained during development is used to set the screw speed on the commercial-scale roller compactor based on screw design and roller width and diameter. The force generated by the powder propelled by the feeder screw into the slip region is counterbalanced by the hydraulic pressure applied to the rollers. Based on the material properties, the balance between these two forces determines the gap between the rollers.

4.5.1.3 Mill Screens and Speeds The velocity at which the milling blade rotates affects the milling process in two ways: the force of the instantaneous impact on the ribbon and the residence time in the milling chamber. These two aspects have two different potential consequences. The higher impact could result in the ribbon shattering into smaller granules. Residence time in the chamber affects the amount of attrition the ribbon undergoes, which results in finer particles. Brittle materials are more susceptible to the impact, whereas pliable (viscoelastic) materials are more prone to attrition. The properties of the MCC/lactose active blend in Case Study 4.1 suggested that the ribbons produced would possess both brittle as well as viscoelastic properties. Since ribbons are porous solid bodies, it is plausible that the ribbons fractured into smaller pieces that underwent size reduction by attrition. Hence, it was determined that residence time was the important factor to study and not the instantaneous impact.

A derived parameter was used for mill speed. The effect of milling speed on granule size is due to its effect on residence time of the ribbon in the milling chamber. To more accurately estimate the effect of residence time on granule properties, a ratio of mill speed to roller speed was studied as a derived parameter. Increasing the mill-to-roller speed ratio reduced residence time for the ribbons in the milling chamber by increasing the output rate for a given input rate. Increasing the mill-to-roller speed ratio (shorter residence time) increased the mass flow rate of the granules. Increased granule flow was due primarily to the increased mean particle diameter and reduced fines.

The bulk powder property of interest for tablet compression is the mass flow. Powder mass flow should be governed by the properties of the particles making up the bulk. The two particle properties controlled by the roller compaction process are mean particle size and amount of fines. These two particle properties have an opposite effect on mass flow. Figure 4.17 is a contour plot showing the interactive effect of mean size and fines on flow. To maximize flow, the mean particle size would have to be increased to compensate for a higher amount of fines. The area of maximum flow is bounded approximately by the lower left to upper right diagonal in Figure 4.17.

Due to the statistical design used, the effect of screen size was nested within the mill/roller speed ratio. The effect of the upper and lower screen on the output of the milling step could only be determined for a given mill-to-roller speed ratio. As expected, the effect of the lower screen on the output of the milling step was more pronounced than that for the upper screen. A maximum in mean particle size was observed with a 1 mm lower screen opening. This maximum mean particle size coincided with a similar maximum in mass flow with a 1 mm lower screen opening.

To understand the milling process in more depth, the difference between the upper screen size of openings in the screen material and the lower screen opening size was studied. It was theorized that this difference may be important since the upper screen controls the quality of the input for the lower screen, and because size reduction in this case is partially through attrition. Figure 4.18 depicts the relationship between mass flow and the difference in screen opening size between the upper and lower screen. Equation (4.5) depicts the relationship, where F is the mass flow (g/s), F_{\max} is the maximum predicted flow (g/s), X is the difference in screen opening (mm) between the upper and lower screens, and X_0 is the difference at maximum flow.

$$F = F_{\max} \times e^{-0.5 \times \left(\frac{X - X_0}{b}\right)^2} \quad (4.5)$$

The fitted model strongly suggests an optimum for mass flow at ~2 mm difference in screen opening. The maximum predicted flow is ~4 g/s. The combination of the lower screen

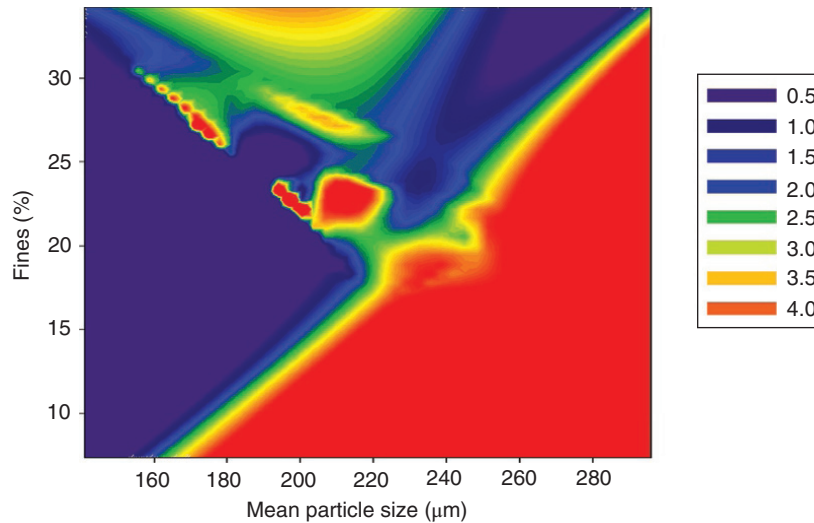


FIGURE 4.17 The interactive effect of mean particle size (μm) and amount of fines (%) on bulk mass flow (g/s).

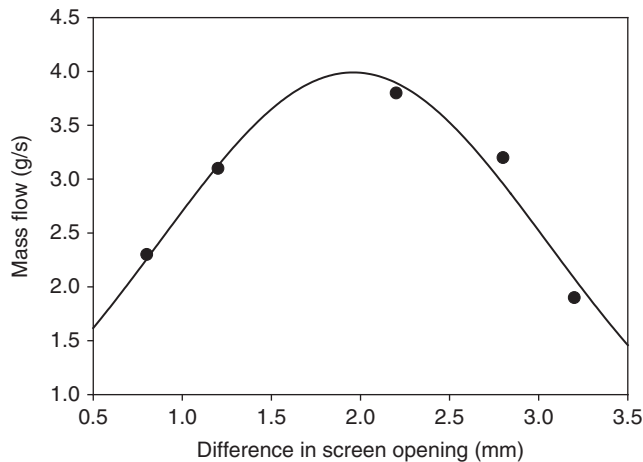


FIGURE 4.18 The effect of difference in opening between upper and lower screens on mass flow.

maximum flow at ~ 1 mm opening size and the screen difference maximum at ~ 2 mm additional opening defines the upper screen optimum at ~ 3 mm opening size. These same screen openings were used for scale-up in Case Study 4.1.

EXAMPLE 4.1

Parametric Scale-Up with a Procedure and Parameters Recommended from the Equipment Manufacturer

The example presented in this section illustrates scale-up recommendations similar to one of the major equipment manufacturers based on the design and experience developed with their equipment. In this case the pilot scale equipment has a roller diameter, D_1 , is $\sim 60\%$ of the size of the

commercial equipment, D_2 , and a roller width, B_1 is $\sim 53\%$ of the commercial equipment, B_2 . It was recommended that the gap for the commercial machine be calculated from the acceptable values from the pilot scale machine according to Eq. (4.6). For this choice of equipment sizes, scale-up leads to the larger machine gap setting about 1.7 times the pilot scale machine.

$$G_2 = \frac{D_2}{D_1} \times G_1 \quad (4.6)$$

The screens on the post-compaction mill for the commercial machine are approximately 5 times the size of the pilot-scale machine so a throughput, T , increase of approximately 4.5–5 times is expected in this example. The suggested scale-up procedure then uses machine parameters to calculate the roller speed based on Eq. (4.7). The throughput of the pilot-scale machine and the desired ribbon density are used as inputs with the remaining parameters coming from the machine geometry. The roller width may have an adjustment factor to the added for some machine designs depending on the ribbon width produced by the roller geometry (e.g. nothing produced at the edge if the roller has rim to contain powder). The roller speed, n , calculation typically has a result with the commercial-scale machine about 1.7 times faster than the pilot-scale machine for the commonly available units this example was based on.

$$n_2 = \frac{T_1 \times 5}{D_2 \times \pi \times (B) \times G_2 \times \rho} \quad (4.7)$$

The next step in the scale-up is to relate the hydraulic pressure settings from the pilot-scale machine to the commercial machine. First, the force per unit distance of roller width

needs to be estimated from the manufacturer's conversion factor by a rearrangement of Eq. (4.8). The force per unit distance of roll width for the commercial machine is then calculated from Eq. (4.9).

$$P = F \times (\text{pressure conversion factor}) \quad (4.8)$$

$$F_2 = \frac{D_2}{D_1} \times F_1 \times t_c \quad (4.9)$$

where

t_c is an adjustment factor related to the dwell time.

The hydraulic pressure setting for the commercial-scale machine is then calculated from Eq. (4.8) using the appropriate conversion factor for the commercial-scale machine. Typical values are approximately an 8% increase in the hydraulic pressure as suggested by the manufacturer for the models in this example.

The screw feed is adjusted to achieve the calculated gap and the calculated pressure. It is recommended by this manufacturer to use a feedback loop to adjust the feed screw to maintain the gap. Typical values for screw speed on the larger unit are about 80% of the value for the pilot-scale unit for two typical models from this manufacturer, but it should be noted that the larger system has twin screws compared to the single screw on the smaller machine.

Finally, the mill speed is set. From the manufacturer's experience a mill speed setting on the commercial machine of about 75% of the pilot-scale machine will yield granules of similar size distribution when using screens with the same aperture.

CASE STUDY 4.2 ATTRIBUTE-BASED SCALE-UP

Case Study 4.2 reports on the roller compaction of a model MCC/lactose-active blend with 2.5% active using the automatic operation. The roller speed for the commercial equipment was set at high speed to maximize operational efficiency. The roller gap and roller pressure were adjusted to produce ribbons with the attributes identified in development as optimal for downstream processing. The feed-screw speed was allowed to float to maintain the powder feed rate sufficient to provide an adequate flow into the nip area.

The optimum ribbon attributes were identified from development results by multiple constraint optimization of the granule properties such as powder flow and compactability.

The variability in ribbon and granules properties depended on roller compaction process stability that is controlled by the powder flow to the feed screw during manufacture.

4.5.2 Ribbon Thickness and Density

Roller speed had a marginal impact on the thickness, recovery (e.g. increase in ribbon thickness after the roller due to material elasticity), or density of ribbons produced. Increasing the roller pressure increased the ribbon density and reduced the ribbon recovery. The augmented load experienced by the powder during its transit through the rollers resulted in increased consolidation producing stronger ribbons. The increased interparticle bonding resisted the relaxation post-compaction, which explains the reduced recovery seen. Increasing the roller gap produced thicker ribbons, while reducing the recovery. The effect of gap on ribbon density was minimal.

There was little interactive effect between gap and pressure on ribbon thickness (Figure 4.19). The desired ribbon thickness could be obtained simply by setting the gap to the necessary setting (under gap control) allowing for the appropriate relaxation. Equation (4.10) depicts this relationship, where h is the ribbon thickness (mm), G is the gap setting (mm), and P is the roll pressure (bar). The intercept accounts for the relaxation post-compaction.

$$h = 0.98 \times G - 0.002 \times P + 0.45 \quad (4.10)$$

There was a marginal two-way interactive effect between gap and pressure on ribbon density (Figure 4.20). Equation (4.11) depicts this relationship, where ρ_a is the

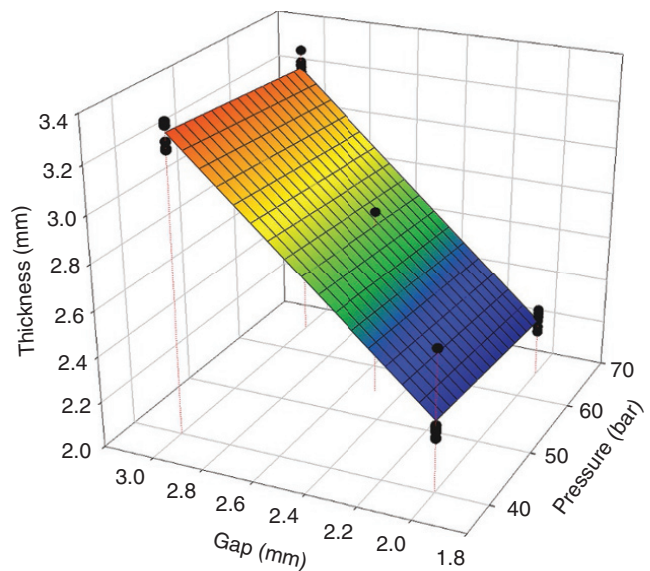


FIGURE 4.19 Effect of roller gap and roller pressure on ribbon thickness.

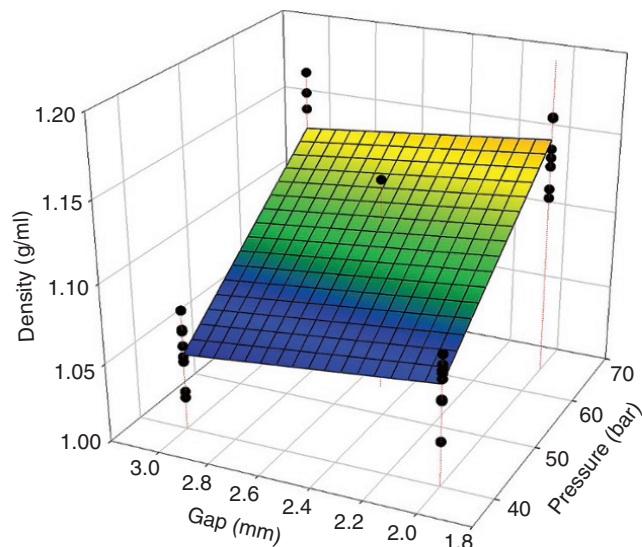


FIGURE 4.20 Effect of roller gap and roller pressure on ribbon density.

apparent ribbon density (g/ml), G is the gap setting (mm), and P is the roll pressure (bar). The intercept probably refers to the density in the powder bed just prior to significant consolidation.

$$\rho_a = 0.94 + 0.004 \times P - 0.008 \times G - 0.0001 \times P \times G \quad (4.11)$$

Ribbon densification was more efficient at lower gap settings, which aligns with the common understanding of force transmission through a powder bed under load. The desired ribbon density could be obtained simply by setting the load to the necessary setting (under gap control).

Figure 4.21 shows that ribbon recovery is determined by roll pressure and roll gap. Equation (4.12) depicts this relationship, where R is the apparent ribbon relaxation (%), ϵ refers to the maximum elastic recovery (19.8%), G is the gap setting (mm), and P is the roll pressure (bar).

$$R = \epsilon - 0.03 \times P - 3.298 \times G \quad (4.12)$$

Both applied pressure and gap influenced the degree of recovery seen post-compaction, with gap being more dominant. The least recovery is seen with high pressure and wide gap, and most recovery is seen with low pressure and narrow gap. The effect of compression pressure on recovery is opposite to that seen with tablets [140]. This is consistent with the view that weaker ribbons recover more (e.g. thickness increases after the roller).

4.5.3 Particle Size and Powder Flow

Increasing roller speed reduced the granule flow. This reduction was mainly attributed to the change in particle size

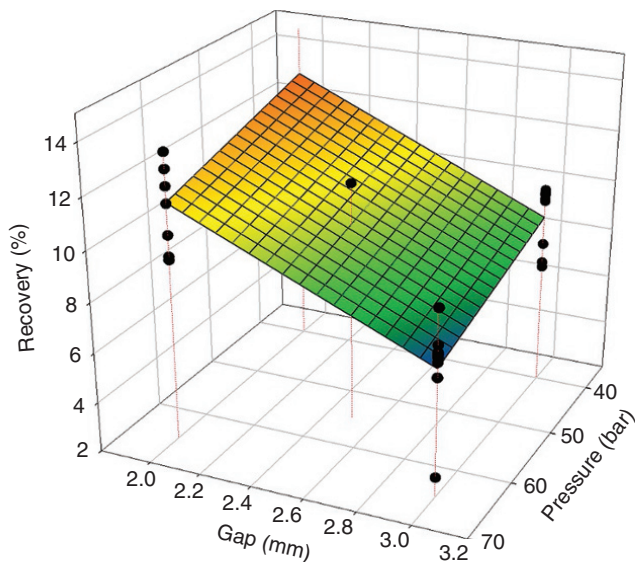


FIGURE 4.21 Effect of roller gap and roller pressure on ribbon recovery.

distribution; the mean particle size decreased while the fines increased. Potentially, this can be the result of increased residence time in the milling chamber. Increasing the input rate into the milling chamber, while keeping the output constant, would increase the residence time. Increased residence time should produce a particle size distribution skewed to smaller sizes (see Case Study 4.1).

Increasing the roller pressure used to produce the ribbons increased the powder flow. This increase is mainly attributed to the change in particle size distribution; the mean particle size increased while the fines decreased. This is consistent with the theory that stronger ribbon (produced at higher loads) would produce larger granules with fewer fines [74, 76].

Increasing the roller gap used to produce the ribbons decreased the powder flow. This reduction is mainly attributed to the change in particle size distribution; the mean particle size decreased while the fines increased. This is consistent with the theory that ribbons have a density gradient decreasing from the surface to the center. The thicker the ribbon is, the lower the density at the center. This low-density center will mimic ribbons made at lower loads and produce smaller particles during the milling process. Using principal component analysis, Soh et al. [29] showed that roller gap is an important predictor of granule properties. Figure 4.22 shows the interactive effect of roller pressure and roller gap on mass flow. The data shows that the ribbons made at opposite extremes of pressure and gap have minimum and maximum flow. Low pressure and large gap produced minimum flow; high pressure and narrow gap produced maximum flow. These granule properties are directly linked to the particle size distribution obtained from the ribbons. The data strongly

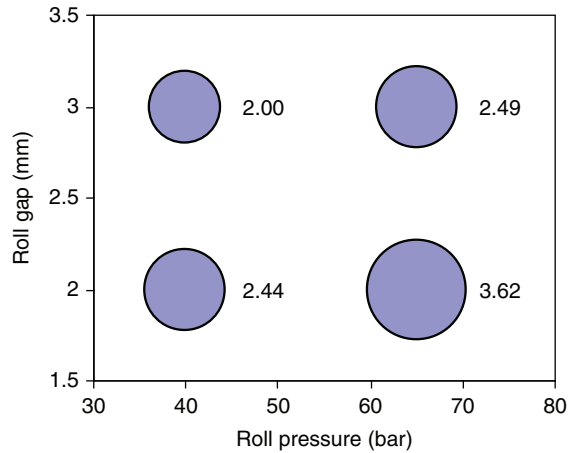


FIGURE 4.22 Effect of roller pressure and roller gap on granule mass flow (g/s).

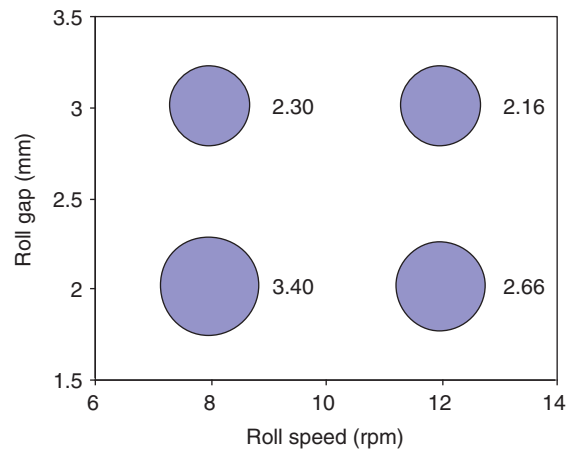


FIGURE 4.23 Effect of roller speed and roller gap on granule mass flow (g/s).

suggests that pressure can be used with some degree of success to offset the effect of increased gap. This clearly illustrates the importance of control over ribbon quality as a scale-up metric.

The interaction between roller speed and roller gap is notable since it affects process efficiency (Figure 4.23). The best flow is seen with a small gap running at slow speed, which coincides with the largest mean particle size and lowest amount of fines. These conditions would be the least efficient. Efficiency gains could be obtained with wider roller gaps at faster speeds. However, this combination yielded the slowest flow, which coincides with the smallest mean particle size and the highest amount of fines.

The effect of ribbon density and thickness on bulk mass flow is illustrated in Figure 4.24. Equation (4.13) depicts this relationship, where F is the mass flow (g/s), ρ_a is the apparent ribbon density (g/ml), and h is the ribbon thickness (mm).

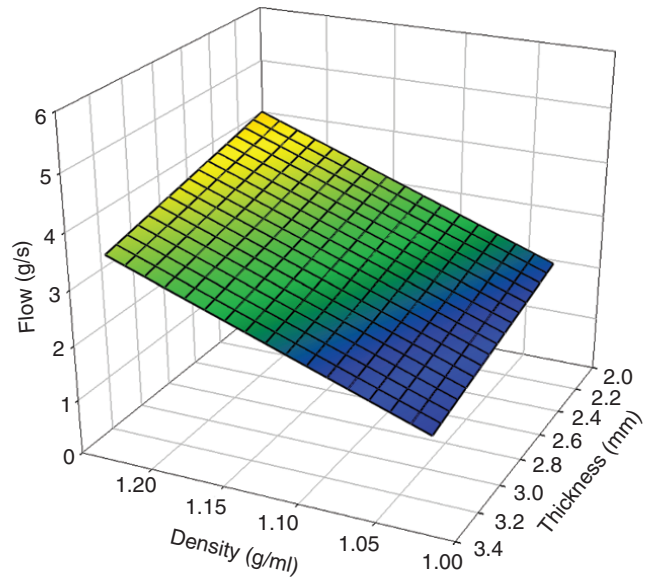


FIGURE 4.24 Effect of ribbon density and thickness on granule mass flow.

$$F = 9.49 \times \rho_a - 0.69 \times h - 0.582 \quad (4.13)$$

This coincided with the effect on granule mean size. Increased ribbon density produces larger granules with enhanced flow. Increased ribbon thickness leads to smaller granules with reduced flow. The effect of ribbon thickness on the amount of fines produced after milling was marginal, but the effect of ribbon density was significant. Higher ribbon densities correlated with higher amounts of fines. This could be due to increased residence time in the mill leading to more attrition by shear. Herting and Kleinebudde [45, 74, 76] reported a similar growth in particle size and improved flow with increased ribbon density.

Fragility of granules can be a main contributor to reduction in mass flow with processing post-roller compaction. This was due to particle attrition during additional mixing of the granules with extra-granular ingredients. To separate out the relative importance of changes in fines from changes in mean size on flow, the relative change in flow (final blend-granule) was evaluated against the relative change in mean size and fines (Figure 4.25). Equation (4.14) depicts this relationship, where ΔF is the change in mass flow (g/s), P is the roll pressure (bar), and G is the roll gap (mm).

$$\Delta F = -0.24 - 0.04 \times G + 0.004 \times P - 0.04 \times G^2 + 0.0004 \times P^2 \quad (4.14)$$

Flow was relatively resistant to change mean size over the range observed. Flow was quite sensitive to changes in the amount of fines. This indicates that efforts to improve flow are better spent reducing the amount of fines rather than

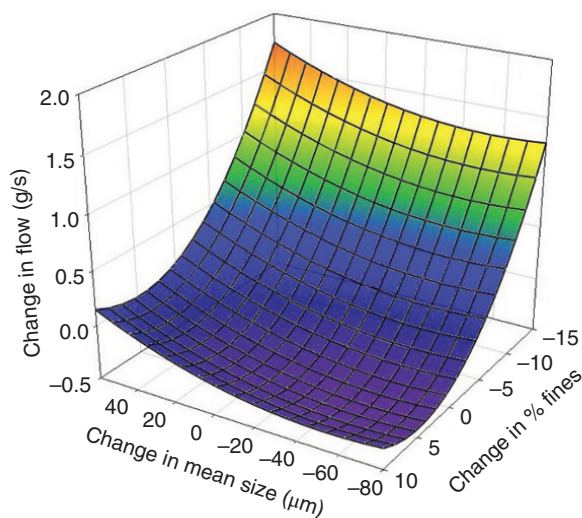


FIGURE 4.25 The interaction between change in mean diameter and fines on change in bulk mass flow.

optimizing the mean size. The bulk powder property of interest for tablet compression is the mass flow. Powder mass flow is governed by the properties of the particles making up the bulk. The two particle properties controlled by the roller compaction process are mean particle size and amount of fines (e.g. material <53 μm). These two particle properties have an opposite effect on mass flow. To maximize flow, the mean particle size would have to be increased to compensate for a higher amount of fines.

4.5.4 Powder Compactability

The performance of the compression operation depends on the characteristics of the input final blend. There are two main metrics that describe the performance of the process: compactability (how hard the press has to work to make tablets) and process stability (how variable the product properties are).

Ribbon production variables affected compactability, whereas ribbon milling conditions did not affect compactability appreciably. There were two populations of ribbons analyzed for compactability, one set corresponding to low roller pressure and one corresponding to high roller pressure. The ribbons compressed at low pressure had a lower apparent density and a higher compactability compared to ribbons compressed at high pressure that had a higher apparent density and a lower compactability. Malkowska and Khan [78] reported a similar observation of increased compactability with a reduction in ribbon density.

In addition to density, ribbons also had a characteristic thickness. The effect of density and thickness on compactability was evaluated. Figure 4.26 suggests that ribbon thickness had only a small impact, while ribbon density had a

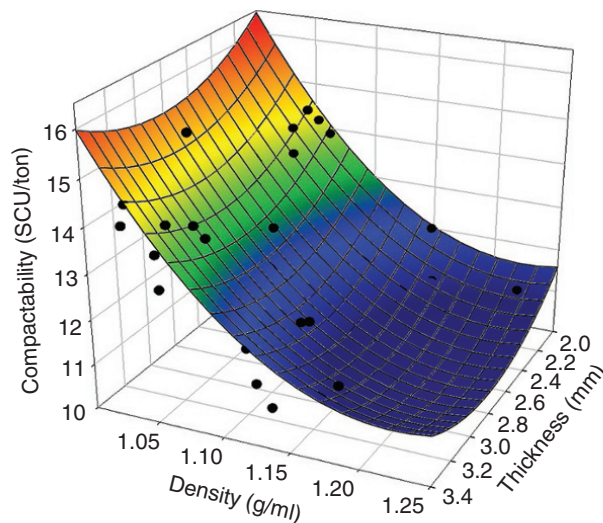


FIGURE 4.26 Effect of ribbon density and thickness on final blend compactability.

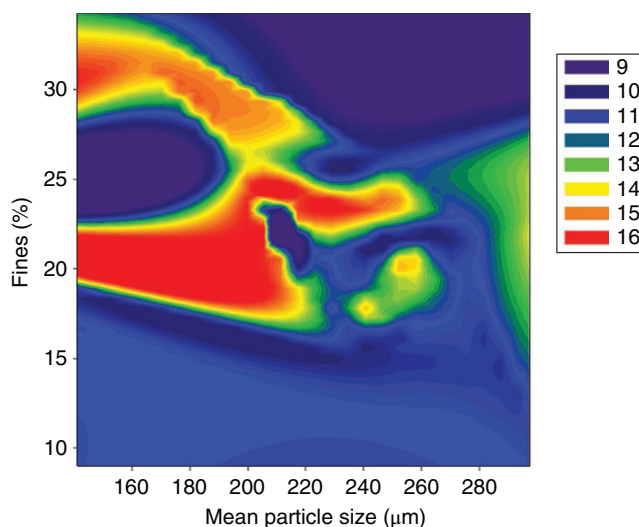


FIGURE 4.27 Effect of mean particle size and amount of fines on final blend compactability.

substantial impact on compactability. Equation (4.15) depicts this relationship, where K is the compactability (SCU/ton), ρ_a is the apparent ribbon density (g/ml), and h is the ribbon thickness (mm).

$$K = 174 - 9.74 \times \rho_a - 245 \times h + 1.74 \times \rho_a^2 + 100 \times h^2 \quad (4.15)$$

In addition to the direct effect of density on compression, the link between density and compactability may also be mediated through particle size (Figure 4.27).

Blends with lower amounts of fines or a small mean particle size had on average higher compactability. There was a direct relationship between ribbon density and mean size, but

an inverse relationship between the ribbon density and the amount of fines. The hypothesis explaining increased compactability with decreased ribbon density includes two parts. The ribbon density (and by extension the particle density) affects the particle strength and hence the force needed for deformation. Ribbon density affects the particle size distribution (mean size and fines) of the resulting final blend. A higher amount of fines and a smaller mean particle size promotes compactability.

4.5.5 Process Stability

Process quality has two measurable quantities, i.e. process stability and process repeatability. Process stability is defined as the variance in the metric of interest during a single batch run (within run variability). Process repeatability or capability is defined as the variance in the metric of interest across a study run (between run variability), also as the capability of a process to meet its purpose. This analysis can be done on unit operations for which the parameters are not varied, or for upstream operation variables under evaluation.

Process repeatability is a metric for the downstream effect of the roller compaction unit operation. Based on the proposed target core tablet attributes (tablet weight, hardness, and thickness), the process capability indices (C_p) [141] were calculated. A C_p value of 2 is generally expected for a process under control (6σ).

The C_p analysis based on tablet weight showed a compression operation unaffected by the roller compaction settings. Process capability (C_p) based on tablet hardness was improved with a decrease in roller gap. A narrower roller gap correlated with a faster mass flow, higher recovery, and lower amount of fines. Of these three coincidental factoids, mass flow variations would have been reflected in the weight-based C_p analysis. In principle, variations in recovery and fines amount affect compactability, and, by reasoning, hardness-based C_p .

4.6 SUMMARY

Dry granulation is an effective means of improving the performance of powders. Improvements in flow and compaction for most pharmaceutical applications aid in the filling of capsules, and compression of tablets. Although several methods are available, roller compaction is the predominant method for the production of granules without solvent. The roller compaction unit operation is comprised of several subprocesses. Powder feeding and milling, the first and last of these subprocesses, have large bodies of knowledge which can be adapted to roller compaction. The compaction step itself is of more interest when designing the roller compaction process. Material properties, in-process and final product measurements, and their interaction with the feed rate, roller speed,

roller pressure, and gap have been reviewed and examples given for process development and scale-up. This information confirms the utility of roller compaction and provides tools for designing the roller compaction process.

Symbols (noted by section)

Principles of Operation

Consolidation and Compaction in Between the Rollers

α = nip angle

δ = angle of internal friction

δ_E = effective angle of internal friction

ϑ = angular roll position

ϑ_h = angular position at which feed pressure is applied

μ = friction coefficient

ν = acute angle roller tangent to normal stress

ρ = density

τ = shear stress

ϕ, ϕ_w = angle of wall stress

χ = position

S = gap

Scale-Up of Roller Compaction

Example 4.1: Parametric Scale-Up with a Procedure and Parameters Recommended from the Equipment Manufacturer

a = outer layer flake (mm), can be neglected on small compactors

G = roller gap (mm)

B = roller width (mm)

D = outer roller diameter (mm)

P = hydraulic pressure (roller pressure)

F = compaction force per unit distance of roller width (kN/cm_{roller width})

R_{TH} = required throughput (kg/h)

n = roller speed (1/min)

ϵ = maximum elastic recovery

R = apparent ribbon relaxation

ρ_0 = density of compacted ribbon (kg/dm³)

h = thickness of compacted ribbon

K = compactability

t_c = correction factor for dwell time

Scale-Up of Roller Compaction

Case Study 4.2: Attribute-Based Scale-Up

h = ribbon thickness (mm)

G = gap setting (mm)

P = roll pressure (bar)

ρ_a = apparent ribbon density (g/ml)

R = apparent ribbon relaxation (%)

ϵ = maximum elastic recovery (%)

F = mass flow (g/s)

K = compactability (SCU/ton)

C_p = process capability index

REFERENCES

- Nishii, K. and Horio, M. (2007). Dry granulation. In: *Granulation*, vol. 11 (ed. A.D. Salmon, M.J. Hounslow and J.P.K. Seville), 289–322. Amsterdam: Elsevier.
- Horio, M. (2003). Binderless granulation – its potential, achievements and future issues. *Powder Technology* 130: 1–7.
- Pietsch, W. (2004). *Agglomeration in Industry*, vol. 1. Weinheim: Wiley-VCH.
- Johnson CE, Production of granular products by roll compaction. In: Beestman GB, Vander Hooven DIB editors. *Pesticide Formulations and Application Systems: 7th Volume*, ASTM Special Technical Publication 968. Philadelphia, PA: American Society for Testing and Materials, 1987; 199–206
- Bakele, W. (2003). New developments in the production of heavy soda-ash via compacting method. *Powder Technology* 130: 253–256.
- Feige, F. (2011). The stress on defined, laterally confined, layers of material between horizontally supported rollers – a review. *Cement International* 9: 30–41.
- Commuri, S., Mai, A.T., and Zaman, M. (2011). Neural network-based intelligent compaction analyzer for estimating compaction quality of hot asphalt mixes. *Journal of Construction Engineering and Management* 137 (9): 634–644.
- Kleinebudde, P. (2004). Roll compaction/dry granulation: pharmaceutical applications. *European Journal of Pharmaceutics and Biopharmaceutics* 58: 317–326.
- Mouro, D., Noack, R., Musico, B., et al. 2006. Enhancement of xcelodose capsule-filling capabilities using roller compaction. *Pharmaceutical Technology*. http://license.icopyright.net/3.74587?icx_id=301460 (accessed April 18 2009).
- Majerova, D., Rysl, T., and Zamostny, P. (2014). Optimizing the properties of blend for hard gelatin capsules filling by incorporating roller compaction in manufacturing process. Proceedings of the 21st International Congress of Chemical and Process Engineering (CHISA 2014), Prague, Czech Republic (23–27 August 2014). <http://toc.proceedings.com/31465webtoc.pdf> (accessed 14 November 2018).
- Bejugam, N.K., Mutyam, S.K., and Shankar, G.N. (2015). Tablet formulation of an active pharmaceutical ingredient with a sticking and filming problem: direct compression and dry granulation evaluations. *Drug Development and Industrial Pharmacy* 41 (2): 331–341.
- Dalziel, G., Nauka, E., Zhang, F. et al. (2013). Assessment of granulation technologies for an API with poor physical properties. *Drug Development and Industrial Pharmacy* 39 (7): 985–995.
- Hadzovic, E., Betz, G., Hadzidedic, S. et al. (2011). Investigation of compressibility and compacted theophylline and its binary mixtures. *International Journal of Pharmaceutics* 416: 97–103.
- Gupta, A., Peck, G.E., Miller, R.W., and Morris, K.R. (2005). Influence of ambient moisture on the compaction behavior of microcrystalline cellulose powder undergoing uni-axial compression and roller compaction: a comparative study using near-infrared spectroscopy. *Journal of Pharmaceutical Sciences* 94 (10): 2301–2313.
- Omar, C., Dhenge, R.M., Osborne, J.D. et al. (2015). Roller compaction: effect of amorphous content of lactose upon product properties. *International Journal of Pharmaceutical Sciences*. 496 (1): 63–74.
- Miller, R.W. (1994). Advances in pharmaceutical roller compactor feed system designs. *Pharmaceutical Technology* March: 154–162.
- Miller RW, Roller compaction technology. In: Parikh D editor, *Handbook of Pharmaceutical Granulation Technology*. 2nd edition New York: Taylor & Francis; 2005. p 159–172.
- Daugherty, P.D. and Chu, J.H. (2007). Investigation of serrated roll surface differences on ribbon thickness during roll compaction. *Pharmaceutical Development and Technology* 12: 603–608.
- Rambali, B., Baert, L., Jans, E., and Massart, D.L. (2001). Influence of the roll compactor parameter settings and the compression pressure on the buccal bio-adhesive tablet properties. *International Journal of Pharmaceutics* 220: 129–140.
- Tan, B.M.J., Chan, L.W., and Heng, P.W.S. (2016). Determination of the Nip angle in roller compactors with serrated rolls. *Journal of Pharmaceutical Sciences* 105: 1967–1975.
- Pietsch, W. (1997). Granulate dry particulate solids by compaction and retain key powder particle properties. *Chemical Engineering Progress* April: 24–46.
- Guangsheng, H., Lingyun, W., Zhongwei, Z. et al. (2005). Manufacturing technique of magnesium alloy sheets by powder rolling. *Materials Science Forum* 488–489: 445–448.
- Sandler, N. and Lammens, R.F. (2011). Pneumatic dry granulation: potential to improve roller compaction technology in drug manufacture. *Expert Opinion on Drug Delivery* 8 (2): 225–236.
- Saarienen, T., Antikainen, O., and Yliruusi, J. (2017). Simultaneous comparison of two roller compaction techniques and two particle size analysis methods. *AAPS PharmSciTech* 18 (8): 3198–3207.
- Bacher, C., Olsen, P.M., Bertelsen, P., and Sonnergaard, J.M. (2008). Granule fraction inhomogeneity of calcium carbonate/sorbitol in roller compacted granules. *International Journal of Pharmaceutics* 349: 19–23.
- Singh, R., Sahay, A., Karry, K.M. et al. (2014). Implementation of an advanced hybrid MPC-PID control system using PAT tools into a direct compaction continuous pharmaceutical tablet manufacturing pilot plant. *International Journal of Pharmaceutics* 473: 38–54.
- Singh, R., Ierapetritou, M., and Ramachandran, R. (2013). System-wide hybrid MPC-PID control of a continuous pharmaceutical tablet manufacturing process via direct compaction. *European journal of Pharmaceutics and Biopharmaceutics* 85: 1164–1182.
- Johanson, J.R. (1965). A rolling theory for granular solids. *Transactions of the American Society of Mechanical Engineers* December: 842–848.
- Soh, J.L.P., Boersen, N., Carvajal, T.M. et al. (2007). Importance of raw material attributes for modeling ribbon and granule properties in roller compaction: multivariate analysis on roll

- gap and NIR spectral slope as process critical control parameters. *Journal of Pharmaceutical Innovation* 2: 106–124.
30. Soh, J.L.P., Wang, F., Boersen, N. et al. (2008). Utility of multivariate analysis in modeling the effects of raw material properties and operating parameters on granule and ribbon properties prepared in roller compaction. *Drug Development and Industrial Pharmacy* 34: 1022–1035.
 31. Gupte, A., DeHart, M., Stagner, W.C., and Haware, R.V. (2017). Comparative binder efficiency modeling of dry granulation binders using roller compaction. *Drug Development and Industrial Pharmacy* 43 (4): 574–583.
 32. Allesso, M., Torstenson, A.S., Bryder, M., and Holm, P. (2013). Presenting a rational approach to QbD-based pharmaceutical development: a roller compaction case study. *European Pharmaceutical Review* 18 (6): 3–10.
 33. Leane, M., Pitt, K., and Reynolds, G. (2015). A proposal for a drug product manufacturing classification system (MCS) for oral solid dosage forms. *Pharmaceutical Development and Technology* 20 (1): 12–21.
 34. Shi, Z. and Hilden, J.L. (2017). Small-scale modeling of pharmaceutical powder compression from tap density testers, to roller compactors, and to the tablet press using big data. *Journal of Pharmaceutical Innovation* 12: 41–48.
 35. Sheskey, P.J. and Dasbach, T.P. (1995). Evaluation of various polymers as dry binders in the preparation of an immediate-release tablet formulation by roller compaction. *Pharmaceutical Technology* 19 (10): 98–112.
 36. Falzone, A.M., Peck, G.E., and McCabe, G.P. (1992). Effects of changes in roller compactor parameters on granulation produced by compaction. *Drug Development and Industrial Pharmacy* 18: 469–489.
 37. Inghelbrecht, S. and Remon, J.P. (1998). Roller compaction and tableting of microcrystalline cellulose/drug mixtures. *International Journal of Pharmaceutics* 161: 215–224.
 38. Mitra, B., Hilden, J., and Litster, J.D. (2015). Novel use of monodisperse granules to deconvolute impacts of granule size versus granule solid fraction on tablet tensile strength. *Advanced Powder Technology* 26: 553–562.
 39. Riepma, K.A., Vromans, H., Zuurman, K., and Lerk, C.F. (1993). The effect of dry granulation on the consolidation and compaction of crystalline lactose. *International Journal of Pharmaceutics* 97: 29–38.
 40. Inghelbrecht, S. and Remon, J.P. (1998). The roller compaction of different types of lactose. *Journal of International Pharmaceutics* 166: 135–144.
 41. Sheskey, P.J., Cabelka, T.D., Robb, R.T., and Boyce, B.M. (1994). Use of roller compaction in the preparation of controlled-release hydrophilic matrix tablets containing methylcellulose and hydroxypropyl methylcellulose polymers. *Pharmaceutical Technology* 18 (9): 132–150.
 42. Heiman, J., Tajarobi, F., Bindhumadhavan, G. et al. (2014). *AAPS PharmSciTech* 16 (2): 267–277.
 43. Freitag, F. and Kleinebudde, P. (2003). How do roll compaction/dry granulation affect the tableting behavior of inorganic materials? Comparison of four magnesium carbonates. *European Journal of Pharmaceutical Sciences* 19: 281–189.
 44. Freitag, F., Runge, J., and Kleinebudde, P. (2005). Coprocessing of powdered cellulose and magnesium carbonate: direct tableting versus tableting after roll compaction/dry granulation. *Pharmaceutical Development and Technology* 10: 353–362.
 45. Herting, M.G., Klose, K., and Kleinebudde, P. (2007). Comparison of different dry binders for roll compaction/dry granulation. *Pharmaceutical Development and Technology* 12 (5): 525–532.
 46. O'Mahony, G.E., Crowley, M.E., Dontireddy, R., et al. (2014). Behaviour of HPMC and microcrystalline cellulose during a roller compaction process. American Institute of Chemical Engineers National Meeting, November 2014, <https://aiche.confex.com/aiche/2014/webprogram/Paper385335.html> (accessed 22 January 2018).
 47. Dave, V.S., Fahmy, R.M., Bensley, D., and Hoag, S.W. (2012). Eudragit RS PO/RL PO as rate-controlling matrix-formers via roller compaction: influence of formulation and process variables on functional attributes of granules and tablets. *Drug Development and Industrial Pharmacy* 38 (10): 1240–1253.
 48. Dave, V.S., Fahmy, R.M., and Hoag, S.W. (2015). Near-infrared spectroscopic analysis of the breaking force of extended-release matrix tablets prepared by roller-compaction: influence of plasticizer levels and sintering temperature. *Drug Development and Industrial Pharmacy* 41 (6): 898–905.
 49. Gangurde, A., Patole, R.K., Sav, A.K., and Amin, P.D. (2013). A novel directly compressible co-processed excipient for sustained release formulation. *Journal of Applied Pharmaceutical Science* 3 (9): 89–97.
 50. Khomane, K.S. and Bansal, A.K. (2014). Differential compaction behaviour of roller compacted granules of clopidogrel bisulphate polymorphs. *International Journal of Pharmaceutics* 472: 288–295.
 51. Matji A, Carvajal I I, Conde F, Pena MA, Donato N, Serrano DR, Torrado JJ, Effect of the characteristics of raw material ibuprofen on roller compaction and dissolution. *Journal of Drug Delivery Science and Technology*, 2017; 42: 237–244.
 52. Lopes, D.G., Garsuch, V., Becker, K. et al. (2016). Improving the granule strength of roller-compacted ibuprofen sodium for hot-melt coating processing. *International Journal of Pharmaceutics* 510: 285–295.
 53. He, X., Seceast, P.J., and Amidon, G.E. (2007). Mechanistic study of the effect of roller compaction and lubricant on tablet mechanical strength. *Journal of Pharmaceutical Sciences* 96 (5): 1342–1355.
 54. Dawes, J., Gamble, J.F., Greenwood, R. et al. (2012). An investigation into the impact of magnesium stearate on powder feeding during roller compaction. *Drug Development and Industrial Pharmacy* 38 (1): 111–122.
 55. Dawes, J., Allenspach, C., Gamble, J.F. et al. (2013). Application of external lubrication during the roller compaction of adhesive pharmaceutical formulations. *Pharmaceutical Development and Technology* 18 (1): 246–256.
 56. Inghelbrecht, S. and Remon, J.P. (1998). Reducing dust and improving granule and tablet quality in the roller compaction process. *International Journal of Pharmaceutics* 171: 195–206.

57. Bultmann, J.M. (2002). Multiple compaction of microcrystalline cellulose in a roller compactor. *European Journal of Pharmaceutics and Biopharmaceutics* 54: 59–64.
58. Mirtic, A. and Reynolds, G.K. (2016). Determination of breakage rate and breakage mode of roller compacted pharmaceutical materials. *Powder Technology* 298: 99–105.
59. Campbell, G.M., Bunn, P.J., Webb, C., and Hook, S.C.W. (2001). On predicting roller milling performance: part II. The breakage function. *Powder Technology* 115: 243–255.
60. Barrasso, D., Oka, S., Muliadi, A. et al. (2013). Population balance model validation and prediction of CQAs for continuous milling processes: toward QbD in pharmaceutical drug product manufacturing. *Journal of Pharmaceutical Innovation* 8: 147–162.
61. Fonteyne, M., Vercruyse, J., DeLeersnyder, F. et al. (2015). Process analytical technology for continuous manufacturing of solid-dosage forms. *Trends in Analytical Chemistry* 67: 159–166.
62. Kona, R., Fahmy, R.M., Claycamp, G. et al. (2015). Application of near-infrared spectroscopy to monitor roller compaction in-process and product quality attributes of immediate release tablets. *AAPS PharmSciTech* 16 (1): 202–216.
63. Qiu, J., Li, N., Pramesti, F.P. et al. (2012). Evaluating laboratory compaction of asphalt mixtures using the shear box compactor. *Journal of Testing and Evaluation* 40 (5): 844–852.
64. Miguelez-Moran, A.M., Wu, C.Y., and Seville, J.P.K. (2008). The effect of lubrication on density distributions on roller compacted ribbons. *International Journal of Pharmaceutics* 362: 52–59.
65. Weyenberg, W., Vermeire, A., Vandervoort, J. et al. (2005). Effects of roller compaction settings on the preparation of bio-adhesive granules and ocular minitables. *European Journal of Pharmaceutics and Biopharmaceutics* 59: 527–536.
66. Freeman, T., Vom Bey, H., Hanish, M. et al. (2016). The influence of roller compaction processing variables on the rheological properties of granules. *Asian Journal of Pharmaceutical Sciences* 11: 516–527.
67. Zhang, J.Q. and Yan, Y. (2003). On-line continuous measurement of particle size using electrostatic sensors. *Powder Technology* 135: 164–168.
68. Frake, P., Greenhalg, D., Grierson, S.M. et al. (1997). Process control and endpoint determination of a fluid bed granulation by application of near infra-red spectroscopy. *International Journal of Pharmaceutics* 151: 75–80.
69. Rantanen, J. and Yliruusi, J. (1998). Determination of particle size in a fluidized bed granulator with a near infrared (NIR) set-up. *Pharmaceutics and Pharmacology Communications* 4: 73–75.
70. McAuliffe, M.A.P., O'Mahony, G.E.O., Balckshields, C.A. et al. (2015). The use of PAT and off-line methods for monitoring of roller compacted ribbon and granule properties with a view to continuous processing. *Organic Process Research and Development* 19: 158–166.
71. Rantanen, J., Rasanen, E., Tenhunen, J. et al. (2000). In-line moisture measurement during granulation with a four-wavelength near infrared sensor: an evaluation of particle size and binder effects. *European Journal of Pharmaceutics and Biopharmaceutics* 50: 271–276.
72. Bordes, C., Garcia, F., Snabre, P., and Frances, C. (2002). On-line characterization of particle size during an ultrafine wet grinding process. *Powder Technology* 128: 218–228.
73. Liao, C.W. and Targ, Y.S. (2009). On-line automatic optical inspection system for coarse particle size distribution. *Powder Technology* 189: 508–513.
74. Herting, M.G. and Kleinebudde, P. (2007). Roll compaction/dry granulation: effect of raw material particle size on granule and tablet properties. *International Journal of Pharmaceutics* 338: 110–118.
75. Patel, C., Kaushal, A.M., and Bansal, A.K. (2008). Compaction behavior of roller compacted ibuprofen. *Eur. J. Pharmaceutics and Biopharmaceutics* 69: 743–749.
76. Herting, M.G. and Kleinebudde, P. (2008). Reduction of tensile strength of tablets after roll compaction/dry granulation. *International Journal of Pharmaceutics* 70: 372–378.
77. Sun, C.C. and Himmilspach, M.W. (2006). Reduced tabletability of roller compacted granules as a result of granule size enlargement. *International Journal of Pharmaceutics* 95: 200–206.
78. Malkowska, S. and Khan, K.A. (1983). Effect of recompression on the properties of tablets prepared by dry granulation. *Drug Development and Industrial Pharmacy* 9: 331–347.
79. Sheskey, P.J. and Cabelka, T.D. (1992). Re-workability of sustained release tablet formulation containing HPMC polymers. *Pharmaceutical Technology* 7: 60–74.
80. Miguelez-Moran, A.M., Wu, C.Y., Dong, H., and Seville, J.P.K. (2009). Characterisation of density distributions in roller-compacted ribbons using micro-indentation and X-ray micro-computed tomography. *European Journal of Pharmaceutics and Biopharmaceutics* 72: 173–182.
81. Alleso, M., Holm, R., and Holm, P. (2016). Roller compaction scale-up using roll width as scale factor and laser-based determined ribbon porosity as critical material attribute. *European Journal of Pharmaceutical Sciences* 87: 69–78.
82. Sun, W.J. and Sun, C.C. (2017). Ribbon thickness influences fine generation during dry granulation. *International Journal of Pharmaceutics* 529: 87–88.
83. Zinchuk, A.V., Mullarney, M.P., and Hancock, B. (2004). Simulation of roller compaction using a laboratory scale compaction simulator. *International Journal of Pharmaceutics* 269: 403–415.
84. Farber, L., Hapgood, K.P., Michaels, J.N. et al. (2008). Unified compaction curve model for tensile strength of tablets made by roller compaction and direct compression. *International Journal of Pharmaceutics* 346: 17–24.
85. Khorasani, M., Amigo, J.M., Sun, C.C. et al. (2015). Near-infrared chemical imaging (NIR-CI) as a process monitoring solution for production line of roll compaction and tableting. *European Journal of Pharmaceutics and Biopharmaceutics* 93: 293–302.
86. Jain, S., Wang, Y.H., and Fredlund, D.G. (2015). Non-contact sensing system to measure specimen volume during shrinkage test. *Geotechnical Testing Journal* 38 (6): 936–949.

87. Sprockel, O.L., Yang, E., Jayawickrama, D., and Li, L. (2008). Ultrasound propagation in porous materials. *AICHE annual meeting* 294.
88. Akseli, I., Iyer, S., Lee, H.P., and Cuitino, A.M. (2011). A quantitative correlation of the effect of density distributions in roller-compacted ribbons on the mechanical properties of tablets using ultrasonics and X-ray tomography. *AAPS PharmSciTech* 12 (3): 834–853.
89. Feng, T., Feng, W., Pinal, R. et al. (2008). Investigation of the variability of NIR in-line monitoring of roller compaction process by using fast fourier transform (FFT) analysis. *Pharmaceutical Science and Technology* 9: 419–424.
90. Acevedo, D., Muliadi, A., Gridhar, A. et al. (2012). Evaluation of three approaches for real-time monitoring of roller compaction with near-infrared spectroscopy. *AAPS PharmSciTech* 13 (3): 1005–1012.
91. Crowley, M.E., Hegarty, A., McAuliffe, M.A.P. et al. (2017). Near-infrared monitoring of roller compacted ribbon density: Investigating sources of variation contributing to noisy spectral data. *European Journal of Pharmaceutical Sciences* 102: 103–114.
92. Austin, J., Gupta, A., Reklatis, G.V., and Harris, M.T. (2013). The use of near-infrared and microwave resonance sensing to monitor a continuous roller compaction process. *Journal of Pharmaceutical Sciences* 102 (6): 1895–1904.
93. Lim, H., Dave, V.S., Kidder, L. et al. (2011). Assessment of the critical factors affecting the porosity of roller compacted ribbons and the feasibility of using NIR chemical imaging to evaluate the porosity distribution. *International Journal of Pharmaceutics* 410: 1–8.
94. Quyet, P.V., Samanta, A.K., Liew, C.V. et al. (2013). A prediction model for monitoring ribbed roller compacted ribbons. *Journal of Pharmaceutical Sciences* 102 (8): 2667–2677.
95. Talwar, S., Nunes, C., Stevens, T. et al. (2016). Understanding the impact of chemical variability and calibration algorithms on prediction of solid fraction of roller compacted ribbons using near-infrared (NIR) spectroscopy. *Applied Spectroscopy* 71 (6): 1209–1221.
96. Ghorab, M.K., Chatlapalli, R., Hasan, S., and Nagi, A. (2007). Application of thermal effusivity as a process analytical technology tool for monitoring and control of the roller compaction process. *Pharmaceutical Science and Technology* 8: 1–7.
97. Sheskey, P.J. and Hendren, J. (1999). The effects of roll compaction equipment variables, granulation technique, and HPMC polymer level on a controlled-release matrix model drug formulation. *Pharmaceutical Technology* 223: 90–106.
98. Guigon, P. and Simon, O. (2000). Interaction between feeding and compaction during lactose compaction in a laboratory roll press. *KONA* 18: 131–138.
99. Guigon, P. and Simon, O. (2003). Roll press design – influence of force feed systems on compaction. *Powder Technology* 130: 41–48.
100. Funakoshi, Y., Asogawa, T., and Satake, E. (1977). The use of a novel roller compactor with a concavo-convex roller pair to obtain uniform compacting pressure. *Drug Development and Industrial Pharmacy* 3 (6): 555–573.
101. Parrott, E.L. (1981). Densification of powders by concavo-convex roller compactor. *Journal of Pharmaceutical Science* 70: 288–291.
102. Romero-Torres, S., Chokshi, R., Huang, J., and Krishan, M. (2011). Concomitant use of near-infrared chemical imaging and Raman spectroscopy during the selection of a solid dosage manufacturing platform. *American Pharmaceutical Review* 14: 10–17.
103. Alleso, M., Holm, P., Carstensen, J.M., and Holm, R. (2016). Quantitative surface topography assessment of directly compressed and roller compacted tablet cores using photometric stereo image analysis. *European Journal of Pharmaceutical Sciences* 87: 79–97.
104. Alexander, A.W. and Muzzio, F.J. (2002). Batch size increase in dry blending and mixing. In: *Pharmaceutical Process Scale Up* (ed. M. Levin), 115–132. New York: Informa Health Care.
105. Prescott, J.K. (2002). Powder handling. In: *Pharmaceutical Process Scale Up* (ed. M. Levin), 133–150. New York: Informa Health Care.
106. Yang, S. and Evans, J.R.G. (2007). Metering and dispensing of powder; the quest for new solid freeforming techniques. *Powder Technology* 178: 56–72.
107. Sander, U. and Schonert, K. (1999). Operation conditions of a screw-feeder-equipped high-pressure roller mill. *Powder Technology* 105: 282–287.
108. Simon, O. and Guigon, P. (2003). Correlation between powder-packing properties and roll press compact heterogeneity. *Powder Technology* 130: 257–264.
109. Lecompte, T., Doremus, P., Thomas, G. et al. (2005). Dry granulation of organic powders – dependence of pressure 2D distribution on different process parameters. *Chemical Engineering Science* 60: 3933–3940.
110. Podczeczek, F. (1998). *Particle-Particle Adhesion in Pharmaceutical Powder Handling*. London: Imperial College Press.
111. Bindhumadhavan, G., Seville, J.P.K., Adams, M.J. et al. (2005). Roll compaction of a pharmaceutical excipient: experimental validation of rolling theory for granular solids. *Chemical Engineering Science* 60: 3891–3897.
112. Hervieu, P., Dehont, F., Jerome, E. et al. (1994). Granulation of pharmaceutical powders by compaction an experimental study. *Drug Development and Industrial Pharmacy* 20 (1): 65–74.
113. am Ende, M.T., Moses, S.K., Carella, A.J. et al. (2007). Improving the content uniformity of a low-dose tablet formulation through roller compaction optimization. *Pharmaceutical Development and Technology* 12: 391–404.
114. Zhao, J., Koo, O., Pan, D. et al. (2017). The impact of disintegrant type, surfactant, and API properties on the processability and performance of roller compacted formulations of acetaminophen and aspirin. *AAPS Journal* 19 (5): 1387–1395.
115. Rajkumar, A.D., Reynolds, G.K., Wilson, D. et al. (2016). Investigating the effect of processing parameters on pharmaceutical tablet disintegration using a real-time particle imaging approach. *European Journal of Pharmaceutics and Biopharmaceutics* 106: 88–96.

116. Rekhi GS, Sidwell R, Sizing of granulation. In: Parikh D editor, *Handbook of Pharmaceutical Granulation Technology*. 2nd ed. New York: Taylor & Francis; 2005: 159–172
117. Sajjia, M., Shirazian, S., Egan, D. et al. (2017). Mechanistic modelling of industrial-scale roller compactor “Freund TF-MINI model”. *Computers and Chemical Engineering* 104: 141–150.
118. Yusof, Y.A., Smith, A.C., and Briscoe, B.J. (2005). Roll compaction of maize powder. *Chem. Eng. Sci.* 60: 3919–3931.
119. Sommer, K. and Hauser, G. (2003). Flow and compression properties of feed solids for roll-type presses and extrusion presses. *Powder Technology* 130: 272–276.
120. Yehia, K.A. (2007). Estimation of roll press design parameters based on the assessment of a particular nip region. *Powder Technology* 177: 148–153.
121. Dec, R.T., Zavaliangos, A., and Cunningham, J.C. (2003). Comparison of various modeling methods for analysis of powder compaction in roller press. *Powder Technology* 130: 265–271.
122. Schonert, K. and Sander, U. (2002). Shear stresses and material slip in high pressure roller mills. *Powder Technology* 122: 136–144.
123. Michrafy, A., Zavaliangos, A., and Cunningham, J. (2017). Dry granulation process modeling. In: *Predictive Modeling of Pharmaceutical Unit Operations* (ed. P. Pandey and R. Bharadwaj), 71–97. New York: Elsevier Ltd.
124. Hilden, J. and Earle, G. (2011). Prediction of roller compacted ribbon solid fraction for quality by design development. *Powder Technology* 213: 1–13.
125. Bi, M., Alvarez-Nunez, F., and Alvarez, F. (2014). Evaluating and modifying Johanson’s rolling model to improve its predictability. *Journal of Pharmaceutical Sciences* 103: 2062–2071.
126. Hein, S., Picker-Freyer, K.M., and Langridge, J. (2008). Simulation of roller compaction with subsequent tableting and characterization of lactose and microcrystalline cellulose. *Pharmaceutical Development and Technology* 13: 523–532.
127. Liu, Z., Bruwer, M.J., MacGregor, J.F. et al. (2011). Scale-up of a pharmaceutical roller compaction process using a joint-Y partial least squares model. *Industrial and Engineering Chemistry Research* 50: 10696–10706.
128. Muliadi, A.R., Litster, J.D., and Wassgren, C.R. (2012). Modeling the powder roll compaction process: comparison of 2-D finite element method and the rolling theory for granular solids (Johanson’s model). *Powder Technology* 221: 90–100.
129. Muliadi, A.R., Lister, J.D., and Wassgren, C.R. (2013). Validation of 3-D finite element analysis for predicting the density distribution of roll compacted pharmaceutical powder. *Powder Technology* 237: 386–399.
130. Inghelbrecht, S., Remon, J.-P., Fernandes de Aguiar, P. et al. (1997). Instrumentation of a roll compactor and the evaluation of the parameter setting by neural networks. *International Journal of Pharmaceutics* 148: 103–115.
131. Turkoglu, M., Aydin, I., Murray, M., and Sakr, A. (1999). Modeling of a roller-compaction process using neural networks and genetic algorithms. *European Journal of Pharmaceutics and Biopharmaceutics* 48: 239–245.
132. Mansa, R.F., Bridson, R.H., Greenwood, R.W. et al. (2008). Using intelligent software to predict the effects of formulation and processing parameters on roller compaction. *Powder Technology* 181: 217–225.
133. Shi, W. and Sprockel, O.L. (2016). A practical approach for scale-up of roller compaction process. *European Journal of Pharmaceutics and Biopharmaceutics* 106: 15–19.
134. Process Systems Enterprise gPROMS. <https://www.psenderprise.com/products/gproms> (accessed 29 January 2018).
135. Nesarikar, V.V., Vatsaraj, N., Patel, C. et al. (2012). Instrumented roll technology for the design space development of roller compaction process. *International Journal of Pharmaceutics* 426: 116–131.
136. Nesarikar, V.V., Patel, C., Early, W. et al. (2012). Roller compaction process development and scale up using Johanson model calibrated with instrumented roll data. *International Journal of Pharmaceutics* 436: 486–507.
137. Kuriyan, K. and Reklaitis, G.V. (2013). Creating shared resources for pharmaceutical technology education – simulation tools. *Education for Chemical Engineers* 8: e124–e131.
138. Morrison, R.D., Shi, F., and Whyte, R. (2007). Modeling of incremental rock breakage by impact – for use in DEM models. *Minerals Engineering* 20: 303–309.
139. Kochhar, S.K., Rubinstein, M.H., and Barnes, D. (1995). The effects of slugging and recompression on pharmaceutical excipients. *International Journal of Pharmaceutics* 115: 35–43.
140. Adolfsson, A. and Nyström, C. (1996). Tablet strength, porosity, elasticity and solid state structure of tablets compressed at high loads. *International Journal of Pharmaceutics* 132: 95–106.
141. NIST/sematech e-Handbook of statistical methods. <http://www.itl.nist.gov/div898/handbook> (accessed 29 January 2018).

5

MODEL-BASED DEVELOPMENT OF ROLLER COMPACTION PROCESSES

GAVIN REYNOLDS

Pharmaceutical Technology & Development, AstraZeneca, Macclesfield, UK

5.1 INTRODUCTION

Granulation is a commonly used size-enlargement process across many industries. Size enlargement improves the handling and flowability of particulate materials. For multicomponent formulations, granulation can also reduce segregation propensity.

Roller compaction is a continuous dry granulation method that relies on particle bonding via a compaction mechanism rather than through liquid binders, such as in the case of wet granulation. Within the pharmaceutical industry, it is used in production of oral solid dosage products, such as granules and tablets.

Figure 5.1 shows a generalized process-flow diagram for pharmaceutical tablet manufacture using roller compaction. Formulation components are initially mixed to make a relatively homogeneous blend and often some lubricant is then added to reduce sticking to the metal surfaces in the roller compactor. This powder blend is then fed into the roller compactor. The powder is compacted between two counter-rotating rollers where it is compressed into a “ribbon,” “briquette,” or “flake.” The compacted material is then milled. The milling process may be part of the same roller-compactor equipment, or using ancillary equipment. The milling process will typically include a screen with apertures that control the maximum granule size. If extragranular formulation components are required, these will be mixed with the granules to produce a homogeneous blend. A final lubrication step is then used before feeding into a tablet press

which will produce tablets. The uncoated tablets may subsequently undergo film coating.

The successful development of a robust roller compaction process requires an understanding of the influence of formulation material properties on the operation of the roller compactor, the properties of the subsequent granules and, in the case of a tablet manufacturing process, the quality attributes of the tablet.

This chapter will give an overview of the roller compaction-unit operation, typical equipment designs and configurations, the key process parameters, and influence on material attributes. The chapter will aim to set these in the context of the overall tablet manufacturing process. Traditional pharmaceutical development has typically treated the various unit operations as a “black box.” Systematic design of experiments has been used to vary process parameters (“factors”) and to measure the output properties (“responses”). Polynomial regression can then be used to understand the relationships, for example, between roller pressure and granule size. Challenges arise in transferring this knowledge when moving between equipment or scales where the equipment design and process parameters inevitably change. The reader will be introduced to modeling techniques and how these can be used to look inside the “black box” and form an integrated part of the design, development, and scale-up of a roller compaction process. In particular, examples of process and property functions will be presented so that models of the roller compaction process can be connected with downstream product attributes. The aim of the chapter is to present the models

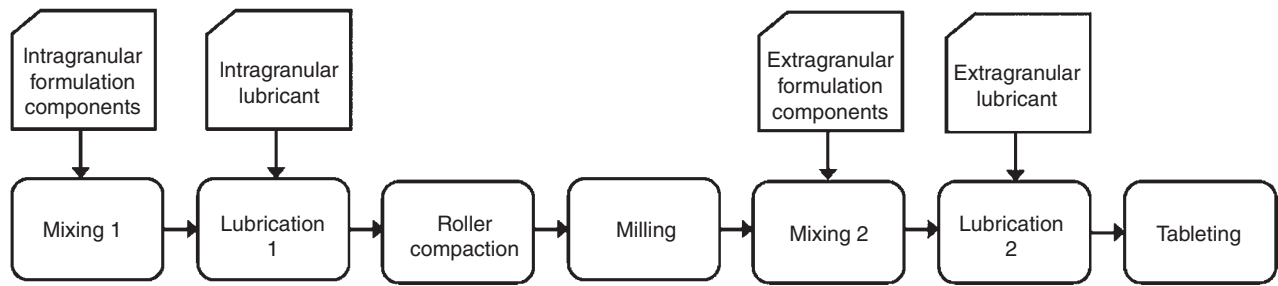


FIGURE 5.1 Process flow diagram of a pharmaceutical tablet manufacturing process with roller compaction.

in such a way that a student can follow and implement some of the models to aid understanding.

5.2 ROLL COMPACTION

5.2.1 Equipment Overview

The key features of a roller compactor are illustrated in Figure 5.2. Blended formulation components are fed into the equipment. Typically, single, twin screws, or a combination of both are used to convey the powder toward the rollers. Ensuring consistent flow of the powder is important for stable operation and to ensure a coherent compacted ribbon. In some roller compactor designs, there may be a deaeration zone prior to the rollers to aid escape of air between the particles as the powder begins to be packed and compressed. The main features of a roller compactor are the two counter-rotating rollers. Depending on the equipment manufacturer, these may be mounted with a vertical, horizontal, or inclined gap. As the rolls rotate, the powder is drawn into the gap and compressed. The rolls may be smooth or patterned to aid in gripping the powder. To avoid material leaking out the sides of the rolls, a cheek plate system may be used either side of the rollers. Alternatively, rim rolls may be used. Here, one roll has a rim that the other can fit into ensuring the sides are contained. In operation, the feed screw speed into the rolls can be controlled directly or the equipment operated in a “gap control” configuration. In the former configuration, the roll gap will vary to accommodate the quantity of powder between the rolls at the desired roll force and roll speed. In the “gap control” mode, a desired gap will be specified and the feed screw speed will vary in response. The powder leaves the rollers as a compressed “ribbon” or “flake.” Typically, there are “scrapers” positioned near the exit of the rollers to ensure the material does not stick to the roll surface. Downstream from the rollers there may be a “ribbon breaker” that rotates and breaks the ribbon into sections. In many roller compactors, there may be an integrated mill. A variety of designs are employed, but the key features are often an impeller and a screen for separation of the granules from the ribbon. In some equipment, multiple milling stages are used, with

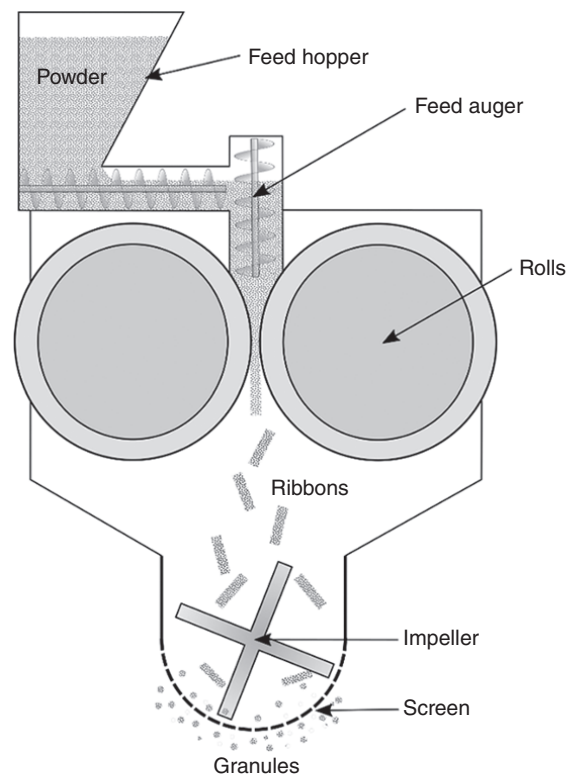


FIGURE 5.2 Diagram illustrating the key features of a roller compactor.

different screen sizes. Although the specific configuration will vary, the primary aim of this part of the process is to generate a free-flowing granule from the roller-compacted ribbon.

5.2.2 Developing a Process Model of a Roller Compactor

Quantitative modeling of roller compaction has been challenging to develop because of the complex behavior of particulate materials. Early work by Johanson [1] using powder mechanics to describe the densification of powders between the rollers has led to a group of models that provide a one-dimensional

description of the compaction region. More sophisticated modeling approaches, such as finite element modeling (FEM), allow two- and three-dimensional analysis of the compaction region and the ability to include more detailed constitutive powder properties and frictional boundary conditions, at increased computational expense. For the purpose of developing a process model that can be used for process design, scale-up, and system modeling, we will focus our attention on the one-dimensional approach, although the deficiencies will be discussed.

An early one-dimensional model of the compaction region between the rolls was developed by Johanson. A schematic diagram of the system in consideration is shown in Figure 5.3. Johanson divided the flow of powder between the rollers into two regions. Initially, the powder occupies a “slip” region, where the powder slips against the roll surfaces. As the gap narrows, the flow changes to the “nip” or “no-slip” region, where the powder is assumed to move at the same speed as the roll periphery. The angular position of the transition between these two regions is termed the nip angle. Using the powder mechanics theory developed by Jenike and Shield [3], Johanson developed an expression for the pressure distribution between the rolls above the nip angle. For the pressure distribution within the nip angle, the powder velocity was assumed to be one-dimensional. The mass of material within a small volume at the nip angle must be compressed to a smaller volume at any other location in the no-slip region. Using the observation that the log density of a powder is a linear function of log pressure, Johanson derived the pressure distribution within the nip angle:

$$P_{\theta} = P_{\alpha} \left[\frac{(1 + S/D - \cos \alpha) \cos \alpha}{(1 + S/D - \cos \theta) \cos \theta} \right]^K \quad (5.1)$$

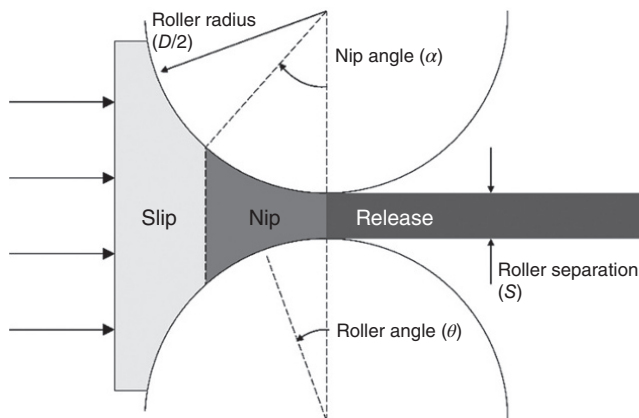


FIGURE 5.3 A schematic diagram of the compaction region within a roller compactor. *Source:* Reproduced from Reynolds et al. [2] with permission from Elsevier.

where

P_{θ} is the pressure at angle θ ,
 α is the nip angle,
 S is the roll gap,
 D is the roll diameter.

K is the power-law exponent for the powder pressure and relative density (γ) relationship as given in the following equation:

$$\frac{P_{\theta}}{P_{\alpha}} = \left(\frac{\gamma_{\theta}}{\gamma_{\alpha}} \right)^K \quad (5.2)$$

Johanson equated the pressure gradient in the slip-region with the pressure gradient in the no-slip region to find the nip angle, by solving the following equation:

$$\frac{4((\pi/2) - \alpha - \nu) \tan \delta_E}{(\cot(A - \mu) - \cot(A + \mu))} = \frac{K(2\cos \alpha - 1 - (S/D)) \tan \alpha}{\cos \alpha} \quad (5.3)$$

where

$$A = \frac{\alpha + \nu + (\pi/2)}{2} \quad (5.4)$$

$$\mu = \frac{\pi}{4} - \frac{\delta_E}{2} \quad (5.5)$$

$$2\nu = \pi - \arcsin\left(\frac{\sin \phi_W}{\sin \delta_E}\right) - \phi_W \quad (5.6)$$

The bulk frictional properties of the powder can be seen in Eqs. (5.3), (5.5), and (5.6), where δ_E is the effective angle of internal friction and ϕ_W is the wall friction angle.

The relationship with the key process parameter of roll force is determined by integrating the pressure distribution from the nip angle to the minimum gap, essentially assuming that the pressure contribution in the slip ($\theta > \alpha$) and release ($\theta < 0$) regions is negligible, resulting in:

$$R_f = P_{\max} WDF/2 \quad (5.7)$$

where P_{\max} is the maximum pressure exerted by the rolls on the powder and F is given by:

$$F = \int_{\theta=0}^{\theta=\alpha} \left[\frac{S/D}{(1 + (S/D) - \cos \theta) \cos \theta} \right]^K \cos \theta d\theta \quad (5.8)$$

Several studies have been made comparing the Johanson approach with experimental measurements. Using an instrumented roll fitted with a miniature piezo-electric pressure transducer, Bindhumadhavan et al. [4] found that the Johanson model generally gave a good description of the pressure

profile, although they acknowledged that the nip pressure, P_α , was essentially used as a fitting parameter. They also found that the measured nip angles were within approximately 15% of that calculated using Johanson's method. Yu et al. [5] compacted microcrystalline cellulose (MCC) and dibasic calcium phosphate dihydrate using the same instrumented roller compactor and demonstrated that the maximum pressures were not obtained at the minimum roll gap, but at a different angular position termed the neutral angle. This observation challenges one of the key assumptions of the Johanson model and is further supported by other experimental studies [6] and numerical FEM simulations [7].

The slab method was originally developed by von Karman [8] for sheet metal forming, but extended for roll compaction of metal powders by Katashinskii [9] by employing the Drucker–Prager rather than the von Mises yield criterion in order to account for compressibility. Dec et al. [10] also developed a slab model and presented a comparison with the Johanson model and FEM. The predictive capability of the slab method is limited due to the requirement of experimentally determined nip and neutral angles as inputs.

Further developments of the Johanson approach have been made to improve its applicability and usability. Reynolds et al. [2] developed an approach based on the Johanson model that utilizes a power-law description of ribbon relative density (γ_R) as a function of the maximum pressure exerted by the rolls, P_{\max} :

$$\gamma_R = \gamma_0 P_{\max}^{1/K} \quad (5.9)$$

where

γ_0 is the powder preconsolidation relative density
 K is the powder compressibility.

An iterative procedure is presented where the model can be calibrated with these material parameters using measured ribbon density from small-scale experiments. The parameters can then be used in the model to predict ribbon densities on different scales of roller compactor. This approach has been demonstrated to provide a good level of prediction for scale-up on a range of different roller compaction equipment [11, 12].

Liu and Wassgren [13] revisit the assumption of one-dimensional flow in the no-slip region. They identify that FEM simulations of powder compaction between the rolls show that the powder speed is faster at the roll surface than it is at the centerline. Neglecting the second dimension explains why the Johanson model will over-predict the powder relative density. Liu and Wassgren include a mass correction factor, f_θ , in the derivation, resulting in the following pressure distribution in the no-slip region:

$$P_\theta = P_\alpha \left[\frac{f_\theta (1 + S/D - \cos \alpha) \cos \alpha}{(1 + S/D - \cos \theta) \cos \theta} \right]^K \quad (5.10)$$

To determine the angular dependency of the mass correction factor, they ran FEM simulations over a range of different boundary conditions and fitted the following functional form:

$$\frac{f_\theta}{f_0} = 1 + \frac{1-f_0}{f_0} \left(\frac{\theta}{\alpha} \right)^n \quad (5.11)$$

where

f_0 is the mass correction factor at the minimum gap
 n is a fitted exponent.

They found that n was independent of the boundary conditions used in the simulations, but dependent on material properties. For practical use of the model for commercial roll compactors, f_0 needs to be fitted by using Eq. (5.10) in the Johanson model. They also present an approach whereby the compressibility constant, K , is derived directly from cap-hardening parameter Drucker–Prager Cap (DPC) model.

Models based on the Johanson approach can form a useful basis for a process model of roller compaction. They can model the nonlinear interaction between roll force and gap on the pressure applied to the material and therefore the extent of densification. This interaction, where ribbon solid fraction increases with increasing roll force, but, to a lesser extent, decreases with increasing roll gap, has been documented experimentally [11]. By capturing this interaction, the model can be useful for process design and optimization. In addition, the models can take into account changes in equipment geometry, specifically roll diameter and width, and are therefore useful when moving processes to different equipment. In this instance, the models become useful for scale-up and process transfer. Although these models can be practically useful, it must be recognized that they are still an abstraction of the real process. Firstly, the model assumes the peak pressure applied to the material is coincident with the minimum separation of the rollers. Measurement of the pressure distribution between the rollers using pressure transducers has demonstrated that this maximum pressure typically occurs in advance of the minimum separation point (for example, Ref. [5]). Souihi et al. [12] demonstrated that although there were differences in the magnitude of the maximum normal stress measured using roller-pressure transducers and P_{\max} calculated using the approach of Reynolds et al. [2], they were strongly correlated. This correlation ensures that once calibrated, the model is able to adequately capture the relative changes in process parameters on the peak pressure between the rollers and therefore the changes in ribbon solid fraction. Secondly, there is a well-documented variation

in pressure across the width of the rolls. There is evidence that this can also depend on the sealing system used. Mazar et al. [14] show that in the case of side-seal systems, there will be a maximum ribbon solid fraction in the center of the width of the ribbon with lower density on the edges. In the case of rim roller systems there is a more uniform distribution of density, with some higher density exhibited on the edges of the ribbon. Again, one-dimensional models do not capture this effect and therefore changes in sealing systems cannot be accounted for directly.

5.2.3 Using a Process Model of a Roller Compactor

The model of Reynolds et al. [2] will be used to demonstrate how a process model for roller compaction can be constructed and then used. In principle, the model can be used in two ways:

1. Before roller compacting a new formulation
2. Moving between roller compaction process settings or equipment

In both approaches, bulk frictional properties of the roller-compactor input powder blend will need to be determined in order to calculate the nip angle using Eq. (5.3). Both the effective angle of internal friction (δ_E) and the wall friction angle (ϕ_w) can be measured using a shear cell. Reynolds et al. [2] illustrate that the calculated ribbon density varies slowly as a function of these properties, except for very low frictional values, suggesting that if measurements of these properties are not available, an estimate can be used when calibrating the other material properties.

The first approach may be useful early in formulation development when no process data are available and some suitable starting process settings on a roller compactor are required. In this case, compressibility data generated using a small quantity of data in a uniaxial test could be used [15]. Due to many of the limitations of the model already discussed, using material data from this test will lead to an overprediction of the ribbon density. Based on the author's experience, the following procedure can still be useful as an initial estimate of process settings to achieve a desired ribbon density. Firstly, the out-of-die compressibility is measured by generating compacts at different compaction pressures and measuring the resultant solid fraction following ejection. These data can be fitted to Eq. (5.9) to determine the compressibility factor (K) and the preconsolidation relative density (γ_0). These parameters can then be used directly in the model to identify the process parameters required to achieve the target ribbon density. Again, based on the author's experience, the measured ribbon porosity resulting from these settings will typically be about 10% higher than the target ribbon porosity (i.e. if aiming for a ribbon porosity

of 30%, the resultant ribbon porosity at those settings will be about 33%). Although this is a relatively empirical way of using the model, it can provide some useful starting parameters for initial roller compaction process studies where ribbons of varying density are required to start studying the effect of the roller compaction process on the formulation performance.

The second approach will provide much more reliable modeling performance and can be used once some roller compaction studies have been performed and the ribbon solid fraction determined. In this case, roller compaction process settings should have been varied in order to achieve a number of different ribbon solid fractions, ideally 3 or more over the range of ribbon solid fractions that represent a reasonable operating range for roller compaction, such as 0.6–0.8 [15]. These data can be fitted to Eq. (5.9), where P_{\max} is calculated from Eq. (5.7) in order to determine the compressibility factor (K) and the preconsolidation relative density (γ_0). The fitting needs to be performed in an iterative manner as P_{\max} depends on the nip angle (α), obtained by solving Eq. (5.3), which also includes the fitting parameters K and γ_0 . Once calibrated in this way, the model can be used to simulate the effect of changes in process parameters (such as roller force and roller separation) [2, 11] and movement between roller compaction equipment [11, 12].

5.2.4 Worked Example: Calibrating and Using the Roller Compactor Process Model

This section will provide an example of calibrating a roller compactor process model using data from trial runs on a lab-scale roller compactor. The aim is to enable the reader to implement the models using simple spreadsheet software, such as Microsoft Excel.

5.2.4.1 Dataset We will use a dataset generated from roller compaction of a pharmaceutical formulation, containing 17.5% (w/w) of proprietary API. The formulation was roller compacted on an Alexanderwerk WP 120 roller compactor (Alexanderwerk, Germany). The frictional properties of the powder blend were measured using a ring shear tester (RST-XS, Dietmar Schultze, Germany), with values of 49.1° and 16.6° for the effective angle of internal friction and wall friction, respectively. Ribbon solid fraction was calculated from measurement of the ribbon envelope density using a Geopyc 1360 (Micromeritics, USA) and measurement of the powder blend true density using a helium pycnometer (Accupyc 1330, Mircomeritics, USA). In practice, ribbon solid fraction (or porosity) can be measured by several methods. For example, a portion of ribbon can be cut and its dimensions and weight measured to calculate the density. If these methods are unavailable, a mass balance approach can be used to provide an estimate of ribbon density.

Performing a mass balance on the minimum separation region of the rollers, yields [2]:

$$\dot{m}_R = \pi \rho_{\text{true}} \gamma_R D W S N_R \quad (5.12)$$

where

- \dot{m}_R is the measured mass flow rate of collected ribbons,
- γ_R is the ribbon solid fraction,
- ρ_{true} is the formulation true density,
- N_R is the rotation rate of the rollers,
- S is the minimum separation,
- W is the roll width
- D the roll diameter.

Rearranging this equation allows the ribbon solid fraction to be calculated from a measure of the mass flow rate of ribbons where all the other variables are known. This approach can provide a useful estimate, but care needs to be taken to ensure the material that does not pass between the rollers (i.e. side-seal leakage powder).

Table 5.1 provides a summary of the measured process settings and corresponding ribbon solid fraction that will be used for model calibration.

5.2.4.2 Setting Up the Model To calibrate the model we will use Microsoft Excel's Solver add-in to find values of K and γ_0 that minimize the sum of square differences between the measured ribbon porosities and the model calculations. In order to do this, we need to construct a spreadsheet that contains our known material and equipment parameters and also our measured experimental data. An example spreadsheet is shown in Figure 5.4. In this spreadsheet, the known values for the material properties and equipment parameters are in rows 1–8. The material property group includes the estimated values for K and γ_0 . There is only one set of these values for all calculations. The known experimental data are shown

TABLE 5.1 Summary of Roller Compaction Data for Model Calibration

Run	Roll Force (kN)	Roll Gap (mm)	Ribbon Solid Fraction (-)
01	10.7	2.0	0.703
02	11.4	3.0	0.689
03	25.8	2.0	0.804
04	11.1	3.0	0.704
05	18.4	2.5	0.783
06	25.4	3.1	0.778
07	25.6	2.0	0.828
08	25.4	3.1	0.796
09	18.4	2.6	0.763
10	11.1	2.0	0.727

in rows 10–12, which contain separate columns for each experimental run. The most challenging part of the calculation procedure is to calculate the force factor (Eq. 5.8), which requires the nip angle to be calculated (Eq. 5.3).

To find the nip angle, α , Eq. (5.3) needs to be solved numerically. This can be achieved iteratively using the Newton–Raphson Method [16] with the following formula:

$$\alpha_{i+1} = \alpha_i - \frac{f(\alpha_i)}{f'(\alpha_i)} \quad (5.13)$$

where:

$$f(\alpha) = \frac{4(\pi/2 - \alpha - \nu) \tan \delta_E}{\left(\cot \left(\frac{\alpha + \nu - \pi/2}{2} - \mu \right) - \cot \left(\frac{\alpha + \nu - \pi/2}{2} + \mu \right) \right)} - \frac{K \left(2 \cos \alpha - 1 - \frac{S}{D} \right) \tan \alpha}{\cos \alpha}$$

$$f'(\alpha) = \frac{1}{D} [K(D+S) \sec^3(\alpha) - 2DK \sec^2(\alpha) + K(D+S) \sec(\alpha) \tan^2(\alpha) - D \operatorname{cosec}(2\mu)(2 \cos(2\mu) + (2\alpha + 2\nu - \pi) \cos(\alpha + \nu) + 2 \sin(\alpha + \nu)) \tan(\delta_E)]$$

When using a reasonable starting value for α , such as 0.1 rad, Eq. (5.13) converges rapidly after only two or three steps. In Figure 5.4 the calculations are shown in rows 18–28. In this case, three iteration steps have been used. Once the nip angle is calculated, the integral to calculate the force factor can be performed. The integration is performed by using 20 steps. Row 31 shows the calculated step size, which is calculated by dividing the nip angle (row 27) by 20. The end points of each step are shown in rows 33–53 and columns A, C, E, etc. The function inside the integral of Eq. (5.8) is then calculated using the midpoint of each step (rows 33–52 and columns B, D, F, ...). The force factor, F , is then calculated by summing the calculated integral function at each step and multiplying by the step size. Finally, P_{max} (row 14) can be calculated from Eq. (5.7) and the ribbon solid fraction, γ_R (row 15), from Eq. (5.9).

In order to calibrate the model by estimating K and γ_0 , the sum of square errors between the measured and calculated ribbon porosities is performed (cell B16). Microsoft Excel's Solver add-in is used to minimize this cell by changing the values of cells B4 (K) and B5 (γ_0), using default method and options. This results in the estimated values of 6.48 and 0.374, respectively, as shown in Figure 5.4.

Calculations can now be performed using this calibrated model for different roller compaction equipment (e.g. different roll diameters and widths) and different roller compaction set-points (e.g. roll force and roll gap). For example, the

	A	B	C	D	E	F	G	H	I	J
1	Material properties									
2	δ_E (°)	49.1								
3	ϕ_W (°)	16.6								
4	K	6.48								
5	γ_0	0.374								
6	Equipment parameters									
7	D (m)	0.12								
8	W (m)	0.04								
9	Experimental data									
10	R_f (kN)	10.7	11.4	25.8	11.1	18.4				
11	S (m)	0.002	0.003	0.002	0.003	0.0025				
12	γ_{expt}	0.703	0.689	0.804	0.704	0.783				
13	Model calculations									
14	P_{max} (MPa)	65.9	58.0	159.0	56.1	101.9				
15	γ_{calc}	0.714	0.701	0.819	0.697	0.764				
16	γ_{SSE}	0.0013								
17	Nip angle calculation									
18	α_1 (rad)	0.100	0.100	0.100	0.100	0.100				
19	$f(\alpha_1)$	0.818	0.823	0.818	0.823	0.821				
20	$df(\alpha_1)$	-12.229	-12.173	-12.229	-12.173	-12.201				
21	α_2 (rad)	0.1669	0.1676	0.1669	0.1676	0.1673				
22	$f(\alpha_2)$	-0.003	-0.003	-0.003	-0.003	-0.003				
23	$df(\alpha_2)$	-12.304	-12.246	-12.304	-12.246	-12.275				
24	α_3 (rad)	0.1666	0.1674	0.1666	0.1674	0.1670				
25	$f(\alpha_3)$	0.000	0.000	0.000	0.000	0.000				
26	$df(\alpha_3)$	-12.304	-12.246	-12.304	-12.246	-12.275				
27	α_{final} (rad)	0.1666	0.1674	0.1666	0.1674	0.1670				
28	α_{final} (°)	9.5	9.6	9.5	9.6	9.6				
29	Force factor calculations									
30	F	0.0676	0.0821	0.0676	0.0821	0.0753				
31	$d\theta$ (rad)	0.008	0.008	0.008	0.008	0.008				
32	θ (rad)	FF int	θ (rad)	FF int	θ (rad)	FF int	θ (rad)	FF int	θ (rad)	FF int
33	0.000	0.997	0.000	0.998	0.000	0.997	0.000	0.998	0.000	0.997
34	0.008	0.971	0.008	0.980	0.008	0.971	0.008	0.980	0.008	0.976
35	0.017	0.921	0.017	0.947	0.017	0.921	0.017	0.947	0.017	0.936
36	0.025	0.851	0.025	0.899	0.025	0.851	0.025	0.899	0.025	0.879
37	0.033	0.768	0.033	0.839	0.033	0.768	0.033	0.839	0.033	0.810
38	0.042	0.677	0.042	0.770	0.042	0.677	0.042	0.770	0.042	0.731
39	0.050	0.584	0.050	0.697	0.050	0.584	0.050	0.697	0.050	0.649
40	0.058	0.493	0.059	0.621	0.058	0.493	0.059	0.621	0.058	0.566
41	0.067	0.409	0.067	0.546	0.067	0.409	0.067	0.546	0.067	0.486
42	0.075	0.333	0.075	0.474	0.075	0.333	0.075	0.474	0.075	0.411
43	0.083	0.268	0.084	0.406	0.083	0.268	0.084	0.406	0.084	0.343
44	0.092	0.212	0.092	0.344	0.092	0.212	0.092	0.344	0.092	0.282
45	0.100	0.166	0.100	0.289	0.100	0.166	0.100	0.289	0.100	0.230
46	0.108	0.129	0.109	0.240	0.108	0.129	0.109	0.240	0.109	0.186
47	0.117	0.099	0.117	0.198	0.117	0.099	0.117	0.198	0.117	0.149
48	0.125	0.076	0.126	0.162	0.125	0.076	0.126	0.162	0.125	0.118
49	0.133	0.058	0.134	0.132	0.133	0.058	0.134	0.132	0.134	0.094
50	0.142	0.044	0.142	0.107	0.142	0.044	0.142	0.107	0.142	0.074
51	0.150	0.033	0.151	0.086	0.150	0.033	0.151	0.086	0.150	0.058
52	0.158	0.025	0.159	0.069	0.158	0.025	0.159	0.069	0.159	0.045
53	0.167		0.167		0.167		0.167		0.167	

FIGURE 5.4 Example Microsoft Excel spreadsheet for roller compaction model calibration.

ribbon solid fraction can be calculated for a roller compactor with roller diameter 25 cm, roller width 5 cm operated at a roll force of 40 kN, and a roll gap of 2 mm by typing these into the relevant cells (B7, B8, B10, and B11, respectively). The solid fraction will then be calculated in cell B15 (and should be found to be 0.799).

5.2.4.3 Scaling Factors for Roll Force We may well ask whether there is a simple scaling factor that can be used to relate the roll force between roller compactors of different dimensions? Typically, when scaling up a roller compaction process, we seek to keep the ribbon solid fraction constant to ensure similar downstream material properties. We can write this as follows for ribbons from roller compactors A and B:

$$\gamma_{R,A} = \gamma_{R,B} \quad (5.14)$$

Using Eq. (5.9), this becomes:

$$\gamma_0 P_{\max,A}^{1/K} = \gamma_0 P_{\max,B}^{1/K} \quad (5.15)$$

When scaling from the same formulation, γ_0 and K are constants and this becomes:

$$P_{\max,A} = P_{\max,B} \quad (5.16)$$

and using Eq. (5.7):

$$\frac{2R_{f,A}}{D_A F_A W_A} = \frac{2R_{f,B}}{D_B F_B W_B} \quad (5.17)$$

Rearranging this equation to solve for the equivalent roll force (per unit width) on roller compactor B, gives:

$$\frac{R_{f,B}}{W_B} = \frac{R_{f,A} D_B F_B}{W_A D_A F_A} \quad (5.18)$$

So, we can see that the new roll force per unit width is related to the current roll force per unit width multiplied by the ratio of roller diameters and a “scaling factor,” represented by the ratio of force factors. Using the approach already shown in Section 5.2.4.2, the force factors can be calculated. For example, using the calibrated model from the previous section and assuming roller compactor A has a diameter of 12 cm and roller compactor B has a diameter of 25 cm, the ratio of force factors as a function of gap size is shown in Figure 5.5. This shows that the ratio is a weak function of gap size and therefore a representative force factor ratio could be selected for use in the scaling relationship. However, the force factor ratio itself is a function of the diameter ratio, as shown in Figure 5.6. Therefore, to use this scaling relationship, a suitable scaling factor would need to be read from Figure 5.6 depending on the diameter ratio for scale-up. Although less accurate than the full model, this

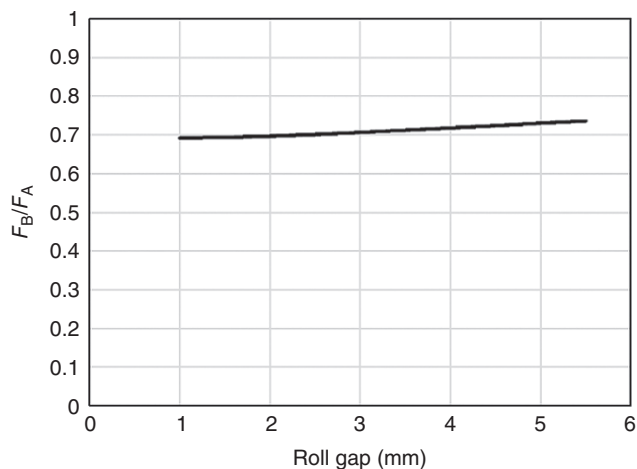


FIGURE 5.5 Ratio of force factors for 25 cm roll diameter and 12 cm roll diameter compactors as a function of gap size.

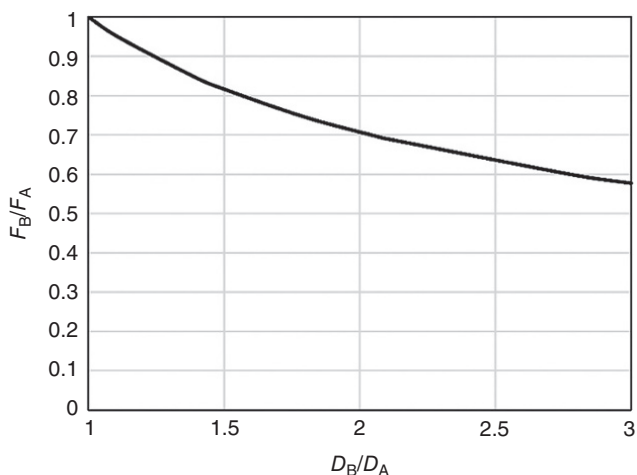


FIGURE 5.6 Ratio of force factors for roller compactors with different diameter ratios.

approach can provide a good estimate of the roll force required to achieve a similar ribbon porosity on scale-up.

A parametric control strategy may include specifying operating ranges for roll force, as this will ensure ribbon porosities will be within a certain range and therefore downstream product quality will be maintained. This can be normalized for different roller compactors by expressing in terms of roll force divided by roller width. Table 5.2 shows a roll force-control range for roller compactor A. If moving to roller compactor B, which has a diameter of 20 cm, the process could still be operated at a set point within the original control range, such as 6 kN/cm. However, if moving to a very different scale roller compactor such as C, with a diameter of 12 cm, the control ranges for the process would have to be changed to ensure similar ribbon density and therefore downstream product quality as there is no overlap with the original operating ranges.

TABLE 5.2 Comparison Between Roller Force Control Ranges for Two Different Scales of Roller Compactor

Roller Compactor	Roll		
	Diameter (cm)	Minimum Roller Force (kN/cm)	Maximum Roller Force (kN/cm)
A	25	5.0	7.0
B	20	4.5	6.3
C	12	3.5	4.9

5.3 MILLING

5.3.1 Equipment Overview

Roller-compacted ribbon size reduction is a key operation to produce a granular product suitable for downstream processing. Coupled with the roller-compaction process settings, this milling operation will result in a granule size distribution that will influence granule performance such as bulk density and flowability. Invariably, size reduction is achieved using a screening mill. Depending on the manufacturer, the mill may be integrated with the roller compactor or it may need to be a separate unit. For integrated designs, there can sometimes be intermediate milling stages, such as an impeller to break ribbons into pieces as they exit the rollers, or multiple cascading screening mills with different screen aperture sizes. The key features of a screening mill are an impeller and a screen. The impeller may operate at relatively low rotational rates and include an oscillating motion, as in the case of an oscillating mill. The design of the impeller will also vary between manufacturers. Potentially a large range of process settings can be manipulated during the milling stage due to the different configuration options available, in addition to settings such as impeller speed and screen size. Furthermore, fines can be incorporated into the final granule size distribution from leakage in the sealing mechanism of the rollers. In some equipment designs, this particular source of fines can be separated and potentially recycled. Perez-Gandarillas et al. [17] conducted an extensive study on the interaction of roller compaction and milling process parameters on the granule properties. Using a Gerteis Mini-Pactor 250/25 they were able to systematically vary roll force, milling speed, and mill oscillating angle. They also investigated the effect of roll sealing system (cheek plates or rimmed-roll) and mill impeller type (star granulator or pocket mold-grooved). In all cases they found a bimodal granule size distribution, typical of the roller compaction process. They found that, of the process configurations, the roller force and sealing system had the most significant effect on the level of fines (characterized using the $D[v,0.1]$), where a higher roller force and a rimmed-roll sealing system resulted in increased $D[v,0.1]$. They found the type of impeller and mill speed were

nonsignificant factors, although the clockwise-oscillating mode resulted in slightly less fines. For screening mills, the choice of screen size will have a significant effect on the granule size distribution. am Ende et al. [18] studied the effect of roll force, roll gap, screen size, and milling speed and found that roll force and screen size had the most significant effect on granule size. Mirtič and Reynolds [19] found that the $D[v,0.9]$ increased proportionally to screen size. They also found that reduced screen sizes increased the mill residence time and resulted in granules with a higher sphericity, which was attributed to increased attrition of the granule edges.

5.3.2 Developing a Process Model for a Mill

Population balance equations (PBE) have been used for many years to describe the dynamic change in particle size distributions as a result of underlying rate processes. During milling, particles will experience breakage due to collision with impellers, screen, mill wall, and abrasion between particles. For a batch system, the PBE for breakage can be written:

$$\frac{\partial n(v,t)}{\partial t} = \int_v^{\infty} S(\varepsilon,t)b(\varepsilon,v)n(\varepsilon,t)d\varepsilon - S(v,t)n(v,t) \quad (5.19)$$

where

n is the number density of particles of size v

S is the selection or breakage rate

b is the breakage function which describes the distribution of the fragments from a breaking particle [20].

This equation generally needs to be solved using numerical methods, such as matrix methods [21], moment methods [22], and sectional methods [23, 24]. The PBE as written assumes a well-mixed system where all particles undergo a consistent set of rate processes. One of the main challenges in utilizing the breakage PBE practically is to identify suitable breakage rate (S) and fragment distribution (b) models that are physically relevant and descriptive of the system. Diemer and Olson [25] provide a useful summary of the range of breakage rate and fragment size distribution models available in the literature. Rather than giving an exhaustive account, we will focus on models that have been used for screen milling of agglomerates and ribbons. In particular, the dominant feature of bimodal granule size distributions from this process needs to be described.

Gavi and Reynolds [26] developed a system model of a tablet manufacturing process based on roller compaction. To model the milling of the ribbons they utilized the model of Vogel and Peukert [27], which was developed for impact mills. In the model, the breakage probability is given as

$$P_B = 1 - \exp[f_{\text{mat}} x k (W_{\text{m,kin}} - W_{\text{m,min}})] \quad (5.20)$$

where

x is the particle size,
 f_{mat} is a material strength parameter,
 $W_{\text{m,kin}}$ is the mass-specific impact energy,
 $W_{\text{m,min}}$ is the mass-specific threshold energy.

The fragment distribution was modeled as:

$$B = \left(\frac{x}{y}\right)^q \frac{1}{2} \left\{ 1 + \tanh\left(\frac{y-y'}{y'}\right) \right\} \quad (5.21)$$

where

y is the fragment size,
 q is a power-law exponent,
 y' is the minimum fragment size.

Gavi and Reynolds [26] included a classifier in their model where granules larger than the screen size were recirculated to the mill model. They estimated the six parameters in the model by fitting to roller-compacted granule size distributions and found that the model failed to predict the bimodal granule size distribution observed experimentally. This is perhaps not surprising, as the model of Vogel and Peukert [27] has been developed for the breakage of continuous material particles, whereas agglomerates, such as the ribbon, tend to exhibit a different mode of breakage.

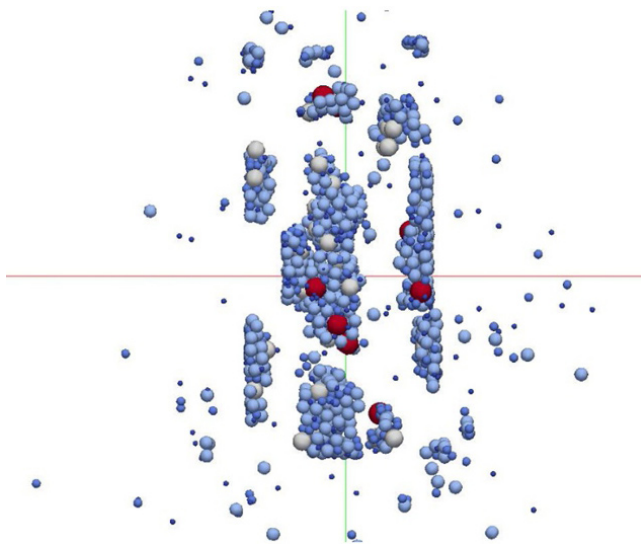


FIGURE 5.7 Pattern of ribbon breakage based on DEM simulations. *Source:* Reproduced from Loreti et al. [29] with permission from John Wiley and Sons.

Subero and Ghadiri [28] studied the breakage modes of agglomerates undergoing impact. They observed localized disintegration around the impact area whereby the mass of the agglomerate is reduced to the constituent fine particles or doublets or triplets of those particles. In addition, large cracks were observed to propagate into the body of the agglomerate, generating several larger fragments. This breakage pattern has been observed in discrete element method (DEM) simulations of ribbons and is illustrated in Figure 5.7. Reynolds [30] utilized this breakage pattern to model granule-size reduction in a conical screening mill. In this case, the breakage function was described using a bimodal distribution as follows:

$$b(v,u) = \frac{uz}{3k_v^{1/3}v^{5/3}} \times \left[\frac{2\zeta \exp\left[-\left(\ln\left(v^{1/3}/k_v^{1/3}\mu_1\right)/\sqrt{2}\ln\sigma_1\right)^2\right]}{\sqrt{2\pi}\ln\sigma_1 \left(1 + \operatorname{erf}\left[\ln\left(u^{1/3}/k_v^{1/3}\mu_1\right)/\sqrt{2}\ln\sigma_1\right]\right)} \right] + \frac{2(1-\zeta) \exp\left[-\left(\ln\left(v^{1/3}/k_v^{1/3}\mu_2\right)/\sqrt{2}\ln\sigma_2\right)^2\right]}{\sqrt{2\pi}\ln\sigma_2 \left(1 + \operatorname{erf}\left[\ln\left(u^{1/3}/k_v^{1/3}\mu_2\right)/\sqrt{2}\ln\sigma_2\right]\right)} \right] + (1-z) \frac{p(v/u)^{q-1}(1-(v/u))^{r-1}}{u B(q,r)} \quad (5.22)$$

where

u is the size of the initial agglomerate, often termed the “mother” particle,
 v is the fragment size,
 k_v is a volumetric shape factor,
 μ is a lognormal mean size,
 σ is a geometric standard deviation,
 z is the volume fraction in the fine mode,
 ζ is a parameter in a bimodal lognormal distribution,
 p is the number of fragments (in the fragmentation mode),
 q controls the fragment-size dependence, $r = q(p - 1)$,
 B is the beta function.

Although this breakage function looks complicated, it is simply the volume fraction in the fine mode multiplied by a distribution describing the size distribution of fines with the remainder of the volume described by a fragmentation model. The additional complexity in the form shown in Eq. (5.22) is that the size distribution of the fines was fitted to a bimodal lognormal distribution. The fragmentation mode

was described using a generalized form of the Hill-*Ng* daughter distribution proposed by Diemer et al. [22], with two parameters, p and q , giving the shape of the larger fragmentation mode. Although the breakage function contains a large number of parameters, the fine-mode parameters were fitted to the ungranulated powder size distribution and p and q were selected to give a multiple fragmentation pattern. This only left z to be estimated by fitting the model to granule size distribution data. In addition to the breakage function, the selection rate needs to be defined. In this case a perfect classification kernel was used, where only granules larger than a critical size were broken. In this case, the population balance model (PBM) can be solved to the limit $t \rightarrow \infty$, making the solution insensitive to the selection rate. It was found that z was a function of impeller tip speed, whereby a higher impeller tip speed resulted in a higher value of z , i.e. a greater proportion of fines.

Loreti et al. [31] performed DEM simulations to model the impact breakage of ribbons. Their simulations showed a bimodal granule size distribution, with a strong correlation between the size of the fines and the initial unagglomerated particles. They used the DEM single-breakage fragment distributions to calibrate a bimodal breakage function, similar to Eq. (5.22) by estimating p and z . Weibull and Gompertz distributions were used to describe p and z , respectively, as functions of surface energy and impact velocity. When using these distribution in a PBM they found that their model showed a strong dependency on impeller tip speed, which was not seen in ribbon milling experiments using an oscillating screening mill. To complement this work, they simulated ribbon breakage in an abrasion test following the same methodology [31]. The results showed a lesser dependency on impeller speed, more consistent with the experimental findings, suggesting that breakage in these types of mills is dominated by attrition rather than impact. This conclusion is consistent with work performed by Yu et al. [32] where they studied the motion of ribbons in an oscillating screening mill using positron emission particle tracking (PEPT) and concluded that ribbons are milled primarily by abrasion.

In summary, ribbons exhibit a bimodal granule size distribution when milled. The fine mode is consistent with the underlying primary powder size distribution and the large mode is produced via fragmentation with a clear upper size determined by the size of the mill screen aperture. A breakage function needs to be able to describe both these modes. The prevalence of oscillating screening mills for the size reduction of ribbons suggests the need for a model that can describe abrasion breakage with a sharp classification effect. For abrasion breakage, the mill impeller speed has a minor effect, but due to observations on the significance of compaction force on granule size distribution, a relationship needs to be established between the ribbon strength and the breakage function. The sharp classification effect leads to a breakage model that is primarily determined by the breakage

function rather than the selection function, simplifying the modeling effort.

5.3.3 Worked Example: Characterizing Bimodal Size Distributions and Exploring a Ribbon Milling Model

For this worked example, we will look at how to characterize granule size distributions from a roller compaction process. Developing a milling model is beyond the scope of what can be covered here as this requires numerical solution of a PBM. Nevertheless, we will explore some features of this type of model and discuss a typical workflow.

As discussed, invariably roller compaction granule size distributions are bimodal. A useful technique to understand these distributions is to parameterize them by fitting to standard distributions. The fine mode will often be qualitatively, and in some cases quantitatively, similar to the underlying ungranulated powder distribution. A lognormal distribution is often a good fit to this mode. The large mode of the distribution usually has a sharp truncation just below the screen size used. A lognormal distribution can be used to fit this, although a normal distribution is often a better choice as it reduces more quickly in the larger sizes. A truncated version of these distributions is also a suitable option, but the added complexity is unnecessary as we are usually more interested in the fine mode as part of building the breakage function. Table 5.3 contains some granule size distributions that we will use for the analysis. To characterize these distributions, we will fit a bimodal distribution consisting of a lognormal mode for the fines and a normal mode for the granules:

$$f_{\text{bimodal}} = z \cdot f_{\text{lognormal}}(\mu_1, \sigma_1) + (1 - z) \cdot f_{\text{normal}}(\mu_2, \sigma_2) \quad (5.23)$$

TABLE 5.3 Roller-Compacted Granule Size Distributions Represented as Percentage Volume Between Size Classes

Size Classes (μm)	Run A	Run B	Run C	Run D
5	0.09	0.11	0.09	0.08
9	0.44	0.51	0.40	0.37
14	1.33	1.52	1.21	1.11
21	3.84	4.34	3.52	3.22
33	6.59	7.44	6.19	5.73
50	7.91	8.91	7.48	6.83
78	6.97	7.83	6.38	5.86
120	5.10	5.52	4.46	4.41
185	4.43	4.63	4.44	4.81
286	6.02	6.19	6.27	7.46
441	12.33	11.84	12.13	14.14
682	31.02	29.21	31.05	32.46
1053	13.78	11.79	16.23	13.40
1627	0.04	0.03	0.06	0.03
2500				

	A	B	C	D	E	F	G	H	I	J	K	L
1	Mode 1 (lognormal)				Mode 2 (normal)							
2	exp[μ ₁]	74.2			μ ₂	848.4						
3	exp[σ ₁]	2.40			σ ₂	274.7						
4	z	0.388			1-z	0.612						
5	SSE	0.0024										
6					Mode 1		Mode 2		Bimodal			
7	Size class (μm)	Volume fraction	Mean size (μm)	q _{3lg, expt}	Cumulative volume	Volume fraction	Cumulative volume	Volume fraction	Volume fraction	q _{3lg, bimodal}		
8	5.0	0.00093	6.7	0.0016	0.0010	0.0070	0.0011	0.0001	0.0027	0.0047		
9	9.0	0.00438	11.2	0.0099	0.0080	0.0205	0.0011	0.0001	0.0080	0.0181		
10	14.0	0.01327	17.1	0.0327	0.0285	0.0464	0.0012	0.0001	0.0180	0.0445		
11	21.0	0.03839	26.3	0.0851	0.0749	0.1026	0.0013	0.0002	0.0399	0.0883		
12	33.0	0.06589	40.6	0.1587	0.1775	0.1485	0.0015	0.0003	0.0578	0.1390		
13	50.0	0.07911	62.4	0.1781	0.3260	0.1964	0.0018	0.0007	0.0766	0.1722		
14	78.0	0.06971	96.7	0.1620	0.5225	0.1856	0.0025	0.0015	0.0728	0.1691		
15	120.0	0.05103	149.0	0.1180	0.7080	0.1431	0.0040	0.0039	0.0578	0.1336		
16	185.0	0.04430	230.0	0.1018	0.8512	0.0869	0.0079	0.0124	0.0413	0.0948		
17	286.0	0.06018	355.1	0.1391	0.9380	0.0409	0.0203	0.0487	0.0457	0.1055		
18	441.0	0.12333	548.4	0.2832	0.9789	0.0154	0.0690	0.2033	0.1305	0.2992		
19	682.0	0.31015	847.4	0.7148	0.9943	0.0045	0.2723	0.4995	0.3076	0.7082		
20	1053.0	0.13783	1308.9	0.3171	0.9988	0.0010	0.7718	0.2259	0.1387	0.3188		
21	1627.0	0.00043	2016.8	0.0010	0.9998	0.0002	0.9977	0.0023	0.0015	0.0034		
22	2500.0	0.00000			1.0000		1.0000					
23												
24												
25												

FIGURE 5.8 Microsoft Excel spreadsheet illustrating a bimodal distribution fitting procedure.

Figure 5.8 shows an example Microsoft Excel spreadsheet that can be used for the fitting procedure. Initial guessed values for the fitting parameters are entered into cells B2, B3, B4, E2, and E3. Experimental data are included in cells A8–B22. For plotting purposes, the geometric mean size (cells C8–C21) and the logarithmic density function (cells D8–D21) are included. The geometric mean size is calculated from:

$$x_{g,i} = \sqrt{x_{0,i} \times x_{0,i+1}} \quad (5.24)$$

where

$x_{0,i}$ and $x_{0,i+1}$ are the lower and upper size classes, respectively.

The logarithmic density function is calculated from:

$$q_{3lg,i} = \frac{p_{3,i}}{\ln(x_{i+1}) - \ln(x_i)} \quad (5.25)$$

where

$p_{3,i}$ is the volume fraction in size class i .

For mode 1, the cumulative distribution is calculated in cells E8–E22 using the Microsoft Excel formula:

$$= \text{LOGNORM.DIST}(x_i, \mu_1, \sigma_1, \text{TRUE}) \quad (5.26)$$

The volume fraction is just the difference in the cumulative distribution between the lower and upper size classes.

In a similar manner, for mode 2, the cumulative distribution is calculated in cells G8–G22 using the Microsoft Excel formula:

$$= \text{NORM.DIST}(x_i, \mu_2, \sigma_2, \text{TRUE}) \quad (5.27)$$

Finally, the bimodal volume fraction for each size class is calculated by multiplying the volume fraction from mode 1 by z and adding to the product of the volume fraction from mode 2 with $1 - z$.

The Microsoft Excel Solver add-in can then be used to minimize the sum of square errors between $q_{3lg,expt}$ and $q_{3lg,bimodal}$ (cell B5) by adjusting the values in cells B2, B3, B4, E2, and E3. Figure 5.8 shows the example spreadsheet containing the fitted parameters to aid the reader. It may be that reasonable estimates need to be included as starting values (i.e. $\exp(\mu_1) = 100 \mu\text{m}$ and $\mu_2 = 1000 \mu\text{m}$). Figure 5.9 shows the fitted bimodal size distribution, showing that it provides a good description of the experimental granule size distribution. Fitted parameters for the other size distributions are shown in Table 5.4. The fitted parameters show a very consistent mean and standard deviation for both modes. The fitted volume fraction of the fine mode (z) shows

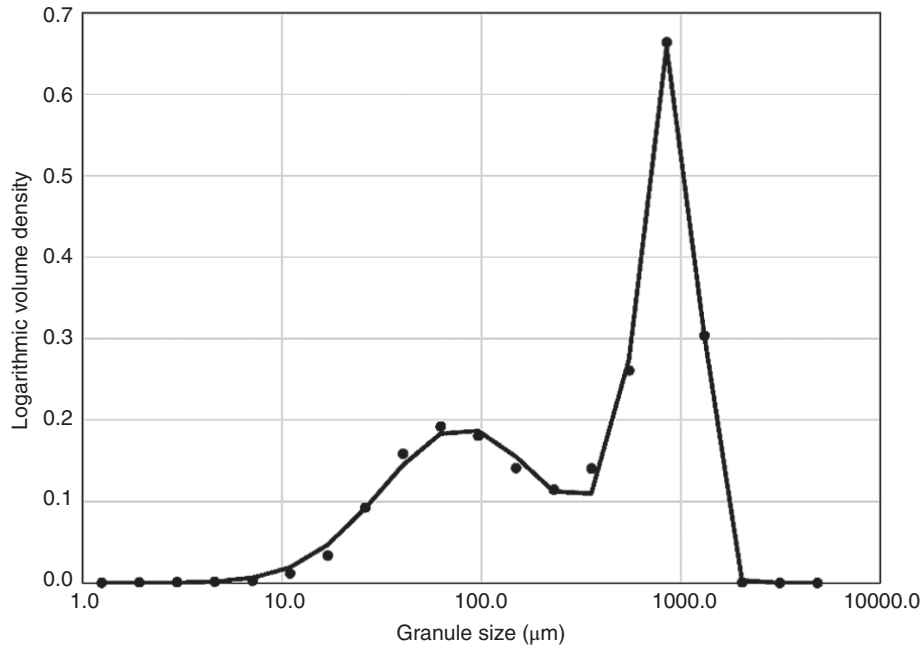


FIGURE 5.9 Experimental granule size distribution (symbols) and fitted bimodal distribution (line).

TABLE 5.4 Fitted Bimodal Distribution Parameters

	Run A	Run B	Run C	Run D
$\exp(\mu_1)$ (μm)	74.2	72.6	73.7	80.2
$\exp(\sigma_1)$	2.40	2.37	2.41	2.60
μ_2 (μm)	848.4	835.9	863.2	826.6
σ_2 (μm)	274.7	270.1	291.2	279.3
z	0.388	0.431	0.357	0.350

some differences. This parameter is plotted against measured ribbon porosity data for the associated granules in Figure 5.10. There is an approximately linear, systematic increase in z with increasing ribbon porosity. It has been found that as the ribbon porosity increases, there is a reduction in the tensile strength [19] and therefore the weaker ribbons are likely to yield a greater proportion of fines on each impact. If proceeding to calibrate a PBM milling model, the parameters for the small mode will also be an excellent starting point for the fine mode of the breakage function if using a bimodal model consisting of both disintegrating and fragmentation as already described.

To illustrate this, a PBM was used. The mean and standard deviation of the small mode was fitted from the bimodal analysis, and the remaining parameters of z (the volume fraction in the disintegration mode of the breakage function) and p and q (from the Hill-Ng fragmentation mode of the breakage function) were estimated. A perfect classification was used at a screen size of 1.4 mm (corresponding to the $D[v,0.95]$ of the granule size distribution). Typically, the measured largest

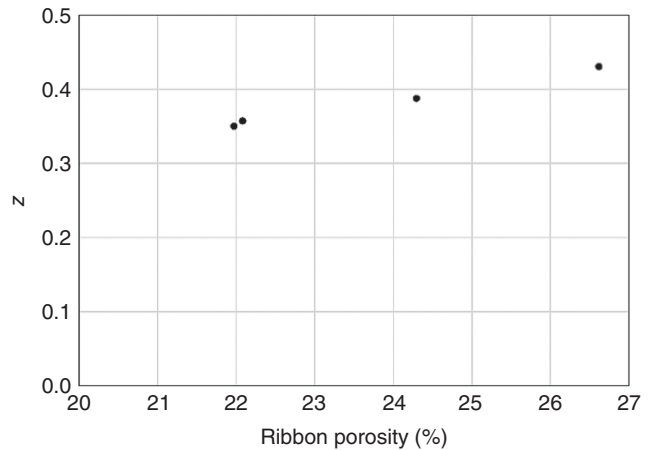


FIGURE 5.10 Fitted fine-mode volume fraction (z) compared with measured ribbon porosity.

granules (e.g. $D[v,0.90]$ or $D[v,0.95]$) will be slightly different to the actual screen size used, due to, in part, the effects of particle shape, depending on the particle size analysis technique used. It is generally better to use this maximum granule size as the screen size in the model, although for these types of screening mills, this size is very strongly correlated with the screen size [19]. The fitted parameters were 0.261, 40, and 3.62 for z , p , and q , respectively. The parameter z in this case is smaller than the volume fraction of the fine mode we fitted when performing the bimodal analysis above. This is not unexpected as the z we fit using the PBM is for the breakage function. Each time a granule breaks, this proportion will

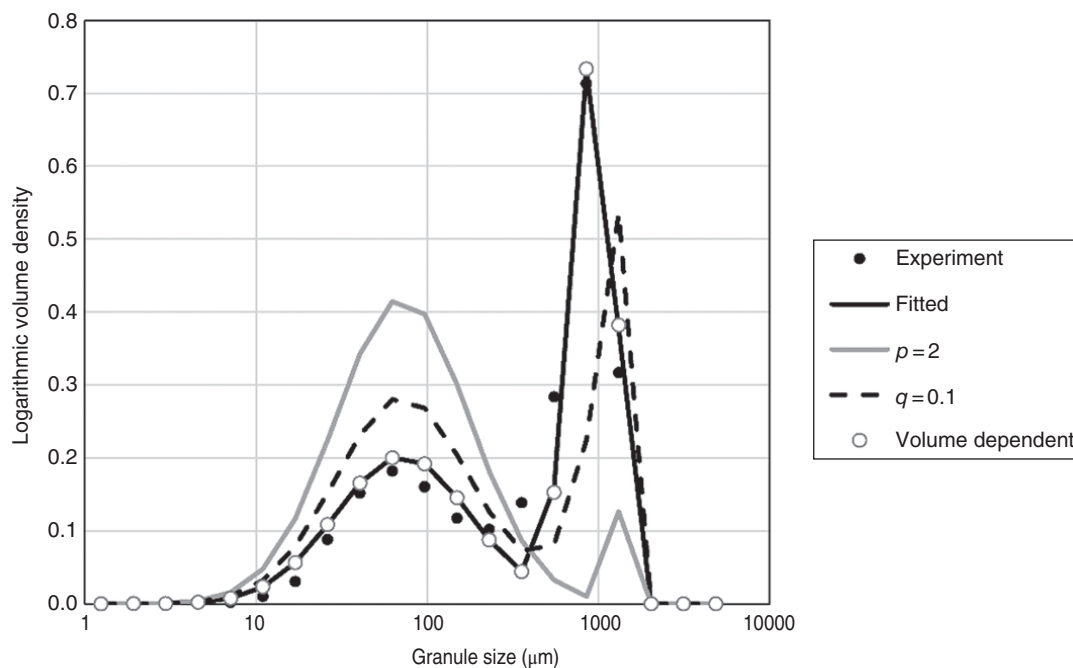


FIGURE 5.11 Measured granule size distribution (symbols) and simulated granule size distributions.

be released as fines and therefore if the ribbon has some residence time in the mill and undergoes multiple breakage events, the proportion of fines in the product will increase. For this reason, although we can use the bimodal analysis to identify information on the size of the disintegration mode, the quantitative volume fraction cannot be used. The fitted values of p and q are consistent with a multiple fragmentation model. A high value of p indicates many fragments generate from each breakage event and a large value of q indicates a fragmentation breakage mode rather than erosion or chipping. The fitted distribution is shown in Figure 5.11. To illustrate the effect of parameters p and q , additional simulations have been performed. Where not indicated in the legend, the parameters were kept the same as the fitted values. Reducing p to 2 results in a dramatic reduction in the large mode and an increase in the small mode. By reducing p , each breakage event produces only a couple of large fragments. These need to undergo many more breakage events to be reduced below the screen size, resulting in a greater proportion of fines. Reducing q results in a greater proportion of fines and less intermediate-size granules as the fragment distribution is shifted to an erosion mode. To illustrate the insensitivity of the model to the choice of breakage rate model, a volume-dependent breakage rate model was simulated. The resultant size distribution is shown in the figure to be identical to the fitted model that used a size-independent breakage rate. With the fitted breakage function parameters, the model can be used to simulate granule size from different milling screen sizes. Similar to the bimodal analysis above, some

correlations may be established between the breakage function parameters and ribbon porosity, allowing the influence of changes of ribbon mechanical strength on the granule size distribution to be taken into account.

5.4 PROPERTY FUNCTIONS FOR MODEL-BASED DEVELOPMENT OF ROLLER COMPACTION

When developing a roller compaction process, there are several product quality attributes that may influence the process design and optimization. In the case of a tablet dosage form, these quality attributes may include hardness (tensile strength), uniformity of dosage unit, and dissolution. Other quality attributes are either related (for example, disintegration is considered a precursor to dissolution) or are unlikely to be directly influenced by the roller compaction process (for example, it is uncommon for degradation products to be formed under the relatively low-energy process of roller compaction and screen milling). There is extensive literature to support the link between roller compaction and tablet tensile strength [33]. Uniformity of dosage unit can be more difficult to ascertain. The main failure modes that contribute to this quality attribute are variation in content uniformity (for example, due to poor mixing of formulation components or segregation) or weight variability during tablet manufacture (for example, due to poor flow into the die). Tablet dissolution will be influenced by a range of factors including API

physicochemical properties, API particle size, and formulation composition. The roller compaction process can have an influence on the tablet disintegration and dissolution [34], although disintegration and dissolution rates tend to be more rapid than wet-granulated formulations as binder is often not required. In this section, we will discuss property functions that can be used to model some of these quality attributes and discuss how they can be linked to process models of roller compaction.

5.4.1 Tablet Tensile Strength

The mechanical integrity of pharmaceutical tablets is often characterized by measuring the breaking force in a diametral loading test. The breaking force is commonly called hardness, but this term is misleading [35]. Tensile strength provides a more fundamental measure and normalizes for differences in tablet geometry. Typically, a tensile strength greater than 1.7 MPa is sufficient to ensure the tablet is robust enough for downstream film coating, packing, and shipping [36]. The phenomenon of “loss in compactability” or “loss in tabletability” has been well documented for powders undergoing roller compaction and milling prior to tableting. In this case compactability and tabletability refer to the relationship of tensile strength with porosity and pressure, respectively. In this case, tablets made from the roller-compacted granules exhibit lower tensile strength than tablets made by direct compression of the ungranulated powder. Sun and Kleinebudde [33] provide a summative review of this phenomenon. They conclude that contributory mechanisms can be explained by considering their influence on either bonding area or bonding strength.

Farber et al. [37] introduced a unified compaction curve (UCC) model, which allows this loss in tablet tensile strength after roller compaction to be taken into account while still maintaining usage of the same tabletability relationship for a given material. For this model, they assumed that the tensile strength of tablets was dependent on the cumulative compaction of primary particles during roller compaction and tableting, that this process was irreversible and that the ribbon milling process destroys any tensile strength generated without affecting the compaction state of the primary particles. Essentially, the tabletability follows the underlying direct compression relationship between tensile strength and pressure. Following roller compaction, the origin is moved to a point on the curve representing the pressure and tensile strength achieved following production of the roller-compacted ribbon. Application of further pressure during tableting results in tablets with only an incremental tensile strength generated beyond that of the ribbons and granules. It was found that this approach worked well for 3:1 and 1:1 MCC:lactose monohydrate mixtures but not for a 1:3 mixture or with mixtures of MCC and dicalcium phosphate. They concluded that the model was generally applicable to

systems containing significant proportions of plastically deforming material.

Gavi and Reynolds [26] represented the UCC model in terms of compactability, as described using the Ryshkewitch–Duckworth equation, resulting in the following equation for tablet tensile strength (T_T) as a function of tablet porosity (ϵ_T):

$$T_T = \bar{T}e^{-k_b\epsilon_T} - \bar{T}e^{-k_b\epsilon_R} \quad (5.28)$$

where

ϵ_R is the ribbon porosity

\bar{T} and k_b are the underlying Ryshkewitch–Duckworth material parameters representing the tensile strength at zero porosity and the bonding capacity, respectively.

It is clear from this equation that, simplistically, the tensile strength of the final tablet is equivalent to the tensile strength that would have been achieved following direct compaction of the ungranulated powder blend less the tensile strength of the ribbon, generated during roller compaction. This form of the model is more useful as a property function as it is related to the properties of the intermediate materials (ribbons/granules) rather than upstream processing conditions. Gavi and Reynolds [26] were able to successfully calibrate this model to granules generated from a formulation containing 17.5% (w/w) of a proprietary API.

In summary, the phenomenon of loss in compactability is complex with clear differences in the mechanisms exhibited by ductile and brittle materials. A framework for a model is available for use as a property function. Although this model has primarily been shown to be suitable for mixtures with significant proportion of plastic material, this approach has also been shown to be suitable for a pharmaceutical formulation. One of the golden rules of roller compaction process development identified by Sun and Kleinebudde [33] is to balance the mechanical properties of the formulation components. If this advice is followed, it is likely that the resultant formulation will contain enough ductility to be amenable to this type of loss in compactability model.

5.4.1.1 Worked Example: Calibrating a Tensile Strength Model

The following workflow can be used to calibrate a loss in compactability model. Similar to calibration of the roller compaction process model, ribbons can be manufactured at different roller forces to generate different ribbon porosities. The ribbons can then be milled into granules and compression profiles generated by creating tablets at different compaction pressures using a tablet press or compaction simulator. The dimensions of the tablets can be measured to calculate the porosity and the tensile strength calculated from diametral compression. Table 5.5 shows some example

TABLE 5.5 Compaction Profiles for Granules Manufactured from Different Ribbon Porosities

Ribbon porosity (%)	19.5		24.6		30.0	
Compaction Pressure (MPa)	Tablet Porosity (%)	Tablet Tensile Strength (MPa)	Tablet Porosity (%)	Tablet Tensile Strength (MPa)	Tablet Porosity (%)	Tablet Tensile Strength (MPa)
100	17.0	0.30	16.9	0.85	19.9	0.94
150	15.0	0.56	13.2	1.34	15.8	1.44
200	12.8	0.88	10.2	1.81	13.2	1.80
250	10.6	1.22	8.1	2.17	10.8	2.16
300	8.8	1.52	6.8	2.40	9.4	2.40

data generated from this procedure using granules from three different roller forces.

The model calibration procedure essentially involves estimating \bar{T} and k_b by fitting Eq. (5.28) to these data. One method of doing this using Microsoft Excel is to guess some starting values for \bar{T} and k_b . These values can be used in Eq. (5.28) along with the measured ribbon and tablet porosity of each tablet. The Microsoft Excel Solver add-in can then be used to minimize the sum of square errors between the calculated and measured tablet porosities by changing the values in the \bar{T} and k_b cells.

If this procedure is followed, the estimated values of \bar{T} and k_b will be found to be 6.3 MPa and 3.86, respectively, giving an R^2 of 0.78. A parity plot of the measured and calculated tablet tensile strengths is shown in Figure 5.12. Although the model is quite simplistic, it provides an adequate description of the change in tablet tensile strength as a result of changes in the roller compaction process making it useful for modeling the effect of changes in upstream process conditions (roller compaction and tableting) on tablet performance. When using the model in practice, some limitations should be considered. In particular, the Ryshkewitch–Duckworth compactability equation, on which the model is based, will only provide a good description of tensile strength as a function of porosity over typical pharmaceutical compaction pressures or porosities (for example, between 100 and 300 MPa, or above 10%, respectively). This is due to over-compression, where, at high compression force flaws are introduced due to localized, intensive shear stresses leading to premature failure of the compact during diametral compression testing resulting in a lower than expected tensile strength.

5.4.2 Granule Flowability

In Section 5.3.2 we described process models that can be used to predict the granule size following roller compaction and milling. Although this is useful, it is often the resultant bulk flow properties that are important for process design and optimization. It is well established that particle size dominates the flow behavior of powders. Leane et al. [38]

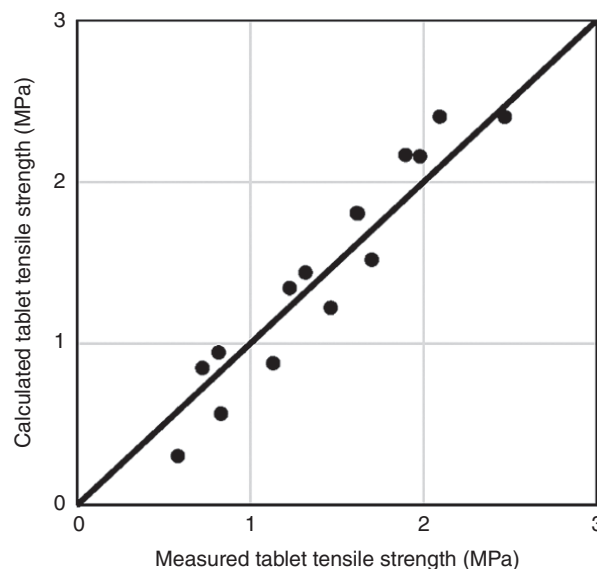


FIGURE 5.12 Parity plot showing the measured and calculated tablet tensile strengths using the calibrated loss in compactability model.

recommend a $D[v,0.5]$ of 50–500 μm for tableting, with smaller sizes prone to problems of flow, electrostatics, and adhesion. This is due to the relatively high surface area to volume ratio of small particles, which causes interparticulate attractive forces to dominate over bulk shear or gravitational forces.

A number of authors have explored multivariate modeling approaches to attempt to correlate particle properties with bulk flow behavior. Yu et al. [39] characterized the particle size and shape distributions of 23 commonly used pharmaceutical excipients and 38 binary blends. They measured the flow behavior of these samples using a shear cell and built a partial least squares (PLS) mathematical model. The PLS model was found to provide excellent predictions of the powder flow function coefficient (FFC) and also revealed that both particle size and shape play an important role in determining the powder flow behavior. Sandler and Wilson [40] also utilized a PLS modeling approach. They characterized

size and shape distributions of granular intermediates and correlated these with the bulk flow and packing behavior. They conclude that this approach should be generally applicable to the flow of granular materials in which the properties are not dominantly cohesive in nature.

In order to provide a more fundamental description of bulk-scale powder behavior, Capece et al. [41] considered particle-scale interactions by using the granular Bond number. The granular Bond number is defined as the ratio of cohesion force to particle weight due to gravity. Particles are considered cohesive with a Bond number greater than 1. They were able to use this approach to predict the flowability of binary and ternary mixtures of pharmaceutical materials. They found that this approach was most applicable to fine cohesive powders, typically including particles with diameters less than 100 μm or with Bond numbers greater than 1.

DEM is a common particle-scale modeling tool that can be used to model bulk behavior as a function of particle properties (for example, Cleary [42] and Ketterhagen et al. [43]). However, due to the significant computational burden of DEM simulations, they are not suitable for use directly as a property function as part of a system model.

5.5 CONNECTING PROCESS AND PROPERTY MODELS

Throughout this chapter we have introduced models to describe some of the key unit operations used in pharmaceutical roller compaction-based tablet manufacture. These models allow us to relate the material properties, equipment design, and operating parameters to changes in the intermediate material properties. We have also introduced some key property function models that can be used to relate our intermediate material properties to product performance. Utilizing all of these models together now allow us to take a holistic systems-based approach to the modeling of the entire process.

Although process system modeling, or flowsheet modeling, is an established approach in the processing industries for fluid-based processes, it is only recently emerging as a tool for solid-based processes. This slow uptake is in part related to the complexity of the behavior and mechanisms governing particulate and granular material in comparison to fluids, where the level of fundamental knowledge is higher. Boukouvala et al. [44] summarize a number of other reasons, including lack of unit-operation process models, complexity and distributed nature of particulate properties, and lack of standardized testing procedures and comprehensive material property databases.

Boukouvala et al. [44] assembled an integrated flowsheet model for a continuous tablet manufacturing process based on roller compaction. Models were implemented for feeders, powder mixing, roller compaction, milling, hopper flow, and

tableting. The focus for their model was to explore dynamic sensitivities in the system as a result of step changes in process settings as a basis for developing a control strategy for a continuous process.

As outlined in the introduction, traditional pharmaceutical development involves generating process understanding through executions of large multivariate experimental designs. Gavi and Reynolds [26] investigated whether developing a system model for a roller compaction-based tablet manufacturing process could yield a similar level of process understanding with reduced experimental effort. They connected process models of roller compaction, ribbon milling, and tableting with property functions for tablet tensile strength and dissolution. They calibrated the models using a subset of data, representing a 70% reduction in data requirements, and found the model was able to describe the multivariate interactions between the process-operating parameters and the product-quality attributes of tensile strength and dissolution. This work highlights some of the benefits of using a systems modeling approach within an industrial setting and identifies some of the potential opportunities.

5.5.1 Worked Example: Connecting a Process Model with a Property Function

As we connect process models with property functions, the complexity of the modeling implementation inevitably increases, making a detailed worked example difficult to provide. For more elaborate examples, the reader is referred to some of the literature examples already discussed, or to the next section which describes a more detailed case study. In this section, a worked example showing how to connect the roller compaction process function with tablet tensile strength will be used to demonstrate some of the concepts.

The key features of the model are shown in Figure 5.13. The relevant process functions for roller compaction, milling, and tablet compression are shown along with the property function for tensile strength. Some example input process parameters for the process functions are shown, such as roll force and gap size for the roller compactor. Intermediate properties that are important for downstream models are identified on the connecting streams, such as granule porosity and tablet porosity.

For the purpose of this worked example, we will exclude the screening mill model. The granule size distribution is not needed for the property function for the tensile strength calculation and we will assume that the granule porosity is equivalent to the ribbon porosity as it is unlikely that low-energy screen milling will cause further densification.

In Section 5.2.4 we set-up and calibrated a model for the roller compactor, which allows us to relate the process settings for this unit operation with the intermediate ribbon properties (in particular, porosity). We have also set-up and calibrated a property function model for tablet tensile

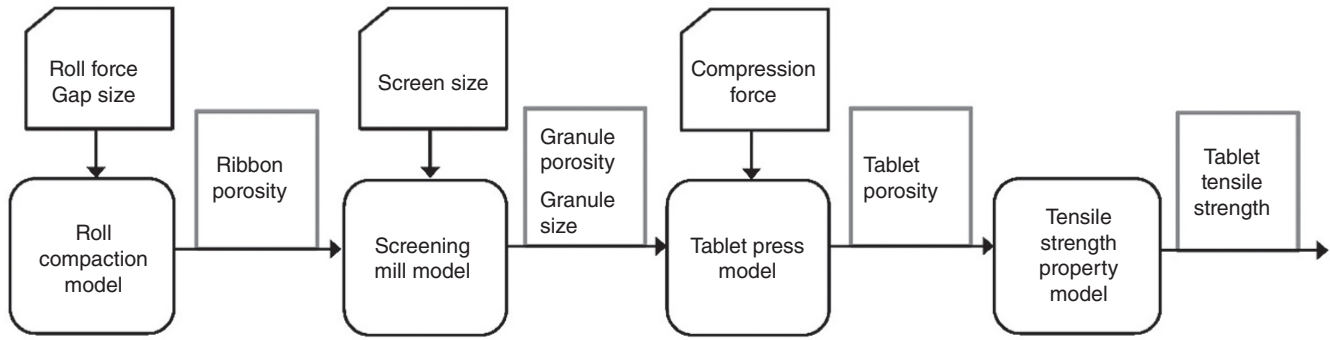


FIGURE 5.13 Schematic overview of simple system model to connect roller compaction with tablet tensile strength.

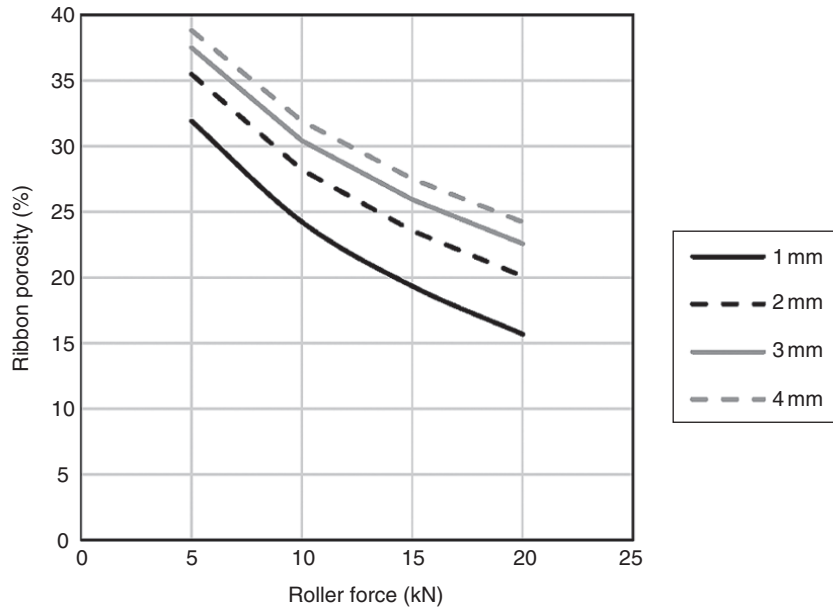


FIGURE 5.14 Calculated ribbon porosities over a range of roller forces and roller gaps.

strength in Section 5.4.1.1. This model requires the tablet and ribbon (or granule) porosity. However, we have not provided any further detail or discussion on a process model for the tablet press. For the purpose of this example, we will use the same relationship between compaction pressure and tablet solid fraction that were used for the roller compaction model (Eq. 5.9):

$$\gamma_T = \gamma_{T,0} P_{TP}^{1/K_T} \tag{5.29}$$

where

γ_T is the tablet solid fraction

P_{TP} is the pressure applied in the tablet press, which can be calculated by dividing the compression force by the punch area.

In this case $\gamma_{T,0}$ and K_T are parameters that can be calibrated from compression profile data. The data in Table 5.5 can be used to estimate values of 0.512 and 9.76 for $\gamma_{T,0}$ and K_T , respectively.

For this worked example, we will explore the influence of process ranges for a roller compactor with roll diameter 25 cm and roll width 2.5 cm. All other material properties and model parameters are as we have previously calibrated. We are interested in exploring the influence of operating the roller compactor with roll forces between 5 and 20 kN and roll gaps between 1 and 4 mm on the ribbon porosity and subsequent tablet tensile strength. The influence of operating the tablet press with compaction pressures between 100 and 300 MPa will also be investigated.

The spreadsheet assembled in Section 5.2.4 can be used to calculate ribbon solid fractions for different combinations of roll force and roll gap within the ranges of interest. The

ribbon porosity, ϵ_R , (required for downstream calculations) can be calculated easily from the solid fraction:

$$\epsilon_R = 1 - \gamma_R \tag{5.30}$$

Figure 5.14 shows the calculated values of ribbon porosity (expressed as a percentage) over the ranges of roll force and roll gap. We can see how the model captures the nonlinear interaction of these operating parameters on the ribbon porosity: increasing roll force results in ribbons of lower porosity and increasing roll gap results in an increased ribbon porosity.

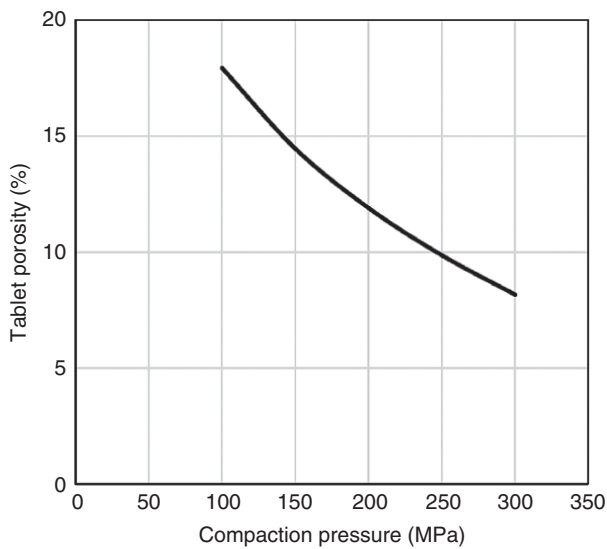


FIGURE 5.15 Tablet compressibility profile.

Having calculated the ribbon porosity, which we are also assuming is equivalent to our granule porosity, we need to use the tablet press model equation (Eq. 5.29) to calculate tablet solid fractions (and therefore tablet porosities) over the compaction pressures of interest. The result of these calculations is shown in Figure 5.15.

The final step is to use the calculated ribbon and tablet porosities in the tensile strength model (Eq. 5.28) to determine the tablet tensile strength at each combination of roll force, roll gap, and tablet compaction pressure. A summary, showing contours of the calculated tensile strengths, is shown in Figure 5.16. These plots show that as roller force is increased, the tensile strength decreases and at higher tablet compression pressures there is an increase in tensile strength. They also show that at higher roll gaps, there is an increase in tablet tensile strength. From these plots, it is possible to start identifying potential parametric controls that could be implemented as part of a control strategy to ensure adequate tablet tensile strength. If we take a minimum tensile strength of 1.7 MPa, based on the recommendation by Pitt and Heasley [36], we find that we might be constrained to quite a narrow region in the top left of the processing space for a roll gap of 1 mm, where we may need to keep roller force below 9kN and tablet compression pressure above 200 MPa. At higher roller gaps, these ranges expand and therefore greater flexibility in operating ranges for roll force and tablet compression pressure may be achieved by increasing the lower roll gap limit. This discussion shows how assembling even a simple system model can aid in beginning to explore potential operating spaces and product-control strategies. However, for this simple example we also need to be mindful of some of the limitations. Over the range of conditions explored here, the model moves outside of regions that may be reliable. Calculated tablet porosities at the highest compaction pressure of

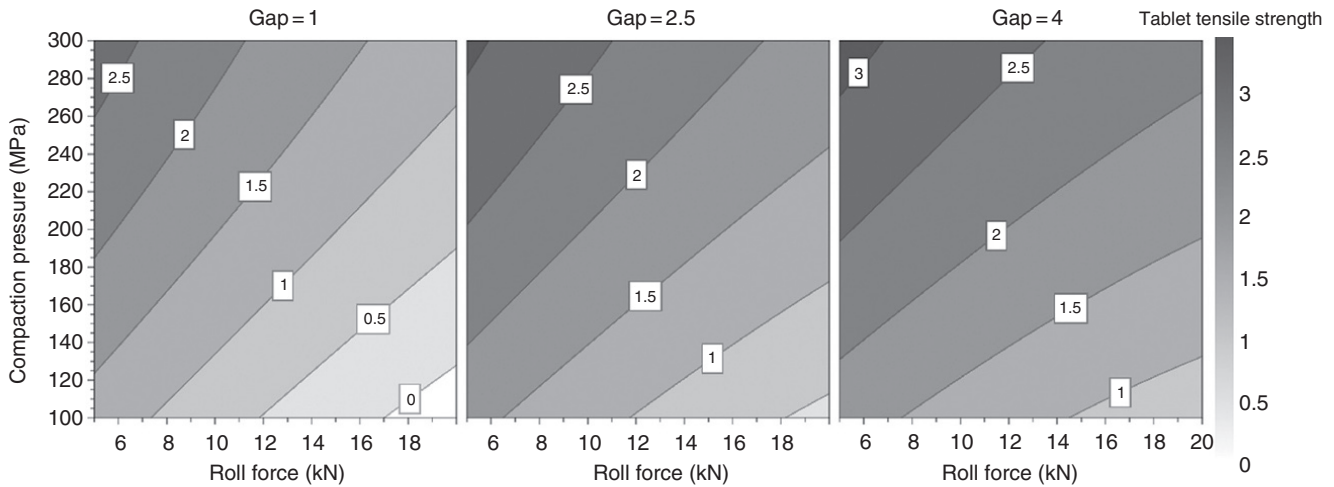


FIGURE 5.16 Contour plots of tablet tensile strength (in MPa) calculated for a range of roller forces and tablet compaction pressures. From left to right, contours are shown for roller gaps of 1, 2.5, and 4 mm.

300 MPa are 8.2%, which is moving into a region of over-compression with its associated problems, so in practice we may not achieve the highest tensile strengths shown at the top of the contour plots. At the lowest tablet compaction pressures and highest roll forces, the ribbon and tablet porosities begin to become very close or overlap, resulting in very low calculated tablet tensile strengths (or even negative in the case of 20 kN/1 mm/100 MPa). In practice, it would be very difficult to make robust compacts that could be satisfactorily characterized using diametral compression testing and therefore the exact profiles of the contours in this region may be less reliable, although it is clearly a region of the potential operating space we would want to stay well away from.

Another important consideration when using these calculations for exploring the operating space is that we have only modeled the tablet tensile strength, and even then have neglected some process-unit operations that could impact on tensile strength. A crude way to interpret the results of our calculations if we wanted to maximize tablet tensile strength is to say we need to minimize roller force and maximize tablet compression pressure, but this is likely to have undesirable consequences. In practice, we will have other product-quality attributes that need to be achieved, such as uniformity of dosage unit and dissolution. We would need to establish the sensitivity of the system to other quality attributes and potentially include property function models for those that are influenced by the process, before the system model could be used to fully explore the operating space and identify control strategies.

5.5.2 Case Study

The worked example in Section 5.5.1 allowed us to explore some concepts of using a system model of a tablet manufacturing process based on roller compaction. In this section, a more complex system model is presented to illustrate further applications of this approach. With greater model complexity, more sophisticated modeling tools need to be utilized. In this case, the model was implemented in gPROMS ModelBuilder 5 (Process Systems Enterprise Ltd.).

The scope of the system model needs to be considered in advance. It may not be possible or even desirable to implement models for every possible unit operation and product-quality attribute. Some unit operations will have a much greater influence on some quality attributes and other product-quality attributes may be very insensitive to the manufacturing process. The scope may be identified through risk assessment based on prior knowledge and current formulation or product knowledge. Early in development, the emphasis may be more on utilizing prior knowledge to guide experimentation to collect information on boundaries of operation or to guide scale-up. For example, the roller compaction modeling presented earlier can be used very early on to aid process development and scale-up as the multivariate

interactions between equipment-operating parameters are already quantitatively understood. Later in development, more understanding will have been generated to identify which product-quality attributes are most sensitive to the manufacturing process. For this case study, although dissolution is a critical product-quality attribute, even extremes of processing have demonstrated that there is a very low risk of the process impacting performance. It has been identified that the process can have a significant effect on product appearance and tablet hardness via insufficient tablet tensile strength. In addition, there is a theoretical risk that significantly poor granule flow could retard consistent flow in the tablet press leading to weight variability and therefore potentially impact on uniformity of dosage unit. This risk assessment process identified that the most important model responses should be granule flowability and tablet tensile strength.

Table 5.6 summarizes the main factors and responses that were modeled. Suitable process and property models were implemented in order to model interactions between the factors and responses. Briefly, the roller compactor model was the same as that used in this chapter and calibrated in the same way. The mill model used a PBM and a bimodal fragment distribution based on a fine mode described by a lognormal distribution matching the ungranulated powder size distribution and a fragmentation mode. The volume fraction of the disintegration mode was found to be a function of ribbon porosity as described in Section 5.3.3. Granule flowability was modeled by implementing a correlation between the granule size and the granule FFC. The blending model was based on the model of Kushner and Schlack [45]:

$$\sigma = \sigma_0[(1 - \beta) + \beta e^{-c \cdot r}] \quad (5.31)$$

where

σ is the lubricated tablet tensile strength,

TABLE 5.6 Main Factors and Responses Investigated in System-Model Case Study

Process Stage	Factors	Responses
Raw materials	Grade of main excipient	
Roller compactor	Roller force Roller gap Side-seal leakage	Ribbon porosity
Mill	Screen size	Granule size Granule flow
Blending/ lubrication	Variability in lubricant	
Tablet press	Compression pressure	Tablet porosity Tablet tensile strength

σ_0 is the unlubricated tablet tensile strength,
 r is the number of revolutions
 the parameters β and c describe the extent and rate of reduction in tensile strength.

Blending granules with a lubricant, such as magnesium stearate, is primarily performed to reduce material adherence to tablet punches and reduce appearance defects. However, as granules become coated with magnesium stearate, the subsequent tensile strength of the tablet will reduce. This model was implemented to incorporate the effect of decreasing tablet tensile strength with increasing extragranular blending time. The tablet press and tablet tensile strength were modeled as described in the previous worked example, with the added complexity of including the effect of side-seal leakage. Side-seal leakage was included as a factor in the study as this can vary between roller compaction equipment and over time in the same equipment for some designs. The effect on the flowability model was captured as the proportion of side-seal

finer was mixed with the granules leaving the mill to create the granule size distribution. To model the effect on tablet tensile strength, a mixture model was implemented to account for the different compactability of the side-seal material (assumed to have the same compactability as the ungranulated powder) and the granules. The compressibility-based model of Reynolds et al. [46] was used, which allows the calculation of the tensile strength of mixtures of pharmaceutical powders based on their constitutive properties.

For a constant set of material properties, the processing factors were varied in a “virtual experimental design.” An automated procedure was used to run over 3000 simulations across the combinations of the process factors. The output of the simulations was analyzed in Modde 11 (Umetrics AB) to identify the most important factors on each response and explore the operating space.

The granule flowability as a function of the upstream process factors is shown in Figure 5.17. Each contour plot shows how granule FFC changes as roll force and roll gap are varied, with higher roll forces and lower roll gaps leading to

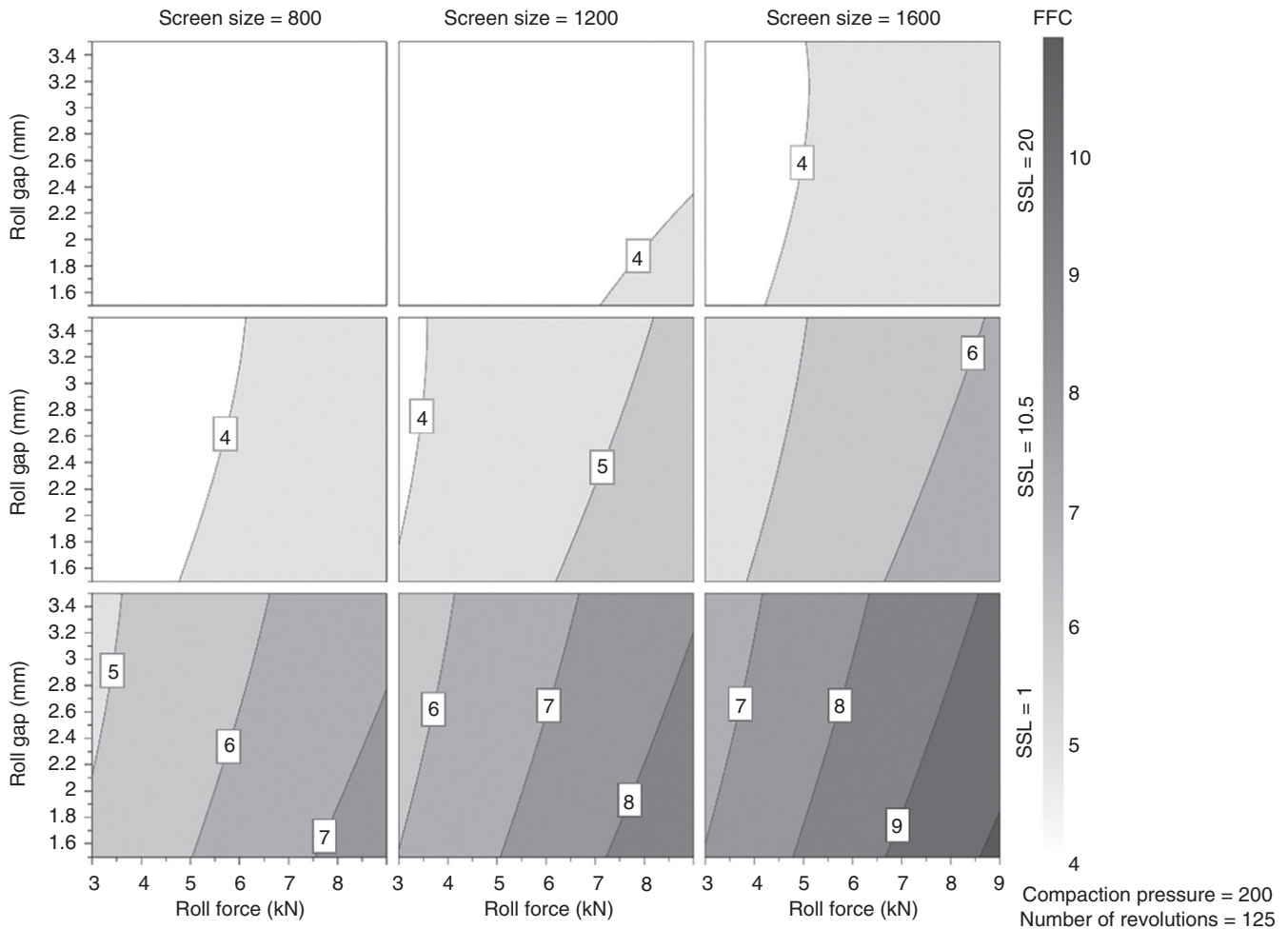


FIGURE 5.17 Contours of granule flowability (FFC) generated from system-model simulations.

increased flowability. As the plots move from left to right it can be seen that increasing screen size leads to an increase in granule flowability. The lower, middle, and upper plots correspond to side-seal leakage of 1, 10.5, and 20%. Increasing side-seal leakage significantly reduces the granule flowability as a result of the increased level of fines in the granule size distribution. In practice, this is unlikely to vary to such an extent for an established roller compactor, but it can help risk assess a change in equipment if that is known to have a different level of side-seal leakage. If our goal was solely to increase granule flowability, this shows that it could be achieved through a combination of maximizing roll force and milling screen size and minimized roll gap and side-seal leakage.

Tablet tensile strength is shown in Figure 5.18. Each contour plot shows that increasing tablet compaction pressure increases tensile strength, but increasing roll force reduces tensile strength. Increasing duration of blending (number of revolutions) leads to a reduction in the tensile strength due to an increased extent of powder lubrication. The effect of increasing side-seal leakage can be seen to marginally increase tensile strength as we move from the lower to the

upper plots. This is explained by an increase in the proportion of powder in the tablet that does not exhibit loss in compactability compared to the granule. Although not shown, an increase in roll gap increases the tensile strength slightly due to an increase in the ribbon porosity, although roll force has a greater influence. The effect of screen size is not shown either, but this has no effect as the influence of granule size on tablet tensile strength was not captured by any of the models used. Based on the output of these simulations, we now have a good idea which process settings and choices will have an influence on tablet strength. If our goal was to increase tablet tensile strength, this could be achieved by maximizing compaction pressure, roll gap, and side-seal leakage and by minimized roll force and blending time.

What we find is that there is a trade-off between the settings required to increase granule flowability and tablet tensile strength. A compromise needs to be reached and an operating space needs to be identified where granule flowability is greater than a certain value consistent with “good flow” and where tablet tensile strength is sufficient to ensure robustness to further processing, transport, and handling. Figure 5.19 shows a region in white where both the granule

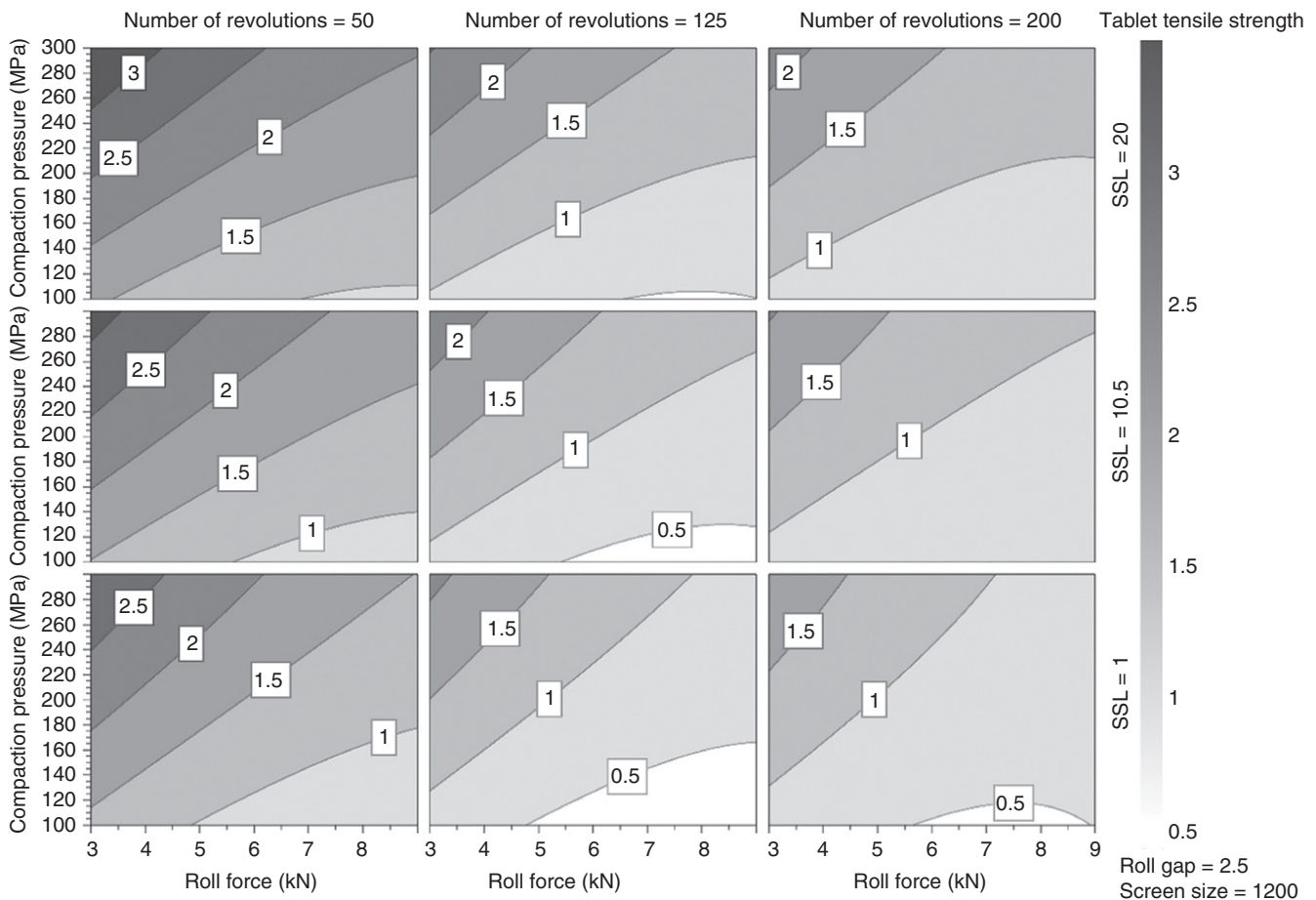


FIGURE 5.18 Contours of tablet tensile strength generated from system-model simulations.

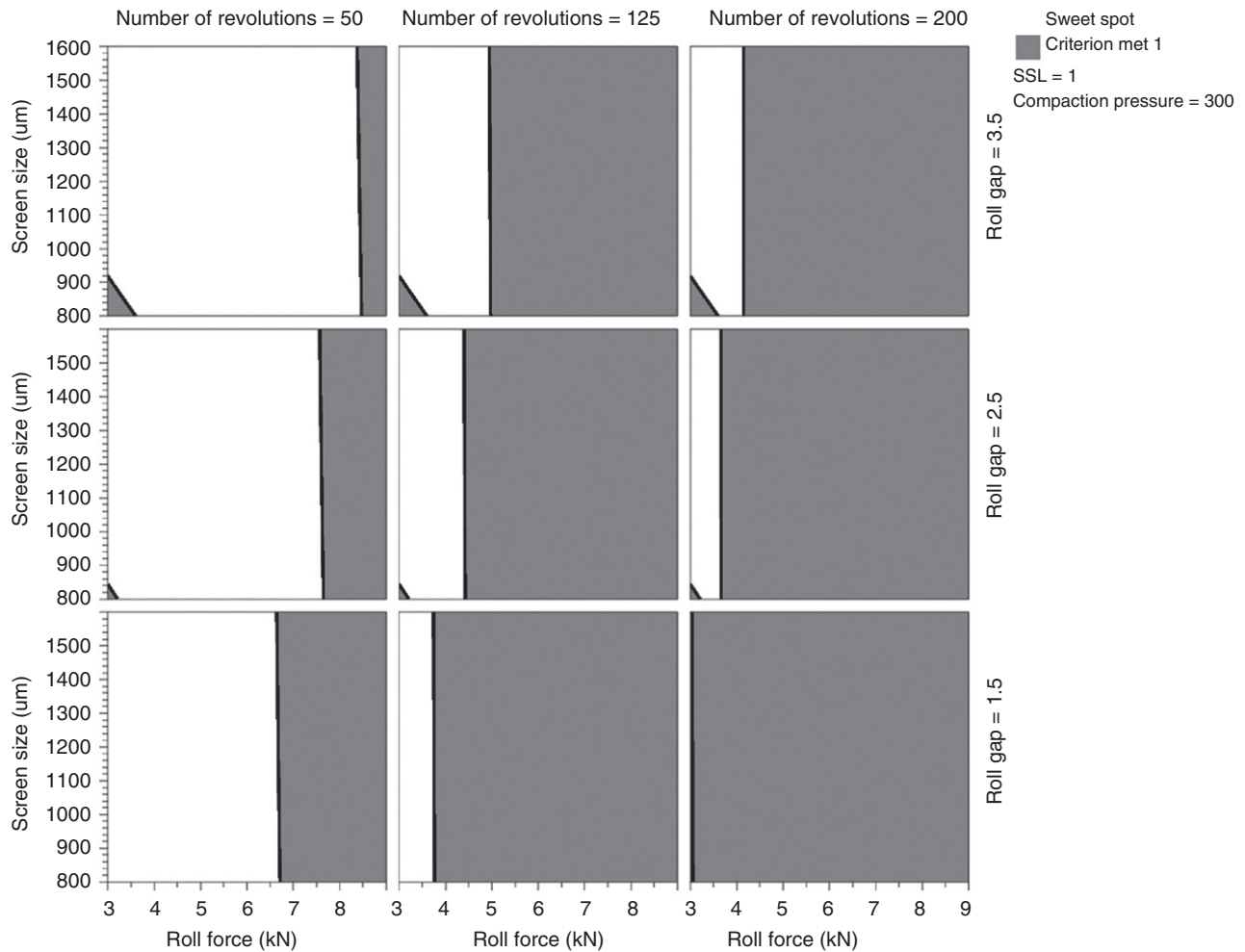


FIGURE 5.19 Operating regions that satisfy granule flowability ($FFC > 5$) and tablet tensile strength (> 1.7 MPa).

FFC is greater than 5 and the tablet tensile strength is greater than 1.7 MPa. We can see that in this case overblending significantly reduces the potential operating space due to the associated reduction in tablet tensile strength.

Used in this way, the system model provides process understanding by identifying the most significant process parameters and interactions. Operating ranges can be explored virtually as a basis for developing the control strategy. With such a model constructed, the impact of material variability and changes can also be explored.

To illustrate this, experiments were performed with a different grade of lubricant and used to calibrate new parameter values describing the rate and extent of reduction in tablet tensile strength in Eq. (5.31) (β and c). Simulations were performed by randomly sampling from a normal distribution of parameter values where the means were the initial values of β and c and the standard deviation was calculated as a third of the difference between the initial and subsequent parameter sets. We do not know the natural batch-to-batch variability, but it allows us to ask the question of what would be the impact if we saw variability that covered the space between

two different grades. The sample parameters are shown in Figure 5.20a and the resultant variation in tablet tensile strength is shown in Figure 5.20b. The coefficient of variation of tensile strength is 4.7%. This type of analysis allows us to quantify the effect of variability on the product performance. In this case we may be concerned as the variability is resulting in some tablet tensile strengths below the target of 1.7 MPa [36].

Another use for this type of system model is to simulate a change in raw materials. This could be a prospective simulation to assess the impact of a change in grade of one of the formulation components. In this case a change in grade of a major formulation filler is simulated. Small-scale compaction data are generated to recalibrate some of the material parameters used in the roller compaction, tablet press, and tensile strength models. Figure 5.21 shows the operating region that satisfies the requirements of granule flowability ($FFC > 5$) and tablet tensile strength (> 1.7 MPa). Comparing this with Figure 5.19 we can see some changes. There is clearly a wider range of roll forces that can be used and the overall operating ranges are wider suggesting a change

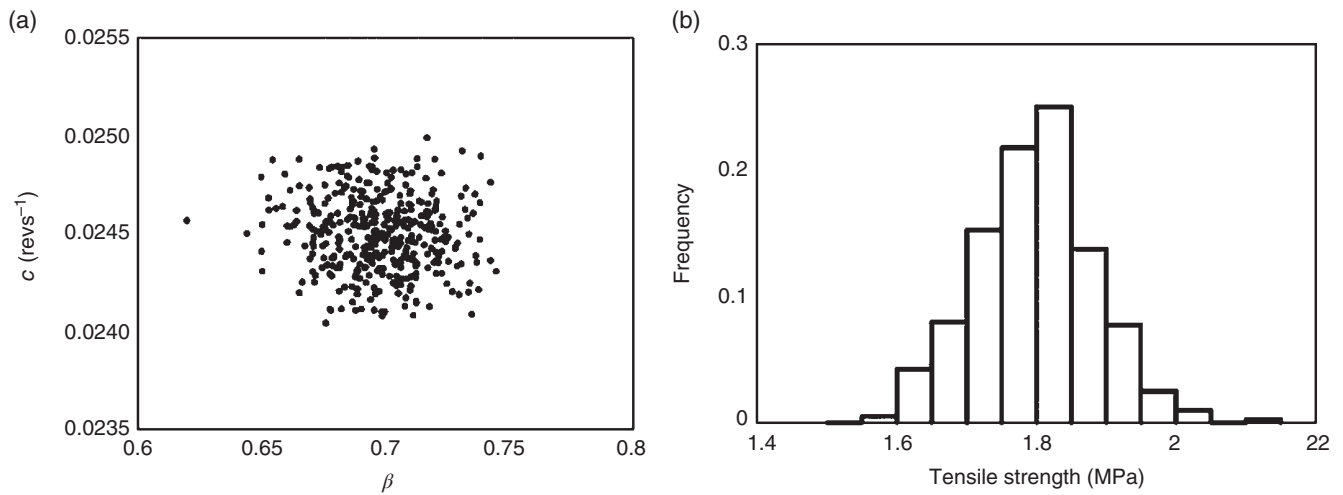


FIGURE 5.20 (a) Sampled variation in lubrication model parameters. (b) Distribution of tablet tensile strength as a result of variability in lubrication.

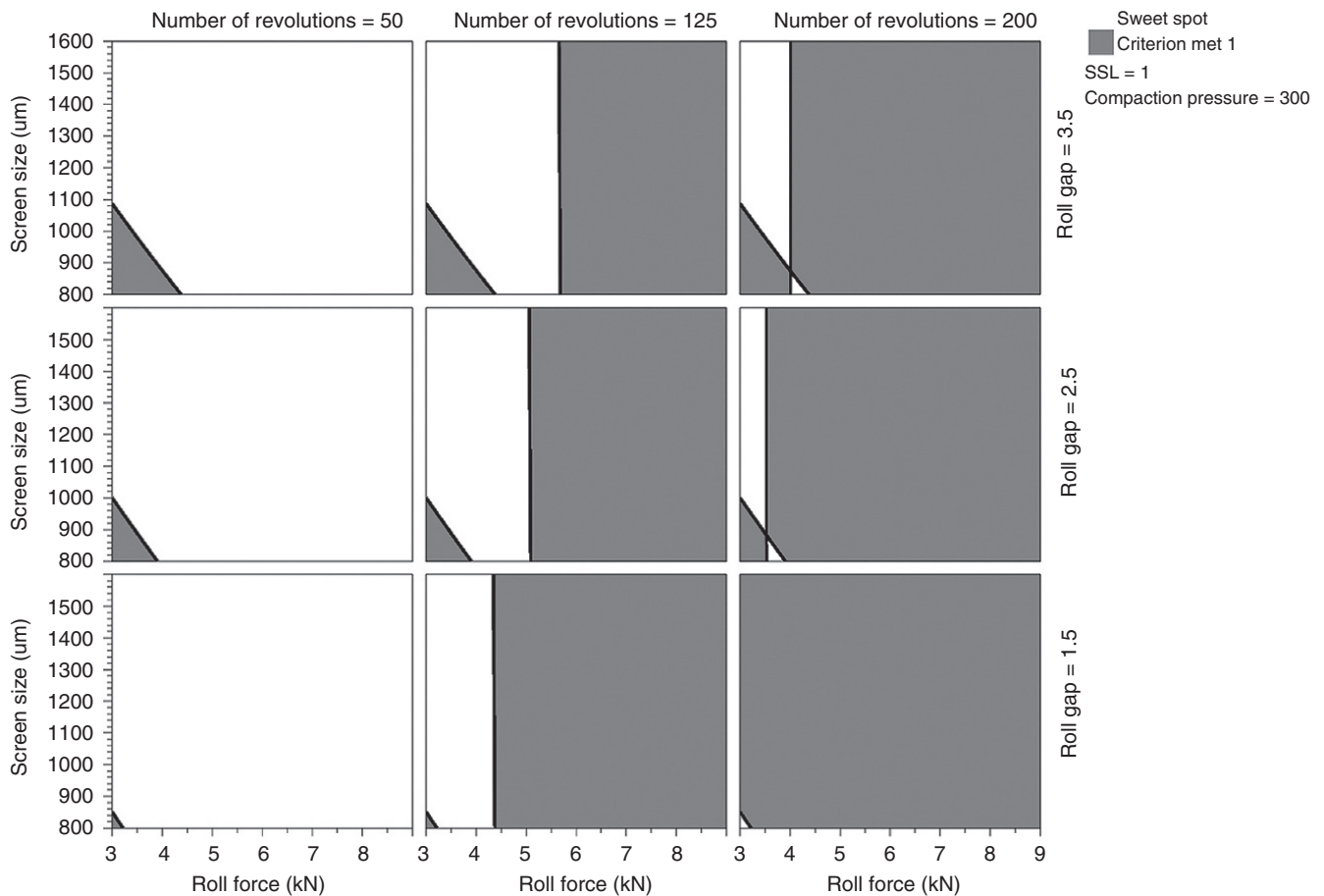


FIGURE 5.21 Operating regions that satisfy granule flowability ($\text{FFC} > 5$) and tablet tensile strength ($> 1.7 \text{ MPa}$) for a change in formulation filler grade.

to this grade of excipient will be low risk and may even improve the robustness of our process.

This case study has been used to highlight how a system model can be assembled and used for a roller compaction-based tablet manufacturing process. Even with relatively simplistic constituent models, the influence and interactions of many process settings and choices on product-quality attributes can be explored. The impact of potential changes and variability can be quantified to aid in risk assessment and decision making. As improved mixture models are developed to predict the bulk properties of powders, this type of model can be used as part of formulation design to identify a composition that will meet the required performance criteria.

5.6 SUMMARY

This chapter has illustrated how process modeling can be integrated into development of roller compaction-based tablet manufacturing processes. Topics have ranged from models of unit operations that can be used to aid in process design and scale-up through to more holistic systems-based models that link process models to product performance models. System modeling presents many opportunities for in-silico design of pharmaceutical processes. Virtual experimental designs can be run to quantify the risks associated with particular process parameter ranges and interactions. The impact of equipment and raw material changes can be explored and risk assessed in advance of implementing those changes. As more fundamental models emerge, these can be implemented and used within this framework to improve the level of prediction and reduce the burden of experimental model calibration. Ultimately, we see a vision emerging where model-based development of powder and particulate processes becomes a routine part of chemical engineering design.

5.7 NOTATION

A	defined in Eq. (5.4)
b	breakage function
B	beta function
c	parameter describing the rate of tensile strength reduction
D	roll diameter
$D[v,X]$	particle size corresponding to the X th volumetric percentile
F	defined in Eq. (5.8)
f_0	mass correction factor at minimum gap
f_{bimodal}	bimodal distribution function
$f_{\text{lognormal}}$	lognormal distribution function
f_{mat}	material strength parameter
f_{normal}	normal distribution function
f_{θ}	mass correction factor

K	power-law exponent for the powder pressure and relative density relationship
K_T	power-law exponent for the powder pressure and relative density relationship (Eq. 5.29)
k_b	bonding capacity
k_v	volumetric shape factor
\dot{m}_R	ribbon mass flow rate
n	fitted exponent (Eq. 5.11) or number density function (Eq. 5.19)
N_R	roll rotation rate
p	number of fragments
$p_{3,i}$	volume fraction in size class i
P_B	breakage probability
P_{max}	maximum pressure exerted by the rolls on the powder
P_{TP}	pressure applied in the tablet press
P_{α}	pressure at nip angle
P_{θ}	pressure at angle θ
q	power-law exponent (Eq. 5.21) or fragment-size dependence (Eq. 5.22)
$q_{3lg,i}$	discrete logarithmic density function in size class i
r	number of blender revolutions
R_f	roll force
S	roll gap or breakage rate
t	time
\bar{T}	tensile strength at zero porosity
T_T	tablet tensile strength
u	mother particle size
W	roll width
$W_{\text{m,kin}}$	mass-specific impact energy
$W_{\text{m,min}}$	mass-specific threshold energy
$x_{0,i}$	lower particle size for size class i
$x_{g,i}$	geometric mean particle size for size class i
y	fragment size
y'	minimum fragment size
z	fine-mode volume fraction
α	nip angle
β	parameter describing the extent of reduction in tensile strength
γ_0	preconsolidation relative density
γ_R	relative density of the ribbon
γ_T	relative density of the tablet
$\gamma_{T,0}$	parameter in Eq. (5.29)
γ_{α}	relative density at nip angle
γ_{θ}	relative density at angle θ
δ_E	effective angle of internal friction
ϵ_R	ribbon porosity
ϵ_T	tablet porosity
ζ	fine-mode fraction of bimodal distribution
θ	angle
μ	defined in Eq. (5.5)
μ_i	lognormal mean of i th mode

ν	defined in Eq. (5.6)
ρ_{true}	true density
σ	lubricated tablet tensile strength
σ_0	unlubricated tablet tensile strength
σ_i	geometric standard deviation of i th mode
v	particle size
ϕ_w	wall friction angle

REFERENCES

- Johanson, J.R. (1965). A rolling theory for granular solids. *ASME Journal of Applied Mechanics, Series E* 32: 842–848.
- Reynolds, G.K., Ingale, R., Roberts, R.J. et al. (2010). Practical application of roller compaction process modeling. *Computers and Chemical Engineering* 34: 1049–1057.
- Jenike, A.W. and Shield, R.T. (1959). On the plastic flow of coulomb solids beyond original failure. *ASME Journal of Applied Mechanics, Series E* 26: 599–602.
- Bindhumadhavan, G., Seville, J.P.K., Adams, M.J. et al. (2005). Roll compaction of a pharmaceutical excipient: experimental validation of rolling theory for granular solids. *Chemical Engineering Science* 60: 3891–3897.
- Yu, S., Gururajan, B., Reynolds, G. et al. (2012). A comparative study of roll compaction of free-flowing and cohesive pharmaceutical powders. *International Journal of Pharmaceutics* 428: 39–47.
- Lecompte, T., Doremus, P., Thomas, G. et al. (2005). Dry granulation of organic powders dependence of pressure 2D-distribution on different process parameters. *Chemical Engineering Science* 60: 3933–3940.
- Cunningham, J.C., Winstead, D., and Zavaliangos, A. (2010). Understanding variation in roller compaction through finite element-based process modeling. *Computers & Chemical Engineering* 34: 1058–1071.
- von Karman, T. (1925). On the theory of rolling. *ZAMM - Journal of Applied Mathematics and Mechanics* 5: 130–141.
- Katashinskii, V.P. (1986). Analytical determination of specific pressure during the rolling of metal powders. *Soviet Powder Metallurgy and Metal Ceramics* 10: 765–772.
- Dec, R.T., Zavaliangos, A., and Cunningham, J.C. (2003). Comparison of various modeling methods for analysis of powder compaction in roller press. *Powder Technology* 130: 265–271.
- Pérez Gago, A., Reynolds, G., and Kleinebudde, P. (2018). Impact of Roll Compactor Scale on Ribbon Density. *Powder Technology* 337: 92–103.
- Souihi, N., Reynolds, G., Tajarobi, P. et al. (2015). Roll compaction process modeling: transfer between equipment and impact of process parameters. *International Journal of Pharmaceutics* 484: 192–206.
- Liu, Y. and Wassgren, C. (2016). Modifications to Johanson's roll compaction model for improved relative density predictions. *Powder Technology* 297: 294–302.
- Mazor, A., Perez-Gandarillas, L., de Ryck, A., and Michrafy, A. (2016). Effect of roll compactor sealing system designs: A finite element analysis. *Powder Technology* 289: 21–30.
- Zinchuk, A.V., Mullarney, M.P., and Hancock, B.C. (2004). Simulation of roller compaction using a laboratory scale compaction simulator. *International Journal of Pharmaceutics* 269: 403–415.
- Press, W.H., Teukolsky, S.A., Vetterling, W.T., and Flannery, B.P. (1992). *Numerical Recipes in C: The Art of Scientific Computing*. Cambridge University Press.
- Perez-Gandarillas, L., Perez-Gago, A., Mazor, A. et al. (2016). Effect of roll-compaction and milling conditions on granules and tablet properties. *European Journal of Pharmaceutics and Biopharmaceutics* 106: 38–49.
- am Ende, M.T., Moses, S.K., Carella, A.J. et al. (2007). Improving the content uniformity of a low-dose tablet formulation through roller compaction optimization. *Pharmaceutical Development and Technology* 12: 391–404.
- Mirtič, A. and Reynolds, G.K. (2016). Determination of breakage rate and breakage mode of roller compacted pharmaceutical materials. *Powder Technology* 298: 99–105.
- Randolph, A.D. and Larson, M.A. (1971). *Theory of Particulate Processes*. New York: Academic Press.
- Berthiaux, H. and Dodds, J. (1997). A new estimation technique for the determination of breakage and selection parameters in batch grinding. *Powder Technology* 94: 173–179.
- Diemer, R.B., Spahr, D.E., Olson, J.H., and Magan, R.V. (2005). Interpretation of size reduction data via moment models. *Powder Technology* 156: 83–94.
- Kumar, J., Peglow, M., Warnecke, G., and Heinrich, S. (2008). An efficient numerical technique for solving population balance equation involving aggregation, breakage, growth and nucleation. *Powder Technology* 182: 81–104.
- Vanni, M. (1999). Discretization procedure for the breakage equation. *Aiche Journal* 45: 916–919.
- Diemer, R.B. and Olson, J.H. (2002). A moment methodology for coagulation and breakage problems: part 3 generalized daughter distribution functions. *Chemical Engineering Science* 57: 4187–4198.
- Gavi, E. and Reynolds, G.K. (2014). System model of a tablet manufacturing process. *Computers & Chemical Engineering* 71: 130–140.
- Vogel, L. and Peukert, W. (2005). From single particle impact behaviour to modelling of impact mills. *Chemical Engineering Science* 60: 5164–5176.
- Subero, J. and Ghadiri, M. (2001). Breakage patterns of agglomerates. *Powder Technology* 120: 232–243.
- Loreti, S., Chuan-Yu, W., Reynolds, G., and Seville, J. (2018). DEM-PBM modeling of abrasion dominated ribbon breakage. *AIChE Journal* 64: 1191–1204.
- Reynolds, G.K. (2010). Modelling of pharmaceutical granule size reduction in a conical screen mill. *Chemical Engineering Journal* 164: 383–392.
- Loreti, S., Wu, C.-Y., Reynolds, G. et al. (2017). DEM-PBM modeling of impact dominated ribbon milling. *Aiche Journal* 63: 3692–3705.
- Yu, S., Chuan-Yu, W., Adams, M.J. et al. (2015). The use of positron emission particle tracking (PEPT) to study milling

- of roll-compacted microcrystalline cellulose ribbons. *Powder Technology* 285: 74–79.
33. Sun, C.C. and Kleinebudde, P. (2016). Mini review: mechanisms to the loss of tableability by dry granulation. *European Journal of Pharmaceutics and Biopharmaceutics* 106: 9–14.
 34. Rajkumar, A.D., Reynolds, G.K., Wilson, D. et al. (2016). Investigating the effect of processing parameters on pharmaceutical tablet disintegration using a real-time particle imaging approach. *European Journal of Pharmaceutics and Biopharmaceutics* 106: 88–96.
 35. USP-40-NF-35. (2017). United States Pharmacopeia and National Formulary. *USP 40–NF 35*, Rockville, MD.
 36. Pitt, K.G. and Heasley, M.G. (2013). Determination of the tensile strength of elongated tablets. *Powder Technology* 238: 169–175.
 37. Farber, L., Hapgood, K.P., Michaels, J.N. et al. (2008). Unified compaction curve model for tensile strength of tablets made by roller compaction and direct compression. *International Journal of Pharmaceutics* 346: 17–24.
 38. Leane, M., Pitt, K., Reynolds, G., and Group Manufacturing Classification System Working (2015). A proposal for a drug product manufacturing classification system (MCS) for oral solid dosage forms. *Pharmaceutical Development and Technology* 20: 12–21.
 39. Yu, W., Muteki, K., Zhang, L., and Kim, G. (2011). Prediction of bulk powder flow performance using comprehensive particle size and particle shape distributions. *Journal of Pharmaceutical Sciences* 100: 284–293.
 40. Sandler, N. and Wilson, D. (2010). Prediction of granule packing and flow behavior based on particle size and shape analysis. *Journal of Pharmaceutical Sciences* 99: 958–968.
 41. Capece, M., Ho, R., Strong, J., and Gao, P. (2015). Prediction of powder flow performance using a multi-component granular bond number. *Powder Technology* 286: 561–571.
 42. Cleary, P.W. (2008). The effect of particle shape on simple shear flows. *Powder Technology* 179: 144–163.
 43. Ketterhagen, W.R., am Ende, M.T., and Hancock, B.C. (2009). Process modeling in the pharmaceutical industry using the discrete element method. *Journal of Pharmaceutical Sciences* 98: 442–470.
 44. Boukouvala, F., Niotis, V., Ramachandran, R. et al. (2012). An integrated approach for dynamic flowsheet modeling and sensitivity analysis of a continuous tablet manufacturing process. *Computers & Chemical Engineering* 42: 30–47.
 45. Kushner, J.t. and Schlack, H. (2014). Commercial scale validation of a process scale-up model for lubricant blending of pharmaceutical powders. *International Journal of Pharmaceutics* 475: 147–155.
 46. Reynolds, G.K., Campbell, J.I., and Roberts, R.J. (2017). A compressibility based model for predicting the tensile strength of directly compressed pharmaceutical powder mixtures. *International Journal of Pharmaceutics* 531: 215–224.

6

WET GRANULATION PROCESSES

KAREN P. HAPGOOD

School of Engineering, Deakin University, Geelong, Victoria, Australia

JAMES D. LITSTER

Department of Chemical and Biological Engineering, The University of Sheffield, Sheffield, UK

6.1 INTRODUCTION

Granulation is a size enlargement process, where individual powder particles, usually of several different components, are aggregated together to form a larger structured particle where the original particles can still be distinguished. Granulation of one or more drugs and excipients is a common first step in the manufacture of tablets or capsules for pharmaceutical drug delivery. Although direct compression is possible, granulation usually ensures good flow properties and uniform bulk density essential for tableting and reduces the risk of segregation by creating multicomponent granules with more uniform composition than the dry blend. Dry granulation, also known as roller compaction, uses compressive forces to form the aggregates and is covered in detail in Chapter 4.

In this chapter, we focus on wet granulation, where a liquid is used to form wet agglomerates, which are subsequently dried. We first provide an overview of the many advances in our understanding of the science underlying wet granulation. In the second half of the chapter, we discuss scale-up approaches and provide a case study of how the mechanistic knowledge can be applied to design the granulation process and scale-up with minimal trial and error. We conclude with some future directions for granulation process development and manufacturing in the pharmaceutical industry.

6.2 MECHANISMS IN WET GRANULATION

Wet granulation is the process of using liquid and a binder material (usually a polymer such as HPC, PVP, and HMPC)

to aggregate the individual particles in the dry mix into particle assemblies. The assemblies contain a mixture of drug and excipients and have a porous structure, which provides improved compression properties due to the rearrangement of particles during the collapse of the granule structure. There are three main stages in granulation [1] (refer to Figure 6.1):

1. Nucleation and wetting: where the spray drops form the initial granules, or nuclei.
2. Growth and consolidation: where agitation leads to granule–granule collisions and granule–particle collisions, resulting in larger granules – this is a size enlargement process. These collisions may also result in a reduction in internal pore space of the granule – this densification process is termed consolidation.
3. Attrition and breakage: agitation forces exceeding granule strength will result in either fracture of the granule into several large pieces or attrition of the outer layer of particles from the granule. These are both size reduction processes, and both may occur simultaneously.

Each mechanism is discussed in more detail below.

6.2.1 Nucleation

Mixing and distribution of the liquid during the nucleation phase (spraying phase) is an important step, and poor initial liquid distribution leads to a heterogeneous nuclei size distribution and increased variability in the granulation process [2]. Nuclei formed from under conditions of poor liquid

distribution will have a broad distributions of size, porosity, and saturations [2]. This in turn will lead to different growth and breakage rates for each granule. Ensuring a controlled nucleation step is the first step toward a controlled granulation process.

Immersion nucleation is where fine powders are engulfed by larger drops to form nuclei [3–5]. There are five steps in immersion nucleation as shown in Figure 6.2. Initially, the

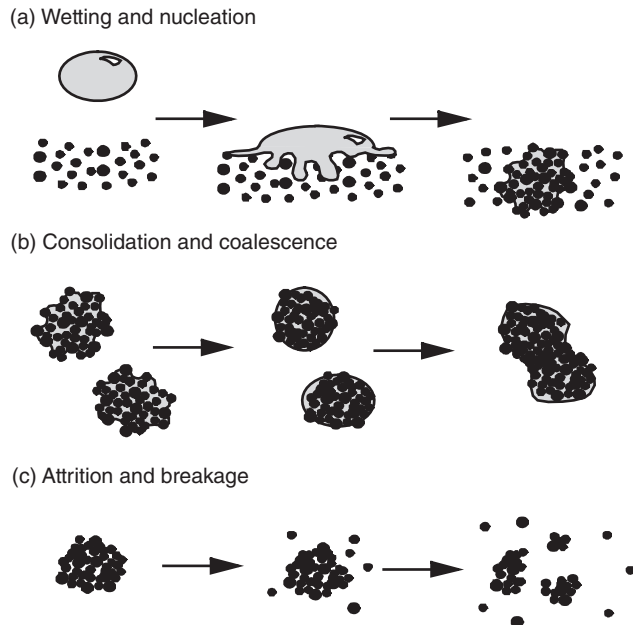


FIGURE 6.1 Rate processes in granulation: (a) wetting and nucleation, (b) consolidation and growth, and (c) breakage and attrition [1]. *Source:* Reproduced with permission of Elsevier.

drop must be formed at the nozzle. After landing on the powder surface, the drop may potentially shatter and break into fragments, as shown experimentally [7], or coalesce with another drop already at the surface of the powder. Once it reaches the powder surface, the drop may penetrate into the powder bed via capillary action (step 4 in Figure 6.2), or it may require mechanical agitation and shear to disperse the fluid through the powder. A similar diagram for drops landing on a fluidized powder bed has also been published [8], which also depicts the nuclei passing through the spray zone multiple times. Distribution nucleation [3–5] can occur in fluid beds operated with very fine drops that are the same size or smaller than the particle size. The mechanisms of distribution nucleation are less well understood.

There are two parameters describing the immersion nucleation process in the spray zone of the granulator – the drop penetration time and the dimensionless spray flux [9, 10]. These parameters have been combined to form a nucleation regime map [11] and can be used for scale-up [12] and to quantitatively model the nuclei size distribution in the spray zone [13, 14].

6.2.1.1 Drop Penetration Time As the powder moves beneath the spray zone, the small droplets will land on the powder surface and begin to mix with the powder (see Figure 6.2). When the powder is easily wetted or hydrophilic, i.e. the contact angle between the powder and fluid is less than 90° , penetration of the fluid into the powder pores will begin to occur. The rate at which a single drop of fluid, with volume V_d , viscosity μ and surface tension $\gamma_{L,v}$, will penetrate into a static porous medium with cylindrical pores of radius R_{pore} , and an overall porosity of ϵ is given by the drop penetration time t_p [15, 16]:

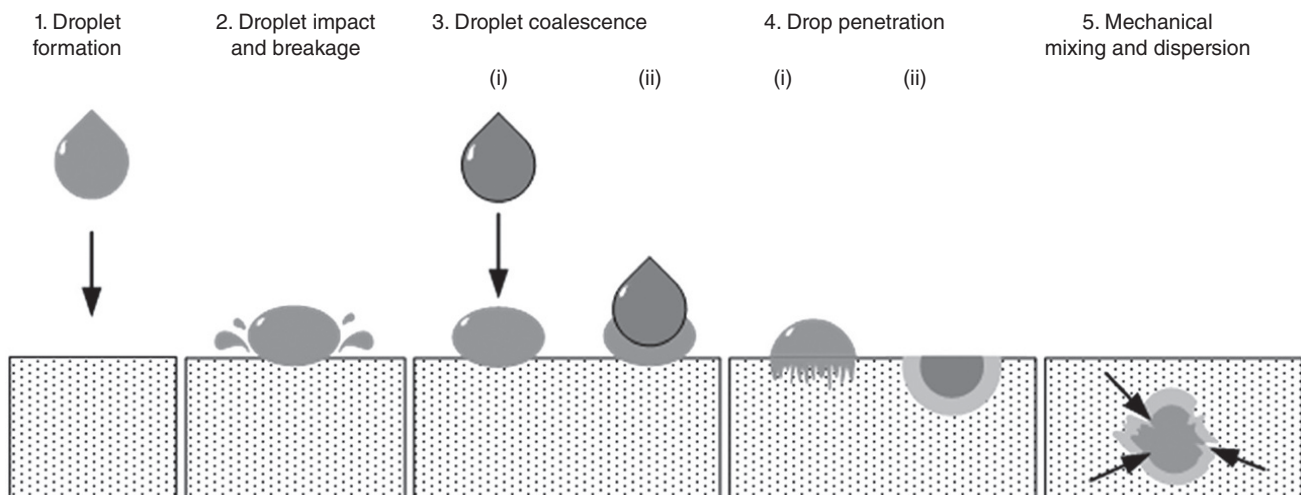


FIGURE 6.2 Five possible steps in immersion nucleation: (1) droplet formation, (2) droplet impact and possible breakage on the powder bed, (3) droplet coalescence upon contact with other droplets at high spray flux, (4) drop penetration into the powder bed (i) to form a nucleus granule (ii), and (5) mechanical mixing and dispersion of the liquid and powder. *Source:* Reproduced from Ref. [6] with permission from Wiley.

$$t_p = 1.35 \frac{V_d^{2/3} \mu}{\varepsilon^2 R_{\text{pore}} \gamma_{LV} \cos \theta} \quad (6.1)$$

In a loosely packed powder, similar to that found during agitation of the powder in the granulator, the voidage and pore size will be fairly heterogeneous. The powder will contain some combination of small pores and much larger pore spaces, or macrovoids. Liquid will flow through the small micropores, but there is no capillary driving force for the liquid to flow into the large macrovoids, as the rapid expansion of the pore radius dramatically reduces the Laplace pressure, which is the driving force for capillary flow. This means that the liquid does not flow through the macrovoids and must instead find a path around the macrovoid, slowing down the penetration [9] (Figure 6.3).

The effective voidage ε_{eff} is the first simple step toward including the effect of nonuniform pore structure on the penetration of a fluid into a real powder bed [9]. At the tap density ε_{tap} , the bed is assumed to contain no macrovoids. As the bed becomes less densely packed, the porosity of the bed increases, and the fraction of the voidage above the initial tapped bed voidage is assumed to form macrovoids. The effective bed voidage thus estimates the amount of pore space that is actually available for capillary driven flow:

$$\varepsilon_{\text{eff}} = \varepsilon_{\text{tap}} (1 - \varepsilon + \varepsilon_{\text{tap}}) \quad (6.2)$$

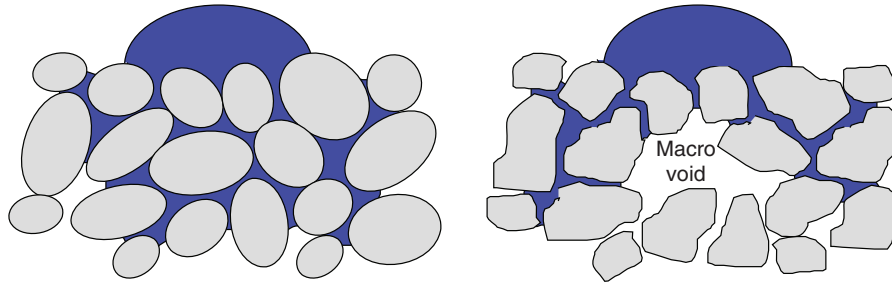


FIGURE 6.3 Liquid will penetrate the micropores driven by capillary action, but wherever a pore rapidly expands into a macrovoid, the fluid has to find a path around the flow obstruction [9]. *Source:* Reproduced with permission from Elsevier.

The effective porosity is used in the Kozeny equation in place of overall bed porosity ε to estimate the effective micro-pore size R_{eff} , which is pore size available for liquid flow:

$$R_{\text{eff}} = \frac{\phi d_{32} \varepsilon_{\text{eff}}}{3 (1 - \varepsilon_{\text{eff}})} \quad (6.3)$$

Thus, the most appropriate equation for estimating the drop penetration time t_p into a loosely packed porous powder bed is

$$t_p = 1.35 \frac{V_d^{2/3} \mu}{\varepsilon_{\text{eff}}^2 R_{\text{eff}} \gamma_{LV} \cos \theta} \quad (6.4)$$

The drop penetration time is an indication of the kinetics of nucleus formation. Equation (6.4) shows that the penetration time depends on several factors, including the drop size V_d , the fluid properties (viscosity μ , surface tension γ_{LV} , and contact angle θ), and the powder packing structure (ε_{eff} and R_{eff}). However, the fluid viscosity has the largest effect, as the fluid viscosity can range over several orders of magnitude. This is also commonly found in pharmaceutical granulation, where low viscosity fluids such as water or ethanol are used as frequently as high viscosity fluids (e.g. a 7% HPC solution has a viscosity of 104 mPa·s [10]). Figure 6.4 shows a water drop penetrating into a lactose powder

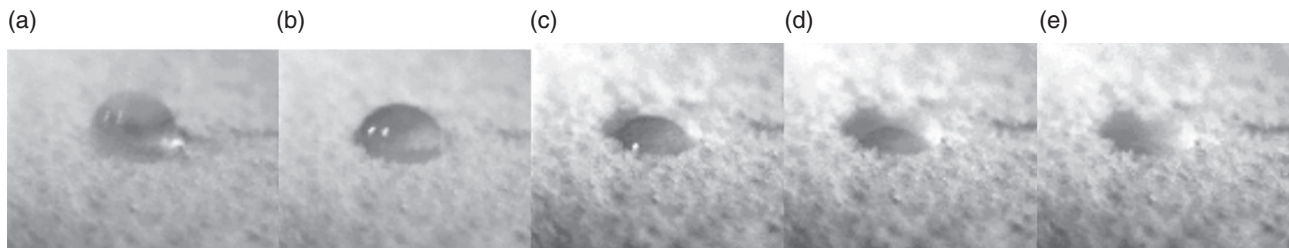


FIGURE 6.4 Water drop (~6 μl volume) penetrating into lactose powder: (a) impact, (b) 0.23 seconds, (c) 0.9 seconds, (d) 1.4 seconds, and (e) 2.3 seconds. *Source:* Adapted from Ref. [9]. Reproduced with permission from Elsevier.

powder bed over 2.3 seconds. In contrast, a similar drop of 7% HPC takes approximately two minutes to penetrate into the powder bed [9]. This time scale is clearly much longer than the time scale of agitation and implies that nucleation via wetting and capillary action (step 4 in Figure 6.2) is not the dominant mechanism for high viscosity systems. Instead, dispersion of the fluid through the powder will need to occur via mechanical mixing of the powder (see step 5 in Figure 6.2).

For a given powder, the drop penetration time is proportional to the liquid properties group $\mu/\gamma_{LV}\cos\theta$. For a given fluid, as the powder becomes finer (i.e. the d_{32} particle size decreases), the drop penetration time increases, primarily due to a decrease in packing efficiency, which creates a higher proportion of macrovoids [9]. Equation (6.4) can be used to estimate the drop penetration time from theory but can have large errors for very fine, cohesive powders [9]. Experimental tests are recommended in this case, with Eq. (6.4) used to scale the experimental results to account for differences in drop size between the experiment and the actual manufacturing process (e.g. see Ref. [17]).

Drop penetration time tends to decrease if the powder is already partially wet [6, 18], and the effect is more pronounced for viscous fluids with long penetration times [6]. For drops with long penetration times that land adjacent to a prewet patch of powder, the drop will tend to migrate over and penetrate into the preexisting wet patch rather than penetrate into the dry powder directly underneath it (see Figure 6.5).

The above discussion is for drop penetration into a stationary powder bed. Drops impacting and penetrating into a moving powders show more complex behavior. In fluid bed granulation, the fluid slowly flows outward, and the powder layer is slowly built up by collisions between the wet outer surface of the drop or wet agglomerate and the agitated dry powder [19]. Fluid flow and nuclei formation kinetics are still

controlled by the same fluid properties as shown in Eq. (6.4), but the rate of powder addition to the exterior surface is an additional factor in the kinetics [19]. Growth will continue until the saturation of the agglomerate decreases to the point that no further liquid can reach the powder surface. This saturation limit was defined as the “wetting saturation” S_w and determines whether additional growth will occur (see Figure 6.6).

6.2.1.2 Dimensionless Spray Flux Dimensionless spray flux [10, 14] considers the granulator spray zone and the flux of drops landing on the moving powder surface. The derivation of spray flux is straightforward [10] and is not intended to be equipment specific, although it has been most frequently applied in mixer granulation.

During the liquid addition stage of the granulation process, the powder surface is moving beneath the spray with a velocity v underneath a spray of width W at the powder surface. Therefore, the area of dry powder passing beneath the spray nozzle per second is vW . Each drop hitting the powder surface in the spray zone will leave a “footprint” as it wets into the powder. The number of drops hitting the powder surface per unit time can be calculated by dividing the total liquid flow rate by the volume of an individual drop. Assuming that each drop lands separately on the surface, without overlapping with any other drops, the total wetted area created per unit time can be estimated by multiplying the number of drops being added per second by the area that each drop will occupy when it lands on the powder surface. Dimensionless spray flux is then defined as the ratio of the rate at which wetted area is created by the incoming droplets compared with the total area of dry powder passing through the spray zone [10]:

$$\Psi_a = \frac{\dot{a}}{\dot{A}} = \frac{3\dot{V}}{2\dot{A}d_d} \quad (6.5)$$

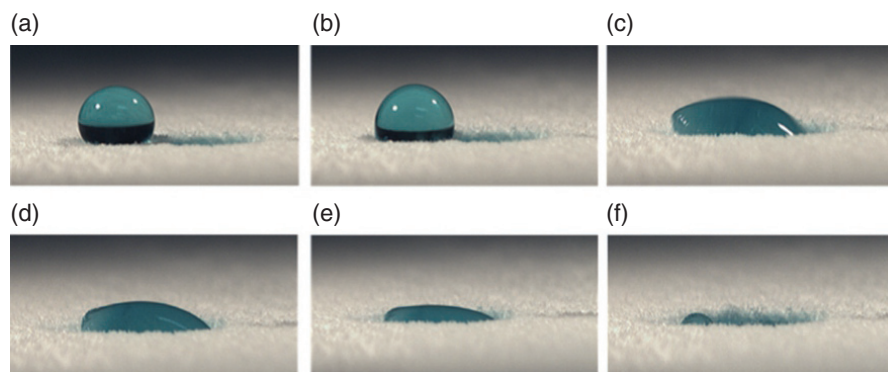


FIGURE 6.5 Penetration of a drop of PEG400 ($\mu = 120$ mPa-s) into prewetted lactose 200 mesh powder. The footprint of the previous drop is visible to the right of the added drop (separation distance of 3 mm). (a) Impact, (b) 4.05 seconds after impact, (c) 5 seconds, (d) 5.02 seconds, (e) 12.06 seconds, and (f) 51.6 seconds [6]. *Source:* Reproduced with permission from Wiley.

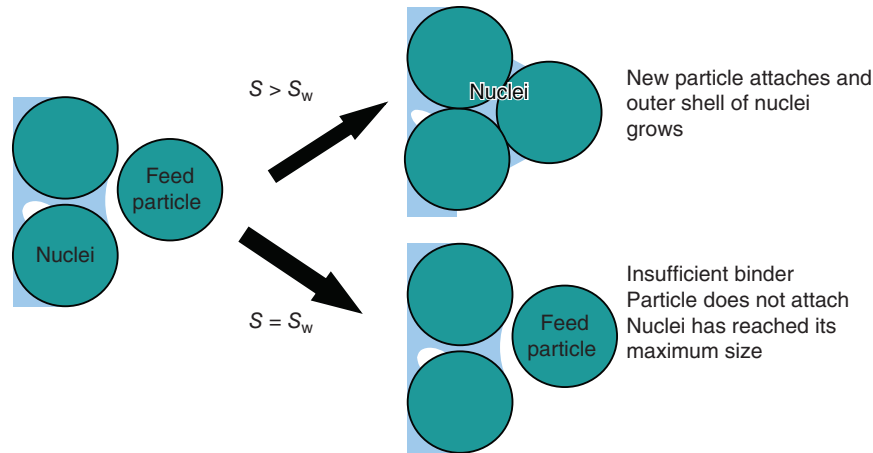


FIGURE 6.6 Nuclei growth in a fluid bed via growth of the outer nuclei shell continues provided the agglomerate saturation is higher than the wetting saturation limit S_w . *Source:* Adapted from Schaafsma et al. [19].

The dimensionless spray flux is a measure of the density of drops falling on the powder surface. Physically, $\Psi_a = 0.1$ means that approximately 10% of the powder surface will be wetted per pass through the spray zone. At low spray flux ($\Psi_a \ll 1$), drop footprints will not overlap and each drop will form a separate nucleus granule. The size of the nuclei will be directly proportional to the drop size [9, 20]. As Ψ_a increases, the fraction powder wetted also increases, although the relationship is linear only at low Ψ_a values. At high spray flux ($\Psi_a \sim 1$), the spray rate is high compared with the rate of dry powder entering the spray zone and the drops significantly overlap each other as they land on the powder bed. Spray flux $\Psi_a \geq 1$ means that the incoming droplets will theoretically cover 100% of the dry powder passing beneath the nozzle, assuming no drop overlap occurs [10]. At this value, the nuclei formed will be fragments of a much larger sheet of wet powder and will bear little relationship to the original drop size. Good granulation can still be achieved provided that the shear forces during granulation are strong enough and uniform enough to be effective. In pharmaceutical granulation, these conditions usually also consolidate the granules leading to lower porosity granules with slower dissolution. The process is illustrated schematically in Figure 6.1.

Note that \dot{A} is a dynamic quantity, defined as the outer perimeter of powder sprayed per second (m^2/s), and is not equivalent to static footprint spray area A . The difference between the static footprint area of the spray and the dynamic area flux is illustrated schematically in Figure 6.7 for several cases of varying powder velocities.

The impact of Ψ_a on nuclei formation has been studied in both mixer granulators [12, 17, 21, 22] and externally on a simplified moving power bed [10, 23]. To eliminate the effect of granule growth, the nuclei size distribution was measured after a single pass of the powder through the spray zone [10]. The results clearly show that at low spray flux ($\Psi_a < 0.2$) the nuclei

size distribution is quite narrow. As spray flux increases, the distribution broadens as agglomerates begin to form. At very high spray flux ($\Psi_a > 1$), the spray zone has become a continuous cake and the nuclei distribution bears no resemblance to the drop distribution. Further, when the spray flux is low, changes to the spray drop distribution are directly mapped onto the nuclei size distribution (Figure 6.8) [10].

The dimensionless spray flux parameter is intended to capture the major effects of drop overlap in the spray zone on the nuclei distribution as simply as possible, to encourage its use as a scale-up parameter. The derivation contains several major simplifying assumptions. Firstly, the spray is assumed to consist of mono-sized drops. Secondly, the spray flow rate is assumed to be uniformly distributed over the entire spray area. This is rarely true in industrial applications, as generally, the flow rate is higher in the center of the spray area than at the sides [8, 24]. The theory has been developed and validated for the case of a completely dry powder entering the spray zone. In practice this is true only for the very initial stages of the granulation process. As spraying progresses, a mixture of dry powder and previously formed nuclei and granules will pass through the spray and will be rewetted, which will affect the final size distribution for some fluids [6].

Finally, the derivation ignores the fact that the nucleus size is always larger than the drop size. During spraying, two drops may land close to each other but without touching. As the liquid penetrates into the powder, the larger nuclei may grow and touch each other, causing coalescence of the nuclei even when the spray drops landed separately. Wildeboer et al. [14] extended the theory and modeling to account for the effects of drop size distribution, for nonuniform spray density, and for probability of coalescence due to nucleus spreading. Nucleus spreading is described by the nucleation ratio K , which can be defined on either a volume basis – i.e.

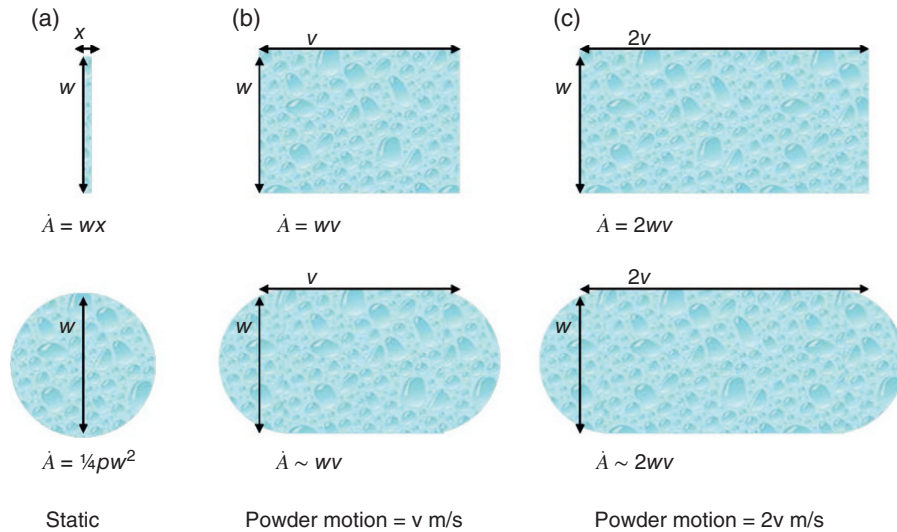


FIGURE 6.7 Relationship between the spray area of the nozzle, A , and the dynamic spray area \hat{A} for a line spray. Upper case shows a line spray and lower case shows a circular spray. (a) represents the static spray case where there is no powder motion, and (b) show the area flux of powder beneath the nozzle for powder moving at v m/s and (c) area wetted by the spray for powder moving at $2v$ m/s.

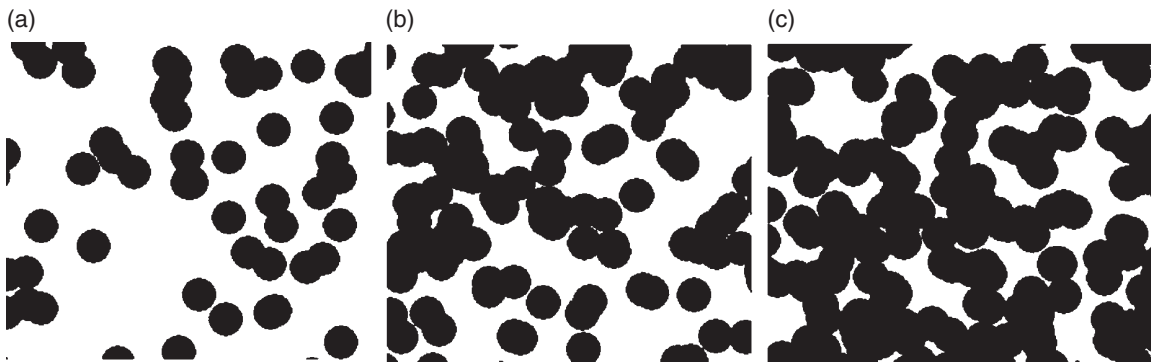


FIGURE 6.8 Monte Carlo simulations of spray flux. (a) $\Psi_a = 0.37, f_{\text{covered}} = 0.31$ (b) $\Psi_a = 0.62, f_{\text{covered}} = 0.46$ (c) $\Psi_a = 1.2, f_{\text{covered}} = 0.70$.

the ratio of the nucleus volume to the drop volume, K_v [19] – or on an area basis where K_a is the ratio of the projected area of the drop (a_d) to the nucleus (a_n) [14] and a is the projected area of the nucleus (a_n) and drop (a_d). Typical values for K_v range between 3 and 30 or higher [18, 19, 23, 25, 26]. The probability of a single drop forming a single nucleus is therefore related to the dimensionless nucleation number, Ψ_n [14]:

$$\Psi_n = K_a \frac{3\dot{V}}{2wvd_d} \quad (6.6)$$

The dimensionless nucleation number Ψ_n differs from the original spray flux Ψ_a by the factor K_a , which accounts for the degree of nucleus spreading. Equation (6.6) has been used to model the nuclei distributions in the spray zone over a range of Ψ_n [14], accounting for nonuniform sprays and nucleus spreading and coalescence. At a given value of the dimensionless nuclei number Ψ_n , the density and size distribution

of the nuclei formed on the surface is constant, i.e. the final value of Ψ_n is the sole determinant of the final nuclei size distribution.

Assuming complete spatial randomness, spatial statistics can be used to derive an analytical solution for both the fraction surface coverage and fraction of agglomerates formed [13]. Under these conditions, the drops landing randomly on the target area are described by a Poisson distribution. The fraction of the surface covered by drops in a single pass through the spray zone is given by [13]:

$$f_{\text{covered}} = 1 - \exp(-\Psi_a), \quad (6.7)$$

The fraction of nuclei formed from n drops is given by:

$$f_n = \exp(-4\Psi_a) \left(\frac{(4\Psi_a)^{n-1}}{(n-1)!} \right), \quad (6.8)$$

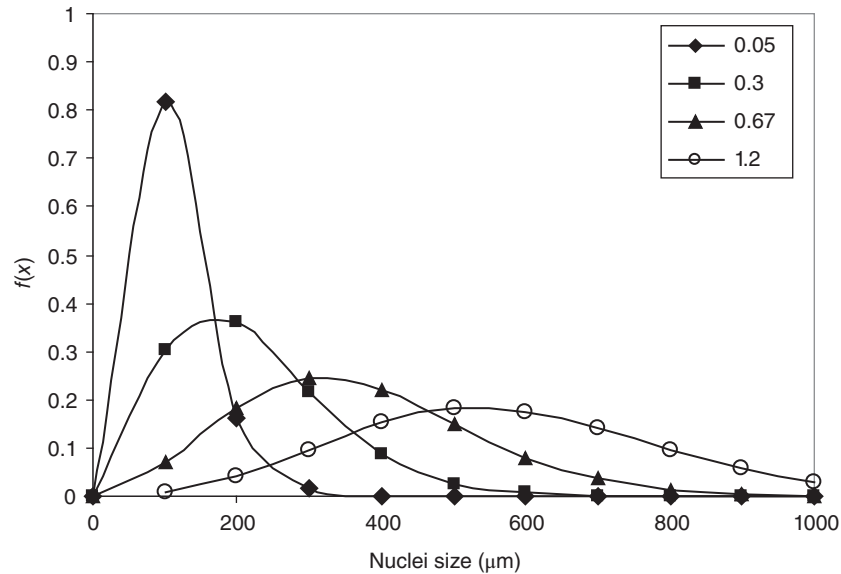


FIGURE 6.9 Nuclei size distributions predicted by Eq. (6.9) at a range of spray flux values assuming 100 μm mono-sized drops.

Thus, we can calculate the number of single drops, not overlapping with any other drops, and by difference, the number of agglomerates [13]:

$$f_{\text{single}} = \exp(-4\Psi_a) \quad (6.9)$$

Equation (6.8) can be used to estimate the initial nuclei size distribution as a function of spray flux for mono-sized drops. Figure 6.9 shows the nuclei distributions predicted by Eq. (6.8) at a range of spray flux values for 100 μm mono-sized drops. As the spray flux increases, a higher percentage of nuclei are formed from multiple drops, creating a larger and broader nuclei size distribution. However, prediction of the bimodal nuclei distributions that form at higher spray flux [27] appears to require a more sophisticated modeling approach incorporating overlapping drop size distributions.

Practical application of dimensionless spray flux requires measurements of the drop size distribution and powder velocity, and spray distribution for Ψ_a . Spray distribution is the easiest of the three parameters to measure (e.g. see Ref. [24]), and laser diffraction and/or doppler analysis of spray nozzles is widely available for drop size characterization. Currently, the most challenging parameter in Ψ_a to measure in real granulators is the powder velocity, and this is discussed separately in Section 6.3.2. Drop size in a spray is a strong function of the distance from the nozzle, so the measurements of drop size must be taken at the approximate distance where the spray intersects the bed. This distance usually increases as the granulator scale increases and varies depending on spray lance design. Increases in the nozzle lance depth cause a reduction in the spray area and hence an increase in

spray flux. Maintaining a constant nozzle height is important in reducing unwanted variation in granulator performance.

6.2.1.3 Nucleation Regime Map The drop penetration time describes the kinetics of nucleation from a single drop as a function of the material properties in the formulation. The spray flux describes the physical interactions of multiple drops in the spray zone. Together, these two parameters form the basis of the nucleation regime map [11], which describes the optimal conditions for uniform liquid distribution and suggests some ideal conditions for controlled nucleation.

When the drop penetration time is short, the fluid will sink quickly into the powder bed to form a nucleus granule. If no other drops land on top of the sinking drop as it passes through the spray zone, a single nucleus granule will be formed with a size equal to 2–3 times the drop volume (see Eq. (6.7)). If this process occurs for all the drops, the nuclei size distribution will be directly proportional to the drop size distribution. This is known as “drop-controlled nucleation” [11] and occurs at low drop penetration time and low spray flux (i.e. low spray density). As a guide, the spray flux needs to be less than $\Psi_a < 01$ for approximately 2/3 of the nuclei to be formed from a single drop (see Eq. (6.11)).

This “one drop produces one granule” mode of nucleation will not occur with a formulation with the same low drop penetration time, but granulated at a high spray flux. At higher spray flux, the spray density will be too high and the vast majority of the drops will coalesce with another drop on the powder surface. The surface of the powder will be wetted by an almost continuous sheet of liquid, rather than discrete drops. As the powder moves due to agitation, the “caked”

powder will be broken down and the fragments will be disperse through the powder. This is the “mechanical dispersion” regime of nucleation [11], where the liquid is dispersed primarily due to powder agitation and shear, rather than by fluid flow and wetting.

If the drop penetration time is much longer, the liquid will remain on the surface of the powder for an extended period of time, in the order of minutes. The constant powder motion means that it is more likely to coalesce with other unpenetrated droplets (see step 3 in Figure 6.2), merge into a section of powder wetted by an earlier drop [6], or roll into depressions in the powder and form rivulets [23]. Even if the drop does not coalesce with other drops, powder agitation and shear will be required to disperse the fluid through the powder to form a nucleus (step 5 in Figure 6.2). Again, this is the “mechanical dispersion” nucleation regime.

These two regimes of nucleation can be summarized using a nucleation regime map (Figure 6.10). The axes of the map are the dimensionless spray flux Ψ_a and the dimensionless penetration time τ_p :

$$\tau_p = \frac{t_p}{t_c} \quad (6.10)$$

where t_c is the circulation time for the droplet or nuclei to return to the spray zone. This is currently unable to be quantified due to insufficient understanding of powder flow and circulation patterns in most industrial granulation equipment. In general, the drop penetration time needs to be much faster

than the circulation time, and an arbitrary limit of 1/10th of the circulation time ($\tau_p < 0.1$) has been set as the upper limit for drop controlled nucleation [11].

In the lower left-hand corner of the map is the drop-controlled regime, which occurs at low dimensionless drop penetration time ($\tau_p < 0.1$) and low spray flux ($\Psi_a < 0.1$). In this corner, each drop will generally land separately without touching any other droplets. As soon as either the spray flux or penetration time increases slightly, the system enters an intermediate region where both wetting and agitation will be equally important in determining liquid distribution and nucleation. As spray flux or penetration time increase further, the mechanical mixing will become the dominant mechanism in nucleation, as the wetting kinetics occur on a much longer time scale.

Understanding which nucleation regime the process is operating in can be extremely useful for understanding how to optimize liquid dispersion and troubleshoot during manufacturing. For formulations with low penetration times, optimizing the spray flux by adjusting the fluid flow rate, drop size, spray area, powder velocity, etc., will have a significant effect on the process, as it moves from the mechanical dispersion regime toward the drop-controlled regime, crossing the regime boundaries (see lower horizontal arrow in Figure 6.10). As the process nears the drop-controlled regime, the nuclei size distribution will approach the drop size distribution of the spray. Atomization of a fluid is fairly well understood (e.g. see Ref. [28]) and is therefore much easier to optimize compared to a granulation process. In contrast, when working with a formulation with a long drop penetration time, optimizing the spray parameters to reduce spray flux will have little effect on the nucleation process. Once the drop penetration time is high, it is extremely difficult to achieve drop-controlled nucleation, as the liquid will form a “puddle” on the powder due to either slowly penetrating drops or high spray density. In addition, atomizing a highly viscous fluid can be difficult, and some industrial processes add the fluid as a steady un-atomized stream or even scoop in extremely viscous pastes. There is little benefit in atomizing these extremely viscous fluids, as although atomization may assist fluid distribution, mechanical dispersion will still control the overall fluid dispersion process and hence control the nuclei distribution.

Note that granulating in the mechanical dispersion regime implies that efficient mixing and agitation of the powder is required to achieve effective liquid dispersion and does not automatically mean “poor” liquid distribution. Most mixer granulation processes operate in the mechanical dispersion regime, unless they have been consciously designed to achieve drop controlled nucleation conditions (see Section 6.2.2.1 on spray flux scale-up). However, powder flow, circulation patterns, and shear forces in industrial granulation equipment are still poorly understood although there are several groups actively researching in this area [29–33]. Increasing the

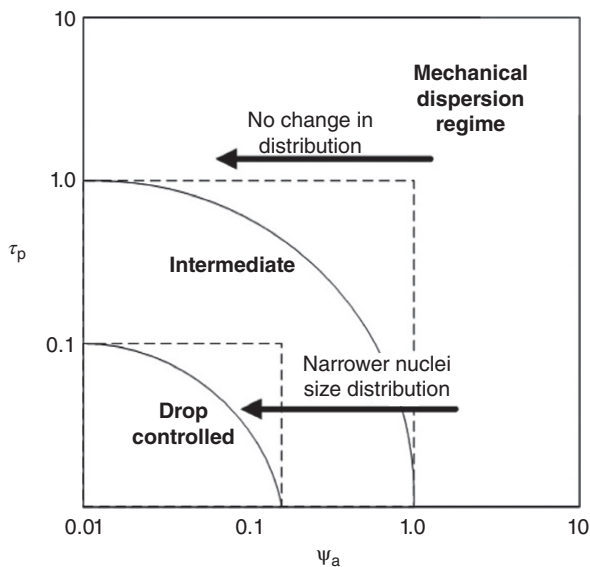


FIGURE 6.10 Nucleation regime map with adjusted boundaries incorporating the drop migration due to rewetting of powder during multiple passes through the spray zone [6]. Dotted lines represent original regime boundaries [11]. Source: Reproduced with permission from Wiley.

fluidizing air or impeller speed to improve agitation seems a simple solution, but commonly cannot be done, either due to equipment or process limitations (e.g. fixed speed impellers in mixers or maximum pressure drop limit in a fluid bed) or due to negative side effects on the granulation process (e.g. change in granule density and/or growth regime – see Section 6.2.1.1). One way to improve mechanical dispersion is to position the spray in a highly agitated region of the granulator. In a mixer granulator, this can be achieved by placing the spray nozzle either directly over the chopper [34] or shortly before the chopper so that the recently sprayed powder flows almost immediately into the turbulent chopper zone, rather than completing a 180–270° rotation before being agitated.

6.2.2 Consolidation and Growth

The second mechanism in wet granulation is granule growth (see Figure 6.1) and is inherently linked to both the liquid level and porosity of the granules. It is well known that different formulations show different growth behavior in the same equipment and that the same formulation can demonstrate changed growth behavior when different granulation conditions or equipment are used. We can think of two extreme models for the process of granule growth:

1. Growth of deformable porous granules: The starting force for granule growth and consolidation is a porous deformable nuclei formed by the processes described in Section 6.2.1.1. The nuclei contains substantial amounts of liquid in the pores but is not necessarily surface wet. This corresponds to a drop size in the spray zone larger than the primary particle size, and this model is a reasonable picture of granule growth in high-shear mixers.
2. Near elastic granules: Here we consider the granule to be a nearly elastic particle coated with a liquid binder after it leaves the spray zone. This model is well suited for processes where granules are dried before they reenter the spray zone and where the liquid drop size is small compared with the primary particle size. Thus, this model is often a reasonable description for fluidized bed granulation.

For both models, if we can answer the question “Will two colliding granules coalesce or rebound?” we will have gone a long way to describing the granule growth behavior. Granule growth regime maps summarize the causes of much of this behavior. Before, describing these models, we first discuss granule strength and deformation, as well as granule consolidation.

6.2.2.1 Granule Strength and Deformation One way to experimentally study granule strength is through the use of

wet granule pellets, which are small wet powder compacts with controlled size, porosity, and liquid content. Pellets are deformed under different conditions (in particular, strain rate $\dot{\epsilon}$), and the effects of many variables, including binder fluid viscosity and surface tension, particle size, pellet porosity, etc., can be investigated [35–38]. The stress versus time is recorded and the peak stress σ_{pk} indicates the point of failure and converted to a dimensionless peak flow stress Str^* [38]:

$$Str^* = \frac{\sigma_{pk} d_p}{\gamma_{LV} \cos \theta} \tag{6.11}$$

The Str^* data for all the experiments was plotted against the capillary number Ca , which is the ratio of viscous forces to surface tension forces, and proportional to the strain rate $\dot{\epsilon}$:

$$Ca = \frac{\mu d_p \dot{\epsilon}}{\gamma_{LV} \cos \theta} \tag{6.12}$$

Figure 6.11 shows this data for spheres. A single relationship can be formed with two distinct regimes [37, 38]. In region 1, the strain rate is low (low Ca) and the peak flow stress is independent of strain rate. In region 2, the higher strain rates applied (high Ca) mean that the viscous resistance forces begin to dominate and the peak flow stress is proportional to the strain rate.

Note that this approach confirms that at high enough values of Ca the peak flow stress (granule dynamic strength) is a strong function of strain rate. As collisions in the granulator are dynamic and collision velocities may vary from of

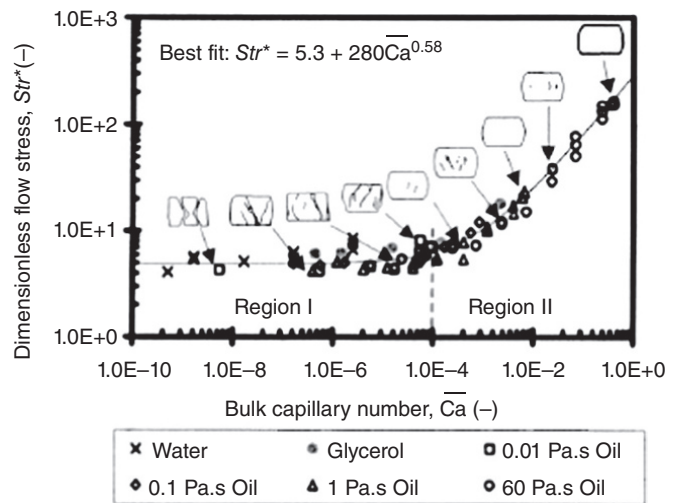


FIGURE 6.11 Dimensionless flow stress versus capillary number [37]. In region 1, the stress is independent of flow rate. At higher strain rates (region 2), viscous forces dominate and the stress is proportional to the strain rate. *Source:* From Iveson and Page [37]. Reproduced with permission from ASME.

order 0.1 ms^{-1} to of order 1 ms^{-1} , this is important. The granule strength should not be measured under static conditions, but rather at strain rates similar to those experienced in the granulator. Granules made from nonspherical primary particles follow a similar behavior to that illustrated in Figure 6.11 but are generally significantly stronger than granules made from spherical model particles. In Section 6.2.3.4, an extended version of the granule strength model is presented that includes effects of primary particle shape and granule liquid saturation.

6.2.2.2 Granule Consolidation A granule is a three-dimensional composite of solid particles, liquid bridges (which covert to solid bridges after drying), and vacant pore space occupied by air. Consolidation is the increase in granule density that results when the primary particles are forced to move closer to each other as a result of collisions between particles. Consolidation can only occur while the binder is still liquid. Consolidation determines the porosity and density of the final granules. Factors influencing the rate and degree of consolidation include particle size, size distribution, and binder viscosity as well as the impeller speed or fluidizing velocity [39–41].

The structure of granules, particularly the proportion and arrangement of the pore space, plays an important role in downstream processing, particularly compaction of the granules into a tablet, and in product performance, especially dissolution of the final solid dosage form [42].

The structure of real granules is complex and has not been able to be studied until the relatively recent development of micro X-ray tomography (XRT) [43–47]. Some examples of the structure of some pharmaceutical granules made in the same equipment from the same formulation under two different mixing conditions are shown in Figure 6.12, and this is an area of continuing research.

Although detailed analysis of the pore size distribution via mercury porosimetry [41, 48, 49] or XRT is possible [42–45], the overall average porosity of the granule has been found to be a very useful parameter in granulation. The overall porosity of a granule ϵ is defined as the volume fraction of air *within* a granule. Care is needed to avoid confusion between the interparticle voidage between granules and the intraparticle porosity of the granules (i.e. the internal porosity), both of which affect the bulk and tap density of an assembly of granules. The porosity of the granules generally begins at a high value ($\sim 50\text{--}60\%$) and decays exponentially as the granulation process proceeds [40] (see Figure 6.12), due to granule deformation and particle rearrangement as a result of collisions with other particles, granulator wall, and the impeller or chopper in a mixer granulator. The rate of densification depends on several parameters, but large particle size, smooth round particles, and low viscosity fluids allow rapid densification [40]. The final porosity reached by the granule, ϵ_{min} , is often used to determine the yield strength of the granule. The rate at which granules densify can be related to the granule peak flow stress and the typical collision velocity in the granulator through the Stokes deformation number St_{def} :

$$St_{\text{def}} = \frac{\rho_g U_c^2}{2Y_g} \quad (6.13)$$

where

ρ_g is the granule density,

U_c is the collision velocity, and

Y_g is the yield strength of the granule.

The yield strength of the granule is a function of the formulation and the extent of consolidation and is usually

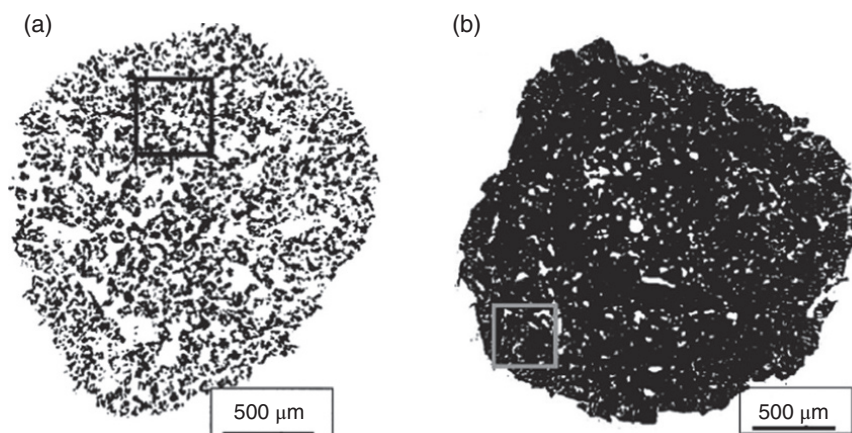


FIGURE 6.12 X-ray tomography images of internal structure of granules produced in a 2-l mixer at different shear conditions (a) 200 rpm impeller and 600 rpm chopper speed $\epsilon \sim 58\%$ (b) 600 rpm impeller and 1800 rpm chopper speed $\epsilon \sim 15\%$ [43]. *Source:* Reproduced with permission from Elsevier.

evaluated as the peak flow stress at the minimum porosity ϵ_{\min} (see Figure 6.12). The granulation consolidation rate constant k_c is then given by:

$$k_c = \beta_c \exp(a \cdot St_{\text{def}}) \quad (6.14)$$

where β_c and a are constant. k_c is the consolidation rate constant for a first order consolidation equation of the form:

$$\frac{\epsilon - \epsilon_{\min}}{\epsilon_0 - \epsilon_{\min}} = \exp(-k_c t) \quad (6.15)$$

Granule porosity is closely coupled with the granule saturation. The granule saturation s is defined as the proportion of pore space that is occupied by liquid:

$$s = \frac{w \rho_s (1 - \epsilon)}{\rho_l \epsilon} \quad (6.16)$$

where

- ϵ is the average granule porosity,
- w is the mass liquid/mass dry powder,
- ρ_l is the density of liquid, and
- ρ_s is the true density of the solid particles.

Four general saturation states have been defined – pendular, funicular, capillary, and droplet (Figure 6.13). There are two ways that the overall saturation of the granule can be increased – the amount of fluid added to the system can be increased, or the granule can be consolidated to reduce the pore space available [50, 51]. During the liquid addition phase of wet granulation, a combination of both processes is most likely occurring. For some formulations, consolidation of the granule will gradually decrease the porosity of the granule until the saturation reaches the droplet state, when the binder fluid will be squeezed to the exterior surface of the granule [24, 52–55]. The sudden presence of fluid at the outer granule surface often induces rapid coalescence and runaway growth of granules [40, 53, 56].

The rate of granule consolidation varies significantly, depending on the properties of the powder and liquid used [39, 40]. The interparticle friction must be overcome so that the granule can consolidate. Interparticle friction is increased by using smaller particle size, as their high surface area and high number of interparticle contacts provide more resistance to consolidation [39, 40, 57]. Granules formed from coarser particles tend to consolidate more quickly (see Figure 6.12). Viscous binder fluids also reduce the rate of consolidation, as the liquid resists flowing during deformation. However, the presence of fluid also acts as a lubricant and higher liquid content (i.e. saturation) causes the rate of consolidation to increase.

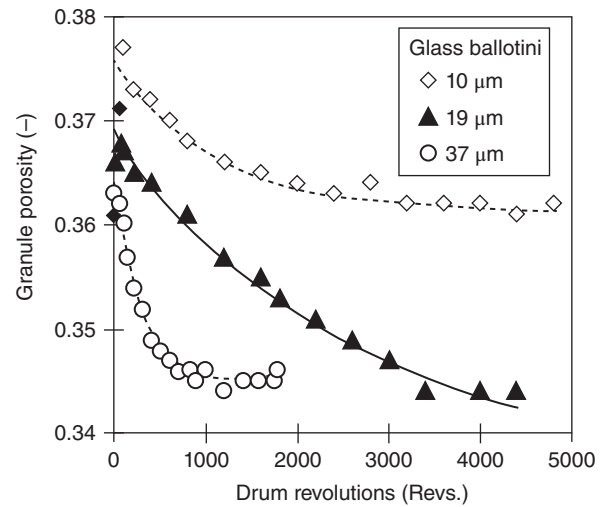


FIGURE 6.13 Exponential decay in granule porosity as granulation proceeds [40]. *Source:* Reproduced with permission from Elsevier.

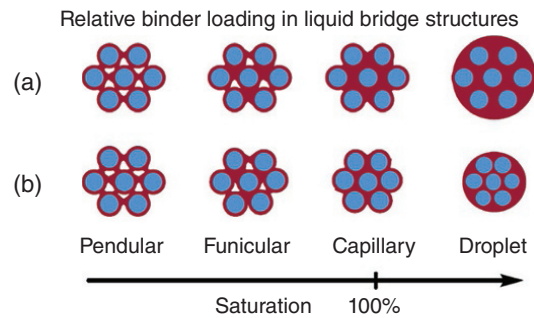


FIGURE 6.14 Granule saturation has four main states, and increases as either the liquid content is increased [50]. (a) Filling pores by binder addition (b) pore space reduction by consolidation. *Source:* Reproduced with permission from Elsevier.

6.2.2.3 Granule Growth Behavior for Deformable Porous Granules

There are two main types of granule growth behavior for porous, deformable granules – steady growth and induction growth (Figure 6.14). “Steady growth” occurs when the rate of growth is constant (at a given liquid level). On a plot of granule size versus granulation time, steady growth behavior produces a linear trend. This type of growth occurs in formulations where the granules are easily deformed by the forces in the granulator and tends to occur when using coarse powders and low viscosity fluids [52].

In contrast, “induction growth” occurs when the initial nuclei remain at a constant size for a long period, before very rapid granule growth occurs, resulting in a sudden increase in granule size (Figure 6.14). During the induction period, the granules consolidate and approach some minimum porosity (see Figure 6.12 for example data) but do not grow in size except through the layering of ungranulated fines.

Eventually, the granule porosity can be reduced enough to squeeze liquid to the surface. If there are no ungranulated fines remaining, this excess free liquid on the granules causes sudden rapid coalescence of many granules and results in the rapid increase in granule size characteristic of an induction formulation.

6.2.2.4 Granule Growth Regime Map for Deformable Granules It is well known that different formulations show different granulation behavior, such as induction versus steady growth (Figure 6.15). The links between formulation properties and granule growth behavior are summarized in the granule growth regime map [52, 54] (Figure 6.16). The

horizontal axis indicates the granule saturation s , which is a function of the weight fraction of liquid w and the granule porosity ϵ (see Eq. (6.13)). On the vertical axis is the Stokes deformation number, St_{def} , which is the ratio of the kinetic energy experienced by the granules during a collision compared to the yield stress or deformation of the granule (see Eq. (6.13)).

To make effective use of the granulation regime map, we need reasonable estimates of the effective collision velocity U_c (controlled by process conditions) and dynamic yield stress Y (a function of formulation properties). Table 6.1 gives estimates of the average and maximum collision velocities for different process equipment. In high-shear mixers,

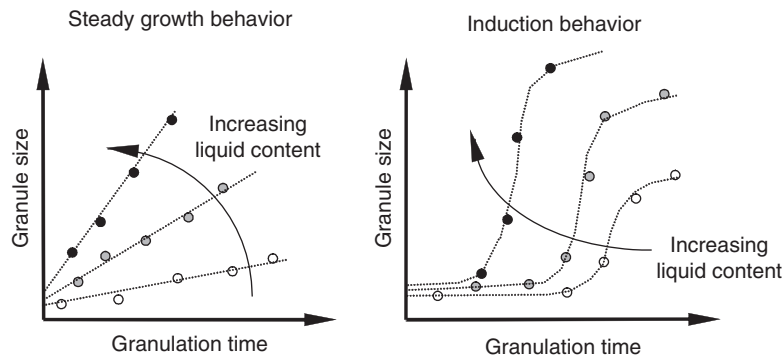


FIGURE 6.15 Two main types of granule growth behavior for deformable granules [52]. Source: Reproduced with permission from Wiley.

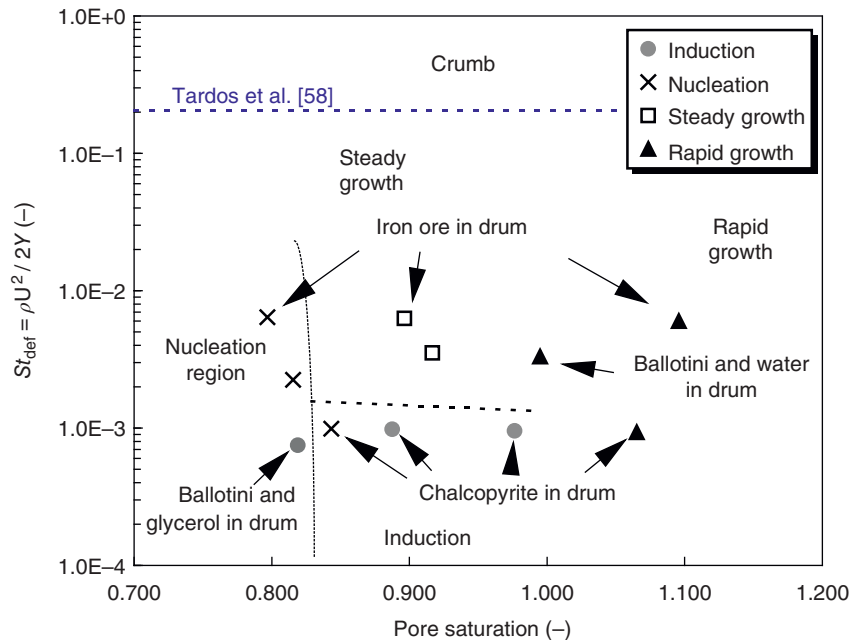


FIGURE 6.16 Granule growth regime map, summarizing the different types of granule growth and the conditions that they will occur [54]. Source: Reproduced with permission from Elsevier.

TABLE 6.1 Estimates of U_c for Different Granulation Processes

Type of Granulator	Average U_c	Maximum U_c
Fluidized beds	$\frac{6U_b d_p}{d_b}$	$\frac{6U_b d_p}{d_b \delta^2}$
Mixer granulators	$0.15\omega_r R$	$\omega_r R$

Source: Adapted from Litster and Ennis [59].

the difference between the average and maximum collision velocities can be very large and the estimates should be taken as indicative.

Let us consider the different growth regimes. First, we should consider a very weak formulation with a low granule yield strength Y_g . For example, for coarse sand and water at the very beginning of the granulation process, the saturation will be low and close to zero and the process in the “dry” free-flowing powder section of the regime map, in the upper left-hand corner. As more fluid is added to the process, the saturation increases and we move from left to right across the map. However, in this case the impact forces from collisions far exceed the granule strength, and the granules shatter as quickly as they are formed, forming a mixture reminiscent of damp sand. Increasing the amount of fluid does not help this, and eventually the system will end up as a slurry, in the upper right-hand corner of the map.

Consider now a formulation that has a slightly higher granule yield stress Y_g , where granules deform rapidly but are not shattered, for example, lactose granulated with water or a low viscosity binder such as PVP. The St_{def} for this type of formulation could be approximately $2/3$ up the St_{def} axis (depending on the granulator conditions, which determine U_c). Initially, at low liquid amounts, the system will be in the nucleation only regime. In this section of the growth map, there is enough liquid to form nuclei but not enough to allow any significant granule growth – the system is water limited. As more fluid is added, the saturation increases and the process crosses the regime boundary to enter the steady growth regime, where the granule size increases steadily with time. If we still add more fluid, the granule growth rate will increase further until we reach the rapid growth regime and eventually a slurry.

Finally, consider a formulation with a yield stress Y_g that is able to resist the impact forces experienced in the granulator. A typical pharmaceutical example would be a formulation containing a very fine powder granulated using a viscous binder fluid such as 5% HPC. After each collision, the granule consolidates slightly and over time approaches a minimum porosity as shown in Figure 6.12. Initially, when only a small amount of fluid has been added, the saturation will be low and the process will be in the nucleation regime. As the amount of fluid increases, we cross into the induction regime. During the induction period, the granules densify and

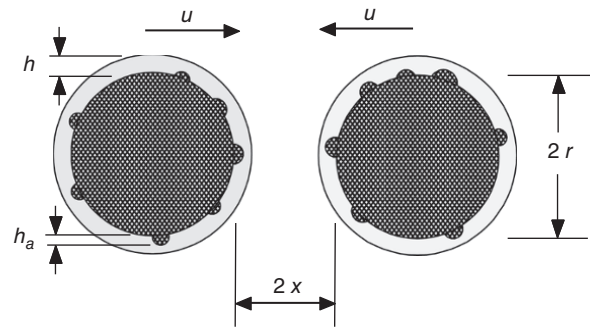


FIGURE 6.17 Two near-elastic granules colliding – the basis for the coalescence/rebound criteria [10]. Source: Adapted from Ennis et al. [60].

approach the minimum porosity. If the level of consolidation is enough for force the saturation to exceed 100% (see case (b) in Figure 6.14), the excess liquid, which is squeezed to the surface of the granule, will cause the granules to coalesce rapidly with the surrounding granules and extremely rapid granule growth occurs.

It is possible to switch the growth behavior from steady growth to induction growth or vice versa. For example, increasing the binder fluid viscosity or decreasing the particle size will slow the rate of consolidation and move the system toward induction behavior.

Pharmaceutical granulation mainly occurs in the nucleation, steady growth and induction regimes, as the final desired granule size is often only 2–4 times the size of the original particles in the formulation. A typical target granule diameter would be 200–400 μm , and the typical size of the particles in the formulation ranges between 50 and 200 μm .

6.2.2.5 Growth Regime Map for Nearly Elastic Granules

For near-elastic granules the conceptual model originally developed by Ennis et al. [60] considers the collision between two near-elastic granules each coated with a layer of liquid (see Figure 6.17). This work has been summarized in several monographs [59, 61, 62]. In this case the key dimensionless group is the viscous Stokes number St_v :

$$St_v = \frac{4\rho_g U_c d_p}{9\mu} \tag{6.17}$$

St_v is the ratio of the kinetic energy of the collision to the viscous dissipation in the liquid layer. Successful coalescence will occur if St_v exceeds some critical value St_v^* and we can define three growth regimes as follows:

1. *Non-inertial growth* [$St_{v,max} < St_v^*$]: The viscous Stokes number for all collisions in the granulator are less than the critical Stokes number. All collisions lead to sticking and growth by coalescence. In this regime,

changes to process parameters will have little or no effect on the probability of coalescence.

2. *Inertial growth* [$St_{v,av} < St^*$]: Some collisions cause coalescence, while others lead to rebound. There will be steady granule growth by coalescence. The extent and rate of growth will be sensitive to process parameters that will determine the proportion of collisions that lead to coalescence. Varying process parameters and formulation properties can push the system into either the non-inertial or coating regimes.
3. *Coating regime* [$St_{v,min} < St^*$]: The kinetic energy in most or all collisions exceeds viscous dissipation in the liquid layer. There is no coalescence. Granule growth will only occur by the successive layering of new material in the liquid phase (melt, solution, or slurry) onto the granule.

This simple model predicts that granules will grow to a maximum size by coalescence and such behavior is commonly seen in fluidized bed granulation. Note that the key formulation parameters are binder viscosity and particle size, as for deformable granule growth case.

6.2.3 Breakage

Of the three granulation mechanisms, granule breakage is the least understood. There is no fully general regime map for breakage although is an active research effort to develop one. A thorough review of breakage research was published recently [63]. Breakage of wet granules is only important in mixer granulators. Attrition and breakage of dry granules can also occur during fluidized bed drying (and during fluid bed granulation) or during later handling.

The two main approaches to understanding wet granule breakage are to conduct breakage studies of single granules (both wet and dry) and to conduct studies of breakage during granulation within the granulator and granular motion in granulators (including the forces and velocities experienced by the granules). It is important to understand how an individual granule will deform and break under certain conditions. Therefore single granules studies, both experimental and theoretical, can be very useful. Breakage during granulation is usually studied either by analyzing the change in granule size distribution with time or by using colored tracers.

Granule breakage is a function of the strength of the granule compared to the impact velocity and shear forces experienced by the granules within the granulator. Granules will respond differently under different conditions, and a given granule may break very differently, or not at all. Therefore granular flow has a large impact on the breakage behavior in the granulating system but is relatively poorly understood.

It is important to note that the breakage behavior of dry granules is completely different to the breakage behavior

of wet granules. We discuss here the breakage of wet granules during the granulation process, which may also be applicable to a wet milling process, but cannot be extrapolated to a dry milling process or any other dry granule breakage process. Dry granule breakage is discussed in detail elsewhere [59].

6.2.3.1 Deformation and Breakage of Single Granules

Deformation and breakage of single wet granules has been studied with high velocity impacts [64], in controlled powder shear [65], and in unconfined compression at varying strain rates [66]. Tardos and coworkers were one of the first groups to study breakage of single wet granules [65]. They performed experiments with individual pellets in a fluidized coquette device, where the shear field applied to the granules was known and carefully controlled. Breakage occurs when St_{def} is greater than some critical value St_{def}^* of the order 0.2. Here, the granule mechanics were modelled as a Herschel–Bulkley fluid and the granule strength is taken as the plastic yield stress under shear. Smith and coworkers [66, 67] extended Iveson's work on granule strength under uniaxial and diametrical compression [35, 37, 38, 68] and showed that in fact the deformation and breakage behavior of single granules is complex. Some formulations show very plastic behavior, while other fail in semi-brittle fashion, with propagation of single large cracks through the granule (see Figure 6.18). It is likely that plastic granules may fail by shear and extensional flow in the granulator, while semi-brittle granules may break by impact in the impeller zone.

While plastic behavior is more likely to be seen at higher strain rate, there is no general relationship between capillary number Ca and breakage mode. However, granules made from non-spherical particles are both much stronger, and more likely to fail in semi-brittle fashion by major crack propagation.

6.2.3.2 Breakage of Granules Within the Granulator: Effect of Formulation Properties and Process Parameters

Wet granule breakage has only been identified as a significant mechanism in high-shear granulators [26, 63, 66, 69–74]. The problem with fundamental studies using measurements in the granulator is that it is very difficult to deconvolute the effects of breakage from those of nucleation and growth simply by analyzing granule size distributions. To get around this problem, two approaches have been taken:

1. Using a small population of well-formed granules or pellets in a flow field of a cohesive, but non-granulating powder such as sticky sand [66, 70, 71].
2. Using marked tracer granules in a flow field of the same material [26, 69, 73–76].

Figure 6.19 shows examples of the breakage of well-formed pellets with time within a vertical axis mixer

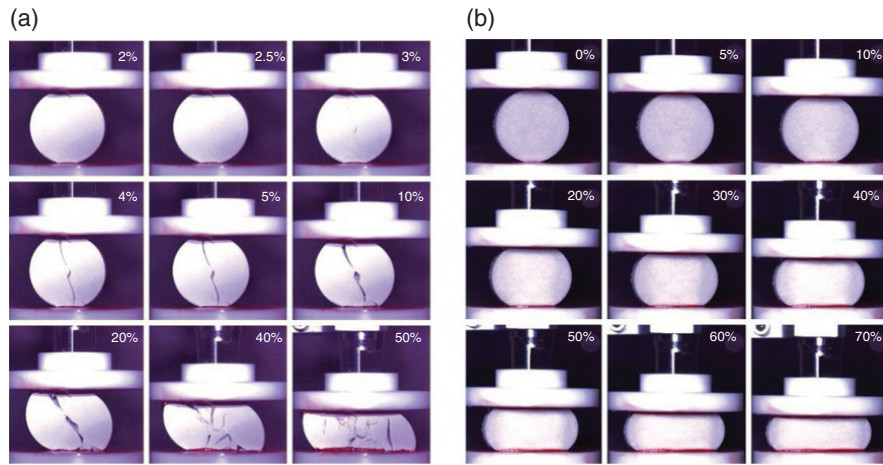


FIGURE 6.18 Extremes of wet granule deformation and breakage behavior in diametrical compression [67]. (a) Semi brittle behavior (b) plastic deformation. *Source:* Reproduced with permission from Elsevier.

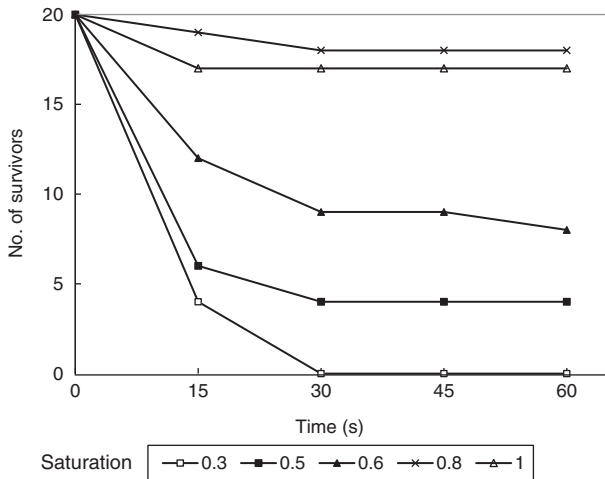


FIGURE 6.19 Breakage of wet granule pellets in a flow field of cohesive sand in a vertical axis mixer granulator (broad size distribution lactose with 1 Pa·s silicone oil binder at different liquid saturation values) [71]. *Source:* Reproduced with permission from Elsevier.

granulator [71]. In this case, pellet survival is a strong function of liquid saturation. In general, strong granules (as measured by peak flow stress in uniaxial compression) do not break as readily in the granulator. We can postulate that in the granulator, granules will break if they experience stresses that exceed their peak flow stress. We can express this as a Stokes deformation number criterion. Breakage will occur if the Stokes deformation number on impact exceeds a critical value [71]:

$$St_{def} = \frac{\rho_g v_c^2}{2\sigma_p} > St^* \quad (6.18)$$

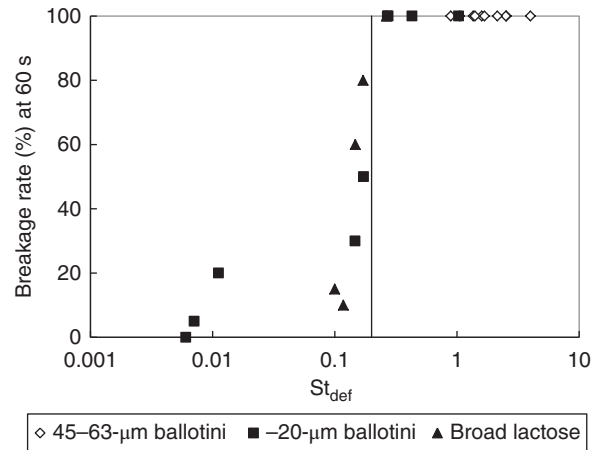


FIGURE 6.20 The extent of breakage at steady state versus Stokes deformation number St_{def} for all the formulations. The vertical line is the St_{def} value of 0.2 [71]. *Source:* Reproduced with permission from Elsevier.

This approach is similar to that used by Tardos and co-workers in the coquette flow rheometer [65]. Figure 6.20 shows that treating breakage data in this way leads to a surprisingly sharp transition from no breakage to breakage at $St^* = 0.2$ for a wide range of formulations.

In similar studies, van der Vries et al. [69] proposed a critical value of $St^* = 0.01$. Closer inspection of the analysis shows that differences in value for St^* are most likely due to differences in mixer geometry and in measuring or estimating the granule strength σ_p and the collision velocity v_c . In their work, Liu et al. [71] combined data from single granule strength measurements [66] with Rumpf's expression for granule strength where both viscous and capillary forces are important, to develop the following expression for granule strength:

$$\sigma_p = AR^{-4.3} S \left[6 \frac{1-\varepsilon}{\varepsilon} \frac{\gamma \cos \theta}{d_{3,2}} + \frac{9(1-\varepsilon)^2}{8 \varepsilon^2} \frac{9\pi\mu v_p}{16d_{3,2}} \right] \quad (6.19)$$

where

AR is the aspect ratio of the primary particles,

S is the granule pore saturation,

ε is the porosity of the granule,

$d_{3,2}$ is the specific surface area diameter of the particles, and

v_p is the relative velocity of the moving particle inside a granule after impact.

The expression explicitly accounts for the effects of primary particle size and shape, liquid binder properties, and liquid saturation. The collision velocity was assumed to 15% of the impeller tip speed. Note this correlation is similar to, although not identical to, the Iveson correlation for granule strength (see Section 6.2.1.2). Both predict granule strength increasing with increasing strain rate, increasing binder viscosity, and decreasing primary particle size.

Note that the mode of breakage as well as the granule strength may vary with formulation properties. Analysis of granule fragments from breakage shows evidence of significant elongation for some formulations (plastic deformation) and sharp angular fragments for others (semi-brittle behavior) [70]. Breakage behavior is similar to that predicted from diametrical compression of single granules.

While this relatively simple approach is fairly general but able to predict the effect of formulation parameters, predicting the effect of changes in equipment parameters is more difficult. In general, increasing impeller speed increases the extent of breakage [26, 72, 74] and this is accounted for by the collision velocity term in the Stokes deformation number. However, breakage is very sensitive to changes in mixer and impeller geometry in ways we cannot yet predict a priori. Figure 6.21 shows an example of breakage of single drop formed granules in the same mixer with (Figure 6.21a) a two-blade beveled edge impeller and (Figure 6.21b) a frictional flat plate at the same impeller speed. For the beveled impeller, there is a reasonable correlation between breakage and granule strength consistent with Eq. (6.19) above. (The degree of scatter is due to the difficulty in keeping granule porosity and liquid saturation constant with this method of granule formation.) In contrast, there is *no* breakage using the frictional flat plate.

These results emphasize that breakage is not occurring uniformly in the bed, but rather in a narrow zone near the impeller. This is the reason that granules take time to break in the granulator. Figure 6.19 shows that the pellet granules take up to one minute to break, even in a small granulator. Figure 6.22 shows an example of velocity fields in roping flow in a two-blade impeller with a vertical axis mixer. Powder velocity (and therefore applied stress) are very nonuniform with the highest velocities in a small zone near the impeller. As granules circulate in roping flow in the granulator, they will often bypass the impeller zone. Only those granules that enter this zone of high impact and shear stress are

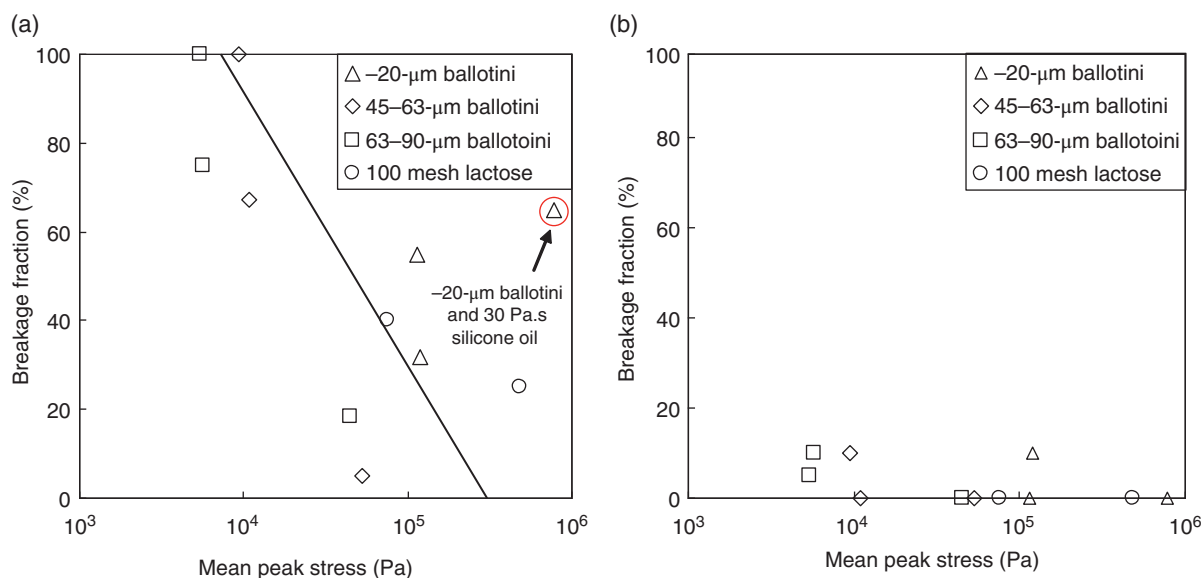


FIGURE 6.21 Breakage fraction of drop-formed granules as a function of granule mean peak stress: (a) 11° bevelled blade at 500 rpm; (b) frictional flat plate impeller at 500 rpm. *Source:* Unpublished data related to Ref. [70].

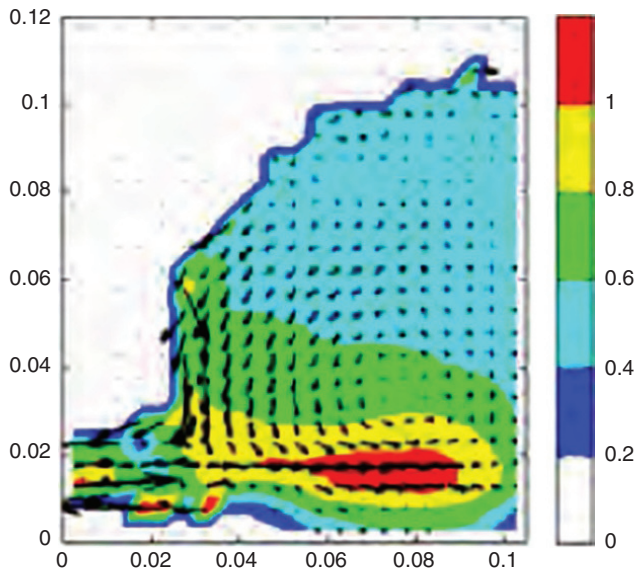


FIGURE 6.22 Powder velocity field during pseudo-rope flow in as two-blade impeller vertical axis mixer measured using positron emission particle tracking (PEPT) [77].

likely to break. The size of this breakage zone, and the maximum stresses and collision velocities seen by the granules in this zone will be a strong function of the impeller design.

In mixer granulators, a “chopper” mounted either in the side wall or in the granulator lid, rotates at high speed (e.g. 3000 rpm). This generates a small localized area with very intensive agitation, where breakage could be expected to dominate. The chopper is commonly thought to break up large lumps and granules, particularly at the powder surface (where they are generally located). Although this seems logical, there is no work to demonstrate this – the few pharmaceutical studies that have been performed [48, 78–80] find that the chopper had a very small effect and that the overall granulation response (e.g. granule size and porosity) is dominated by the impeller speed, liquid level, etc.

6.2.3.3 Breakage of Granules Within the Granulator: Effect of Granule Size and Density There are several studies which show that the breakage probability is proportional to granule size. [71, 73, 81]. These studies are consistent with the literature on particle size effects on crushing and grinding. Larger granules are more likely to have large pores or flaws, which increases their probability of semi-brittle fracture [81]. Consolidated, dense granules are well known to be stronger. An analogy between the granule growth regime map and breakage behavior has been postulated, where three “exchange mechanisms” have been proposed to resemble three growth regimes, and linked to the type and rate of material transfer between granules [76].

Colored tracer granules have been used to follow breakage as a function of granule size. Three sizes of colored tracer

granules (~200, 500, and 1000 μm) were added to a granulator while it was running. Samples were taken as a function of time, and the proportion of colored material in each size fraction was analyzed. The largest granules break at a higher rate than smaller granules [73]. They also investigated the breakage rate of tracer granules that had been granulated for different times, before being removed and added to the running granulator. Younger, newly formed tracer granules broke at a faster rate than tracer granules that had been granulated for a longer period, allowing plenty of time for granule consolidation [73]. Granule strength and hence granule breakage rates have been shown to be quite heterogeneous [75]. In some cases, the color distribution of the granules has become almost uniform within approximately two minutes [26, 69, 73] although this will not be true for all formulations or all granulation conditions.

6.2.3.4 Aiding Controlled Granulation via Breakage Breakage of wet granules during the granulation process is not necessarily detrimental. In pharmaceutical granulation, the formation of large granules is generally undesired, and size-preferential breakage of coarse granules [73] helps keep the proportion of coarse granules low.

Breakage can also occur early in the granulation process, in parallel with nucleation, and can assist in distributing the liquid evenly throughout the powder. The mechanical dispersion nucleation regime (see Section 6.2.1.3) requires breakage to disperse the wet clumps of binder fluid through the powder. Newly formed granules (i.e. nuclei) are easier to break than older granules, due to their relatively high porosity [73]. This mechanism of liquid dispersion via breakage of nuclei is called “destructive nucleation” [26] (Figure 6.23). The initial interaction of the drop either in a fluid bed (via layering of the powder on the exterior of the drop) or during the drop penetration process (more relevant to mixer granulation) forms a primary nuclei with a saturation gradient – the saturation decreases as you move from the inner core of the granule to the exterior surface. The large, low porosity, weak primary nuclei is broken into smaller and stronger secondary nuclei, which form starting materials for coalescence [26]. Tracer studies showed that the proportion of primary nuclei that survive decreases as the impeller speed increases [26, 69], and consequently the colored tracer fluid became more uniformly distributed both between granules and within granules. This effect may be smaller for other formulations and equipment.

The final stage in destructive nucleation shows a balance between coalescence and breakage, which implies a stable maximum granule size. This idea was applied, together with several other ideas, to produce a well-controlled granulation, which required only a single parameter – the impeller speed – to be adjusted on scale-up [82]. Note that because of the kinetics of breakage in time scale to a typical pharmaceutical granulation (one to five minutes), this “steady-state”

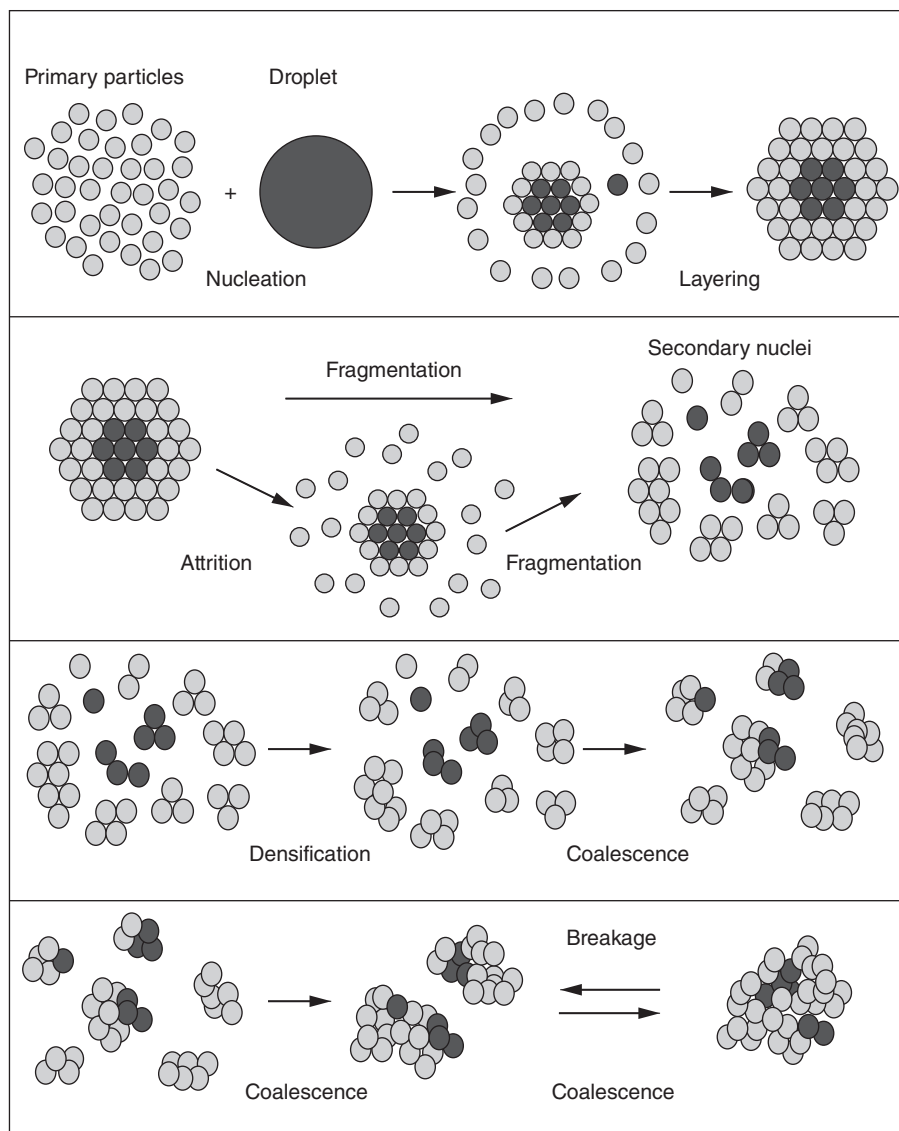


FIGURE 6.23 “Destructive nucleation” where nuclei undergo breakage [26]. *Source:* Reproduced with permission from Elsevier.

approach will require longer granulation times for stable results. This work is described in more detail in Section 6.2.2.4.

6.2.3.5 Summary Comments While much is still to be done in the area of wet granule breakage, we can draw some useful conclusions:

1. Approaches used to measure or estimate the effect of formulation properties on granule strength for the purposes of understanding granule growth are also applicable to breakage studies. Remember, however, that changing the formulation can change the mode of breakage (plastic or semi-brittle) as well as the yield stress.
2. In the granulator, granules will break due to some combination of shear and impact in a relatively small region near the impellers. Granulator and impeller geometry have a very strong, but difficult to predict, effect on the rate and extent of breakage.
3. A simple Stokes deformation number criterion (Eq. 6.18) can be used to predict whether breakage will occur in a granulator of well-defined geometry and to predict the effect of formulation effects and impeller speed on wet granule breakage in the granulator.
4. Granule breakage will occur more easily for large, low density granules.
5. Breakage can sometimes have a positive effect in wet granulation to aid liquid distribution and limit maximum granule size.

6.3 SCALE-UP

6.3.1 Spray Flux Scale-up

Dimensionless spray flux provides a good basis for scale-up of the spray zone conditions to maintain good nucleation and equivalent liquid distribution conditions. However, maintaining constant spray flux on scale-up can be quite challenging (see also discussion in Ref. [50]). The most common scale-up approach is to maintain the same solution addition time, which means that the flow rate through the nozzle increases and therefore the spray flux also increases dramatically. The larger spray width and larger drop sizes at higher flow rates are not usually enough to compensate for this. An alternate approach of maintaining the same flow rate (and drop size, if an identical nozzle tip), but this causes the total batch time to increase in proportion to the batch size. Assuming that liquid level is high enough to produce a granule saturation higher than the nucleation regime limit [52, 54], then the growth and consolidation will be affected due to the longer processing time.

In addition, the powder surface velocity in a mixer granulator generally decreases as the granulator scale increases [21], particularly if the impeller has only two fixed “high” and “low” settings, which are usually scaled to maintain equivalent impeller tip speed, rather than equivalent powder agitation and mixing. Lab-scale mixer granulators display vigorous mixing and tend to operate in the “roping” regime, while at manufacturing scale the powder agitation is much less efficient and spends at least part of the time in “bumping” flow [12, 21, 33]. Increasing the impeller speed to maintain an equivalent powder surface velocity also increases the impact and shear forces experienced by the granules. The higher force increases the consolidation rate and shifts the system upward on the granule growth map (i.e. St_{def} increases). If this upward shift results in the system crossing a regime boundary, a fundamental shift in granulation behavior will occur as the system moves into a different granulation regime.

Table 6.2 below gives an example of spray flux changes during scale-up from a 10 kg to a 50 kg batch, where 25%

TABLE 6.2 The Effect of Scale-up on Spray Flux Ψ_a in Fielder Mixer Granulator

Scale-Up Approach	Base Case	Constant Spray Time	Constant Spray Rate
Batch size (kg)	10	50	50
Flow rate (l/min)	0.5	2.5	0.5
Spray time (min)	5	5	25
Drop size (μm)	200	350	200
Spray width (m)	0.25	0.3	0.3
Impeller speed (rpm)	216	108	108
Powder velocity (m/s)	0.7	0.55	0.55
Spray flux Ψ_a	0.36	1.08	0.38

of granulating fluid is added at 0.5 l/min over five minutes. Initially, the calculated spray flux $\Psi_a = 0.36$, which is above the drop-controlled nucleation regime. In the first approach, the spray rate increases to 2.5 l/min to maintain the equivalent granulation time of five minutes and the impeller speed of 108 rpm is scaled using the common “tip speed” scaling rule [21]. The decrease in powder velocity and the increase in spray rate creates an increase in spray flux to $\Psi_a = 1.08$. In the second case, the spray rate is maintained at 0.5 l/min, but the spray delivery time is increased from 5 to 25 minutes. This is not commonly done, due to fears of significant changes in granule growth and consolidation, which may be unfounded (see Section 6.3.4). In this case, the spray flux remains almost constant at $\Psi_a = 0.38$. The spray flux could be reduced further by increasing the impeller speed or by adding multiple nozzles.

Multiple nozzles are the only way to maintain spray flux independently of granule growth and consolidation rates. This is well known in fluid bed granulation scale-up [83–85] but has not been applied to commercial mixer granulators. A four nozzle spray manifold was designed for a 300 l mixer granulator [86]. The increased liquid distribution area and reduced the flow rate per nozzle reduced the spray flux and can result in reduced lump formation [86]. An alternate method to maximize liquid dispersion in the spray zone is to place the spray nozzle directly over the chopper, where the turbulent powder flow and strong localized shear forces disperse the fluid effectively, even in the mechanical dispersion regime [34].

6.3.2 Scale-up of Powder Flow Patterns in Mixer Granulators

Mixer granulators are often called “high-shear” granulators, and until recently it was assumed that the impeller was able to effectively agitate the powder bed during operation. Powder velocity measurements can be performed using high-speed video cameras and image analysis [12, 21, 22], which generally measures the tangential component of the powder velocity, although the velocity also varies radially. More sophisticated analysis using positron emission particle tracking (PEPT) technology has also been performed in several mixers [31, 87–90] and shows these trends in much greater detail although the data currently available is limited to lab-scale granulators. Experimental PEPT data for powder flow in larger-scale equipment is currently being generated.

Two distinct types of powder flow have been observed in high-shear mixer granulators [12]. At low impeller speeds, the powder surface remains horizontal and the bed “bumps” or “shunts” [91] up and down as the impeller passed underneath. The surface velocity was approximately an order of magnitude lower than the impeller tip velocity. As the impeller speed increases, the powder surface velocity increases linearly although there was still little vertical interchange

of material. After increasing the impeller speeds above a critical point, a vortex appears and spiraling “roping” [12] or “torroidal” powder flow is observed [74]. In the roping regime, the surface velocity is independent of impeller speed [12, 74] and material from the bottom of the powder bed is forced up the vessel wall before tumbling down the vortex in the center of the powder flow. In all observed powder flows in mixer granulators, the powder velocity is at least one order of magnitude lower than the impeller tip speed [10, 21, 74, 89].

An industrial study of powder flow patterns and surface velocity was performed in a series of Fielder mixers (25 l, 65 l, and 300 l) as a function of impeller speed [21]. When running the 300 l granulator at the standard “low” speed setting of 180 rpm, the powder was stagnant approximately 1/3 of the time [21]. Changes in mixer geometry and fill level could significantly change the powder velocity at a given impeller speed. They also measured powder velocity during granulation and found that the surface velocity gradually increased as granulation proceeded. The powder flow pattern also changed during granulation, shifting from bumping flow during dry mixing to roping flow during granulation [21]. Powder velocity during the dry mix stage has also been shown to vary between lots of API for a high drug load formulation [22]. The measured powder surface velocity for three batches containing different lots of the API varied between 0.64 and 0.95 m/s. This variation was presumed to be due to lot-to-lot differences in drug properties, although establishing the causal link between drug properties and powder flow is an area requiring further investigation.

In fluidized beds, the fluidizing airflow is always adjusted to maintain adequate fluidization of the powders. It would be unthinkable to attempt to scale-up a fluidized bed and select a set of operating conditions that did not fully fluidize the bed – yet we routinely scale-up mixer granulators in exactly this way. There are three main approaches to scaling powder flow in a mixer granulator that can be summarized by the following equation [33]:

$$ND^n = \text{constant} \quad (6.20)$$

where

N is the impeller speed (rpm),
 D is the impeller diameter (m), and
 n is a scaling index.

The most common impeller scaling approach is to maintain tip speed, where $n = 0.5$ [47, 92–94]. An alternate approach is to use Froude number ($Fr = N^2D/g$), which is commonly used to scale up fluid mixing by maintaining the ratio of centrifugal to gravitational forces [12, 92, 95, 96]. In this case, the scaling index in Eq. (6.13) is $n = 1$. More recent work used calibrated tracer pellets with a known yield

stress to measure the average shear stress experienced by a granule during granulation. Scale-up studies showed that for a series of geometrically similar Fielder granulators, the scaling index n varied between $n = 0.8$ and 0.85 [33] depending on the height to diameter ratio, fill level, and impeller style used in each case. The “equal shear” scale-up criterion is the subject of ongoing research [97, 98].

In addition to the impeller speed criteria outlined above, other scaling criteria include maintaining swept volume [80], constant energy per unit mass [99], and power number [100].

6.3.3 Granule Growth Scale-up

It is important to stay within the same granule growth regime (see Figure 6.15) during scale-up to avoid dramatic changes in growth behavior. This is unfortunately easier said than done. The forces applied to the powder mass must remain similar; otherwise the system may shift vertically on the regime map, typically into or out of the induction growth regime. For fluid beds this implies maintaining similar fluidizing conditions and maintaining an equivalent excess gas velocity. For mixer granulators the maximum impact (e.g. impeller and/or chopper tip speed) may need to be maintained if the granulation is controlled by direct impeller impacts, or perhaps the overall roping flow field to maintain equivalent shear. In some cases, these requirements may directly contradict the requirements needed to maintain equivalent nucleation conditions.

Shifting vertically on the regime map due to a change in the overall force applied to the granulation also implies that the granule porosity and/or structure will also shift, and this is usually undesirable as granule porosity is often shown to be directly linked to the dissolution rate of the granules, capsules, or tablets [42, 101]. Changing the porosity also changes the overall saturation of the granules, which also means that a change in the granule porosity moves the process both vertically and horizontally on the growth regime map. A change in porosity ϵ and the subsequent change in granule saturation s mean that the granule size will change, even though the amount of liquid added to the batch (w in Eq. (6.11)) remains constant. Typically, pharmaceutical process engineers concentrate on keeping the amount of liquid added to the batch constant. In fact, the granule saturation is the key factor in controlling granule growth, but in the future we hope to see saturation being calculated at each stage of the scale-up process. Currently, the lack of knowledge of porosity changes means that the amount of liquid added to the batch is frequently adjusted from compensate for the changes in growth after scale-up.

6.3.4 Scale-up Case Study: Steady-State Granulation

“Pseudo-steady-state granulation” (also called one-dimensional granulation) is a recent approach to wet

granulation, which resulted in improved control of granule size, properties, and scale-up without any loss in product performance [82]. During a typical 5–25 minutes pharmaceutical granulation process, there are multiple dynamic sub-processes occurring including: liquid distribution; dissolution and hydration of excipients such as lactose, MCC, and dry binders; granule growth; granule consolidation; granule breakage; and the overall granulator flow pattern and shear (particularly the transition between bumping and roping flow). Generally when the batch is stopped, each of these sub-processes is stopped abruptly well before an equilibrium is reached. Each sub-process has its own characteristic time scale, and it is *impossible* to halt in *all* of these dynamic sub-processes at the *same* point at each scale, and this is why the granulation performance shifts during scale-up.

An ideal granulation process would allow complete control of liquid sizes in the granulator, without the need for dry milling [82]. The ideal process would allow all the transient sub-processes to reach a repeatable, controllable equilibrium end point and produce a narrow size distribution of granules between 200 and 500 μm with complete control of the granule size by adjusting only the liquid level. This would involve operating at a low spray flux by using a slow spray rate, using a long granulation time to ensure that all complete dissolution/hydration of all the excipients, ensuring roping flow behavior during the entire process, and using some wet massing time to ensure that an equilibrium can be achieved. Vonk et al. [26] showed that an equilibrium granule size and saturation should exist where the rate of granule growth is exactly

balanced by the rate of granule breakage (see lower diagram in Figure 6.18). As the entire granulation batch converges toward the equilibrium point, the granule size distribution and saturation distribution will also converge. This “steady-state” granulation point should also be scale independent [82].

To demonstrate steady-state granulation, Michaels et al. [82] granulated a standard lactose-MCC-based formulation in a 21 high-shear mixer granulators using “conventional” conditions (40% fluid sprayed over five minutes) and “steady-state” conditions (28% fluid sprayed over 15 minutes plus up to 20 minute wet massing time). The long granulation time caused the initial batches to heat up, creating new transients in evaporation and rheology, so the standard mixer cooling jackets were used to minimize the temperature rise in the batch. The granules produced by the steady-state process were typically 200–300 μm with a narrow size distribution with no granules larger than 1 mm. The final particle size distribution was a function only of the final liquid saturation and shear stress in the agitated wet mass. The granulating fluid level (ratio of liquid added to dry powder ingredients) became a material variable rather than a process variable, i.e. scale independent. In contrast, the conventional granules had a very broad distribution with more than a third of the granules above 1 mm, thus requiring the use of a dry mill.

Scale-up of the steady-state granulation process involves only one process variable: scale-up of the shear stress, controlled by the main impeller speed. Scale-up trials were

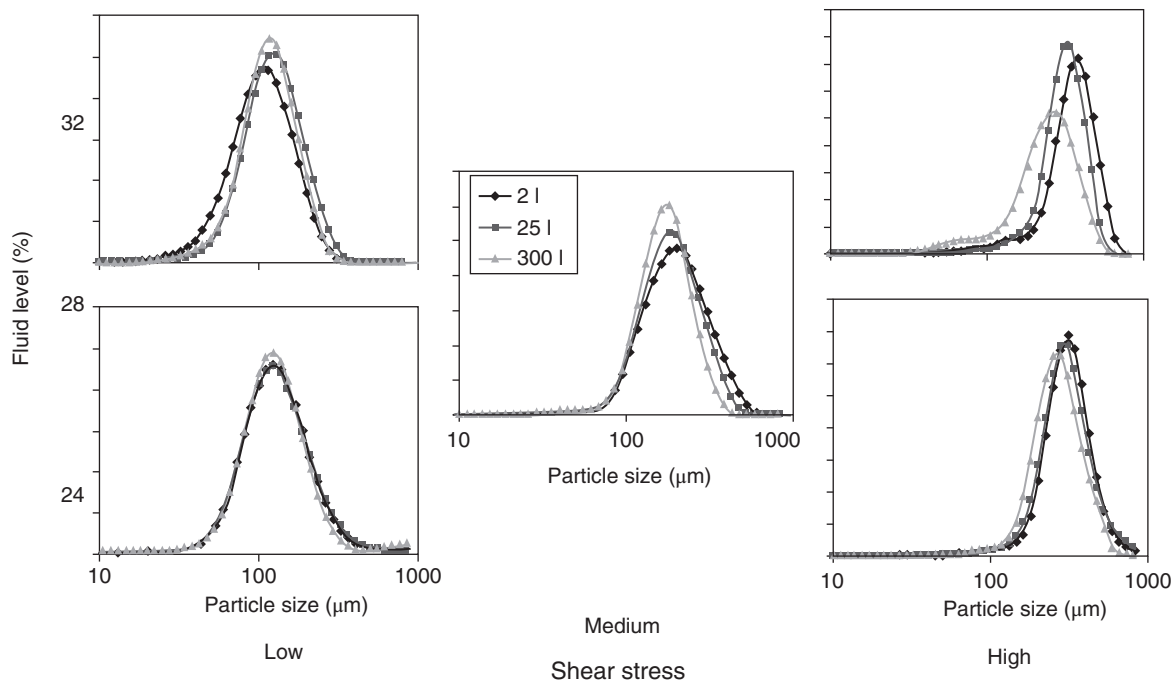


FIGURE 6.24 Particle size of dry, unmilled granules manufactured at 2, 25, and 300 l scale. Data shown are for three fluid levels (24, 28, and 32%) and three shear stresses (low, medium, and high impeller speeds).

conducted at 2, 25, and 300 l scale, using the “equal shear” scale-up correlation ($n = 0.8$ in Eq. (6.15)) between shear stress and main impeller speed [33]. The granule size distributions were unimodal and centered at 200–300 μm depending on the liquid level, and the need for a milling step to control the granule size was eliminated [82]. The entire size distribution was matched “right first time” with only a single batch performed at 25 and 300 l (see Figure 6.24). This exact matching of the entire size distribution was repeated multiple times, over three different liquid levels (24, 28, and 32%) and at two different impeller speed (shown as low, medium, and high shear on the x -axis) at each scale. Although the total granulation process time was far longer than normal (15 minutes solution delivery plus 20 minutes wet massing time), the lower liquid level (~28% compared to 45% for the same formulation granulated using a standard approach) and the small, uniformly size granules meant that the drying time was significantly shorter, resulting in the same overall cycle time at production scale. In addition, the dissolution and compaction of these granules was unaffected, contrary to conventional opinion that such dense granules would fail to meet physical tablet specifications and drug release specifications. This highlights the surprisingly simple opportunity of using the steady-state granulation approach to scale up pharmaceutical formulations. The approach has also been successfully used for granulation of detergents in a fluid bed granulator [102].

6.4 FUTURE DIRECTIONS

There are several areas where the existing knowledge of granulation is currently insufficient – powder flow including the shear rates and collision rates experienced by the granules is one area, and a better understanding of the granule breakage mechanism is another. Both of these areas are the focus of several international groups. Areas particularly relevant to the pharmaceutical industry are the granulation behavior of multicomponent powder blends and understanding how to obtain the most uniform drug distribution possible across all granule size fractions. The behavior of hydrophobic drugs in granulation can also be quite surprising [46, 103] and warrant further effort.

Eventually, the granulation process – i.e. operating conditions for a given formulation to produce granules with a pre-specified set of properties – will be designed entirely in advance using dimensionless groups and regime maps. Selection of robust process conditions based solely on theoretical considerations has already been demonstrated [82]. The existing knowledge provides valuable guidance for troubleshooting process problem and estimating the process risk of atypical events.

Pharmaceutical granulation is also evolving – continuous granulation is now under serious development [104, 105]

and is clearly a strong future direction of pharmaceutical granulation. Continuous granulation will also require a new effort to develop an online process control and analytical technology – particularly for granule size and porosity. The technology for online and at-line granule size measurement already exists but is currently used only sparingly [82, 106], in part due to industry conservatism about applying a novel particle size measurement. The development of real process control will most likely also impact on more traditional batch granulation. Other improvements, including the use of foam to distribute the binder fluid instead of a spray [107–109], are also likely to expand the range of process options available during pharmaceutical wet granulation. The new engineering-focused approach to pharmaceutical granulation is leading to vast leaps in our understanding of the granulation mechanisms, and the future direction is the gradual merging of granulation science and industrial know-how.

REFERENCES

- Iveson, S.M., Litster, J.D., Hapgood, K.P., and Ennis, B.J. (2001). *Nucleation, growth and breakage phenomena in agitated wet granulation processes: a review. Powder Technology* 117 (1–2): 3–39.
- Knight, P.C., Instone, T., Pearson, J.M.K., and Hounslow, M.J. (1998). *An investigation into the kinetics of liquid distribution and growth in high shear mixer agglomeration. Powder Technology* 97: 246–257.
- Schäfer, T. and Mathiesen, C. (1996). *Melt pelletization in a high shear mixer IX. Effects of binder particle size. International Journal of Pharmaceutics* 139: 139–148.
- Schäfer, T. and Mathiesen, C. (1996). *Melt pelletization in a high shear mixer VIII. Effects of binder viscosity. International Journal of Pharmaceutics* 139: 125–128.
- Scott, A.C., Hounslow, M.J., and Instone, T. (1999). *Direct evidence of heterogeneity during high-shear granulation. Powder Technology* 113: 215–213.
- Hapgood, K.P., Nguyen, T.H., Hauw, S. et al. (2009). *Rewetting effects and droplet motion on partially wetted powder surfaces. AIChE Journal* 55 (6): 1402–1415.
- Chouk, V., Reynolds, G., Hounslow, M., and Salman, A. (2009). *Single drop behaviour in a high shear granulator. Powder Technology* 189 (2): 357–364.
- Schaafsma, S.H., Vonk, P., Kossen, N.W.F., and Hoffmann, A. C. (2006). *A model for the spray zone in early-stage fluidized bed granulation. AIChE Journal* 52 (8): 2736–2741.
- Hapgood, K.P., Litster, J.D., Biggs, S.R., and Howes, T. (2002). *Drop penetration into porous powder beds. Journal of Colloid and Interface Science* 253 (2): 353–366.
- Litster, J.D., Hapgood, K.P., Michaels, J.N. et al. (2001). *Liquid distribution in wet granulation: dimensionless spray flux. Powder Technology* 114 (1–3): 32–39.

11. Hapgood, K.P., Litster, J.D., and Smith, R. (2003). *Nucleation regime map for liquid bound granules*. *AIChE Journal* 49 (2): 350–361.
12. Litster, J.D., Hapgood, K.P., Michaels, J.N. et al. (2002). *Scale-up of mixer granulators for effective liquid distribution*. *Powder Technology* 124 (3): 272–280.
13. Hapgood, K.P., Litster, J.D., White, E.T. et al. (2004). *Dimensionless spray flux in wet granulation: Monte-Carlo simulations and experimental validation*. *Powder Technology* 141 (1–2): 20–30.
14. Wildeboer, W.J., Litster, J.D., and Cameron, I.T. (2005). *Modelling nucleation in wet granulation*. *Chemical Engineering Science* 60: 3751–3761.
15. Denesuk, M., Smith, G.L., Zelinski, B.J.J. et al. (1993). *Capillary penetration of liquid droplets into porous materials*. *Journal of Colloid and Interface Science* 158: 114–120.
16. Middleman, S. (1995). *Modelling Axisymmetric Flows: Dynamics of Films, Jets, and Drops*, 299. San Diego: Academic Press.
17. Ax, K., Feise, H., Sochon, R. et al. (2008). *Influence of liquid binder dispersion on agglomeration in an intensive mixer*. *Powder Technology* 179 (3): 190–194.
18. Hapgood, K.P. (2000). *Nucleation and binder dispersion in wet granulation*. PhD thesis. University of Queensland, Brisbane.
19. Schaafsma, S.H., Vonk, P., Segers, P., and Kossen, N.W.F. (1998). *Description of agglomerate growth*. *Powder Technology* 97: 183–190.
20. Waldie, B. (1991). *Growth mechanism and the dependence of granule size on drop size in fluidised bed granulation*. *Chemical Engineering Science* 46 (11): 2781–2785.
21. Plank, R., Diehl, B., Grinstead, H., and Zega, J. (2003). *Quantifying liquid coverage and powder flux in high-shear granulators*. *Powder Technology* 134 (3): 223–234.
22. Hapgood, K.P., Amelia, R., Zaman, M.B. et al. (2009). *Improving liquid distribution by reducing dimensionless spray flux in wet granulation: a pharmaceutical manufacturing case study*. *Proceedings of the 9th International Symposium on Agglomeration* (June 2009). Sheffield, UK: Sheffield University.
23. Wildeboer, W.J., Koppendraaier, E., Litster, J.D. et al. (2007). *A novel nucleation apparatus for regime separated granulation*. *Powder Technology* 171 (2): 96–105.
24. Wauters, P.A.L., Jakobsen, R.B., Litster, J.D. et al. (2002). *Liquid distribution as a means to describing the granule growth mechanism*. *Powder Technology* 123: 166–177.
25. Schaafsma, S.H., Kossen, N.W.F., Mos, M.T. et al. (1999). *Effects and control of humidity and particle mixing in fluid-bed granulation*. *AIChE Journal* 45 (6): 1202–1210.
26. Vonk, P., Guillaumea, C.P.F., Ramaker, J.S. et al. (1997). *Growth mechanisms of high-shear pelletisation*. *International Journal of Pharmaceutics* 157: 93–102.
27. Hapgood, K.P., Tan, M.X.L., and Chow, D.W.Y. (2009). *A method to predict nuclei size distributions in models of wet granulation*. *Advanced Powder Technology* 20: 293–297.
28. Marshall, W.R.J. (1954). *Atomization and spray drying*. *Chemical Engineering Progress, Monograph Series* 50 (2).
29. Stewart, R.L., Bridgwater, J., and Parker, D.J. (2001). *Granular flow over a flat-bladed stirrer*. *Chemical Engineering Science* 56: 4257–4271.
30. Laurent, B.F.C. (2006). *Scaling factors in granular flow: analysis of experimental and simulations results*. *Chemical Engineering Science* 61 (13): 4138–4146.
31. Forrest, S., Bridgwater, J., Mort, P.R. et al. (2003). *Flow patterns in granulating systems*. *Powder Technology* 130: 91–96.
32. Laurent, B.F.C., Bridgwater, J., and Parker, D.J. (2000). *Motion in a particle bed agitated by a single blade*. *AIChE Journal* 46 (9): 1723–1734.
33. Tardos, G.I., Hapgood, K.P., Ipadeola, O.O., and Michaels, J. N. (2004). *Stress measurements in high-shear granulators using calibrated test particles: application to scale-up*. *Powder Technology* 140 (3): 217–227.
34. Hapgood, K., Jain, S., Kline, L. et al. (2003). *Application of spray flux in scale-up of high shear wet granulation processes*. *Proceedings of the AIChE Annual Meeting* (November 2003). San Francisco: AIChE.
35. Iveson, S.M., Beathe, J.A., and Page, N.W. (2002). *The dynamic strength of partially saturated powder compacts: the effect of liquid properties*. *Powder Technology* 127: 149–161.
36. Iveson, S.M. and Page, N. (2002). *Applied mechanics – progress and applications*. *The Third Australasian Congress on Applied Mechanics* (February 2002), 383–388.
37. Iveson, S.M. and Page, N. (2004). *Brittle to plastic transition in the dynamic mechanical behavior of partially saturated granular materials*. *Transactions of the ASME* 71: 470–475.
38. Iveson, S.M. and Page, N. (2005). *Dynamic strength of liquid-bound granular materials: the effect of particle size and shape*. *Powder Technology* 152: 79–89.
39. Iveson, S.M., Litster, J.D., and Ennis, B.J. (1996). *Fundamental studies of granule consolidation. Part 1: Effects of binder content and binder viscosity*. *Powder Technology* 88: 15–20.
40. Iveson, S.M. and Litster, J.D. (1998). *Fundamental studies of granule consolidation. Part 2: Quantifying the effects of particle and binder properties*. *Powder Technology* 99: 243–250.
41. Zoglio, M.A. and Cartensen, J.T. (1983). *Physical aspects of wet granulation III. Effect of wet granulation on granule porosity*. *Drug Development and Industrial Pharmacy* 9 (8): 1417–1434.
42. Ohno, I., Hasegawa, S., Yada, S. et al. (2007). *Importance of evaluating the consolidation of granules manufactured by high shear mixer*. *International Journal of Pharmaceutics* 338 (1–2): 79–86.
43. Farber, L., Tardos, G., and Michaels, J.N. (2003). *Use of X-ray tomography to study the porosity and morphology of granules*. *Powder Technology* 132 (1): 57–63.
44. Ansari, M.A. and Stepanek, F. (2006). *Formation of hollow core granules by fluid bed in situ melt granulation: modelling and experiments*. *International Journal of Pharmaceutics* 321 (1–2): 108–116.
45. Ansari, M.A. and Stepanek, F. (2006). *Design of granule structure: computational methods and experimental realization*. *AIChE Journal* 52 (11): 3762–3774.

46. Hapgood, K.P., Farber, L., and Michaels, J.N. (2009). *Agglomeration of hydrophobic powders via solid spreading nucleation. Powder Technology* 188 (3): 248–254.
47. Rahmanian, N., Ghadiri, M., Jia, X., and Stepanek, F. (2009). *Characterisation of granule structure and strength made in a high shear granulator. Powder Technology* 192 (2): 184–194.
48. Jøgerskou, A., Holm, P., and Schæfer, T. (1984). *Granulation in high speed mixers. Part 3: Effects of process variables on the intragranular porosity. La Pharmacie Industrielle* 46 (3): 310–314.
49. Berggren, J. and Alderborn, G. (2001). *Effect of drying rate on porosity and tableting behaviour of cellulose pellets. International Journal of Pharmaceutics* 227: 81–96.
50. Mort, P.R. (2005). *Scale-up of binder agglomeration processes. Powder Technology* 150 (2): 86–103.
51. Iveson, S.M. (1997). *Fundamental studies of granulation: granule deformation and consolidation.* PhD thesis. University of Queensland, Brisbane.
52. Iveson, S.M. and Litster, J.D. (1998). *Growth regime map for liquid-bound granules. AIChE Journal* 44 (7): 1510–1518.
53. Wauters, P.A.L., van de Water, R., Litster, J.D. et al. (2002). *Growth and compaction behaviour of copper concentrate granules in a rotating drum. Powder Technology* 124 (3): 230–237.
54. Iveson, S.M., Wauters, P.A.L., Forrest, S. et al. (2001). *Growth regime map for liquid-bound granules: further development and experimental validation. Powder Technology* 117: 83–97.
55. Tu, W.-D., Hsiau, S.-S., Ingram, A., and Seville, J. (2008). *The effect of powder size on induction behaviour and binder distribution during high shear melt agglomeration of calcium carbonate. Powder Technology* 184 (3): 298–312.
56. Hoornaert, F., Wauters, P.A.L., Meesters, G.M.H., and Pratsinis, S.E. (1998). *Agglomeration behaviour of powders in a Lödige mixer granulator. Powder Technology* 96: 116–128.
57. Mackaplow, M.B., Rosen, L.A., and Michaels, J.N. (2000). *Effect of primary particle size on granule growth and endpoint determination in high-shear wet granulation. Powder Technology* 108 (1): 32–45.
58. Tardos, G.I., Khan, M.I., and Schaeffer, D.G. (1998). *Forces on a slowly rotating, rough cylinder in a Couette device containing a dry, frictional powder. Physics of Fluids*. 10 (2): 335–341.
59. Litster, J.D. and Ennis, B.J. (2004). *The Science and Engineering and Granulation Processes.* Dordrecht: Kluwer Academic.
60. Ennis, B.J., Li, J., Tardos, G.I., and Pfeffer, R. (1990). *The influence of viscosity on the strength of an axially strained pendular liquid bridge. Chemical Engineering Science* 45 (10): 3071–3088.
61. He, Y., Liu, L.X., Litster, J.D., and Kayrak-Talay, D. (2009). *Scale up considerations in granulation.* In: *Handbook of Pharmaceutical Granulation* (ed. D.M. Parikh), 459–490. Boca Raton: Taylor & Francis.
62. Ennis, B.J. and Litster, J.D. (1997). *Size reduction and size enlargement.* In: *Perry's Chemical Engineers' Handbook* (ed. D. Green). New York: McGraw-Hill.
63. Reynolds, G.K., Fu, J.S., Cheong, Y.S. et al. (2005). *Breakage in granulation: a review. Chemical Engineering Science* 60 (14): 3969–3992.
64. Fu, J.S., Reynolds, G.K., Adams, M.J. et al. (2005). *An experimental study of the impact breakage of wet granules. Chemical Engineering Science* 60 (14): 4005–4018.
65. Tardos, G.I., Khan, M.I., and Mort, P.R. (1997). *Critical Parameters and limiting conditions in binder granulation of fine powders. Powder Technology* 94: 245–258.
66. Smith, R. (2008). *Wet granule breakage in high shear mixer granulators.* PhD thesis, University of Queensland, Brisbane.
67. Smith, R.M. and Litster, J.D. (2012). *Examining the failure modes of wet granular materials using dynamic diametrical compression. Powder Technology* 224: 189–195. 2010.
68. Iveson, S.M., Page, N.W., and Litster, J.D. (2003). *The importance of wet-powder dynamic mechanical properties in understanding granulation. Powder Technology* 130: 97–101.
69. van den Dries, K., Vegt, O.M.d., Girard, V., and Vromans, H. (2003). *Granule breakage phenomena in a high shear mixer: influence of process and formulation variables and consequences on granule homogeneity. Powder Technology* 113: 228–236.
70. Smith, R.M., Liu, L.X., and Litster, J.D. (2010). *Breakage of drop nucleated granules in a breakage only high shear mixer. Chemical Engineering Science* 65: 5651–5657.
71. Liu, L.X., Smith, R., and Litster, J.D. (2009). *Wet granule breakage in a breakage only high-shear mixer: effect of formulation properties on breakage behaviour. Powder Technology* 189 (2): 158–164.
72. Knight, P.C., Johansen, A., Kristensen, H.G. et al. (2000). *An investigation of the effects on agglomeration of changing the speed of a mechanical mixer. Powder Technology* 110: 204–209.
73. Pearson, J.K.M., Hounslow, M.J., and Instone, T. (2001). *Tracer studies of high-shear granulation I: experimental results. AIChE Journal* 47 (9): 1978–1983.
74. Ramaker, J.S., Jelgersma, M.A., Vonk, P., and Kossen, N.W.F. (1998). *Scale-down of a high shear pelletisation process: flow profile and growth kinetics. International Journal of Pharmaceutics* 166: 89–97.
75. Hounslow, M.J., Pearson, J.M.K., and Instone, T. (2001). *Tracer studies of high-shear granulation: II. Population Balance Modeling. AIChE Journal* 47 (9): 1984–1999.
76. Bouwman, A.M., Visser, M.R., Meesters, G.M.H., and Frijlink, H.W. (2006). *The use of Stokes deformation number as a predictive tool for material exchange behaviour of granules in the 'equilibrium phase' in high shear granulation. International Journal of Pharmaceutics* 318 (1–2): 78–85.
77. Tran, A. (2010). *Powder flow in mixer granulators, in chemical engineering.* PhD thesis, University of Queensland, Brisbane.
78. Holm, P., Jungersen, O., Schæfer, T., and Kristensen, H.G. (1983). *Granulation in high speed mixers. Part 1. Effects of process variables during kneading. La Pharmacie Industrielle* 45 (8): 806–811.
79. Holm, P., Jungersen, O., Schæfer, T., and Kristensen, H.G. (1984). *Granulation in high speed mixers. Part 2. Effects of process variables during kneading. La Pharmacie Industrielle* 46 (1): 97–101.
80. Holm, P. (1987). *Effect of impeller and chopper design on granulation in a high speed mixer. Drug Development and Industrial Pharmacy* 13 (9–11): 1675–1701.

81. Johansen, A. and Schæfer, T. (2001). *Effects of interactions between powder particle size and binder viscosity on agglomerate growth mechanisms in a high shear mixer*. *European Journal of Pharmaceutical Sciences* 12: 297–309.
82. Michaels, J.N., Farber, L., Wong, G.S. et al. (2009). *Steady states in granulation of pharmaceutical powders with application to scale-up*. *Powder Technology* 189 (2): 295–303.
83. Aulton, M. and Banks, M. (1978). *The factors affecting fluidised bed granulation*. *Manufacturing Chemistry and Aerosol News* 49: 50–56.
84. Capes, C.E. (1980). *Particle Size Enlargement*. Amsterdam, Oxford, New York: Elsevier Scientific.
85. Rambali, B., Baert, L., and Massart, D.L. (2003). *Scaling up of the fluidized bed granulation process*. *International Journal of Pharmaceutics* 252: 197–206.
86. Hapgood, K., Plank, R., and Zega, J. (2002). Use of dimensionless spray flux to scale up a wet granulated product. Proceedings of the World Congress on Particle Technology 4, Sydney, Australia.
87. Stewart, R.L., Bridgwater, J., Zhou, Y.C., and Yu, A.B. (2001). *Simulated and measured flow of granules in a bladed mixer: a detailed comparison*. *Chemical Engineering Science* 56: 5457–5471.
88. Laurent, B.F.C. and Bridgwater, J. (2002). *Performance of single and six-bladed powder mixers*. *Chemical Engineering Science* 57: 1695–1709.
89. Nilpawar, A.M., Reynolds, G.K., Salman, A.D., and Hounslow, M.J. (2006). *Surface velocity measurement in a high shear mixer*. *Chemical Engineering Science* 61 (13): 4172–4178.
90. Reynolds, G.K., Nilpawar, A.M., Salman, A.D., and Hounslow, M.J. (2008). *Direct measurement of surface granular temperature in a high shear granulator*. *Powder Technology* 182 (2): 211–217.
91. Hiseman, M.J.P., Bridgwater, J., and Wilson, D.I. (1997). Positron emission particle tracking studies of powder mixing in a planetary mixer. In: *Control of Particulate Processes IV*. Delft, Netherlands: Engineering Foundation.
92. Horsthuis, G.J.B., van Laarhoven, J.A.H., van Rooij, R.C.M. B., and Vromans, H. (1993). *Studies on upscaling parameters of the Gral high shear granulation process*. *International Journal of Pharmaceutics* 92: 143–150.
93. Litster, J.D., Hapgood, K.P., Michaels, J.N. et al. (1999). Scale-up of mixer granulators for effective liquid distribution. In: *Control of Particulate Processes 6*. Fraser Island: Engineering Foundation.
94. Rahmanian, N., Ghadiri, M., and Ding, Y. (2008). Effect of scale of operation on granule strength in high shear granulators. *Chemical Engineering Science* 63 (4): 915–923.
95. Knight, P.C., Seville, J.P.K., Wellm, A.B., and Instone, T. (2001). *Prediction of impeller torque in high shear powder mixers*. *Chemical Engineering Science* 56: 4457–4471.
96. Landin, M., York, P., Cliff, M.J. et al. (1996). *Scale-up of pharmaceutical granulation in fixed bowl mixer-granulation*. *International Journal of Pharmaceutics* 133: 127–131.
97. Hassanpour, A., Antony, S.J., and Ghadiri, M. (2007). *Modeling of agglomerate behavior under shear deformation: effect of velocity field of a high shear mixer granulator on the structure of agglomerates*. *Advanced Powder Technology* 18: 803–811.
98. Fu, J., Chan, E.L., Jones, M.R. et al. (2009). *Characterisation of impact stress from main impeller on granules during granulation processes in a high shear mixer*. In: *9th International Symposium on Agglomeration* (ed. A.D. Salman). Sheffield: Sheffield University.
99. Mort, P.R. (2009). *Scale-up and control of binder agglomeration processes: flow and stress fields*. *Powder Technology* 189 (2): 313–317.
100. Faure, A., Grimsey, I.M., Rowe, R.C. et al. (1999). *Applicability of a scale-up method for wet granulation processes in Collette Gral high shear mixer-granulators*. *European Journal of Pharmaceutical Sciences* 8: 85–93.
101. Stepanek, F., Rajniak, P., Mancinelli, C., and Chern, R. (2006). *Determination of the coalescence probability of wet granules by mesoscale modeling*. In: *AICHE Annual Meeting*. San Francisco: AIChE.
102. Boerefijn, R., Juvin, P.Y., and Garzon, P. (2009). *A narrow size distribution on a high shear mixer by applying a flux number approach*. *Powder Technology* 189 (2): 172–176.
103. Hapgood, K.P. and Khanmohammadi, B. (2009). *Granulation of hydrophobic powders*. *Powder Technology* 189 (2): 253–262.
104. Vervaet, C. and Remon, J.P. (2005). *Continuous granulation in the pharmaceutical industry*. *Chemical Engineering Science* 60 (14): 3949–3957.
105. Djuric, D. and Kleinebudde, P. (2009). *Impact of screw elements on continuous granulation with a twin-screw extruder*. *Journal of Pharmaceutical Sciences* 97 (11): 4934–4942.
106. Hu, X., Cunningham, J.C., and Winstead, D. (2008). *Study growth kinetics in fluidized bed granulation with at-line FBRM*. *International Journal of Pharmaceutics* 347 (1–2): 54–61.
107. Cantor, S.L., Kothari, S., and Koo, O.M.Y. (2009). *Evaluation of the physical and mechanical properties of high drug load formulations: wet granulation vs. novel foam granulation*. *Powder Technology* 195 (1): 15–24.
108. Keary, C.M. and Sheskey, P.J. (2004). *Preliminary report of the discovery of a new pharmaceutical granulation process using foamed aqueous binders*. *Drug Development and Industrial Pharmacy* 30 (8): 831–845.
109. Tan, M.X.L., Wong, L.S., Lum, K.H., and Hapgood, K.P. (2009). *Foam and drop penetration kinetics into loosely packed powder beds*. *Chemical Engineering Science* 64 (12): 2826–2836.

TOWARD A GENERIC MODEL FOR TWIN-SCREW WET GRANULATION

DAAN VAN HAUWERMEIREN, MICHAËL GHIJS, AND INGMAR NOPENS

BIOMATH, Ghent University, Ghent, Belgium

MAXIM VERSTRAETEN AND THOMAS DE BEER

Laboratory of Pharmaceutical Process Analytical Technology, Ghent University, Ghent, Belgium

KAI LEE AND NEIL TURNBULL

Drug Product Design, Pfizer Ltd, Sandwich, UK

MARY T. AM ENDE*

Process Modeling & Engineering Technology Group, Pfizer Inc., Groton, CT, USA

PANKAJ DOSHI

Process Modeling & Engineering Technology Group, Pfizer Inc., Groton, CT, USA

DAVID WILSDON

Analytical R&D, Pfizer Ltd, Sandwich, UK

7.1 INTRODUCTION

Traditionally, the solid oral dosage manufacturing in the pharmaceutical industry is composed of a sequence of unit batch operations. In recent years, the industry is moving away from these inefficient and expensive batch processes to more efficient continuous manufacturing techniques. This chapter focuses on one unit operation, i.e. continuous wet granulation by means of a twin-screw granulator (TSG).

Previous studies of twin-screw granulation in the literature usually focus on the granules collected from the outlet of the TSG: linking factors such as process parameters or process variables [1, 2] or formulation properties [3] to responses (e.g. granule properties). Other studies [4–6] assessed the effect of screw elements on the resulting granule properties at the end of the granulator. In yet another study, Lee et al. [7] tried to gain insight in the mixing behavior in TSG by using positron emission particle tracking. However, these

studies lack fundamental knowledge on the mechanisms of granule formation and growth/breakage occurring inside the TSG, which is hard to obtain when only experimental data is available at the system inlet and outlet.

Since the ultimate goal is a predictive population balance model (PBM) for TSG, granule properties should be measured along the barrel to clearly distinguish between the different physical phenomena occurring at different spatial locations. In this chapter, we will address this aspect by gathering experimental data of granules (e.g. size) at different compartments in the TSG.

Mechanistic mathematical models are powerful tools to assist the development of process knowledge. In the case of twin-screw wet granulation (TSWG), where granule size is an important critical quality attribute (CQA), population balance models (PBMs) are a suitable framework as they allow the prediction of the dynamic granule size distribution (GSD). Many PBMs can be formulated, where the number of internal coordinates (i.e. particle properties) determines its dimensionality. Both one-dimensional [8] as well as

*Current address: Lyndra Therapeutics, Watertown, MA, USA

multidimensional PBMs [9] have been defined in the literature. However, PBMs for TSG are often being calibrated using assumptions and/or best-guessed values without any knowledge or experimental validation. A major reason was already mentioned earlier, i.e. the response availability is solely measured at the outlet. Furthermore, this applies to other granule properties such as moisture content, porosity, and other material properties. Therefore, multidimensional PBMs, which, for instance, track the properties of granule size, moisture content, and porosity, become problematic in the current usage with respect to lack of calibration data and initial conditions. For example, what is the initial moisture or porosity distribution when the granulation liquid hits the powder bed?

Although significant improvements in data gathering have been made, it is currently still not feasible to measure such initial conditions. For this reason, we will limit ourselves to a one-dimensional PBM with granule size as internal coordinate. The availability of data at different locations does, however, allow the PBM to be extended in a spatial way, leading to a so-called compartmental PBM. We will, moreover, focus on a tedious calibration procedure and use the knowledge available in the experimentally collected data to formulate the model. In this way, a bottom-up approach is used (starting from the available data) as a first step toward a predictive PBM for TSG.

7.2 EXPERIMENTAL DATA COLLECTION

The experimental setup is described by Verstraeten et al. [10]. Here, only a brief summary with the relevant details is given. Further details can be found in this reference.

The granules in the experimental study were produced using the ConsiGma™-25 system (GEA Pharma Systems, Collette, Wommelgem, Belgium), which is a high-shear TSWG module and is composed of two 25 mm diameter self-wiping, co-rotating screws with a length-to-diameter ratio of 20 : 1 (Figure 7.1).

The preblend and the granulation liquid (demineralized water) were introduced into the system by a gravimetric twin-screw loss-in-weight feeder (KT20, K-Tron Soder, Niederlenz, Switzerland) and two out-of-phase peristaltic pumps, located on top of the granulator (Watson Marlow, Cornwall, UK), connected to 1.6mm nozzles, respectively.

A hydrophobic model drug was used for which the preblend contains 60% (w/w) hydrochlorothiazide (UTAG, Almere, Netherlands), 16% (w/w) lactose monohydrate (Lactochem® Regular, DFE Pharma, Goch, Germany), 16% (w/w) microcrystalline cellulose (Avicel® PH 101, FMC, Philadelphia, USA), 3% (w/w) hydroxypropylcellulose (Klucel® EXF, Ashland, Covington, Kentucky, USA), and 5% (w/w) croscarmellose sodium (Ac-Di-Sol®, FMC, Philadelphia, USA).

A three-level full-factorial design of experiments (DoE) was performed to study the influence of the granulation process parameters screw speed (450, 675, and 900 rpm), material throughput (5, 12.5, and 25 kg h⁻¹), and liquid-to-solid (L/S) ratio (0.3, 0.45, and 0.6) upon local and final GSD, liquid distribution, and porosity distribution per granule size class. Local and final terminologies refer to measurements inside and at the outlet of the granulator, respectively. The process parameters of the experimental conditions of the DoE are provided in Table 7.1.

The DoE was performed with a fixed screw configuration: two kneading compartments, each composed of six kneading elements (length = diameter/6 for each element) with a 60° stagger angle, separated by a conveying element with the same length equal to 1.5 times the diameter.

The barrel's jacket temperature was set at 25 °C. Samples were collected at four locations inside the barrel denoted by white numerals in Figure 7.1. The unique sampling procedure is extensively described in Verstraeten et al. [10]. After collection, the samples were oven dried to ensure that the granule properties did not change before measurement of the GSD.

The size and shape distribution of the collected, oven-dried, granule samples were analyzed using a QICPIC

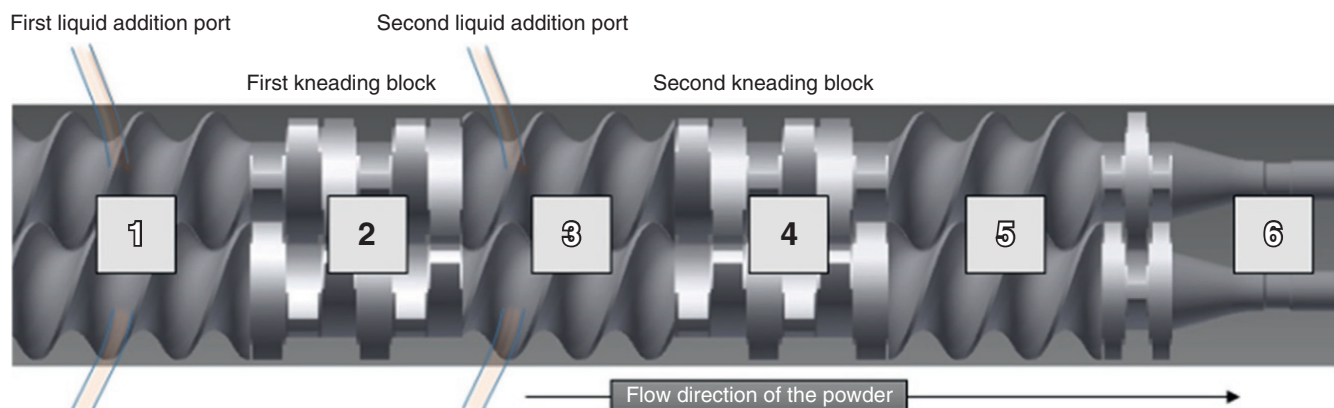


FIGURE 7.1 Visual representation of the different compartments or zones along the length of the twin-screw granulator. The different locations in the granulator are marked with numbers, the locations where samples were taken are marked in white. *Source:* Reprinted with permissions from Verstraeten et al. [10]. Copyright (2017) Elsevier.

TABLE 7.1 Process Parameters of the Experimental Conditions of the DoE

Experiment	Throughput (kg/h)	Screw Speed (rpm)	Liquid Addition (g/min)	L/S Ratio (–)	Residence Time (s)
N1	5	450	25	0.30	11.51
N2	5	900	25	0.30	10.83
N3	25	450	125	0.30	9.71
N4	25	900	125	0.30	8.68
N5	5	450	50	0.60	10.46
N6	5	900	50	0.60	13.73
N7	25	450	250	0.60	12.29
N8	25	900	250	0.60	6.86
N9	12.5	675	93.75	0.45	7.05

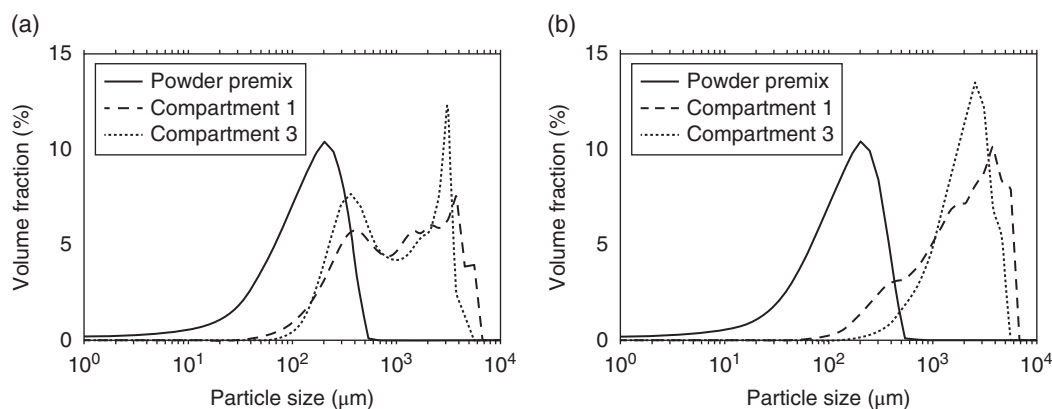


FIGURE 7.2 Particle size distributions (volume fraction in function of granule size) for powder premix, granules produced during wetting and after the first kneading zone with the hydrophobic formulation using a low (a) and high (b) L/S ratio. *Source:* Reprinted with permissions from Van Hauwermeiren et al. [11]. Copyright (2018) Elsevier.

particle size analyzer with WINDOX 5.4.1.0 software (Sympatec, GmbH, Clausthal-Zellerfeld, Germany). The dry premix was analyzed with laser diffraction using a Mastersizer[®] S (Malvern Instruments, Malvern, UK). The liquid distribution was assessed by adding a 0.5% (w/w) colored dye, Congo Red (Sigma-Aldrich, Darmstadt, Germany) to the granulation liquid. The dye recovery of the oven-dried and sieved granules was measured using a UV spectroscopy-based method as described by Verstraeten et al. [10].

7.3 GRANULE SIZE DISTRIBUTIONS AT DIFFERENT LOCATIONS

The L/S ratio was found to be the factor with the greatest impact on the granule properties. After the wetting zone, a bimodal GSD was observed at low L/S ratios. This can be explained by the formation of loose, wetted agglomerates as granulation liquid is dripped into the granulator barrel. Due to this immersion nucleation, the size of the agglomerates is approximately comparable to the size of the granulation liquid droplets. The wetting zone consists solely of conveying elements and is thus a low-shear environment. That is why a significant part of the preblend remains ungranulated after the wetting zone, and this results in the observed bimodal GSDs at low L/S ratios (Figure 7.2a).

At higher L/S ratios, more granulation liquid is available for granulation, leading to a diminished height of the first peak (Figure 7.2b). In the subsequent kneading zones, the granules are reshaped due to the high-shear environment: particles both aggregate (as can be seen in a shift toward larger particles) and break up (Figure 7.2a and b). For low L/S ratio, particles elongate more, whereas for high L/S ratio, the granules are more spherical.

The proof for the explanation of the behavior in the GSD in the wetting zone is found in the liquid distribution among the particle sizes. In this zone, a poor liquid distribution is seen, where the liquid is more likely to be found in the larger granules. Due to the low-shear environment (only conveying elements), the liquid is unable to be distributed homogeneously over the entire GSD. It is only after the kneading zones (high-shear environment) that the liquid is distributed evenly across all granules [10].

7.4 PBM MODEL DEVELOPMENT

7.4.1 PBM Formulation

Traditionally, a single combined aggregation–breakage PBM is used to model TSG. Such a PBM is given in Equation (7.1).

$$\begin{aligned} \frac{\delta n(t,x)}{\delta t} = & \frac{1}{2} \int_0^x \beta(t,x-\varepsilon,\varepsilon)n(t,x-\varepsilon)n(t,\varepsilon)d\varepsilon \\ & - n(t,x) \int_0^\infty \beta(t,x,\varepsilon)n(t,\varepsilon)d\varepsilon \\ & + \int_x^\infty b(t,x,\varepsilon)S(t,\varepsilon)n(t,\varepsilon)d\varepsilon - S(t,x)n(t,x) \end{aligned} \quad (7.1)$$

where

$n(t,x)$ is the number distribution.

$\beta(x,x')$ the aggregation frequency.

$S(x)$ the breakage rate.

$b(x,\varepsilon)$ the daughter size distribution.

The latter functions are termed “kernel” in PBM terminology and will be addressed like this in the remainder of this chapter. The first two terms on the right-hand side represent, respectively, the birth and death of a particle of size x due to aggregation. The third and fourth terms represent birth and death terms due to breakage.

However, from the available data, this hypothesis seems not to hold throughout the twin screw granulator as it behaves spatially inhomogeneous along the barrel length due to the specific configuration of the liquid addition ports and kneading blocks. For a sound mathematical description of this process, the barrel hence needs to be subdivided into four well-mixed zones (three compartments and the initial blend at zone 0). The hypothesized dominant granulation processes are aggregation in the wetting zone and combined aggregation and breakage in the kneading zones. A visual representation of the compartmentalization of the PBM is given in Figure 7.3.

7.4.2 Specifics of PBM Numerical Solution

Equation (7.1) is an integro-differential equation for which, only for a few specific kernels, an analytical solution exists [12]. In all other cases, numerical techniques have to be used to solve this equation: the method of successive approximations [13], the finite elements methods [14–16], the method

of moments [17–19], Monte Carlo simulation methods [20–23], the finite volume Scheme [24–29], and the cell average technique (CAT) [30]. Here, the latter technique is chosen, as it can tackle both aggregation and breakage problems. In a numerical analysis of the CAT, it was denoted to be efficient and accurate, as it prevents overprediction of number density for the larger particles and limits a diverging behavior of higher moments.

The CAT is consistent with the first two moments, if required, it is possible to generalize this to any two moments [30]. The consistency of the first two moments was opted as good modeling practice as they represent the total number of particles (zeroth moment) and the total volume (first moment).

The number of bins in the PBM were chosen as such that experimental data could be loaded in swiftly without the need for interpolation. By doing so, possible interpolation errors are avoided. Two measurement techniques were used: QICPIC for the granules and Mastersizer S for the dry premix (size of the premix powder was too small to be analyzed with image analysis [QICPIC]). The grid is a combination of the grid of the aforementioned techniques with a total of 58 bins: 14 from the Mastersizer S (logarithmically spaced between 0.9 and 7.72 μm and 44 from QICPIC (logarithmically spaced between 8.46 and 34 519.36 μm). The residence times in the different zones of the granulator were chosen to be 1/6, 1/3, and 1/2 of the experimentally determined total residence time for the wetting zone, first and second kneading zones, respectively, as was determined earlier [7, 8].

7.4.3 Kernel Description for the Wetting Zone

As observed earlier, the GSD often exhibits a bimodal shape after the wetting zone. Experiments at a low L/S ratio exhibit more bimodality in the GSD than experiments performed at high L/S ratio. However, only a low-shear environment is present. This leads to the hypothesis that breakage is not likely the mechanism leading to the bimodality, but rather a difference in aggregation rate of different sized particles, which is confirmed by the liquid distribution over the different size classes, where a lower liquid content in the lowest

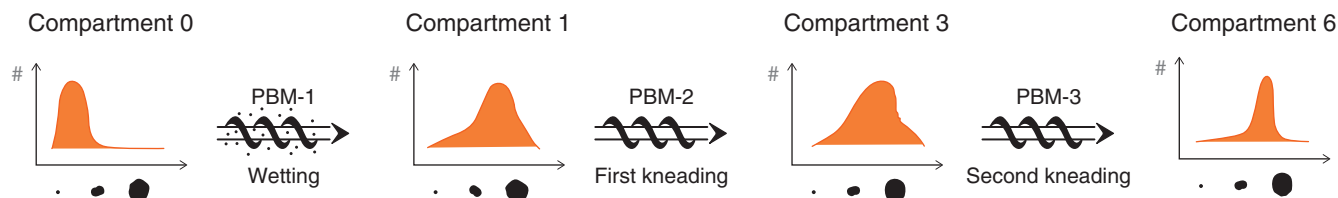


FIGURE 7.3 Visual representation of the compartmentalization of the PBM: these figures link the zones in Figure 7.1 to the compartments used in this study. A total of three compartments are used coinciding with data availability: one for the wetting compartment between zone 0–1 and one for each kneading compartment between zone 1–3 and zone 3–6, respectively. *Source:* Adapted from Van Hauwermeiren et al. [11]. Reprinted with permissions. Copyright (2018) Elsevier.

size classes indicates the existence of ungranulated material. Hence, we hypothesize that a difference in aggregation rate of different sized particles is responsible for the bimodality. Wetted particles aggregate faster than unwetted: the fines have a lower dye recovery than the largest particles.

Nevertheless, a one-dimensional PBM with granule size as internal coordinate was chosen, so a direct incorporation of liquid distribution is not possible. As stated in the introduction, the one-dimensional model was chosen to limit both the number of parameters needed for calibration and the required data collection. With the current availability of data, it was not possible to construct a well-formed multidimensional PBM, with, for instance, liquid content as one of its internal coordinates. However, if measurement methods improve, a sound multidimensional PBM might become possible in the future.

Obtaining bimodality with a pure aggregation kernel is not obvious as most kernels described in literature are monotonous in nature and can never lead to bimodality. In order to obtain bimodality with a pure aggregation process, a new type of kernel hence needs to be developed. From the data, a kernel with a steplike behavior as a function of granule size is proposed. Granules larger than a critical diameter will aggregate significantly faster than those of smaller size, driven by their higher liquid content. Indeed, it was observed that larger size particles still have liquid available to further aggregate, while smaller ones do not and, hence, are unlikely to further aggregate. Introducing such a kernel shape leads to a dip in the GSD, hereby creating a bimodal distribution. The reasoning behind this is that the introduced droplets of granulation liquid will wet the powder, and almost instantaneously form large initial nuclei of about 300 μm (first peak in GSD; see Figure 7.2a and b). These wet large granules will scoop up other particles and granulate further, but due to unequal liquid distribution, they will aggregate faster than the less-wetted or even still completely dry small granules. For

the wetting zone, the newly developed collision frequency is given by Equation (7.2).

$$\beta(x, \epsilon) = \left(\frac{\text{top}_1}{2} \left(1 + \tanh \left(\frac{R_1^3 - (x^2 + \epsilon^2)^{1/2}}{\delta_1} \right) \right) - \frac{\text{top}_1 - \text{top}_2}{2} \left(1 + \tanh \left(\frac{R_2^3 - (x^2 + \epsilon^2)^{1/2}}{\delta_2} \right) \right) \right) \times \left(\left(x^{1/3} + c \right) \left(\epsilon^{1/3} + c \right) \right) \tag{7.2}$$

As can be noted from Equation (7.2), the collision frequency is in fact a convolution of two parts: a two-dimensional stepping function (to induce the bimodality from an initial unimodal distribution) and a monotonously increasing part that drives the aggregation to larger sizes (similar to traditional kernels). A visual representation of the kernel is shown in Figure 7.4.

R_2 represents the particle size where the step is located. Physically, it represents the critical diameter: particles larger than this size will aggregate significantly faster than those who are smaller. In Figure 7.4, this is at 1500 μm. R_1 represents the particle size where the step goes to zero. In this study R_1 is chosen to be at 7000 μm based on the dimensions of the TSG and the measured GSD. The physical interpretation is that granules cannot aggregate further, which was observed from the experimental data. The heights of the steps are determined by parameters top_1 and top_2 .

The smoother (δ) is used to make the step not too abrupt. Its usage is mainly for numerical reasons: the integrator used to solve this ordinary differential equation can have

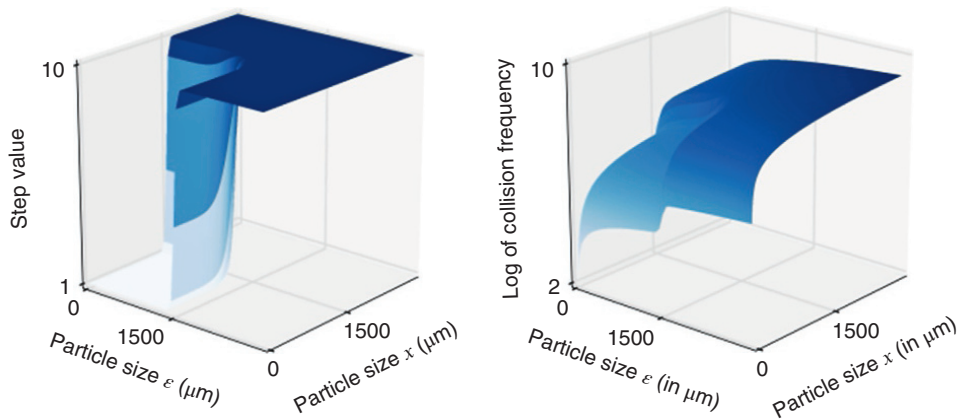


FIGURE 7.4 Visual representations of the collision frequency $\beta(x, \epsilon)$: an example of the steplike function (left) and an example of the full equation (right). *Source:* Reprinted with permissions from Van Hauwermeiren et al. [11]. Copyright (2018) Elsevier.

difficulties with the sudden increase in value of the aggregation kernel and suffer from instabilities (i.e. oscillations).

7.4.4 Kernel Description for the Kneading Zones

The kneading zones each consists of one kneading element and one conveying element (Figure 7.1). As observed in the dye recovery tests, the liquid is redistributed and mixed thoroughly in the kneading zones [10]. Further, due to the kneading elements, there exists an obstruction for larger particles to move through the granulator. Hence, breakage is likely to occur. Therefore, it was hypothesized that a combination of aggregation and breakage to model the dynamics in these compartments is required. The aggregation kernel in this kneading zone is the first part of the previously established kernel (the stepping function) from Equation (7.2). As to the breakage, the frequently employed power law selection function was used. The daughter size distribution was chosen to be a combination of attrition and binary breakage [31] as given in Equation (7.3).

$$b(m, d) = f_{\text{prim}} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(d^{1/3} - \mu)^2}{2\sigma^2}} \frac{m}{\mu^3} \frac{1}{3d^3} + (1 - f_{\text{prim}}) \frac{2}{m} \quad (7.3)$$

Here, m and d are the volume of the mother and daughter particle, respectively. σ and μ are the standard deviation and the mean of the Gaussian normal distribution corresponding to the particles that result from erosion, in particle size. f_{prim} is the fraction of granules selected to form small fragments. Here, σ and μ are chosen to be constant at 50 and 300 μm based on the observed trends in experimental data.

7.5 PBM MODEL CALIBRATION

To calibrate the different models, a global optimization method was used as it is more robust with regard to local minima. The whole parameter space was sampled using Latin hypercube sampling (LHS) [32].

The boundaries of the parameter space were defined as follows: $\beta_0 \in [1e-2, 1e6]$, $R_2 \in [100, 4000]$, $\text{top}_1 \in [1, 1000]$, $c \in [0.01, 100]$, $f_{\text{prim}} \in [0.001, 1]$, and $S_0 \in [1e-7, 1e-3]$. A total of 1 250 000 simulations or, on average, approximately 46 000 for each experiment and location were run. These simulations were run in different rounds of calibration, until the results were deemed to be sufficiently accurate. To speed up the calculation time, the different simulations were run in parallel on a Linux server with eight octa-core processors. The obtained GSDs from the simulations were subsequently analyzed. First, a general check was made whether the simulation was run successfully: volume conservancy, achieved integrator tolerance, and absence of integrator errors and warnings. Afterward, the parameter

set that gives the best fit to the experimental data was determined by searching for that parameter combination that yields the minimal value of a user-defined objective function (i.e. a quantification of the mathematical distance between measurements and model prediction). The objective function used in this study is the sum of squared errors (SSE) between the measured and simulated GSD on the volume fraction ($\text{SSE} = \sum (\text{simulated} - \text{measured})^2$). It is noteworthy that the global optimization method actually allows the use of different objective functions as it is only computed in the data analysis posterior to the model simulations. This is not the case for local optimization methods. For each calibration round, a number of samples was chosen in the parameter space using LHS. After simulation and calculation of the objective function, it was determined whether another iteration was necessary by checking the value of the objective function. If the value of the objective function was sufficiently low, or did not significantly decrease anymore, the calibration procedure was considered to be finished. It was estimated that a relative, Gaussian noise level on the data of about 25–33% would be realistic. This corresponds to an SSE of 45 and 80, respectively.

These values were obtained by adding relative Gaussian noise to the data sets and calculating the SSE value (10 000 repetitions for each data set). If the obtained SSE value from the simulations lies within the aforementioned boundaries, the simulations were deemed to be sufficiently converged. Calibrations are performed separately for each compartment, resulting in 27 calibrations. The optimal parameter sets from the different zones (wetting, first and second kneading) are afterward checked to evaluate the quality of the compartmental model from preblend straight to the end of the granulator.

7.6 RESULTS FOR THE WETTING ZONE 1

A selection of the resulting GSD of the calibrated experiments is visualized in Figure 7.5a–d. An overview of the calibrated parameters and the value of the objective function is given in Table 7.2.

As can be seen in the aforementioned figures, GSDs with both low and high amounts of bimodal behavior can be well predicted using a model with one single kernel and by only modifying a limited number of parameters that have the physical meaning of driving the amount of observed bimodality. Figure 7.5a and b denotes the experiments with low L/S ratio. In these experiments, bimodality is much more present in the GSD. The model is able to predict the location of the two peaks accurately without significant overprediction. However, there is some underprediction in the region between the two peaks, but overall, the model is performing well. Note that the kernel function could be further fine-tuned to improve the prediction in this area, but this was deemed

TABLE 7.2 Calibrated Parameters for the Wetting Zone 1 for the Performed Experiments (Hydrophobic Model Drug)

Experiment	β_0	c	R_2	top_1	SSE
N1	0.0856	0.10	775	16.71	79.15
N2	0.1752	0.10	1185	6.98	73.82
N3	0.0854	8.00	752	40.00	145.38
N4	0.0623	3.81	503	40.00	36.22
N5	0.4926	0.82	2234	3.11	61.64
N6	0.2506	0.50	2038	2.89	34.73
N7	0.1618	0.82	982	9.65	116.04
N8	0.2200	1.81	615	14.00	53.98
N9	0.7766	0.01	336	1.00	71.61

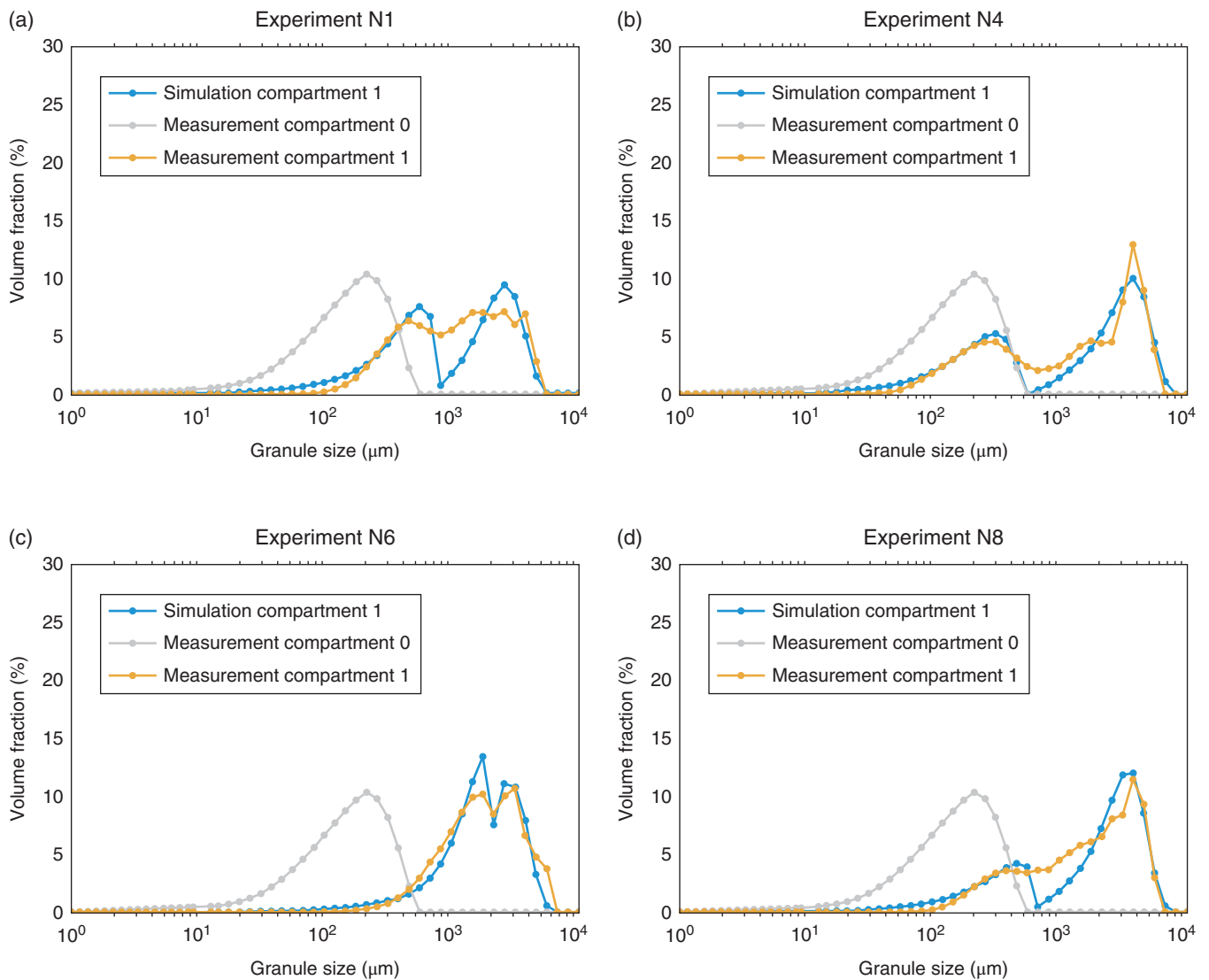


FIGURE 7.5 Calibrated simulation results for the wetting zone 1 (hydrophobic model drug). GSD are visualized in volume fraction (%) as a function of granule size (μm). In this case, granule size is defined as the equivalent diameter of the granule. The process parameters can be found in Table 7.1, while the calibrated model parameters are provided in Table 7.2. *Source:* Adapted from Van Hauwermeiren et al. [11]. Reprinted with permissions. Copyright (2018) Elsevier.

outside the scope of the current work and will be further researched. Regarding the experiments with high L/S ratio (Figure 7.5c and d), the model is performing slightly better. The underprediction as seen at the low L/S experiments is also present, albeit to a smaller extent. Overall, the calibrated experiments fall below the previously established acceptance limit of the objective function. The values for parameter top_1 are generally higher for the low L/S experiments (N1–N4). This was to be expected as a large value of top_1 is needed to achieve a bimodal distribution. This trend can be linked to the liquid availability: lower liquid availability leads to a difference in collision efficiency of the particles due to the unequal liquid distribution [10].

The difference in collision efficiency is obtained in the model with larger values of the parameter top_1 . No clear trend is observed for parameter c , and it could be considered in the future to omit this parameter from the calibration or even from the model. The total aggregation, which can be described as a convolution of MRT, β_0 and top_1 , is, as expected, in the same order of magnitude for all the experiments. This is due to the fact that the largest particles for each experiment are all in the same order of magnitude. For all experiments except N5 and N6, the values of R_2 are more or less in the size range of their expected value, i.e. the dip between the two peaks in the GSD. Considering experiment N6, the cause of the high value of R_2 can be seen in Figure 7.5c: some overfitting of the model can be noticed here.

7.7 RESULTS FOR THE FIRST KNEADING ZONE 3

The resulting GSD of the calibrated experiments are visualized in Figure 7.6a and d. An overview of the calibrated parameters and the value of the objective function is given in Table 7.3.

In this zone, there is some discrepancy in the achieved results between low and high L/S ratio. The mid (Figure 7.6b) and high L/S ratio experiments (Figure 7.6c and d) are predicted better, which is attributed to lower complexity of the GSD shape. The SSE values in Table 7.3 are for more than half of the experiments out of the predefined range. However, it should be noted that for experiment N6 (Figure 7.6d) the SSE is high (88.41), but the obtained fit is nevertheless acceptable. The prediction of the fines in the 100–500 μm granule diameter range could be improved in future work. The parameter values for β_0 and top_1 are higher compared with the simulations of the wetting zone 1 because of the lower value of the collision frequency function, $\beta(x, \epsilon)$, as only the first factor of Equation (7.2) was used for the kneading zone. The values for β_0 are higher for high L/S experiments: this indicates a higher level of aggregation, which was also observed in data. This is explained by a

higher amount of available granulation liquid during granulation. The parameters values of f_{prim} , and to a lesser extent S_0 are within the same order of magnitude for all the experiments. This illustrates that breakage behavior in this first kneading zone appears, to some degree, to be equal.

7.8 RESULTS FOR THE SECOND KNEADING ZONE 6

The resulting GSDs of the calibrated experiments are visualized in Figure 7.7a and d. An overview of the calibrated parameters and the value of the objective function is given in Table 7.4.

In this second kneading zone 6, the calibration results for all experiments is satisfactory: for mid and high L/S ratios the fit is good. The low L/S experiments suffer from the same shortcomings as the experiments in the first kneading zone 3: the predictions in the 100–500 μm granule diameter range could be improved. Overall, the obtained results for the second kneading zone 6 are better than those of the first kneading zone 3. This is reflected in the objective function values (Table 7.4); with exception of experiments N7 and N8, they are all within the predefined range. The same side note as in the first kneading zone should be made: experiment N7 can be simulated with a good precision (Figure 7.7d), but the resulting SSE value is high (341.55). Regarding the parameter values, compared to the parameters of the first kneading zone, the values of top_1 are more within the same range. This could indicate that in the second kneading zone, the aggregation mechanism is more equal for all tested experimental conditions compared to the wetting zone and the first kneading zone. Considering parameters β_0 , f_{prim} and S_0 , the changes are not equal in all the experiments and some experiments show some more generation of fines, whereas, in other experiments, the granules aggregate further. No clear trend is observed for parameter R_2 .

7.9 RESULTS OF DIRECT SIMULATION FROM PREBLEND TO FINAL GRANULES AT TSG OUTLET (ZONES 1–6)

As mentioned in section earlier, the compartmental model is composed of the different calibrated zones of the granulator. Here, the three submodels from the previous sections are put in series to mimic the overall system. The calibrated parameter sets for each submodel obtained in the previous sections are used; hence, no further calibration was performed when constructing the full compartmental PBM. This latter model is important in an industrial context: given an initial powder premix, what GSD is to be expected at the TSG outlet? It is not sufficient that separate compartments can be predicted accurately; the model as a whole needs to predict GSDs at

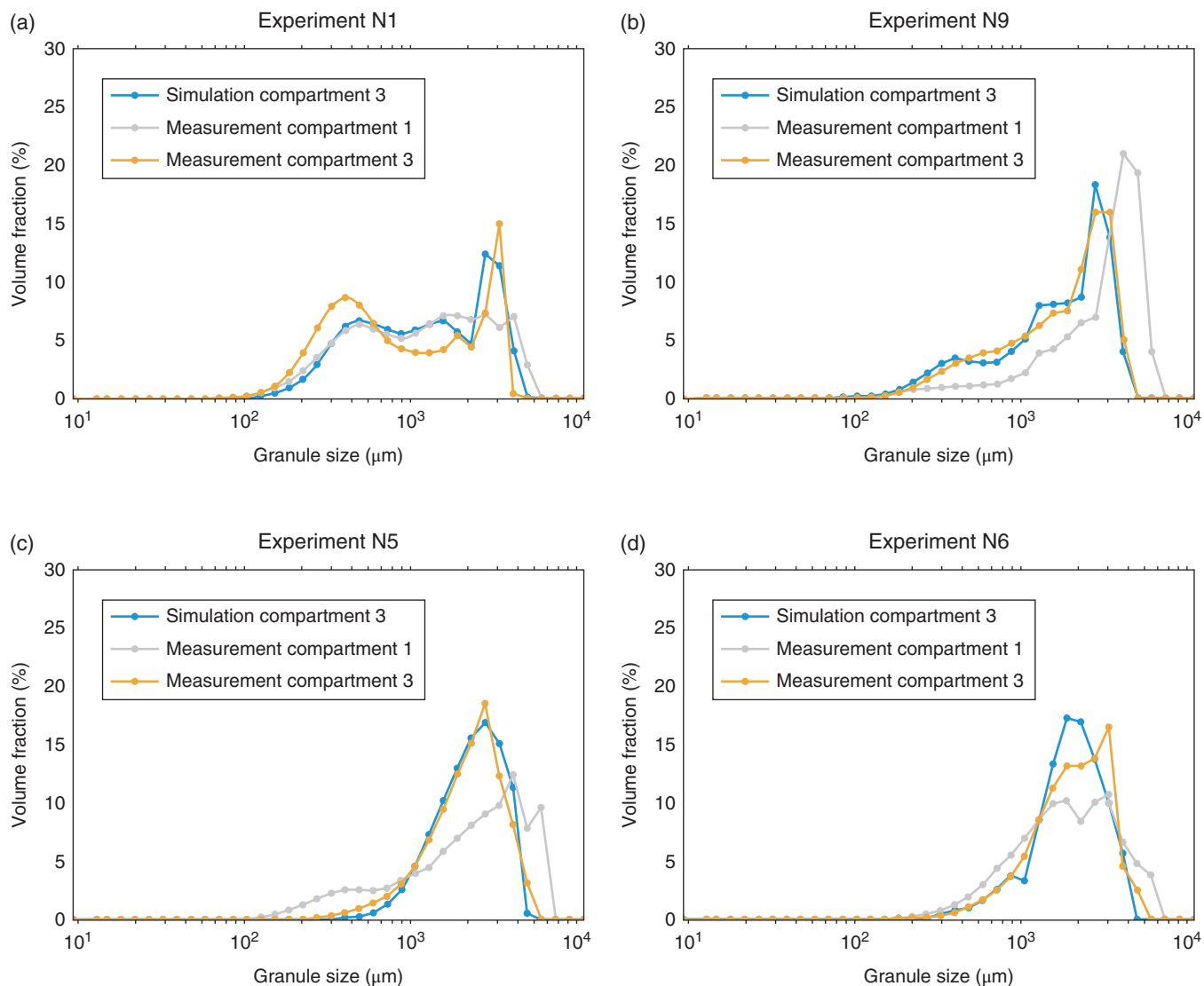


FIGURE 7.6 Calibrated simulation results for the first kneading zone 3 (hydrophobic model drug). GSD are visualized in volume fraction (%) as a function of granule size (μm). In this case, granule size is defined as the equivalent diameter of the granule. The process parameters can be found in Table 7.1, the calibrated model parameters are provided in Table 7.3.

TABLE 7.3 Calibrated Parameters for the First Kneading Zone 3 for the Performed Experiments (Hydrophobic Model Drug)

Experiment	β_0	R_2	top_1	S_0	f_{prim}	SSE
N1	720	1845	3.73E+02	1.48E-04	1.75E-02	106.04
N2	100	1519	1.00E+03	1.00E-07	5.00E-01	103.78
N3	193	1135	3.73E+02	4.39E-05	1.64E-01	262.01
N4	100	965	1.00E+03	1.00E-04	7.07E-02	207.52
N5	100	100	1.00E+03	1.00E-04	1.00E-02	30.87
N6	10 000	965	5.62E+00	1.00E-04	1.00E-2	88.41
N7	100 000	3000	3.98E+00	2.51E-05	4.78E-02	157.38
N8	10 000	311	1.00E+03	1.00E-05	1.00E-02	389.86
N9	2683	1845	1.39E+02	1.48E-04	1.75E-02	24.02

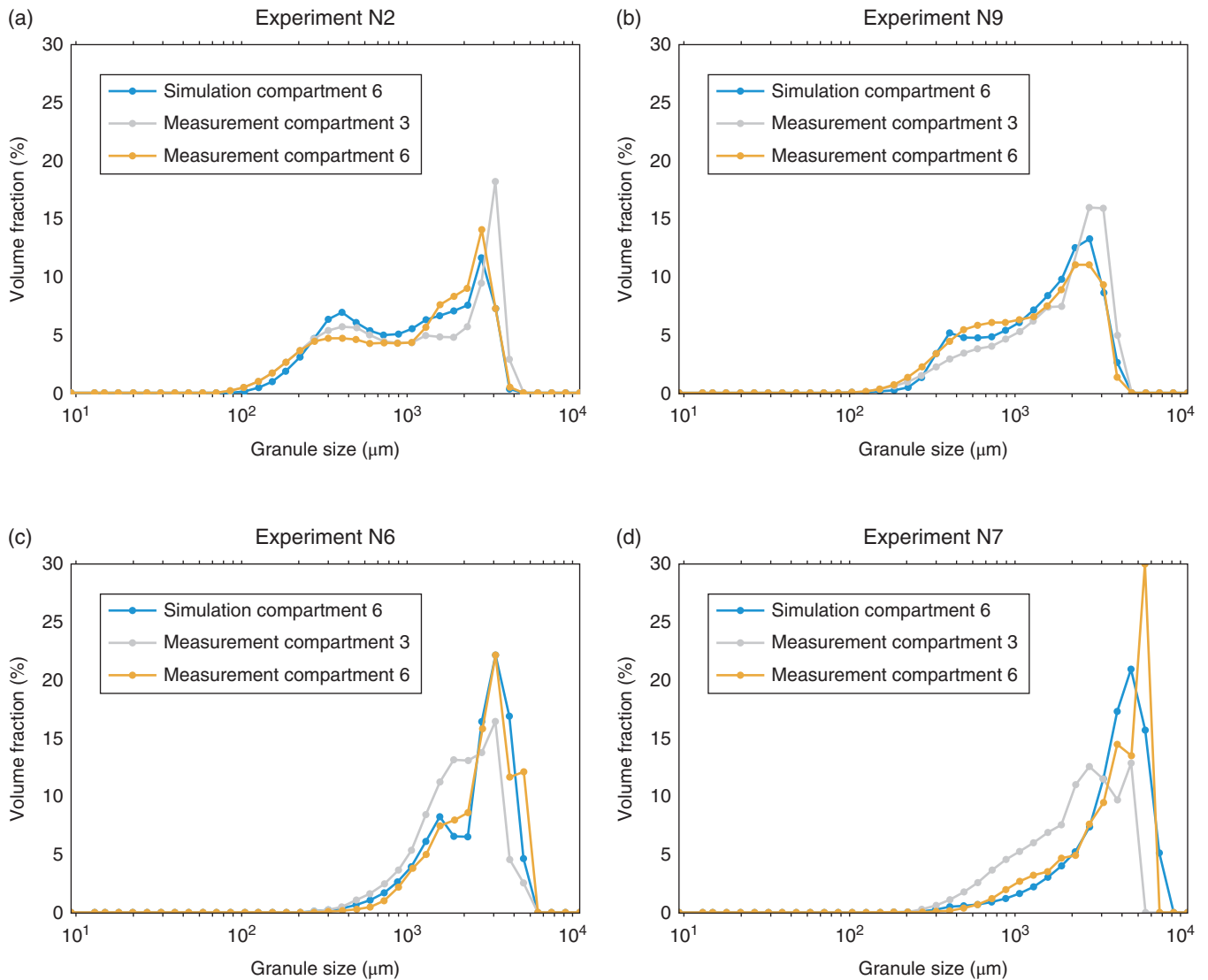


FIGURE 7.7 Calibrated simulation results for the second kneading zone 6 (hydrophobic model drug). GSD are visualized in volume fraction (%) as a function of granule size (μm). In this case, granule size is defined as the equivalent diameter of the granule. The process parameters can be found in Table 7.1, the calibrated model parameters are provided in Table 7.4.

TABLE 7.4 Calibrated Parameters for the Second Kneading Zone 6 for the Performed Experiments (Hydrophobic Model Drug)

Experiment	β_0	R_2	top_1	S_0	f_{prim}	SSE
N1	1	548	1.00E+02	1.00E-04	4.73E-03	36.56
N2	193	1845	3.73E+02	1.48E-04	1.75E-02	26.52
N3	14	698	3.73E+02	4.39E-05	1.75E-02	45.25
N4	193	1845	3.73E+02	1.30E-05	2.86E-01	64.33
N5	10 000	3000	1.00E+01	1.00E-05	5.00E-01	84.30
N6	2683	1845	5.18E+01	1.30E-05	1.75E-02	90.50
N7	10 000	14	1.00E+02	4.64E-05	5.00E-01	303.41
N8	1000	1282	1.00E+02	1.00E-03	1.06E-01	341.55
N9	10 000	1282	1.00E+01	1.00E-07	1.00E-03	17.89

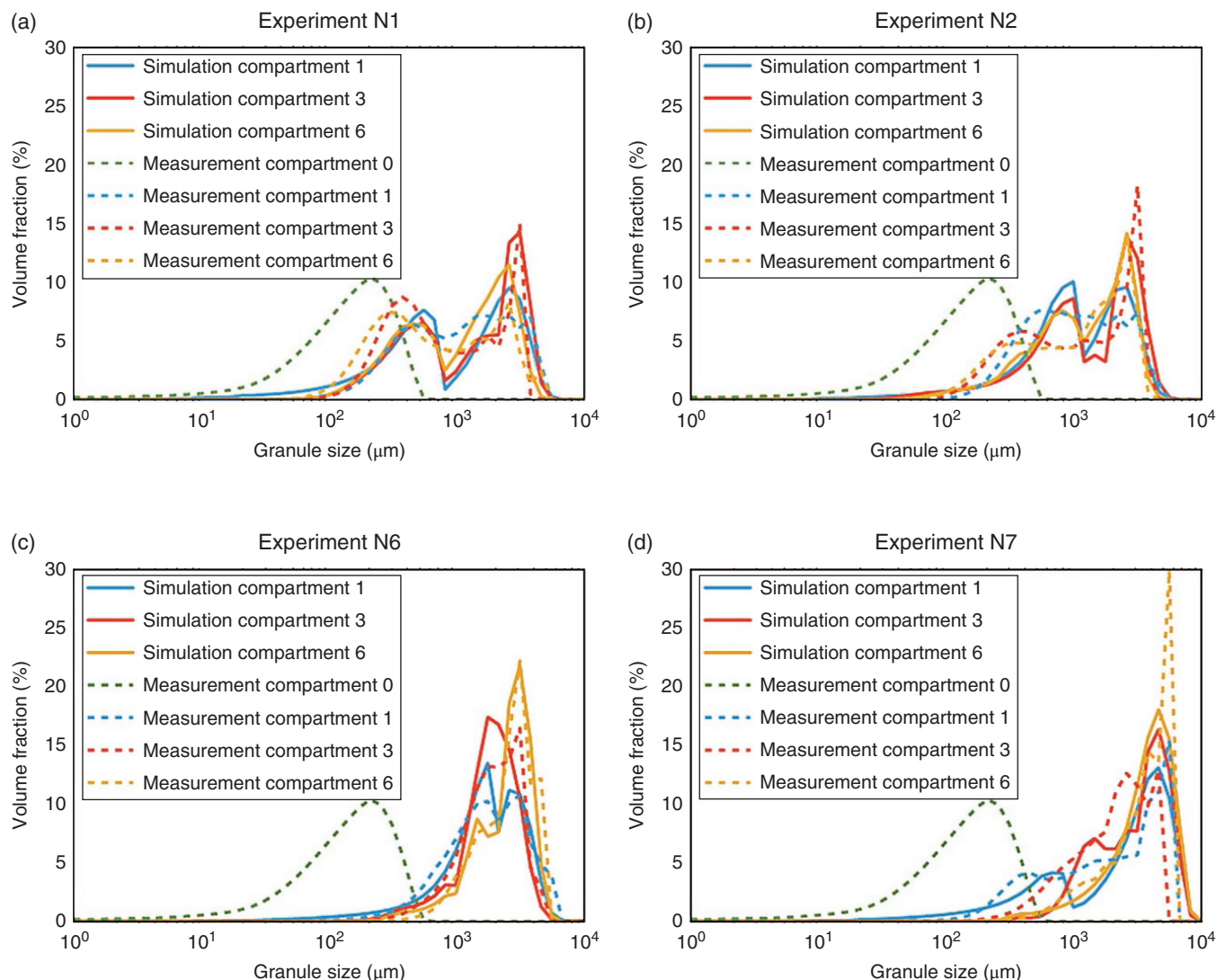


FIGURE 7.8 Simulation results for the whole compartmental model (hydrophobic model drug). The three single calibrated models (wetting zone 1, first kneading zone 3, and second kneading zone 6) are put in series without recalibration. GSD are visualized in volume fraction (%) as a function of granule size (μm). In this case, granule size is defined as the equivalent diameter of the granule. The process parameters can be found in Table 7.1, while the calibrated model parameters are provided in Tables 7.2–7.4. The table with the overview of the resulting SSE values (comparison of the output of the compartmental model and the measurements) can be found in Table 7.5.

the outlet of the granulator in a sufficient manner. A selection of the results is visualized in Figure 7.8a and d.

The overall performance of the model is quite satisfactory. The high L/S ratio experiments are predicted more accurately than low L/S counterparts. It can clearly be seen that imperfect calibration of single zones leads to a propagation of simulation errors through the whole model. However, the model appears to behave quite robust and manages to yield good predictions. The values of the objective function, the comparison of the simulated and the measured GSD at the outlet of the TSG, can be found in Table 7.5. The values in this table

support the claim that the model is quite robust: the error propagation is not excessive, as the SSE values are still in the same range as the values from the individual calibrated compartments.

7.10 OUTLOOK

This modeling effort is the first step toward a generalized mechanistic PBM for TSWG. The main advantage of this work is that it was able to use high quality data, obtained

TABLE 7.5 Objective Function Values for the Comparison of the Granules at the Outlet of the TSG for the Full Three-Compartment PBM (Hydrophobic Model Drug)

Experiment	SSE
N1	102.37
N2	50.00
N3	124.54
N4	88.67
N5	81.63
N6	303.18
N7	389.60
N8	121.67
N9	390.87

by [10], of in-barrel properties, whereas others only use inlet and outlet measurements. This high quality data was essential in the search of a representative kernel for the PBM that would not be possible to attain otherwise. After this first model iteration, some remarks can be made: first, the current kernel formulations can be improved in future work. The kernel structure could be adjusted to better mimic the behavior of the TSG, or some simplifications could be made to reduce the number of parameters, leading to a more straightforward calibration procedure. Indeed, in an industrial context, the GSD data is often only available for the preblend and at the end of the granulator. The possibility of calibrating the model using only this limited data and the obtained knowledge on the mechanisms in the different compartments needs to be investigated. Building on this, a clear link between the model parameters and operational conditions needs to be in place to attain this goal. The next step is to experimentally generate a validation data set and test the predictive power of the model. This would point out that the model structure is valid and is able to predict GSD in the whole design space within certain confidence intervals. Further it is essential to test this model with experimental data from a different formulation. As in this work, a hydrophobic formulation containing hydrochlorothiazide was used, a next logical step would be to calibrate a hydrophilic formulation and make the comparison between these two formulations. After validation with another formulation, the goal would be to use the model at the stage of data collection: predicting which experiments will yield the most information on the granulation behavior and thus reducing the number of required experiments.

REFERENCES

1. Vercruyse, J., Cordoba Diaz, D., Peeters, E. et al. (2012). Continuous twin screw granulation: influence of process variables on granule and tablet quality. *European Journal of Pharmaceutics and Biopharmaceutics* 82 (1): 205–211.
2. Fonteyne, M., Vercruyse, J., Diaz, D.C. et al. (2013). Real-time assessment of critical quality attributes of a continuous granulation process. *Pharmaceutical Development and Technology* 18 (1): 85–97.
3. Dhenge, R.M., Washino, K., Cartwright, J.J. et al. (2013). Twin screw granulation using conveying screws: effects of viscosity of granulation liquids and flow of powders. *Powder Technology* 238: 77–90.
4. Vercruyse, J., Burggraeve, A., Fonteyne, M. et al. (2015). Impact of screw configuration on the particle size distribution of granules produced by twin screw granulation. *International Journal of Pharmaceutics* 479 (1): 171–180.
5. Djuric, D. and Kleinebudde, P. (2008). Impact of screw elements on continuous granulation with a twin-screw extruder. *Journal of Pharmaceutical Sciences* 97 (11): 4934–4942.
6. Thompson, M. and Sun, J. (2010). Wet Granulation in a TwinScrew Extruder: implications of Screw design. *Journal of Pharmaceutical Sciences* 99 (4): 2090–2103.
7. Lee, K.T., Ingram, A., and Rowson, N.A. (2012). Twin screw wet granulation: the study of a continuous twin screw granulator using Positron Emission Particle Tracking (PEPT) technique. *European Journal of Pharmaceutics and Biopharmaceutics* 81 (3): 666–673.
8. Kumar, A., Vercruyse, J., Mortier, S.T.F.C. et al. (2016). Model-based analysis of a twin-screw wet granulation system for continuous solid dosage manufacturing. *Computers and Chemical Engineering* 89: 62–70.
9. Barrasso, D., Samjit, W., and Rohit, R. (2013). Multi-component population balance modeling of continuous granulation processes: a parametric study and comparison with experimental trends. *Powder Technology* 241: 85–97.
10. Verstraeten, M., Van Hauwermeiren, D., Lee, K. et al. (2017). In-depth experimental analysis of pharmaceutical twin-screw wet granulation in view of detailed process understanding. *International Journal of Pharmaceutics* 529 (1–2): 678–693.
11. Van Hauwermeiren, D., Verstraeten, M., Doshi, P. et al. (in press). On the modelling of granule size distributions in twin-screw wet granulation: calibration of a novel compartmental population balance model. *Powder Technology*, <https://doi.org/10.1016/j.powtec.2018.05.025>.
12. Kumar, J. (2006). Numerical approximations of population balance equations in particulate systems. Thesis. Magdeburg University.
13. Ramkrishna, D. and Fredrickson, A. (2000). *Population Balances*, vol. 1. Academic Press.
14. Mahoney, A.W. and Ramkrishna, D. (2002). Efficient solution of population balance equations with discontinuities by finite elements. *Chemical Engineering Science* 57 (7): 1107–1119.
15. Nicmanis, M. and Hounslow, M.J. (1996). A finite element analysis of the steady state population balance equation for particulate systems: aggregation and growth. *Computers & Chemical Engineering* 20 (96): S261–S266.
16. Rigopoulos, S. and Jones, A.G. (2003). *Finite-Element Scheme for Solution of the Dynamic Population Balance Equation*. Wiley.
17. Barrett, J.C. and Jheeta, J.S. (1996). Improving the accuracy of the moments method for solving the aerosol general dynamic equation. *Journal of Aerosol Science* 27 (8): 1135–1142.

18. Marchisio, D.L. and Fox, R.O. (2005). Solution of population balance equations using the direct quadrature method of moments. *Journal of Aerosol Science* 36 (1): 43–73.
19. Madras, G. and McCoy, B.J. (2004). Reversible crystal growth-dissolution and aggregation-breakage: numerical and moment solutions for population balance equations. *Powder Technology* 143–144: 297–307.
20. Kruis, F.E., Maisels, A., and Fissan, H. (2000). Direct simulation Monte Carlo method for particle coagulation and aggregation. *AIChE Journal* 46 (9): 1735–1742.
21. Lee, K. and Matsoukas, T. (2000). Simultaneous coagulation and break-up using constant-N Monte Carlo. *Powder Technology* 110: 82–89.
22. Lin, Y., Lee, K., and Matsoukas, T. (2002). Solution of the population balance equation using constant-number Monte Carlo. *Chemical Engineering Science* 57 (12): 2241–2252.
23. Smith, M. and Matsoukas, T. (1998). Constant-number Monte Carlo simulation of population balances. *Chemical Engineering Science* 53 (9): 1777–1786.
24. Bennett, M.K. and Rohani, S. (2001). Solution of population balance equations with a new combined Lax- Wendroff/ Crank-Nicholson method. *Chemical Engineering Science* 56 (23): 6623–6633.
25. Bove, S., Tron, S., and Hjertager, B.H. (2005). A novel algorithm for solving population balance equations: the parallel parent and daughter classes. Derivation, analysis and testing. *Chemical Engineering Science* 60 (5): 1449–1464.
26. Chang, S.-C. (1995). The method of space-time conservation element and solution element a new approach for solving the Navier-Stokes and Euler equations. *Journal of Computational Physics* 119 (2): 295–324.
27. Filbet, F. and Laurencot, P. (2004). Numerical simulation of the Smoluchowski coagulation equation. *SIAM Journal on Scientific Computing* 25 (6): 2004–2028.
28. Motz, S., Mitrović, A., and Gilles, E.D. (2002). Comparison of numerical methods for the simulation of dispersed phase systems. *Chemical Engineering Science* 57 (20): 4329–4344.
29. Verkoeyen, D., Pouw, G.A., Meesters, G.M.H., and Scarlett, B. (2002). Population balances for particulate processes – a volume approach. *Chemical Engineering Science* 57 (12): 2287–2303.
30. Kumar, J., Peglow, M., Warnecke, G. et al. (2006). Improved accuracy and convergence of discretized population balance for aggregation: the cell average technique. *Chemical Engineering Science* 61 (10): 3327–3342.
31. Tan, H.S., Salman, A.D., and Hounslow, M.J. (2004). Kinetics of fluidised bed melt granulation: IV. Selecting the breakage model. *Powder Technology* 143–144: 65–83.
32. Hardin, R.H. and Sloane, N.J.A. (1993). A new approach to the construction of optimal designs. *Journal of Statistical Planning and Inference* 37: 339–369.

MODELING A DOSATOR CAPSULE FILLING PROCESS FOR HARD-SHELL CAPSULES

PETER LOIDOLT

Institute for Process and Particle Engineering, Graz University of Technology, Graz, Austria

JOHANNES G. KHINAST

*Institute for Process and Particle Engineering, Graz University of Technology, Graz, Austria
Research Center Pharmaceutical Engineering GmbH, Graz, Austria*

EVA FAULHAMMER

Research Center Pharmaceutical Engineering GmbH, Graz, Austria

8.1 OVERVIEW OF THE DOSATOR CAPSULE FILLING PROCESS

Capsules are a widely used oral delivery system, which have several advantages over other dosage forms, such as tablets. While soft-shell capsules typically contain liquids, such as oils, hard-shell capsules are usually used to deliver powders or granules to patients. Hard-shell capsules consist of two halves: a lower-diameter body that is filled and a higher-diameter cap to subsequently seal it [1–3]. They mainly contain dry, powdered ingredients or small pellets. In addition, in clinical studies mini-tablets, pastes, and various combinations of materials can be filled into capsules to separate incompatible products or to achieve specific goals in terms of stability and modified release. Since the filling material must be compatible with the shell, no deliquescent or hygroscopic materials can be used. A range of capsule filling systems based on various technologies and principles are used in the pharmaceutical industry, including tamping-pin and dosator systems.

Depending on the powder density, various capsule fill weights can be obtained. The capsule sizes and the maximum fill volume range from 5 (smallest) to 000 (largest) and from 0.1 to 1.37 ml, respectively. Thus, up to 1.5 g of powder or pellets can be filled into capsules. An important current trend is to manufacture small doses (< 50 mg) of pure potent active pharmaceutical ingredient (API) for early research clinical

trials using the so-called drug-in-capsule approach and for inhalation purposes (dry powder inhalers, DPI) to treat respiratory (and other) diseases [4]. Another trend is the filling of high doses greater than 1 g for dietary supplements and nutraceuticals. Formulations that are relevant for capsule filling typically contain a combination of materials in order to achieve accurate dosing, good bioavailability, ease of filling and production, and product stability. Moreover, the product needs to appeal to patients. Powders must have good flowability and be nonadhesive yet compressible enough to form plugs if required [5–8]. Hard capsules can be filled in several ways, from manual preparation in the lab or pharmacy to fully automated industrial line.

Although capsule filling machines vary widely in their engineering design, the major difference between them is the dosing technique. This chapter focuses on the dosator nozzle filling process. In general, the major challenge is that doses have to be specified by weight, while the filling systems work based on a volumetric principle [3, 9]. Thus, while always the same volume is dosed into a capsule, the fill weight may vary due to natural variations in the bulk density of the powder. Dosator nozzle machines are comparable with rotary tablet press machines in terms of output (maximum output of 250 000 capsules per hour can be achieved). However, the force applied to the powder mass is much lower in a capsule filling process [9] than in tablet presses (a few

hundred Newtons). Easy to clean and simple in design, modern instrumented continuous production machines (e.g. by Harro Höfliger, IMA, Macofar, Matic, MG2, Zanasi, etc.) can fill a variety of dosage forms accurately into capsules.

8.1.1 Operating Principle of the Dosator Process

In dosator nozzle machines the powder bowl is a rotating container periodically fed with powder from a supply hopper. The powder bed height is kept constant by leveling via the metal blades. The dosator nozzles (1 in lab-scale machines; 2–32 in industrial-scale machines) are fixed on the turret, which has a smaller diameter than the powder bowl. Both the powder bowl and the turret rotate in the same direction. The dosator moves into the powder bed and collects the desired volume of powder from the powder layer (the dosing step is described in more detail in Figure 8.1). During dosing, (gentle) compaction is necessary to form a stable plug. Especially for inhalation products, the capsules are filled using a low degree of compaction or even without compaction to ensure that the plug readily disintegrates into a powder for efficient drug delivery. To achieve higher doses, more compaction is applied. The cylindrical volume (dosing chamber) is determined by the dosing chamber length and the dosator diameter. Inside the dosing chamber, there is a piston that can be used to control the level of compaction during the dosing step.

The critical quality attributes (CQA) of capsules include drug potency and variability, fill weight, and fill weight variability. Fill weight of capsules is controlled by adjusting the dosing chamber length and by varying the powder bed height

in the bowl. Even if the powder density is known, the fill weight can only be determined based on experience or statistical predictive models [10, 11]. The dosed mass of drug substance is the most important quality attribute, and a certain stability of the powder plug is required to avoid the loss of powder during the nozzle's movement from the powder bed to the capsule. Too strong plugs, however, may stick to the dosator nozzle during the ejection or disintegrate inadequately during administration. Unfortunately, the powder density is highly variable and depends on many factors, including the processing history, material properties, environmental conditions (such as humidity and vibrations), and process parameters during compression. Even the surface texture of the dosator tools has to be considered [12]. In order to save time and material, an accurate model of the powder properties and its interaction with the tools is required to predict the dosed mass.

A single dosing event consists of multiple steps, as depicted in Figure 8.1. First, the dosator is lowered into the rotary container to reach the top of the product layer. At this point, no stress is imposed on the layer except that due to gravitational forces (1). The piston inside the dosator nozzle is pushed up until the desired volume of the dosing chamber is achieved (volumetric principle). The volume of dosing chamber is constant while sampling (2). Second, the dosator nozzle dips into the powder bed and powder enters the dosing chamber until the nozzle reaches a prescribed minimum distance (the minimum gap) from the bottom (3). Due to (complex) compression of the powder during dipping, particles form a plug and can be lifted by the nozzle and filled into the capsule. If additional compression inside the dosator

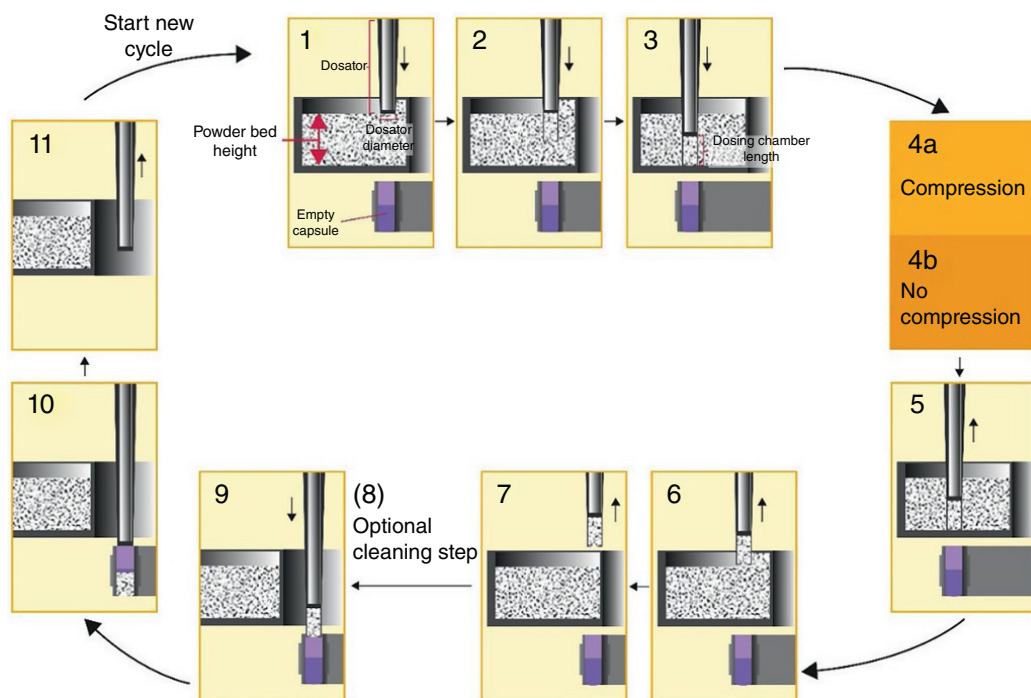


FIGURE 8.1 Powder dosing steps.

nozzle is required for a stable plug, compaction (i.e. inducing compressive stress by the piston) may be performed (4a) [13] by moving the internal piston downwards. For inhalation powders or very low amounts of high potent actives, no additional compression is performed (4b). The compression force should be low to minimize the effort during the subsequent plug ejection [14], when the dosator nozzle is withdrawn from the powder bed (5 and 6). The nozzle is then lined up with the empty capsule body (7). According to the dosing principle, the powder has to be retained in the dosing chamber while in motion. This relies on the ability to form a stable arch at the outlet, which in turn depends on the powder properties and the compression level [7, 8, 14–16]. Hence, the requirements for powders and granules to be used in dosator nozzle devices vary significantly from those for other machines. Many researchers showed that powder retention in the dosing chamber is influenced by cohesiveness of a powder [3, 12, 15, 17]. In low-density and highly cohesive powders, the arch formation occurs easily. In contrast, arch formation is problematic for dense and free-flowing powders [18, 19]. To minimize fill weight variability, a cleaning step can be added to remove the powder adhering to the outer wall or around the dosator nozzle tip (8), which is mainly required when highly potent APIs or inhalation powders are dosed. The downstroke of the piston ejects the powder or plug into the body of the capsule (9). Here, the wall friction has to be overcome [20, 21]. Finally, the upper part of the shell is fitted on the body and the closed capsule is ejected. The dosator rises (10) and a new capsule filling cycle begins (11).

8.1.2 Material Behavior Relevant for the Dosator Filling Process

During dosator filling the material to be dosed undergoes several stress situations. After feeding from the hopper into the

bowl, the powder is leveled (and thus compacted) by metal blades to ensure a uniform and stable powder layer. Since the dosator filling technique is based on volumetric principle, the powder layer and its inhomogeneity is one of the most critical factors for fill weight variability [10, 22]. Powder bed inhomogeneities will introduce variability in a volumetric filling process because fill weight will depend on the position of the nozzle sampling from the powder bed, thus significantly impacting the quality of the final product. Unknown or undetected density variations (channels, holes, compacted, loosened, or segregated areas) within the powder bed can significantly impact the CQAs of the product in terms of fill weight and fill weight variability, as well as reproducibility. Typical powder inhomogeneities encountered in capsule filling are highlighted in Figure 8.2.

In each dosing step, a material is periodically sampled via the dosator nozzle and filled into empty capsule bodies. This will create holes or channels in the powder bed, which need to be filled with the powder prior to the next dosing event. For this to happen, the powder needs to have a minimum flowability. It is highly important to control the density variation to achieve product specification compliance for each final product. Free-flowing powders typically lead to homogeneous powder layers and are more or less self-adjustable but are challenged by stable arch formation. However, for cohesive powders the powder bed surface is more uneven and can form cracks easily. Moreover, agglomerate formation occurs often. Poor flowability and high cohesion of such powders make the feeding of the bowl and the powder bed formation more challenging. High cohesivity may also lead to possible densifications, and, thus, capsule fill weight may increase over time for such powders.

Recently, it was shown that the fill weight correlates with the measured tapped densities of the powders, indicating that the powder layer densifies during the filling operations.

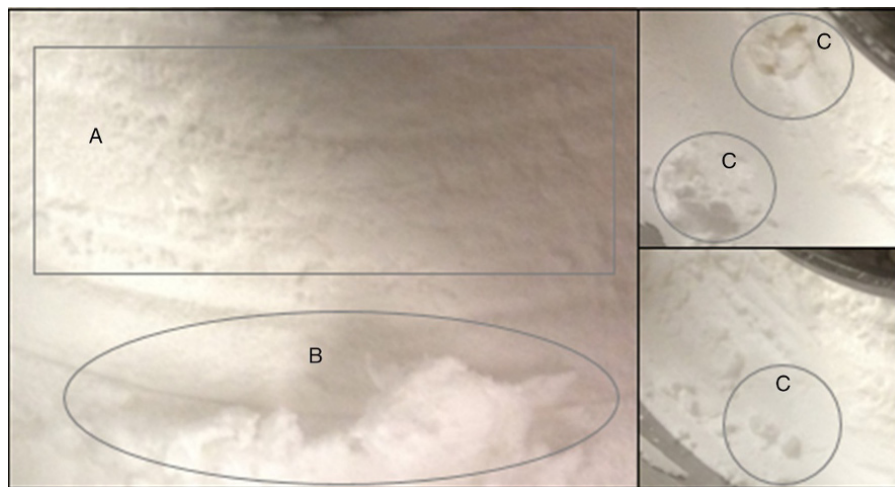


FIGURE 8.2 Top view of an inhomogeneous powder layer. Area A shows a typical layer of cohesive powder with surface roughness. Area B exhibits accumulation close to the container walls. Areas C depict holes or channels created by the dosator nozzle, which are not completely refilled due to poor powder flowability.

Moreover, a larger powder bed height leads to less homogeneous beds than a shallower one because of a significant amount of air entrapped in the powder layer, resulting in larger fill weight variability [22].

For cohesive materials, powder adhesion is also problematic. When filling capsules at low speed, the powder, especially a low-density and sticky one, can adhere to the outer dosator wall and several other machine parts. This powder is occasionally transferred into empty capsule bodies while dosing, thereby causing higher fill weight variations. Another phenomenon to be considered during capsule filling is electrostatic (tribo-)charging of the powder, which is generated when two different materials are brought together into contact and then are separated. Since drug and excipients consist of insulating material, they are unable to dissipate charges, which are consequently accumulated in any process where the powder comes into contact with surfaces [23].

Moreover, segregation can occur over longer processing time. During a dosator capsule filling process, the powder blend rotates in the powder container. This can possibly lead to segregation due to differences in powder density, particle mobility induced by the different shapes of particles, and, above all, the particle size [24]. Segregation results in poor blend uniformity and high dosage variability. Especially, when excipients with larger particle sizes are filled together with APIs with small particle sizes, segregation must be monitored throughout the process to ensure a high product quality.

Another important process parameter to be considered is the gap between the dosator tip at the end of the stroke and the bottom of the powder container. To determine the optimum gap for filling, significant experimental effort is required. Thus, computational approaches (e.g. Loidolt et al. [25]) are of great interest, as shown below.

Future research should be focused on PAT solutions that allow monitoring of the powder bed density over time. Especially, the identification of inhomogeneities within the powder bed, not detectable through visual inspection, could be of tremendous importance to reduce fill weight variability. Currently, no technology exists for such a problem.

8.1.3 Modeling Approaches of the Dosator Filling Process

Modeling the powder used in the dosator process is as important as modeling the process. To that end, in addition to the process parameters, the powder properties have to be considered. Due to the complexity of powder modeling, the influence of powder properties and process parameters on the capsule fill weight has been investigated mainly experimentally in the past [6, 22, 26–30]. Experimental observations are a good basis for model development since they provide an idea of the effect of the process parameters and powder properties on the filling behavior (e.g. that the densification of

powder during the dipping step increases with the powder compressibility).

One of the most important process parameters is the pre-compression ratio, which is defined as the ratio between the powder bed height and the dosator chamber length. The plug density has been observed to increase with the pre-compression ratio. Phenomenological models [31, 32] take the pre-compression densification into account. Other process parameters are difficult to investigate experimentally since the mechanism is not clear. For example, with regard to the influence of filling speed, it is uncertain if it is due to an increase in the machine vibrations densifying the powder bed, an increase in the dosator nozzle dipping speed, or a combination thereof. Likely, for different powders different effects dominate. The example above shows how difficult it is to understand all mechanisms involved in the dosator filling process only by performing experiments. Thus, detailed and mechanistic models are useful to separate the various influencing factors. For instance, it is possible to develop a model that only considers the dipping speed of the dosator nozzle and neglects any influence of machine vibrations.

To date, only a few modeling and simulation studies of the dosator capsule filling process have been performed. However, since powder flow and dosing of powder are important in many applications, they have been extensively studied. The discrete element method (DEM) discussed below is used for modeling the processes mechanistically. For example, there are numerous DEM models developed for die filling prior to tablet compaction or the production of compacted metal powders before sintering, called green bodies [33–35]. During die filling, the volume of the die is fixed in the same manner as the volume of the dosator chamber in the dosator process. Although the filling mechanism is different, a constant filling mass is crucial in both processes. Another process that involves the powder flow and dosing is coil feeding. With that regard, DEM can be used to model a coil feeder with a cohesive powder [36]. An important detail in the dosator process is how particles are packed inside the dosator nozzle. Especially for relatively big particles, such as pellets, packing behavior in the small dosator chamber is important. A similar problem was investigated by Rowe et al. [37], who simulated packing behavior of various pellet shapes inside a hard capsule via a Monte Carlo method. This study was later expanded by Ali et al. [38].

As mentioned above, stability of the powder plug during and after dosing is essential. In their theoretical work, Jolliffe et al. [14] established a criterion for minimum compression strength $\sigma_{c, \min}$ of a powder plug that is required to form a stable arch in a cylindrical chamber (Eq. 8.1). During steps 5–7 (see Figure 8.1), gravity is responsible for the collapse of the powder plug. In the case of rapid vertical acceleration of the nozzle, additional forces may occur. Jolliffe et al. [14] neglected these inertial forces. The minimum compression strength increases with the increasing radius of the dosator

nozzle (r) and the bulk density of powder (ρ_{bulk}). The minimum compression strength of the plug decreases when the angle of wall friction (φ_w) between the dosator wall and the powder increases.

$$\sigma_{c,\min} = \frac{r\rho_{\text{bulk}}g}{\sin(2\varphi_w)} \quad (8.1)$$

Unfortunately, neither the compression strength nor the bulk density is constant in a specific powder. During processing, the powder is compacted and the bulk density increases. Moreover, the compression strength increases during compaction. The flow function, ffc , (Eq. 8.2) is often used to determine the compression strength after compaction based on the main principal stress (σ_1) during compaction [39].

$$\sigma_c = \frac{\sigma_1}{ffc} \quad (8.2)$$

If the main principal stress during dosing and ffc of the powder are known, the compression strength can be estimated. A DEM model, which can predict the compaction pressure inside the dosator chamber in addition to the dosed powder mass, was proposed by Loidolt et al. [25]. It offers an opportunity to arbitrarily modify process parameters and powder properties, making it possible to observe quantities that are not measurable experimentally. Details of this model are provided below.

8.2 DEM MODEL OF THE DOSATOR PROCESS

This section describes the model proposed by Loidolt et al. [25], applying the DEM introduced by Cundall and Strack [40] to investigate the filling of a dosator nozzle. DEM is based on the soft-sphere approach, where particles are represented by (overlapping) spheres. The (hypothetical) overlap between two spheres (or other shapes) is used to calculate the contact forces in normal and tangential directions. The contact law describing the functional relationship between particle overlap and contact force is characteristic for every implementation of DEM model. Based on the contact force, the particle position and velocity are computed at every time step by solving Newton's second law. The particle rotation is not considered in our simulation since it is of minor importance in the densely packed beds applied in the dosator process. This way, the position of every particle in the powder can be tracked during the entire simulation. The open source DEM software LIGGGHTS[®] [41] was used for that purpose.

8.2.1 Model Geometry

Since the powder bed of a dosator machine contains too many particles (many billions) to be considered in a DEM model,

only a small section of the powder bed, 12 mm wide both in the x - and y -directions, was taken into account (the cross section is shown in Figure 8.3). To compensate for this limited size, the boundaries of the domain were set to be periodic, meaning that a particle leaving one side of the boundary enters on the opposite side. To simulate the dosator filling process, particles with nominal diameters of 500, 300, 200, 100, and 75 μm were allowed to settle on the fixed base in order to create a powder bed height of 8 mm. In this geometrical setup, up to about 3.5 million particles with a nominal diameter of 75 μm can be simulated. To avoid modeling artifacts due to monodisperse particles, a mass-based particle size distribution was applied (see Table 8.1). A rather narrow distribution precludes additional computational costs due to a high number of small particles. Since no ambient air is considered in DEM and the particles could accelerate indefinitely, a viscous damping force was applied to achieve a terminal velocity of 0.1 m/s. The limited terminal velocity avoids unrealistically high impact velocities, which could lead to artificial compaction of the powder bed. The diameter of the dosator nozzle d_D was 3.4 mm, which is a standard diameter also used in the experimental study of Faulhammer et al. [6]. The dosator chamber length l_D was set to 4 mm in all

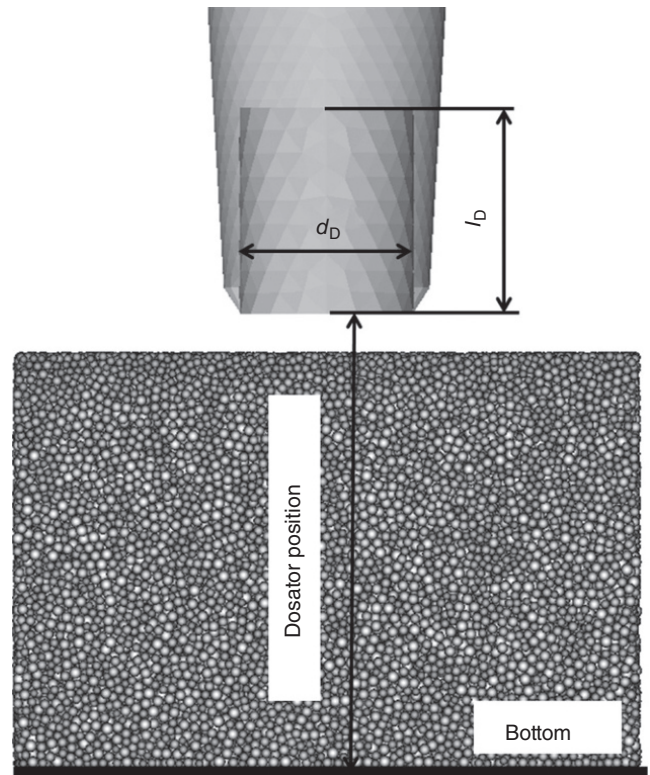


FIGURE 8.3 Cross section of the powder bed and the dosator nozzle. The position of nozzle is reported as distance to the powder bed floor.

TABLE 8.1 Particle Size Distribution

Relative Particle Diameter	0.8	0.9	1.0	1.1	1.2
Probability	10%	25%	30%	25%	10%

The diameters of five particle classes are given as fractions of the nominal particle diameter.

simulations. The distance from the nozzle to the powder bed floor was used to describe the dosator position (Figure 8.3).

In the simulation, the dosator nozzle moves vertically into the powder bed until it touches the bottom. As the nozzle moves, the amount of powder and the pressure inside are analyzed and plotted as a function of distance to the bottom. The pressure is defined as one-third of the trace of the stress tensor. To reduce the amount of data, the pressure inside the dosator is reported as the volume-averaged pressure inside the dosator chamber.

8.2.2 Contact Model

To model powders adequately, the mechanical particle properties have to be defined. Since lactose is one of the most common materials in pharmaceutical capsule filling applications, it was chosen as a reference material. Accordingly, the particle density was chosen as 1538 kg/m^3 . A Poisson ratio of 0.3 and a Young's modulus of 25 MPa were used as elastic constants. As in many DEM simulations, a comparably low Young's modulus was applied rather than the actual Young's modulus of lactose of about 25 GPa [42], reducing the high contact stiffness and, as a consequence, increasing the minimum simulation time step. To determine the effect of this modification on the dosator filling, the Young's modulus was varied, and it was concluded that the influence was insignificant in the case of small and moderate particle contact forces (for details see Ref. [25]).

The material properties directly influence the contact model. A well-known Hertz contact model (a standard model implemented in LIGGGHTS) can be used to compute contact normal force F_n as a function of particle normal overlap δ_n , effective particle radius R^* , and effective Young's modulus E^* (Eq. 8.3). The tangential contact force is a function of tangential overlap δ_t and effective shear modulus G^* (Eq. 8.4). It is limited by the Coulomb friction force, which is a function of coefficient of friction μ_s and contact normal force. The effective properties can be computed as a function of particle diameter and elastic constants of the two particles in contact (Eqs. 8.5–8.7). To enable energy dissipation, damping forces in the normal and tangential directions are computed based on the coefficient of restitution e (equations not shown here). More detailed information about the contact model can be found in [43] the work of Kloss.

$$F_n = \frac{4}{3} E^* \sqrt{R^*} \delta_n^{\frac{3}{2}} \quad (8.3)$$

$$F_t = \min \left\{ \begin{array}{l} 8G^* \sqrt{R^*} \delta_n \delta_t \\ \mu_s F_n \end{array} \right. \quad (8.4)$$

$$\frac{1}{E^*} = \frac{1-\nu_1^2}{E_1} + \frac{1-\nu_2^2}{E_2} \quad (8.5)$$

$$\frac{1}{R^*} = \frac{1}{R_1} + \frac{1}{R_2} \quad (8.6)$$

$$\frac{1}{G^*} = \frac{2(2-\nu_1)(1+\nu_1)}{E_1} + \frac{2(2-\nu_2)(1+\nu_2)}{E_2} \quad (8.7)$$

Powder cohesion is relevant for fine powders present in many pharmaceutical applications. To model cohesive powders, cohesion forces have to be incorporated into the DEM contact model. Powders are considered cohesive if the attractive cohesion forces between the particles or the particles and the walls are large compared with the body forces of the particles. A dimensionless number that describes the relative effect of cohesion is the granular Bond number Bo_g , which is defined as the ratio between the contact cohesion force and the particle gravity force (Eq. 8.8). To model powders with various Bond numbers, in addition to the default cohesionless contact model, a cohesion model was implemented in LIGGGHTS, which adds a constant attractive normal force to the contact if two particles touch each other. By adding various cohesion forces to the particle contacts, the Bond number can be varied by several orders of magnitude. Consequently, powder properties ranging from free-flowing (low granular Bond number) to very cohesive (high granular Bond number) can be adjusted, regardless of the particle diameter.

$$\text{Bo}_g = \frac{\text{cohesion force}}{\text{gravity force}} = \frac{F_c}{\frac{d^3 \pi \rho g}{6}} \quad (8.8)$$

Table 8.2 summarizes the reference model parameters, showing the particle density and the Poisson ratio of lactose. The chosen particle diameter and the Young's modulus are a compromise between the simulation time and the quality of the result. In the reference case, no cohesion is considered and the Bond number is zero. All other contact parameters for the reference case are chosen such that they represent typical properties of lactose particles.

A series of simulations using various particle diameters was performed to investigate the packing behavior inside the dosator nozzle as a function of particle size. To that end, the nominal diameter of particles filled into the nozzle was varied in the range from 75 to 500 μm . Studying particle size has several objectives. On the one hand, dosing big particles or granules most likely involves different mechanisms than those that are relevant for fine powders. On the other

TABLE 8.2 Model Parameters in the Reference Case

Particle Diameter d	200	μm
Particle density ρ	1538	kg/m^3
Young's modulus E	25	MPa
Poisson ratio ν	0.3	—
Coefficient of restitution e	0.45	—
Coefficient of friction (all contacts) μ_s	0.1	—
Bond number Bo	0	—

hand, the particle size is a critical model parameter with respect to computational effort. In case of very fine powders, to fill a certain volume, a very high number of particles have to be considered. Simulating the dosator process is only possible on a realistic time scale if the number of particles in the simulation is reduced without affecting the result by using bigger particles than those in the real process.

One benefit of the investigated model is the opportunity of arbitrarily adjusting the parameters. To gain knowledge about the underlying mechanisms, the coefficient of friction can be varied and, based on the simulations, additional insights into the process may be obtained. The friction dependence may be especially relevant for regions with high shear rates inside the powder bed or particles sliding close to the wall of the dosator bed floor.

The flow behavior of various powders can be modeled based on the relative importance of cohesion forces versus gravity forces. The cohesion force can be adjusted to obtain Bond numbers in the range from 0 to 10 000. A Bond number of $Bo_g = 0$ refers to powders with negligibly small cohesive forces compared to the particle volume forces (e.g. dosing of big particles, such as granules). A Bond number of $Bo_g = 10\,000$ indicates fine powders in which cohesion forces dominate. To estimate which particle size range corresponds to which Bond number, cohesion force measurements can be performed. The cohesion force of microparticles was investigated by Baldelli and Vehring [44]. To obtain a Bond number of 10 000, the particle diameters has to be approximately $10\ \mu\text{m}$. By studying the range of low to high Bond numbers, the transition from free-flowing powders to very cohesive powders can be investigated.

Table 8.3 shows bulk density ρ_{bulk} and relative density ρ_R inside the powder bed after settling (poured density) as a function of Bond number. The relative density of the powder bed is defined as the ratio between the bulk density and true density ρ_T of dense particle material (see Eq. 8.9). At low Bond numbers, the relative density is 0.625, which is close to the relative density (0.64) of randomly close-packed monodisperse spheres. At the highest Bond number considered, the relative density is 0.322. In the study of Bond number, the nominal particle diameter was maintained constant at $200\ \mu\text{m}$ and only the cohesion force was adjusted to obtain the desired Bond number.

TABLE 8.3 Bulk Density and Relative Density of the Powder Bed as a Function of Bond Number Bo_g

Bond Number Bo_g	Bulk Density (kg/m^3)	Relative Density
0	961	0.625
10	955	0.621
100	908	0.590
500	780	0.507
1000	693	0.451
2000	614	0.399
5000	537	0.349
10 000	495	0.322

$$\rho_R = \frac{\rho_{\text{bulk}}}{\rho_T} \quad (8.9)$$

8.3 MODELING RESULTS FOR THE DOSATOR FILLING PROCESS

8.3.1 Evolution of Powder Mass and Pressure Inside the Dosator Nozzle

A cross section of the dosator nozzle at three positions as it dips into the powder bed with various powder properties is shown in Figure 8.4. The first row represents the reference case of cohesionless particles and a $200\ \mu\text{m}$ nominal diameter. The second row highlights the case with the same cohesionless contact properties as in the reference case, but with a nominal diameter of $500\ \mu\text{m}$. Rows 3 and 4 depict the filling of cohesive powders, as discussed below. Particle diameters are reported as well in Figure 8.5. For all graphs, the contact parameters of the reference case are used, except for the particle diameter, which is provided in the inset.

When plotting the dependence of the dosed mass as a function of distance from the bottom, the dosator filling process can be divided into three stages. These three stages mechanistically describe step 2 and step 3 in Figure 8.1. The filling of dosator nozzle begins at a distance of $8\ \text{mm}$ from the powder bed floor (powder bed height) and ends when the dosator nozzle touches the bottom of the powder bed floor. Hence, the feeding process always commences on the right side at the maximum distance and ends at zero distance from the bottom (left side of each graph). During Stage I, the dosator chamber is filled at a constant rate as a function of the traveling path regardless of the particle diameter. There is practically no pressure acting on the powder (and thus no compaction), and thus the first stage may be considered volumetric filling, i.e. the mass is directly proportional to the filled volume. At (intermediate) Stage II, the nozzle moves further into the powder bed and the particles touch the piston on top of the chamber. The amount of powder inside the

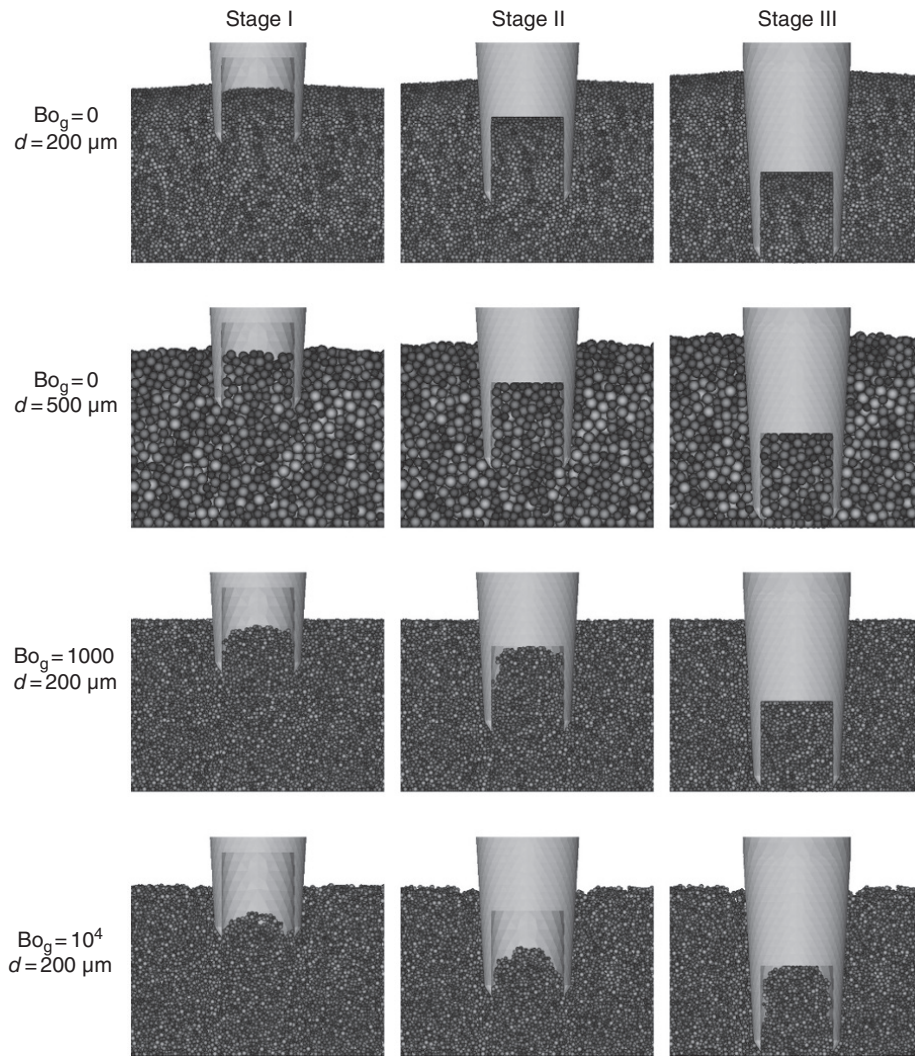


FIGURE 8.4 The three stages of dosator filling for various nominal particle diameters and Bond numbers. Stage I and Stage II describe step 2 in Figure 8.1, while Stage III corresponds to step 3.

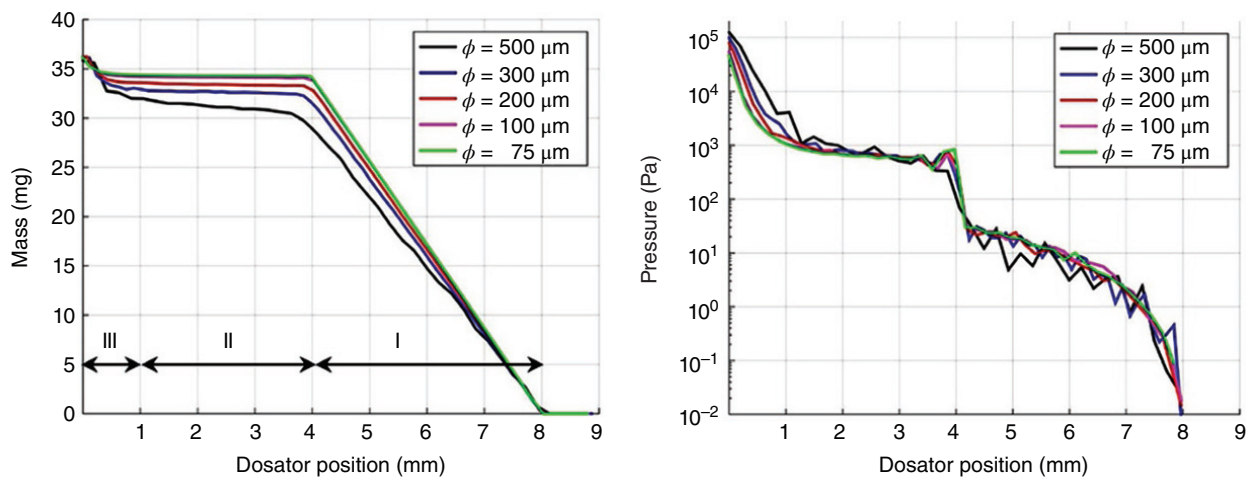


FIGURE 8.5 Powder mass and average pressure inside the dosator as a function of distance to the bottom of powder bed for various nominal particle diameters.

nozzle is nearly constant. The pressure inside the nozzle increases sharply as soon as the piston inside the nozzle contacts the dosed powder. Finally, at Stage III the nozzle moves closer than 1 mm to the bottom and the mass sharply increases again. Since the pressure inside the nozzle rises by a factor of up to 100 compared to the second stage, the third stage is clearly the relevant compression step. Thus an important conclusion is that very high pressure inside the dosator nozzle (i.e. a high level of compression) is only observed at the last stage of dosing. Note that such information cannot be obtained (or only with excessive effort) via experiments. Consequently, the minimum gap formed during the deepest penetration of the nozzle has a very strong impact on the plug stability and dispersibility and is one of the most sensitive (yet easily controllable) process parameters. If a process produces powder plugs with insufficient plug stabilities, the nozzle should simply be moved closer to bottom to enhance the powder compression. Alternatively, additional compression could be applied via the movable piston inside the nozzle. If the powder plug is too stable, the penetration depth has to be reduced.

Concerning the effect of particle diameter, the three snapshots of 500 and 200 μm particles are not much different (Figure 8.4). Moreover, the qualitative progression of the powder inside the nozzle (Figure 8.5) is very similar for all particle diameters regardless of significant quantitative differences. Although the initial relative density of all particle diameters is approximately 0.62 inside the powder bed, after the volumetric filling stage bigger particles result in a lower powder mass inside the nozzle. Since particle packing close to and far away from the walls is quite different, this can be attributed to the wall effect, which is more relevant if the ratio between the particle size and the size of geometry is high. Similarly, a study by Zou and Yu [45] showed the porosity

increased with the increasing particle diameter for particles packed into a cylindrical container.

In Stage II, the mass inside the dosator remains constant regardless of the nominal particle diameter, although there is a small increase in the case of 500 μm particles. On average, the pressure inside the nozzle is nearly the same for all particle diameters during the first and second stages of filling. The fluctuation of pressure is higher for bigger particle diameters since there are fewer of bigger particles inside the nozzle. At compression Stage III, the pressure is significantly higher for bigger particles and more powder is squeezed into the nozzle, which explains why the total amount of powder inside the nozzle at the end of the process is nearly the same for all particle diameters despite being lower for bigger particles after the volumetric filling phase. However, with a small gap of 0.5–1 mm at the end of filling, the powder with larger particles results in a reduced fill weight (up to 5–10%).

Various stages of dosator filling were studied experimentally via force measurements by Pinzon [46]. In these experiments, the same nozzle diameter was used as in the current study and a sharp increase in the force acting on the nozzle was measured when it was within 1–2 mm from the powder bed floor. This indicates high pressure acting on the powder inside the nozzle.

Interestingly, the fill masses converge for small particle diameters (Figure 8.5), i.e. below a certain particle-to-dosator-size ratio, the size effects are no longer noticeable. What matters is the ratio between the particle and the dosator nozzle size, as also reported by Zou and Yu [45] and Suzuki et al. [47]. Since the filling behavior of 200 μm particles (used in the reference case) is similar to that of 75 μm particles, 200 μm was a good choice for the reference case.

The compression stage and the effect of particle diameter should be considered in more detail. Figure 8.6 (left) shows

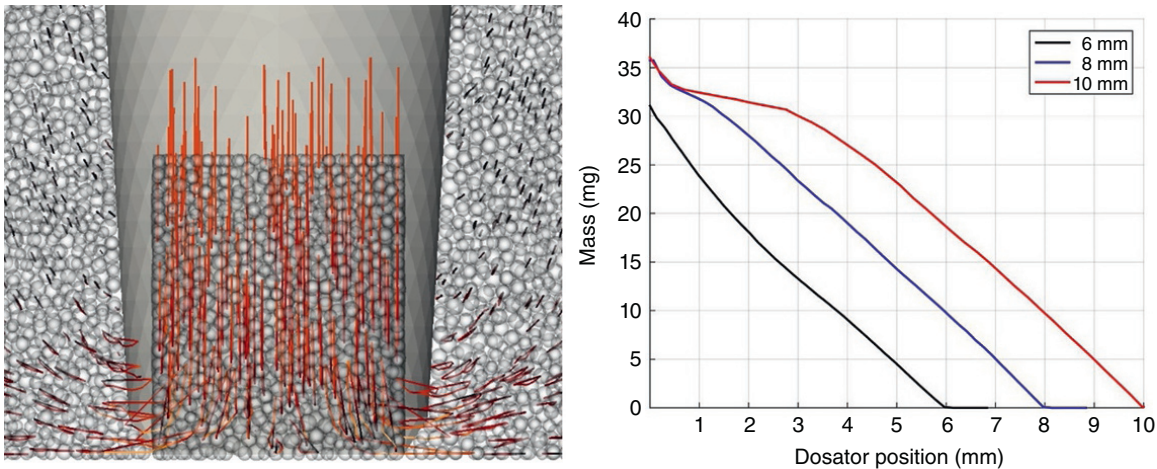


FIGURE 8.6 Streamlines of particles during the filling process in the reference case (left). The velocity is proportional to the length of streamlines. In the case of cohesive powder ($Bo_g = 1000$), the three stages of dosing no longer exist and the powder bed height becomes important (right).

the streamlines of the particles during Stage III. Clearly, the particles can exit the space below the nozzle if their diameter is smaller than the gap between the dosator nozzle and the floor. If the particle diameter is bigger than the gap, the particles are trapped inside the nozzle. If the nozzle moves closer to the bottom, the particles are highly compressed, as it can be seen at the end of Stage III. Since this effect is more pronounced for bigger particles, they are more compressed at Stage III. However, there is another mechanism causing compression at Stage III. In fact, compression begins with the gap size of 1 mm, which is far bigger than any particle diameter. At a higher coefficient of friction, this effect is even more pronounced due to resistance to the shear flow, which occurs when the powder is squeezed through a small gap close to the bottom for frictional particles (Figure 8.6 (left)). In real systems, both compression mechanisms described above are relevant. The mechanism of trapped particles is exclusively responsible for compression at Stage III only in a frictionless contact model.

The three stages of dosator filling do not exist in the case of cohesive powder (Figure 8.6 (right)) at a Bond number of 1000 and 3 powder bed heights. The mechanisms responsible for this behavior are discussed in the Section 8.3.2.

8.3.2 Transition from Free-Flowing to Very Cohesive Powder

The filling behavior of cohesive particles with various Bond numbers is shown in Figure 8.7. The relative density inside the dosator is plotted instead of total mass of the powder. Hence, the bulk density can be computed for various true densities of the particles and the graph is valid for various particulate materials. The inset shows the relative density of initial powder bed. Table 8.3 demonstrates how the relative density correlates to the Bond number. The highest relative density of 0.625 refers to a Bond number of zero (cohesionless), while

the lowest relative density of 0.322 corresponds to a Bond number of 10 000. The relative density of initial powder bed can be used to characterize various powders since it can easily be determined experimentally based on the bulk density and the true density, unlike the Bond number that is difficult to determine for real powders. The contact properties are the same as in the reference case (i.e. $\rho_R = 0.625$). With the increasing cohesion, which decreases the initial relative density, the rate of mass accumulation inside the dosator changes dramatically. At $\rho_R = 0.590$, the plots look similar to the those for cohesionless particles, i.e. three distinct phases can be observed. However, at a relative density of 0.507, the three stages of dosator filling are less pronounced and completely vanishing at lower initial bed densities when the mass inside the dosator increases continuously during the entire movement of the nozzle through the powder bed. The amount of powder during volumetric filling decreases due to looser packing of cohesive powder beds, which clearly increases the compressibility of cohesive powders and is responsible for the different nozzle filling behavior. In addition, another effect shown in the third and fourth row of Figure 8.4 emerges. At the Bond number 10 000, the dosator chamber is not completely filled even if the upper end of the chamber is lower than the level of powder bed. The powder cannot enter the dosator nozzle and (similarly to arches inside hoppers preventing the hopper outflow) an arch is formed hindering the dosator filling. The dosator can only be filled when the nozzle comes closer to the powder bed floor and the powder compressed into it. As such, the total mass inside the nozzle at the bottom of powder bed is much lower in the case of very strong cohesion (the Bond number of 10 000) and the pressure in the nozzle is significantly lower.

The effect of arching is not as pronounced at a Bond number of 1000 ($\rho_R = 0.451$), when the dosator is filled volumetrically with some cavities and a lower relative density than in the cohesionless case (Figure 8.4, Stage II). After volumetric

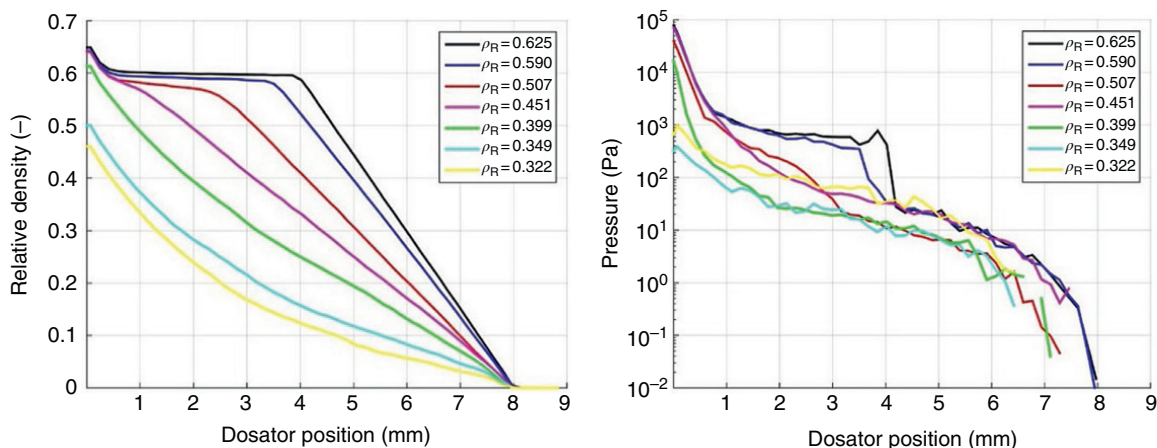


FIGURE 8.7 Dosed mass and average pressure as a function of distance from the bottom for 200 μm particles with various Bond numbers.

filling, the powder is continuously compressed until the nozzle touches the powder bed floor. In this way, the relative density inside the nozzle at the bottom of powder bed and the pressure inside the powder for the Bond number 1000 and the cohesionless powder are identical (Figure 8.7). The important consequences of differences in the evolution of dosed mass for cohesive and non-cohesive powders are as follows:

- For non-cohesive powders, the total powder bed height is of minor importance. If the powder bed height increases, Stage II expands and the dosed mass remains nearly constant.
- For cohesive powders, Stage II does not really occur and the dosed mass increases with an increase in the powder bed height. This is a mechanistic explanation of why the influence of powder bed height is more important for cohesive powders, as experimentally observed, e.g. in Faulhammer et al. [6]. Figure 8.6 (right) shows the evolution of dosed mass for the Bond number 1000 and various powder bed heights (the powder bed height is provided in the inset). If the nozzle stops 2 mm above the bottom, the dosed mass is different in all three cases. When it is closer than 1 mm to the bottom, the result is the same for 8 and 10 mm powder bed heights. A powder bed height of 6 mm is not sufficient to completely fill a 4 mm long dosator.

8.3.3 Rational Design of the Dosator Process Based on DEM Simulations

The model described above generates a basic understanding of the capsule filling process via a dosator system and uses the computational method DEM, which can simulate the dosing process and track the amount of powder inside the nozzle as a function of the nozzle's penetration depth (or the distance to the bottom of the bowl) during dosing for different process and material parameters. The filling behavior of coarse and fine powders and cohesive and non-cohesive powders can be compared. While the wall effects can explain the influence of particle size, the effect of cohesion is more complex. Non-cohesive and cohesive powders have different filling mechanisms, which can mechanistically explain the influence of powder bed height on the fill weight of cohesive and non-cohesive powders observed in the experimental studies.

For non-cohesive powders, since the amount of dosed powder is not affected by the total powder bed height, keeping the powder bed height constant in an industrial process is not required. Unlike in a simulation, in the real process the nozzle cannot be moved all the way to the powder bed floor. Rather, a certain gap must be preserved to avoid damaging

the machine. Since the minimum distance to the powder bed floor is important for achieving the maximum pressure acting on the powder inside the nozzle, it is essential to keep the gap constant. Only at constant pressure, a constant density and strength of the dosed plug can be expected.

For cohesive powders, the amount of dosed powder depends on the total powder bed height and the smallest gap between the nozzle and the powder bed floor. They form a powder bed with a lower relative density and are easier to compress. If the nozzle moves through the powder bed, the powder is continuously densified. Very cohesive powders may form arches, and it is therefore difficult to fill the nozzle. To avoid arching, a dosator chamber with bigger diameter and lower height may be used (which is similar to a wider hopper opening used to prevent arching during the hopper outflow). A smaller diameter may be applied if the powder plug is not stable enough, as indicated by Eq. (8.1).

For a rational design, at the very minimum the dosed mass and the stability of the dosed plug have to be considered. Although the dosed mass of drug substance is the most important quality attribute (and its variability), a certain stability of the powder plug is required to avoid the loss of powder during the nozzle's movement from the powder bed to the capsule. However, an overly stable plug may stick to the dosator nozzle during the ejection or disintegrate inadequately during application, i.e. in a DPI. The powder plug strength depends on both the powder being dosed and the process conditions. The dosed mass can directly be determined based on Figure 8.7 for a given relative density of powder bed and the smallest gap between the dosator nozzle and the powder bed floor (since a stable powder plug is assumed, the powder that enters the nozzle remains in the nozzle). Evaluating the powder plug stability requires an additional step. According to Eq. (8.2), the compression strength can be determined based on the main principal stress during compression and flow function ffc determined via shear experiments. The pressure during dosing determined in the DEM simulations is a lower bound of the main principal stress and a good approximation of the existing problem. Figure 8.7 (right) shows the maximum pressure during dosing as function of the smallest gap between the nozzle and the powder bed floor for various initial relative densities of the powder bed. The minimum compression strength can be determined via Eq. (8.1) for a certain diameter of the nozzle, the bulk density of powder and the angle of wall friction.

8.4 CASE STUDY: PREDICTION OF OPTIMAL INSERTION DEPTH INTO THE POWDER BED DURING A DOSATOR FILLING PROCESS

In the following case study, a dosator filling process should be designed to achieve appropriate powder plug strengths for various powder properties (Table 8.4). Moreover, since

TABLE 8.4 Powders with Different Flow Properties and Bulk Densities to be Dosed Using the Same Dosator System

Powder	Bulk Density ρ_{bulk} (kg/m ³)	True Density ρ_{T} (kg/m ³)	Flow Function Coefficient, ffc (-)	Angle of Wall Friction φ_{w} (°)
1	960	1530	200	8
2	975	1700	40	10
3	725	1450	10	12
4	720	1800	3	19

experiments were not possible due to a limited amount of powder, the dosed mass should be estimated based on the simulation data. The results of shear experiments, indicating the rheological behavior of the powder and the poured bulk density, are provided. The dosator nozzle has a diameter of 3.4 mm. For calculating the dosator volume, a dosator chamber length of 4 mm should be considered. The powder bed has a height of 8 mm to ensure sufficient filling of the dosator nozzle during its movement through the powder bed.

8.4.1 The Task

Based on the powder properties provided and the simulation results in Figure 8.7, the following problems should be solved:

- Determine the minimum compression strength for achieving stable powder plugs for the powders in Table 8.4. As an approximation, take the initial powder bed density into account.
- Determine the main principal stress σ_1 required for every powder to achieve the minimum compression strength.
- How deep does the nozzle have to penetrate into the powder bed to compact the various powders sufficiently? The pressure inside the nozzle can be accepted as the main principal stress since the exact main principal stress is not known and the pressure is the lower bound of the main principal stress. Choose which powder bed density curve in Figure 8.7 is closest to the individual bulk density.
- Report the dosed mass if the dosator nozzle stops in the position that yields the minimum compression strength on the powder plug.
- Is it possible to obtain a stable powder plug for every powder if a minimum gap of 0.5 mm is required to avoid any contact between the dosator nozzle and the bottom of dosator bowl? In other words, is it necessary to perform additional powder compaction using the internal piston of the dosator nozzle?

- How much will the dosed mass be changed if the dosator nozzle is moved 0.5 mm further than required for a stable powder plug? Is there a difference in the sensitivity of the dosator traveling path for various powders?
- How close can the nozzle be to the powder bed floor to avoid a compression strength higher than 500 Pa?
- Which powder is easiest to dose based on the considerations above?

8.4.2 Solution

Table 8.5 summarizes the results for the case of minimum compaction strength required for stable powder plugs for the powders 1–4. Relative density of initial bed $\rho_{\text{R, bed}}$, minimum compression strength $\sigma_{\text{c, min}}$, and minimum compaction pressure p_{min} are obtained from Eqs. (8.1), (8.2), and (8.9). Figure 8.8 (top) can be used to determine the dosator position required for minimum pressure z_{pmin} . Therefore, the curves with the most similar relative density to the four powders can be applied to evaluate the position required for their minimum compaction pressure. Based on this position, Figure 8.8 (bottom) can be used to determine the relative density inside the dosator nozzle $\rho_{\text{R, pmin}}$, based on which bulk density $\rho_{\text{bulk, pmin}}$, and the mass inside the dosator at the minimum compaction pressure can be computed. For powder 1, the nozzle should be moved very close (0.33 mm) to the bottom of the powder bed, which is closer than the permitted 0.5 mm. In this case, the nozzle should stop 0.5 mm away from the bottom and the piston inside the dosator has to compress the powder to achieve a sufficient powder strength.

Table 8.6 shows the results of sensitivity analysis. Position z_2 is 0.5 mm below the position for the minimum pressure. Since this position is below the bottom in the case of powder 1, it is omitted in the table. For the other powders, the relative density is obtained based on the dashed lines in Figure 8.7 (bottom). The corresponding dosed mass is compared to the case with the minimum strength. Powders 3 and 4 have the smallest and biggest variations, respectively. This can be concluded based on the different slopes of the curves in the position where the variation is tested. Obviously, a small variation can be obtained if the lowest position of the nozzle can be chosen at the plateau of the curve.

Table 8.7 provides a summary of the results for the maximum allowed compression strength. The maximum pressure during dosing is very different for the different powders and cannot be reached for powder 1. All other powders can be compressed beyond the maximum strength if the nozzle is moved closer than 0.25–0.35 mm to the bottom. The graphical evaluation of the maximum pressure position can be applied to the minimum pressure and is not shown.

TABLE 8.5 Results for the Minimum Compression Strength of the Powder Plug

Powder	$\rho_{R,bed}$ (-)	$\sigma_{c,min}$ (Pa)	p_{min} (Pa)	Position $z_{p_{min}}$ (mm)	$\rho_{R,p_{min}}$ (-)	$\rho_{bulk,p_{min}}$ (kg/m ³)	$m_{dosator,p_{min}}$ (mg)
1	0.63	58	11 617	0.33	0.605	926	33.6
2	0.57	48	2377	0.65	0.595	1012	36.7
3	0.50	30	297	1.7	0.57	827	30.0
4	0.40	20	59	1.55	0.445	801	29.1

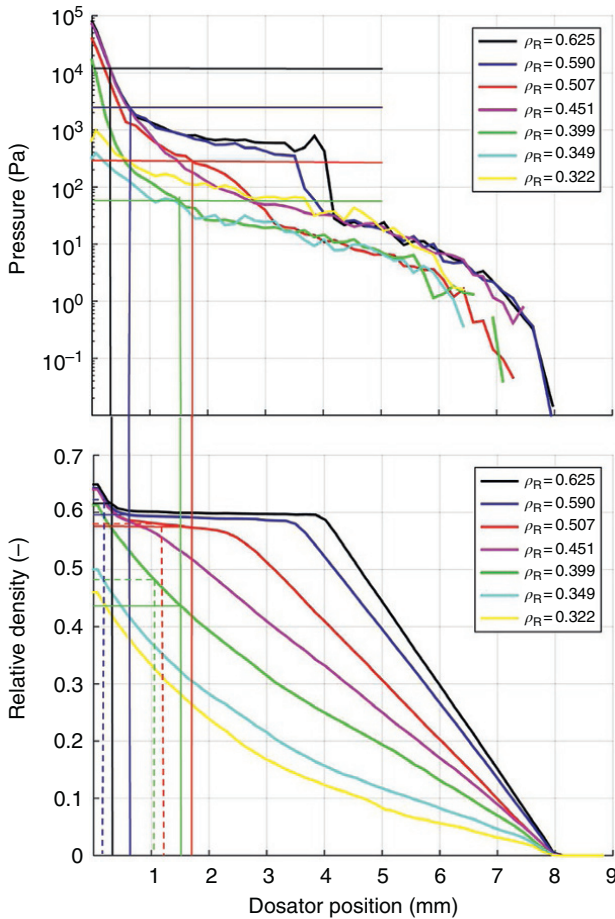


FIGURE 8.8 Graphical evaluation of dosator position and relative density inside the dosator for the minimum compaction pressure.

Concerning the challenge of dosing the various powders, a measure describing how easy or difficult dosing of a certain powder is should be defined. The goal is to achieve a stable process that produces powder plugs of constant mass with a limited powder strength. Hence, since the variation in the dosed mass is smallest, powder 3 is the easiest to dose. Moreover, the widest range of the possible dosator nozzle positions for producing stable but not overly stable plugs is beneficial.

TABLE 8.6 Results of the Sensitivity Analysis

Powder	Position z_2 (mm)	$\rho_{R,2}$ (-)	$\rho_{bulk,2}$ (kg/m ³)	$m_{dosator,2}$ (mg)	$\frac{m_{dosator,2}}{m_{dosator,p_{min}}}$
2	0.15	0.62	1054	38.3	1.04
3	1.2	0.58	841	30.5	1.02
4	1.05	0.48	864	31.4	1.08

TABLE 8.7 Results for the Maximum Compression Strength of 500 Pa

Powder	p_{max} (Pa)	Position $z_{p_{max}}$ (mm)
1	100000	—
2	25000	0.25
3	5000	0.35
4	1500	0.3

8.5 SUMMARY

The dosator filling process is widely used in the pharmaceutical industry to fill powders into hard capsules. The dosed mass and its variability are the most important quality attributes of the capsule products. The process was mainly investigated experimentally in the past. In our contribution, a novel model for the dosator filling process was introduced to describe the filling of the dosator nozzle during moving through the powder bed. The DEM was used for that purpose. Different contact properties of the DEM model describe free-flowing and cohesive powders. Different bulk densities of the powder bed also can be achieved. Different filling behaviors were observed for free-flowing and cohesive powders during simulation. As we could show, the minimum distance of the dosator nozzle to the bottom of the powder bed floor is a very important process parameter. It influences the mass, as well as the stability, of the dosed powder plug. A case study was designed to find the optimal minimum distance of the dosator nozzle to the bottom for different powder properties.

REFERENCES

1. Augsburger, L.L. (2009). Hard- and soft-shell capsules. In: *Modern Pharmaceutics Volume 1 Basic Principles and Systems* (ed. A.T. Florence and J. Siepmann), 499–565. Informa Healthcare USA, Inc.

2. Cole, G.C. (1999). The design and operation of a facility for filling hard shell gelatin capsules. Capsugel library [online].
3. Jones, B.E. (2001). The filling of powders into two-piece hard capsules. *International Journal of Pharmaceutics* 227: 5–26. [https://doi.org/10.1016/S0378-5173\(01\)00781-5](https://doi.org/10.1016/S0378-5173(01)00781-5).
4. Ashurst, I., Malton, A., Prime, D., and Sumbly, B. (2000). Latest advances in the development of dry powder inhalers. *Pharmaceutical Science & Technology Today* 3: 246–256.
5. Allen, L.V., Popovich, N.G., and Ansel, H.C. (2011). Capsules. In: *Ansel's Pharmaceutical Dosage Forms and Drug Delivery Systems* (ed. D.B. Troy), 203–224. Lippincott Williams & Wilkins a Walters Kluwers Business.
6. Faulhammer, E., Llusa, M., Radeke, C. et al. (2014). The effects of material attributes on capsule fill weight and weight variability in dosator nozzle machines. *International Journal of Pharmaceutics* 471: 332–338. <https://doi.org/10.1016/j.ijpharm.2014.05.058>.
7. Felton, L.a., Garcia, D.I., and Farmer, R. (2002). Weight and weight uniformity of hard gelatin capsules filled with microcrystalline cellulose and silicified microcrystalline cellulose. *Drug Development and Industrial Pharmacy* 28: 467–472. <https://doi.org/10.1081/DDC-120003008>.
8. Guo, M., Muller, F.X., and Augsburger, L.L. (2002). Evaluation of the plug formation process of silicified microcrystalline cellulose. *International Journal of Pharmaceutics* 233: 99–109.
9. Podczec, F. and Jones, B.E. (2004). Dry filling of hard capsules. In: *Pharmaceutical Capsules* (ed. B.E. Jones and F. Podczec), 119–138. London: Pharmaceutical Press.
10. Faulhammer, E., Fink, M., Llusa, M. et al. (2014). Low-dose capsule filling of inhalation products: critical material attributes and process parameters. *International Journal of Pharmaceutics* 473: 617–626. <https://doi.org/10.1016/j.ijpharm.2014.07.050>.
11. Faulhammer, E., Llusa, M., Wahl, P.R. et al. (2015). Development of a design space and predictive statistical model for capsule filling of low-fill-weight inhalation products. *Drug Development and Industrial Pharmacy* 1–10. <https://doi.org/10.3109/03639045.2015.1040416>.
12. Jolliffe, I.G. and Newton, J.M. (1983). The effect of dosator wall texture on capsule filling with the mG2 simulator. *Journal of Pharmacy and Pharmacology* 35: 7–11.
13. Tan, S. and Newton, J. (1990). Minimum compression stress requirements for arching and powder retention within a dosator nozzle during capsule filling. *International Journal of Pharmaceutics* 63: 275–280. [https://doi.org/10.1016/0378-5173\(90\)90134-P](https://doi.org/10.1016/0378-5173(90)90134-P).
14. Jolliffe, I.G., Newton, J.M., and Walters, J.K. (1980). Theoretical Considerations of the filling of pharmaceutical hard gelatin capsules. *Powder Technology* 27: 189–195.
15. Jolliffe, I.G. and Newton, J.M. (1983). Extension of theoretical considerations of the filling of pharmaceutical hard gelatin capsules to the design of dosator nozzles. *Powder Technology* 35: 151–157. [https://doi.org/10.1016/0032-5910\(83\)87003-X](https://doi.org/10.1016/0032-5910(83)87003-X).
16. Jolliffe, I.G. and Newton, J.M. (1982). An investigation of the relationship between particle size and compression during capsule filling with an instrumented mG2 simulator. *Journal of Pharmacy and Pharmacology* 34: 415–419.
17. Britten, J.R., Barnett, M.I., and Armstrong, N.a. (1996). Studies on powder plug formation using a simulated capsule filling machine. *Journal of Pharmacy and Pharmacology* 48: 249–254.
18. Tan, S.B. and Newton, J.M. (1990). Influence of compression setting ratio on capsule fill weight and weight variability. *International Journal of Pharmaceutics* 66: 273–282. [https://doi.org/10.1016/0378-5173\(90\)90408-V](https://doi.org/10.1016/0378-5173(90)90408-V).
19. Tan, S.B. and Newton, J.M. (1990). Observed and expected powder plug densities obtained by a capsule dosator nozzle system. *International Journal of Pharmaceutics* 66: 283–288. [https://doi.org/10.1016/0378-5173\(90\)90408-V](https://doi.org/10.1016/0378-5173(90)90408-V).
20. Tan, S.B. and Newton, J.M. (1990). Capsule filling performance of powders with dosator nozzles of different wall texture. *International Journal of Pharmaceutics* 66: 207–211. [https://doi.org/10.1016/0378-5173\(90\)90401-O](https://doi.org/10.1016/0378-5173(90)90401-O).
21. Tan, S.B. and Newton, J.M. (1990). Influence of capsule dosator wall texture and powder properties on the angle of wall friction and powder-wall adhesion. *International Journal of Pharmaceutics* 64: 227–234. [https://doi.org/10.1016/0378-5173\(90\)90273-7](https://doi.org/10.1016/0378-5173(90)90273-7).
22. Stranzinger, S., Faulhammer, E., Calzolari, V. et al. (2017). The effect of material attributes and process parameters on the powder bed uniformity during a low-dose dosator capsule filling process. *International Journal of Pharmaceutics* 516: 9–20. <https://doi.org/10.1016/j.ijpharm.2016.11.010>.
23. Yu, Z.-Z. and Watson, K. (2001). Two-step model for contact charge accumulation. *Journal of Electrostatics* 51–52: 313–318. [https://doi.org/10.1016/S0304-3886\(01\)00071-7](https://doi.org/10.1016/S0304-3886(01)00071-7).
24. Cullen, P.J. (2009). *Food Mixing: Principles and Applications*. John Wiley & Sons.
25. Loidolt, P., Madlmeir, S., and Khinast, J.G. (2017). Mechanistic modeling of a capsule filling process. *International Journal of Pharmaceutics* 532: 47–54. <https://doi.org/10.1016/j.ijpharm.2017.08.125>.
26. Llusa, M., Faulhammer, E., Biserni, S. et al. (2014). The effects of powder compressibility, speed of capsule filling and pre-compression on plug densification. *International Journal of Pharmaceutics* 471: 182–188. <https://doi.org/10.1016/j.ijpharm.2014.04.073>.
27. Llusa, M., Faulhammer, E., Biserni, S. et al. (2013). The effect of capsule-filling machine vibrations on average fill weight. *International Journal of Pharmaceutics* 454: 381–387. <https://doi.org/10.1016/j.ijpharm.2013.07.029>.
28. Patel, R. and Podczec, F. (1996). Investigation of the effect of type and source of microcrystalline cellulose on capsule filling. *International Journal of Pharmaceutics* 128: 123–127.
29. Podczec, F. and Newton, J.M. (2000). Powder and capsule filling properties of lubricated granulated cellulose powder. *European Journal of Pharmaceutics and Biopharmaceutics* 50: 373–377. [https://doi.org/10.1016/S0939-6411\(00\)00100-4](https://doi.org/10.1016/S0939-6411(00)00100-4).
30. Podczec, F. and Newton, J.M. (1999). Powder filling into hard gelatine capsules on a tamp filling machine. *International Journal of Pharmaceutics* 185: 237–254. [https://doi.org/10.1016/S0378-5173\(99\)00169-6](https://doi.org/10.1016/S0378-5173(99)00169-6).

31. Khawam, A. (2011). Modeling powder encapsulation in dosator-based machines: I. Theory. *International Journal of Pharmaceutics* 421: 203–209. <https://doi.org/10.1016/j.ijpharm.2011.10.021>.
32. Khawam, A. and Schultz, L. (2011). Modeling powder encapsulation in dosator-based machines: II. Experimental evaluation. *International Journal of Pharmaceutics* 421: 210–219. <https://doi.org/10.1016/j.ijpharm.2011.09.027>.
33. Bierwisch, C., Kraft, T., Riedel, H., and Moseler, M. (2009). Die filling optimization using three-dimensional discrete element modeling. *Powder Technology* 196: 169–179. <https://doi.org/10.1016/j.powtec.2009.07.018>.
34. Guo, Y. and Wu, C.Y. (2016). Computational modeling of pharmaceutical die filling processes. *Predictive Modeling of Pharmaceutical Unit Operations* 253–271. <https://doi.org/10.1016/B978-0-08-100154-7.00011-9>.
35. Tsunazawa, Y., Shigeto, Y., Tokoro, C., and Sakai, M. (2015). Numerical simulation of industrial die filling using the discrete element method. *Chemical Engineering Science* 138: 791–809. <https://doi.org/10.1016/j.ces.2015.09.014>.
36. Imole, O.I., Krijgsman, D., Weinhart, T. et al. (2016). Experiments and discrete element simulation of the dosing of cohesive powders in a simplified geometry. *Powder Technology* 287: 108–120. <https://doi.org/10.1016/j.powtec.2015.07.051>.
37. Rowe, R.C., York, P., Colbourn, E.A., and Roskilly, S.J. (2005). The influence of pellet shape, size and distribution on capsule filling – a preliminary evaluation of three-dimensional computer simulation using a Monte-Carlo technique. *International Journal of Pharmaceutics* 300: 32–37. <https://doi.org/10.1016/j.ijpharm.2005.05.007>.
38. Ali, A.M., de Matas, M., York, P., and Rowe, R.C. (2009). Influence of pellet aggregate populations on the variability of pellet filling into hard shell capsules: a comparison of experiment and computer simulation. *European Journal of Pharmaceutical Sciences* 38: 197–205. <https://doi.org/10.1016/j.ejps.2009.07.001>.
39. Schulze, D. (2010). Flow properties of powders and bulk solids (fundamentals). *Powder Technology* 65: 321–333.
40. Cundall, P.A. and Strack, O.D.L. (1979). A discrete numerical model for granular assemblies. *Géotechnique* <https://doi.org/10.1680/geot.1979.29.1.47>.
41. Kloss, C., Goniva, C., Hager, A. et al. (2012). Models, algorithms and validation for opensource DEM and CFD-DEM. *Progress in Computational Fluid Dynamics* 12: 140–152. <https://doi.org/10.1504/PCFD.2012.047457>.
42. Roberts, R.J. and Rowe, R.C. (1991). The relationship between Young's modulus of elasticity of organic solids and their molecular structure. *Powder Technology* 65: 139–146. [https://doi.org/10.1016/0032-5910\(91\)80176-J](https://doi.org/10.1016/0032-5910(91)80176-J).
43. Kloss, C. (2016). LIGGGHTS(R)-PUBLIC documentation, Version 3.X [WWW Document].
44. Baldelli, A. and Vehring, R. (2016). Analysis of cohesion forces between monodisperse microparticles with rough surfaces. *Colloids and Surfaces A: Physicochemical and Engineering Aspects* 506: 179–189. <https://doi.org/10.1016/j.colsurfa.2016.06.009>.
45. Zou, R.P. and Yu, A.B. (1995). The packing of spheres in a cylindrical container: the thickness effect. *Chemical Engineering Science* 50: 1504–1507. [https://doi.org/10.1016/0009-2509\(94\)00483-8](https://doi.org/10.1016/0009-2509(94)00483-8).
46. Pinzon, A., (2012). PhD thesis Modelling of Dosator Filling and Discharge. Greenwich Acad. Lit. Arch.–Univ. og Greenwich open access Repos. University of Greenwich.
47. Suzuki, M., Shinmura, T., Imura, K., and Hirota, M. (2008). Study of the wall Effect on particle packing structure using X-ray micro computed tomography. *Advanced Powder Technology* 19: 183–195. <https://doi.org/10.1163/156855208X293817>.

9

POWDER COMPACTION: PROCESS DESIGN AND UNDERSTANDING

DAVID WILSON, RON ROBERTS, AND JOHN BLYTH

AstraZeneca, Macclesfield, UK

9.1 INTRODUCTION

The compressed tablet is the most common pharmaceutical dosage form on the planet, and consequently tablet compaction is the second most common unit operation in pharmaceutical product process engineering after dry powder blending [1–3]. With numerous high value products reliant upon the success of this unit operation for the production of patient critical medicines, understanding the equipment and mechanics of this process should be a key skill for the process engineer in the pharmaceutical industry.

The success of the compressed tablet in pharmaceutical settings can be attributed to a number of reasons: (i) it presents a uniform and discrete dosage to the patient; (ii) it is available in commercially desirable images with ease; (iii) it reduces safety handling concerns as compacted tablets versus powders where the active pharmaceutical ingredient may be toxic; (iv) it takes advantage of the superior chemical stability of the solid state; and (v) it is cheap and fast to manufacture with the potential to be scaled up to a production rate of many 100 000s of dosage units per hour.

Due to their ubiquity across the entire pharmaceutical industry, the design of a well-controlled and scalable tablet compression unit operation is of vital importance to the chemical engineer working in the pharmaceutical industry. This chapter seeks to provide insights into the technical detail of compaction and to provide a useful guide to process design. Furthermore, this chapter seeks to summarize all current thinking as to the material changes occurred during

compaction and review in short the latest modeling techniques that can be used to approach the unit operation.

9.1.1 Duty of the Compaction Step and Success Criteria

In order to successfully design the tableting unit operation, thought must be given to the overall design brief and specification requirements of the solid dosage form. Broadly speaking this can be split into two categories: regulatory requirements and manufacturability concerns.

Regulatory specifications are different depending on the phase of the candidate in development, tablet dose, and active pharmaceutical ingredient (API) physical properties, but typical commercial requirements are given in Table 9.1 and manufacturability concerns are given in Table 9.2

The ability of a tablet unit operation to be optimized for production rate or tablet tensile strength is highly related to the success of any prior unit operation such as granulation, lubrication, and indeed excipient selection. Consequently, the optimization of a tableting unit operation cannot occur in isolation from the rest of the production process train.

9.2 MECHANICS OF TABLET COMPRESSION: SIMPLE BEGINNINGS

The basic act of making a tablet has remained relatively unchanged for over 150 years. It is accomplished by applying

TABLE 9.1 Typical Regulatory Specifications or Critical Quality Attributes for Commercial Supply

Regulatory Specification	Comment
Assay	Should contain an acceptable level of API
Uniformity of content	Should demonstrate a minimal amount of API content variation
Uniformity of weight	Can be used as a substitute for uniformity of content measurements if the API % in the tablet is high
Appearance	Minimal defects (cracks, surface imperfections, etc.) should be apparent
Dissolution performance	Should give assurance that the required biopharmaceutical performance will be met

TABLE 9.2 Typical Tablet Manufacturability Concerns and Interdependencies

Manufacturability Concern	Comment
Tablet tensile strength	Tablets must be mechanically robust enough to survive onward coating/packaging – often linked to appearance specification
Tablet porosity	Linked to the disintegration rate of the tablet and consequently may be rate limiting for dissolution
Tablet production rate	Should be a commercially viable process. A function of the required batch size. 100 000+ tablets/h is not an uncommon target
Required compression force	Excessive force may result in bent tablet punches/slow compression

sufficient force to the formulated blend or granules inside a fixed die between two punches to achieve an intact compact. This technique was first utilized by British inventor William Brockedon who adapted his pencil lead compaction equipment to produce compacts of other powders, without the use of liquid gum or adhesive. His patent in 1843 for “Shaping Pills, Lozenges and Black Lead by Pressure in Dies” was hailed as an important moment for the pharmaceutical industry by review authors [3].

9.3 TYPES OF COMPRESSION EQUIPMENT

Due to the widespread use of tableting processes, there is a wealth of manufacturers and equipment to choose from. The types of tablet press available cover a wide range of applications from small-scale research and development (R&D) to large-scale operations environments, and consequently the choice of equipment depends on a number of

factors, such as budget, output, quality requirements, instrumentation, and product containment. A number of the most typical categories of presses are discussed further below, arranged in order from small to large scale.

9.4 COMPACTION SIMULATION

During early formulation development, material constraints generally mean it is unlikely that products can be tested on high-speed commercial tablet presses. For this reason, scale-up from a low-speed to a high-speed rotary tablet press, particularly for materials of higher strain rate sensitivity, may be of high risk. Improved understanding of some material properties important to compaction (such as yield pressure or tensile behavior, but crucially not bulk properties such as powder flow and other powder handling metrics) may be achieved using a relatively small amounts of material. These machines can accurately control a set of punches at high speeds and can mimic the movement of a punch through a compression cycle of a much faster tablet press. Tablet production is typically in the range of 10–20 tablets/min. Most simulators are able to emulate complex compression cycles, including pre-compression, main compression, ejection phases, and even multilayer tablets. Compaction simulation models are available from Huxley Bertram and Phoenix.

9.5 SINGLE PUNCH TABLET PRESS

These machines are ideally suited for research organizations. They enable early screening of formulation prototypes, e.g. product characterization and small manufactures for assessment of product stability. Typical speeds of approximately 30–60 tablets/min mean that output is not a key requirement for this type of press. The main benefit is the speed of assembly and changeover, as well as the use of small amounts of material, which is often a key consideration in early stages of development. Due to the very low speed of compaction, interpretation of compression profile data derived from single punch presses may not be completely applicable to the performance of a formulation on a high-speed rotary machine for production use. Typical manufacturers and models include the “minipress” by Riva and the XP1 by Korsch.

9.6 SINGLE-SIDED AXIAL ROTARY TABLET PRESS

This category of machine is still by far the most common type of tablet press available today and meets the requirements of most products. Within this there are perhaps further subcategories that relate to production output. Typically low-speed rotary presses might have outputs between 200 and 2500

tablets/min, whereas high-speed rotary presses may easily produce 2000–5000 tablets/min or more. This increase in output is achieved by simply increasing the turret diameter and number of punches.

The slower models are ideal for product development, scale-up, and small batch production for clinical trials. The faster models are generally utilized in commercial environments where output and efficiency become more important for larger batches. Typical models include numerous machines by Korsch, Fette, and Riva.

9.7 ALTERNATIVE ROTARY TABLET PRESS CONFIGURATIONS

Variations on the above rotary press such as multiple layer presses are relatively common in certain industries. The production of bilayer and trilayer tablets is achieved by the introduction of multiple powder inputs and duplicated compression rollers positioned equidistantly around the press. In addition to giving the capability of producing multilayer tablets when filling with multiple products, these presses also give the advantage of significantly increasing output when using a single product.

The traditional tablet press configurations described above account for the majority of tablet presses available on the market, with gravity feeding of material into the die pocket and/or force feed via a series of rotating star wheels. An additional family of tablet presses utilizes the centrifugal force generated by the tablet press itself to fill the dies radially from an internal chamber within the turret through channels connected to the die. This innovative technique may offer benefits for some products both in terms of flow and output [4]. Such machines are commercially available from the IMA company, who suggest a production rate of 5000 tablet/minute for their “Comprima 300” machine.

9.8 THE COMPRESSION CYCLE

For the purposes of this section, the following description follows the journey of material step by step through a compression cycle within a typical single-sided axial rotary machine, depicted in Figure 9.1; equipment specific variations will apply.

1. Material is introduced to the supply hopper of the compression equipment either manually or via an intermediate bulk container (IBC) or a blending vessel mounted above.
2. The lower punch is drawn down by the fill cam as it moves under the feeding assembly – due to tight tolerances in punch and die design, this process is vacuum

assisted. The fill cam determines the volume of powder contained in the die cavity by lowering the bottom punch to a particular depth. Choice of fill cam should be made to ensure an overage of material enters the die relative to the weight required by the dosage form taking into account the bulk density of the material.

3. Product flows from the supply hopper into the feeding assembly. This may be a simple gravity feeder, which restricts the flow of material sufficiently to allow metering of the die holes without flooding the compression chamber, or a more complex force feeder assembly, which draws material in using a series of wheels and gears and distributes material in a controlled way into the die hole. The material is retained by the lower punch tip, which always remains within the die table. The set of an upper punch, lower punch, and a die is called a “station” and they are positioned equidistantly around the turret. It is common practice to use blank dies to reduce the number of active stations. This is utilized especially in R&D environments or for small batches of high value to reduce losses and maximize yield.
4. The weight control cam then lifts the lower punch to the required height to achieve the correct mass of dosage unit. Excess powder is scraped off to a central recirculation channel and guided around the turret before returning to the feed frame (excess should be minimized by correct choice of the fill cam). The dosed material is often contained by a “tail over die” or cover that is held in place against the die table to prevent any material from leaving the die before the upper punch enters the die.
5. The lower punch is then further lowered slightly to a position ready for compression, and the upper punch enters the die to fully enclose the material.
6. The punches pass between the pre-compression rollers (or ramp on some models) to remove entrapped air and consolidate material before the main compression event.
7. The punches then pass between the main compression rollers, and a given force is applied to the material to achieve acceptable tablet properties. The upper and lower rollers may be fixed or adjustable, depending on equipment model.
8. After the main compression event, the upper punch is raised to clear the feeder ready for the next cycle. The lower punch is then pushed up by the ejection cam until the tip is slightly proud of the die table. This enables the tablet to be swept from the lower punch and down the ejection chute for subsequent collection and testing.
9. The cycle returns to the start and tablets are continuously produced.

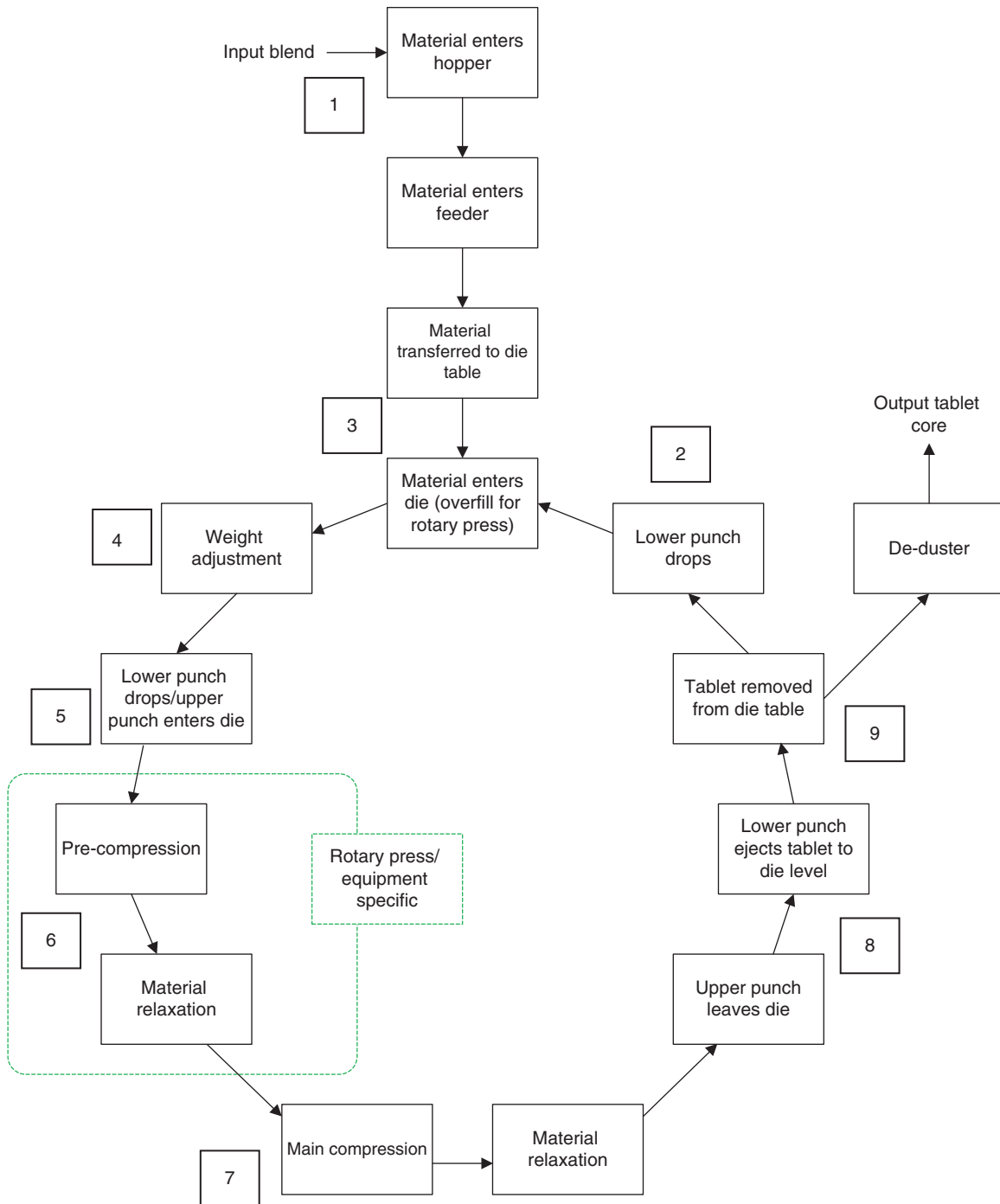


FIGURE 9.1 Compression cycle of an axial rotary tablet machine.

9.9 MAIN FEATURES OF A TABLET PRESS

Given that the above compression cycle describes the production of a single tablet from just one station, it is clear that when you multiply this by the number of stations and the speed of the machine, a lot is happening at any one time. It is therefore useful to know what to do, when to do it, and why you should do it. Here we look in more detail at some

of the key considerations and measurable features of a conventional rotary tablet press.

9.9.1 Punches

The tablet punch is perhaps the most important equipment item when designing and manufacturing a solid dosage form. While the tablet weight may be fixed by the required dose, the

actual tablet geometry can vary dramatically; therefore an appropriate tablet shape should be selected. For example, larger tablets may require elongated or oval dimensions to aid in patient swallowing; increased concavity may be required for older patients for ease of handling.

Punch material of construction can be important for some powders or formulations that are either abrasive or prone to sticking, picking, or filming [5, 6]. In these instances the surface finish of the hardened steel punch tip may be modified with an applied coating, of which there are a wide array, including titanium and chromium nitrides.

9.9.2 Dwell Time

A common measure of compaction processes is dwell time, which is the amount of time that the punch head flat is in contact with the compression rollers, where the compression force applied is greater than 90% of the peak value. This measure is critical for scale-up where press speed and linear punch velocities increase, and therefore dwell times decrease. Critical factors that can increase dwell time for problematic materials are increasing the punch head flat diameter, introducing pre-compression force, and reducing press speed.

9.9.3 Fill Cam

An appropriate cam must be selected to ensure sufficient fill in the die to give the correct weight tablet. Too shallow a cam means that sufficient material cannot fill the die to give the correct tablet weight. Too deep a cam means that excessive material is returned back to the feeder after weight adjustment; this has the potential to cause over lubrication of material in the feeder. In general a fill cam that provides a reasonable overage of volume, e.g. 20%, should give enough flexibility to accommodate any changes in material density throughout the batch.

9.9.4 Turret Speed

This is the speed that dictates how many revolutions the die table travels, and therefore the tablet output, in a given period of time. Knowing the batch size and the number of active stations (taking into account the number of blanks if utilized), an appropriate speed should be chosen to maximize efficiency while trying to minimize losses generated at set-up.

9.9.5 Feeder Speed

In an ideal world the feeder should draw material in at the same rate at which material leaves the press. Too slow and the material does not have time to fill the moving die; too fast the material may be retained in the feeder and become over lubricated, a potential risk for some products. In practice this balance is not necessarily that simple to achieve and small adjustments may be made to the speed while observing the

effect on material flow, force uniformity, and excess material in and around the press particularly in the recirculation channel. In this way the flow in and out of the feeder may be optimally balanced.

9.9.6 Weight Adjustment

This is essentially dictated by the lower punch depth at the final point of filling and is one of the first parameters to establish during the compression process. This is what dictates the correct fill volume for material at start-up and therefore the critical quality attribute of tablet weight. Adjustment to the lower punch depth should be made until the correct weight is achieved. This process can be greatly assisted with measurement of the bulk density of the powder to be compressed and the geometry of the punches. Knowledge of these parameters will allow for the correct fill volume to be set prior to any tablet production. For some presses this depth may be self-adjusted by the machine as the compression run progresses depending on main force feedback.

9.9.7 Pre-compression Force

Pre-compression may be used for strain rate sensitive material that requires extended duration of compression, or dwell time, to strengthen internal bonds and improve tensile strength. In this event, the addition of a relatively small amount of pre-compression, e.g. 1–2 kN, is often enough to eliminate capping issues from problematic materials at a given press speed.

9.9.8 Main Compression Force (Pressure)

The force registered by the press is highly correlated with the amount of material present in the die, e.g. low fill weight exhibits a low compression force and vice versa. This is therefore used by most commercial presses as a surrogate for weight. In turn enabling the press to self-adjust the lower punch depth to maintain constant force, e.g. a drop in force, suggests the weight of material present has reduced; therefore the depth of fill is increased to compensate.

In general the force required to produce a robust tablet will increase as the tablet size increases. For example, a 5 mm diameter tablet may only require 5 kN force, whereas a 10 mm diameter tablet might require 20 kN. For this reason, it is often more sensible to convert compaction force into compaction pressure, by dividing the given force by the surface area of the upper punch. In the example above both the 5 mm and the 10 mm tablets undergo a pressure of approximately 250 MPa therefore enabling a comparison independent of tablet size.

9.9.9 Ejection Force (Shear Stress)

Ejection force is routinely reported on most modern instrumented tablet presses and measures the force required to push the tablet out of the die. For an optimized formulation with good lubrication, this force is typically much less than 1 kN, but again as for main compression force this generally increases with tablet size due to the larger surface area in contact with the die wall. For this reason, it makes more sense to convert ejection force to ejection shear stress, by dividing the given peak ejection force by the tablet wall area, which can be calculated by the tablet perimeter (πD) multiplied by the tablet wall height (W) as depicted in Figure 9.2, where t is the overall tablet thickness.

As a general guideline, typical ejection shear stress values of less than 3 MPa should produce tablets with minimal propensity for capping and/or lamination (due to excessive ejection force). Values up to 5 MPa may be acceptable depending on subsequent mechanical stresses; however values exceeding 5 MPa are far more likely to experience these failure modes [7].

9.10 MATERIAL PROPERTY CHANGES DURING THE COMPACTION PROCESS

The compaction process involves deformation of a loose powder bed, which consists of particles that can range in sizes from 1 to 1000 μm containing air and is by its very nature a complex three-dimensional (3D) nonlinear process. On application of a load to the powder bed, a number of mechanisms occur during the formation of an intact compact. The following processes occur in series or parallel and are difficult to separate into different deformation phases. Furthermore, occurring during these processes is the removal of air:

1. Particle rearrangement.
2. Elastic deformation of particles.
3. Plastic deformation of particles.

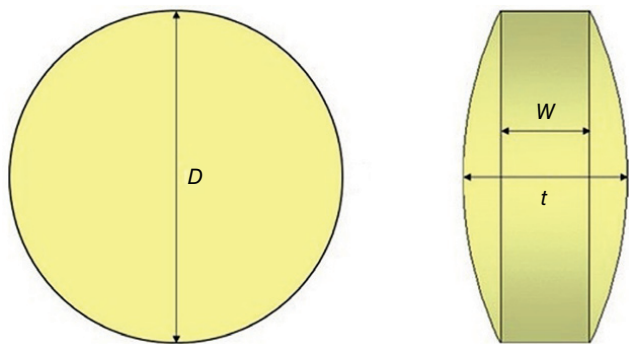


FIGURE 9.2 Definition of typical tablet dimensions for calculation of ejection shear stress.

4. Particle breakage or fragmentation.
5. Formation of interparticulate bonds.
6. Elastic recovery of particles and compact.

To characterize the compaction behavior of powders, both triaxial equipment and uniaxial die compaction equipment have been used [8, 9]. Triaxial equipment allows testing with different loading paths by introducing both the shear and compressive stresses in the compact and can characterize more complete mechanical properties of powders, but it is more complex, expensive, and difficult to use. By comparison, uniaxial die compaction equipment is much cheaper and easier to use. Moreover, it mimics the industrial process of tableting better. Therefore, uniaxial die compaction equipment, also known as compaction simulators in the pharmaceutical industry, have been widely used to investigate the compaction behavior of pharmaceutical powders [10]. To measure the radial die wall pressure and investigate the powder friction at the die wall during compaction, instrumented dies are generally fitted to equip to compaction simulators for tableting R&D [10].

The axial upper/lower punch forces and displacements, and the radial die wall pressure are measured during compaction (Figure 9.3), the full use of these measurements will be described later.

Figure 9.4 shows the typical quasi-static compaction behavior of microcrystalline cellulose (MCC) Avicel PH101 measured by a compaction simulator. The axial stress, σ_z^u , was calculated from the upper punch force (F_z^u) divided by the cross-section area of the compact; the axial strain, ϵ_z , was calculated from the powder height change (ΔH) divided by its initial filling height (H_0) in the die. Both the loading and unloading curves exhibit nonlinear behavior. Moreover, the unloading curves at different compaction densities (or heights) are not parallel, i.e. the unloading properties are density dependent. The axial to radial stress (σ_r) transmission during unloading (Figure 9.4, right), related to Poisson's ratio, also shows nonlinear behavior. Note that the density

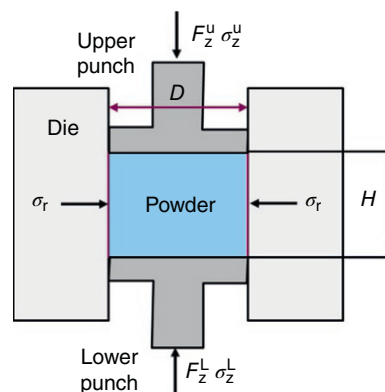


FIGURE 9.3 2D diagram of an instrumented die.

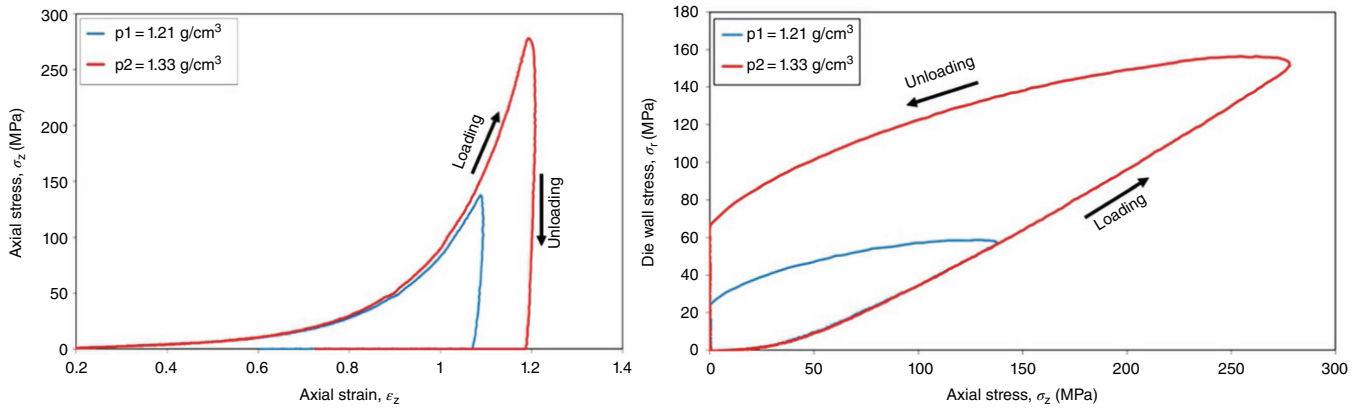


FIGURE 9.4 The typical quasi-static compaction behavior of MCC Avicel PH101 measured by a compaction simulator.

value shown in Figure 9.4 is the compaction density calculated from the powder weight divided by the in-die powder volume.

During compaction, the powder friction at the die wall induces nonuniform axial stress and produces density gradients within the compact. The friction effect can be quantified by the wall friction coefficient during compaction based on Janssen–Walker theory [11].

The results are usually plotted as porosity–axial stress functions. In general, the axial stress is determined by measuring the axial force, whereas the porosity is determined from measuring the tablet dimensions and mass or other methods, such as helium pycnometry. Porosity–axial stress functions can be used for the comparison of materials. For example, in the analysis of the relationship between porosity and compaction pressure through the use of “Heckel plots [12],” a steeper slope suggests a “greater degree of plasticity” in the material; in other words, it indicates ease of densification, which is related to the hardening characteristics of the powder.

The Heckel equation [12] (Eq. (9.1)) is a widely used equation that describes the mechanism of reducing volume during compaction and that powder compression follows first-order kinetics with the interparticulate voids as the reactant and the densification of the powder as the product. Then the degree of compaction with increasing compressing pressure is proportional to the porosity in Eq. (9.1):

$$\frac{d\rho_r}{dP} = k_p \varepsilon \quad (9.1)$$

where

ρ_r is relative density at applied pressure P .

ε is porosity.

k_p is a material pressure-related constant.

The relative density is defined as the ratio between the density of the compact at pressure and the true density of the solid particles. The porosity is given by Eq. (9.2),

where

ε is porosity.

ρ_A is the apparent density.

ρ_T is the true density.

$$\varepsilon = \left(1 - \frac{\rho_A}{\rho_T}\right) \quad (9.2)$$

Replacing the porosity in Eq. (9.1) with the expression in Eq. (9.2) and solving the differential equation gives Eq. (9.3):

$$\ln \left[\frac{1}{1 - \rho_r} \right] = K_H P + A \quad (9.3)$$

Plotting the value of $\ln[1/(1 - \rho_r)]$ against applied pressure P leads to the plot with a linear section with a slope K_H and intercept A . The reciprocal of the parameter K_H [13] expresses the constant dependent on the material known as yield pressure P_y , which is inversely related to the ability of the material to deform plastically under pressure. Lower values indicate the onset of plastic deformation at lower pressures. Intercept A expresses the parameter dependent on the original compact volume during die filling and the initial particle rearrangement. Heckel [12] indicated that the reciprocal of K_H is equal to the indentation hardness of metals. Roberts et al. [14] showed that if the material particle size was close or below the brittle–ductile transition, the reciprocal of K_H will be numerically equal to the yield stress of the particles. Heckel plot data for common pharmaceutical materials are shown in Figure 9.5.

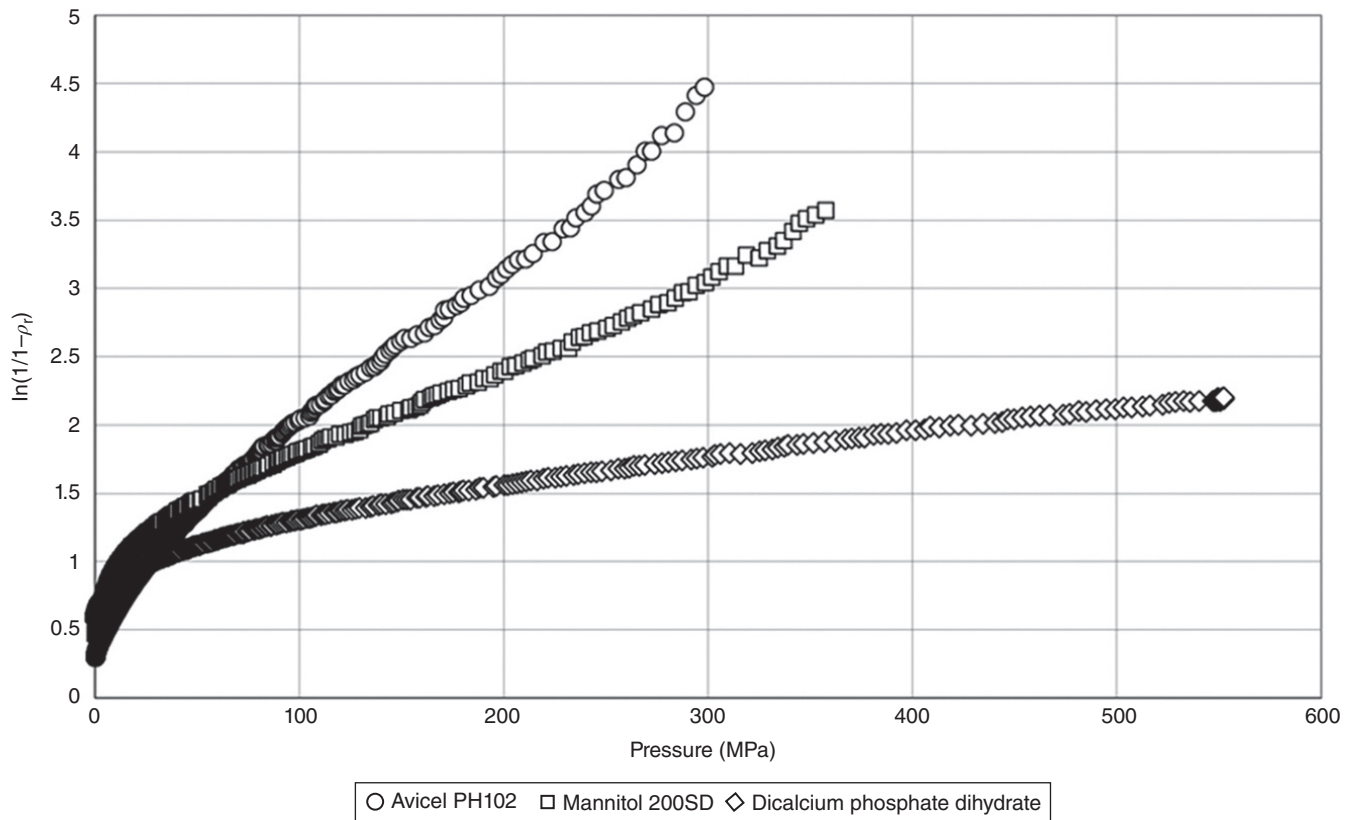


FIGURE 9.5 Heckel plots for Avicel PH102, mannitol 200SD, and dicalcium phosphate dihydrate.

The results in Figure 9.5 indicate that MCC (Avicel PH102) deforms more easily than mannitol (200SD) and dicalcium phosphate dihydrate (Calipharm D). A region below about 70 MPa has considerable curvature consistent with particle rearrangement. For all the materials there are linear regions above this pressure where one can carry out linear regression to obtain the yield pressure ($P_Y = 1/K$). This is in line with published values of the yield pressure from Mahmoodi et al. [15] and Rowe and Roberts [16], which are shown in Table 9.3.

Data in Table 9.3 indicate some agreement between the two sets of workers. Where the agreement is poorer, this might be due to differences such as particle size or testing process parameters such as lubrication, which are known to have a major effect on the measured yield pressure [16].

The Kawakita equation [17] was developed to study powder densification using the degree of reduction in volume C , expressed as:

$$C = \frac{V_0 - V_p}{V_0} = \frac{abP}{(1 + bP)} \quad (9.4)$$

TABLE 9.3 Comparison of the Yield Pressures of Several Materials from Mahmoodi et al. [15] and those from Rowe and Roberts [16]

Material	Yield Pressure ^a (MPa) P_Y	Yield Pressure ^b (MPa) P_Y
Dicalcium phosphate dihydrate	439	431
α -Lactose monohydrate	165	178
Mannitol	144	90
Paracetamol	116	102
Microcrystalline cellulose (Avicel PH102)	90	49
Sodium chloride	79	89
Maize starch	81	40
Aspirin	31	73

Source: Adapted from Mahmoodi et al. 2013 [15].

^aYield pressure data from Mahmoodi et al. [15] was determined by linear regression by locating the minimum value of the first derivative and extending the linear region until the derivative increased by 25%.

^bYield pressure data from Rowe and Roberts [16] was determined by linear regression by locating the minimum value of the second derivative and extending the linear region until the derivative increased by 10%.

Equation (9.4) can be rearranged to give Eq. (9.5):

$$\frac{P}{C} = \frac{P}{a} + \frac{1}{ab} \quad (9.5)$$

where

V_0 is the initial volume of the powder.

V_p is the volume of the powder at pressure P .

Constants a and b are obtained from the slope and intercept of the P/C versus P plot. The constant a expresses the minimum porosity of the powder before compression and the value of $1 - a$ indicates initial relative density $\rho_{r,0}$. The constant b , also known as the coefficient of compression, is related to the plastic material properties. Its reciprocal expresses the pressure parameter P_k , which is the pressure required to reduce the volume of the powder by 50%. For plastic materials, the value of P_k is inversely proportional to the degree of the plastic deformation during the compression with lower values of indicates a higher degree of plastic deformation.

More recently, Adams et al. [18] suggested that the Heckel model did not work satisfactorily for agglomerates and proposed a theoretical model for the confined uniaxial compression of agglomerates. In their model it is assumed that the elastic energy stored in the system is negligible and the energy input to the system is mostly consumed for failure. Based on the assumption of parallel columns for force transmission, Adams et al. [18] obtained the following equation for the relation between pressure on the bed and natural strain:

$$\ln P = \ln \frac{\tau'_0}{\alpha'} + \alpha' + \ln \left(1 - e^{(-\alpha' \epsilon_n)} \right) \quad (9.6)$$

where

P is the applied pressure.

ϵ_n is the natural strain, which is shown in Equation 9.7:

$$\epsilon_n = \ln \left[\frac{H_0}{H_p} \right] \quad (9.7)$$

where

H_0 and H_p are the initial and current height of the powder bed at pressure P , respectively.

The quantity τ'_0 is the apparent granule agglomerate strength, which is related to the actual strength, τ_0 , as Eq. (9.8):

$$\tau'_0 = k_1 \tau_0 \quad (9.8)$$

where

k_1 is a constant.

The quantity α' is related to the pressure coefficient, α_a of the agglomerate strength in Eq. (9.9), where k_2 is a constant:

$$\alpha' = k_2 \alpha_a \quad (9.9)$$

At higher values of natural strain, the last term of the Adams equation becomes negligible and can be omitted, leaving a linear function. The intercept and slope of this linear part of the profile were used to calculate the apparent granule agglomerate strength, τ'_0 . Adams et al. [18], Nicklasson [19], Samimi [20], and Nordström et al. [21], compared the Adams equation and the Kawakita equation to analyse the compaction behaviour of granules prepared by wet granulation. Data from these four groups are compiled in Figure 9.6.

Data for granule strength against the parameter derived from Adam's equation (Figure 9.6 left) for a wider range of granule strengths agree with the findings of Adams et al. [18] and Nicklasson [19]. Both workers suggested that the failure mechanisms were probably fracture (crack opening/tensile failure) and shear failure, in the case of granule strength testing and die compaction respectively. In these data (Figure 9.6 right) the slope of the line is 0.74, this is very similar to that found previously and the good agreement between τ'_0 and $1/b$ also reported by Adams et al. [18] and Nicklasson [19].

9.11 COMPUTATIONAL MODELING OF THE COMPACTION PROCESS

9.11.1 Theory

During tableting by uniaxial die compaction, some common defects such as sticking, picking, capping, and lamination can occur. Several simple theories to explain the causes of these failures [3, 22] have been suggested but these are of limited value. Simulation of the tableting process is required in order to quantitatively predict the occurrence of tablet failure and subsequently derive solutions by either reformulation or a change in the compaction process. Simulation of the tableting process will allow further understanding of the influence of tooling properties, lubrication, compaction speed, and press geometry to provide process improvements. Computational modeling of powder compaction has typically been performed by two different methods: the discrete element method (DEM) and the continuum model method (or finite element method [FEM]). The DEM [23–25] treats each powder particle individually and analyzes the contact interaction and deformation of particles. The continuum model method considers the powder as a continuous medium, whereas the

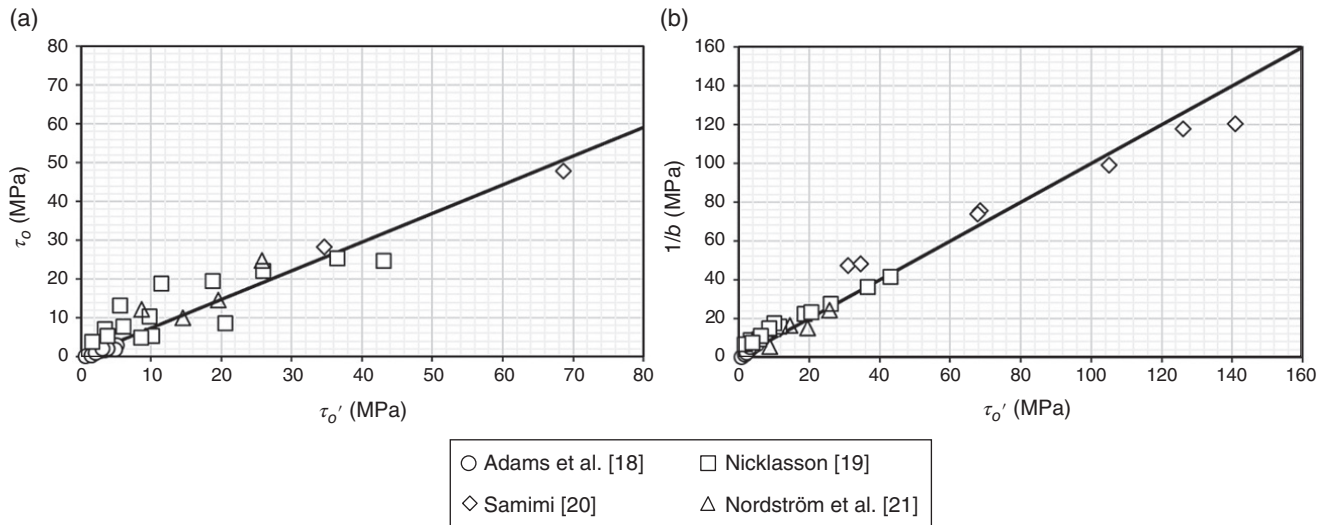


FIGURE 9.6 Plots of (a) Granule strength τ_0 versus Adams compression parameter τ_0' , (b) Kawakita parameter $1/b$ versus Adams compression parameter τ_0' .

discrete model method [26–28] is more useful for understanding the individual processes that occur during powder compaction. Although a pharmaceutical powder is clearly discontinuous at the particle level, as densification proceeds, this becomes irrelevant and therefore ideal for continuum modeling.

The compaction behavior of pharmaceutical powders can be studied using the principles of continuum mechanics or phenomenological models. The phenomenological models, such as critical state models like Cam-Clay plasticity models and Cap plasticity models, originally developed for geological materials in soil mechanics, are well suited for modeling powder compaction, especially in powder metallurgy [29–31]. More recently, Drucker-Prager Cap (DPC) plasticity models have been used for the analysis of compaction of pharmaceutical powders, since they can represent the densification, hardening of the powder, and interparticle friction [32–38]. In these studies, the Young's modulus and Poisson's ratio were assumed constant, which during unloading does not describe the observed nonlinear behavior of pharmaceutical powders. This is particularly important for tablet fracture, since elastic recovery may initiate the cracks within compacts and produce catastrophic flaws, causing the compaction failures such as capping, lamination, and chipping [39, 40]. In order to take account of the decompression phase of compaction, a modified density dependent DPC model was introduced by Han et al. [41].

9.11.1.1 Finite Element Method: Drucker-Prager Cap model The FEM is a reliable and robust modeling tool for analyzing the behavior of powders during compaction. It provides information on the distributions of density, stress, or temperature during compaction. The DPC model [41] was

introduced to model the behavior of pharmaceutical, metallic, and ceramic powders during compaction. The DPC model also considers the hardening, softening, and interparticle friction of powders. The model is assumed to be isotropic, and its yield surface includes three segments: a shear failure surface, providing dominantly shearing flow; a “cap,” providing an inelastic hardening mechanism to represent plastic compaction; and a transition region between these segments, introduced to provide a smooth surface purely for facilitating the numerical implementation.

Since the material parameters of pharmaceutical powders are density dependent [32], a modified density-dependent DPC model is adopted to describe the mechanical behavior of pharmaceutical powders in this work. Further, a nonlinear elasticity law was proposed to describe the unloading behavior. Elastic parameters, the bulk modulus, K , and shear modulus, G , were expressed as functions of the relative density and stress level, rather than constants.

In the DPC model, the yield surface consists of three segments (see Figure 9.7): a shear failure segment F_s , a cap segment F_c , and a transition segment F_t that is introduced primarily to ensure the smooth transition between F_s and F_c , so that computational stability can be maintained [36]. In the p - q plane, these segments are generally defined using two stress invariants, the hydrostatic pressure, p , and the von Mises stress, q (Eqs. 9.10 and 9.11):

$$q = |\sigma_z - \sigma_r| \quad (9.10)$$

$$p = \frac{1}{3}(\sigma_z + 2\sigma_r) \quad (9.11)$$

where σ_z and σ_r are the axial stress and the radial stress, respectively.

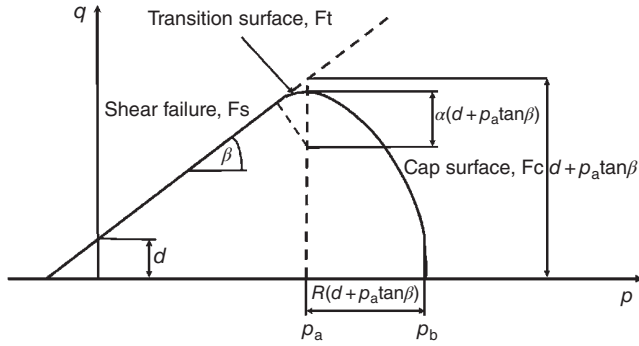


FIGURE 9.7 Drucker-Prager Cap model: yield surface in the p - q plane.

The shear failure segment defines a criterion for the occurrence of shear flow that is a function of the cohesion d and the angle of friction β , as in Eq. (9.12):

$$F_s(p, q) = q - p \tan \beta - d = 0 \quad (9.12)$$

The cap segment is defined in Eq. (9.13):

$$F_c(p, q) = \sqrt{(p - p_a)^2 + \left(\frac{Rq}{1 + \alpha - \alpha / \cos \beta} \right)^2} - R(d + p_a \tan \beta) = 0 \quad (9.13)$$

It intersects the equivalent pressure stress axis and is an elliptical curve with constant eccentricity. In Eq. (9.14), R and α are parameters describing the shape of the cap surface (Figure 9.7). The DPC model in the p - q plane, the cap segment, and the smooth transition surface, p_a , is an evolution parameter describing material hardening or softening:

$$p_a = \frac{p_b - Rd}{1 + R \tan \beta} \quad (9.14)$$

where

p_b is the hydrostatic yield pressure that defines the position of the cap. p_b is generally a function of the volumetric inelastic strain, ϵ_v^p .

This represents the hardening or softening of the cap segment: compaction leads to hardening, while dilation develops softening, where f is a constant.

$$p_b = f \epsilon_v^p \quad (9.15)$$

The transition segment is always very small and its size is controlled by the parameter α , which typically has a value

of 0.01–0.05. The transition segment is mathematically defined as

$$F_t(p, q) = \sqrt{(p - p_a)^2 + \left[q - \left(1 - \frac{\alpha}{\cos \beta} \right) (d + p_a \tan \beta) \right]^2} - \alpha(d + p_a \tan \beta) = 0 \quad (9.16)$$

Inside the yield surface, powder deforms elastically; if the stress state reaches the yield surface, the powder deforms plastically. The density increases if the stress state is on the cap; otherwise it decreases (dilatation) when the stress state reaches the failure line. Dilatation implies softening, so that strain localization and cracking may occur. A more detailed critique of the method to obtain the DPC parameters using uniaxial compression is available in the literature [41].

The Drucker-Prager shear failure surface is determined by using any two of four experiments [42] for measuring tablet strengths: the uniaxial tension, pure shear, diametrical compression (σ_d), and uniaxial compression (σ_c) tests. Mazel et al. [43] discussed two additional failure tests – the three-point bend test and the biaxial test – and showed that care must be taken on the most appropriate test to be selected that best describes the shear failure line. First, the loading paths of the four kinds of tests above are plotted in the p - q space. Under the elastic loading, these four loading paths can be presented by four dashed lines through the origin, point O, whose slopes are -3 , ∞ , $3\sqrt{13}/2$, and 3 , respectively, as shown in Figure 9.8. Then, the four maximum loading points corresponding to four kinds of the tablet strengths are plotted on these four dashed lines, respectively. Finally, the shear failure line can be determined by plotting a straight line through any two maximum loading points. The slope of the line gives the friction angle β , and the intersection with q axis gives the cohesion d . Han et al. [41] used two simple tests: the diametrical compression test (Eq. 9.17) and the uniaxial compression test (Eq. 9.18). The equivalent hydrostatic pressure stress, p , and von Mises equivalent stress, q , are also shown (Eq. 9.19):

$$\sigma_d = \frac{2F}{\pi Dt} \quad (9.17)$$

$$\sigma_c = \frac{4F}{\pi D^2} \quad (9.18)$$

For diametrical compression test, the state of stress in the p - q plane is

$$p = \frac{2\sigma_d}{3} \text{ and } q = \sqrt{13}\sigma_d \quad (9.19)$$

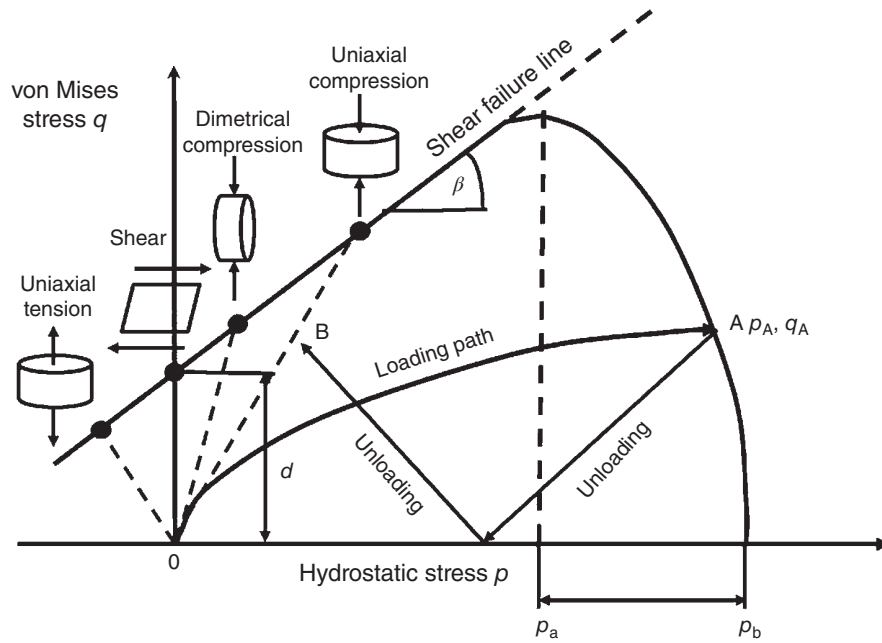


FIGURE 9.8 Determination of the shear failure line from tablet strength test data.

For uniaxial compression test, the state of stress in the p - q plane is

$$p = \frac{\sigma_c}{3} \text{ and } q = \sigma_c \tag{9.20}$$

Consequently, the cohesion d and friction angle β are shown in Eq. (9.21):

$$d = \frac{\sigma_c \sigma_d (\sqrt{13} - 2)}{\sigma_c - 2\sigma_d} \text{ and } \tan \beta = \frac{3(\sigma_c - d)}{\sigma_c} \tag{9.21}$$

Data for the cohesion and internal angle of friction for five excipients measured by LaMarche et al. [44] is reproduced in

Figure 9.9 from fitted data. They found that dicalcium phosphate and MCC had higher cohesion over a range of relative densities when compared to starch and two different grades of lactose. Data for the internal angle of friction were all very similar, suggesting that the slope of shear failure lines were very similar. The anhydrous lactose displays different behavior in the β parameter for low relative densities; the reasons for this are unknown at this time.

Data from compaction experiments on pharmaceutical powders exhibit nonlinear behavior during unloading as shown in Figure 9.10. There exist nonlinear segments at the end of the unloading curves, as is obvious at low compression pressures. The elastic parameters, Young's modulus E and Poisson's ratio, can be estimated from the unloading

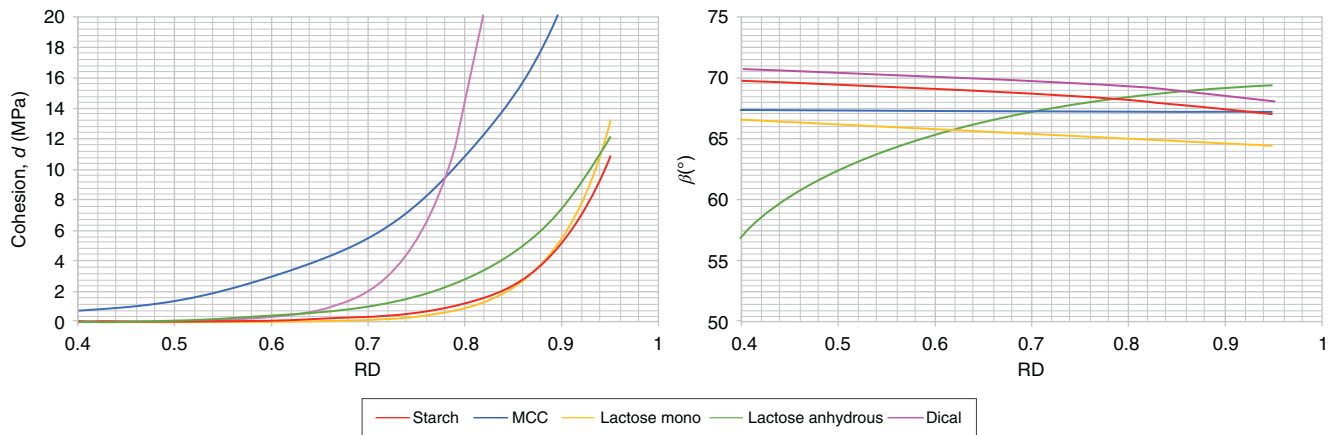


FIGURE 9.9 Cohesion, d , and friction angle, β , versus relative density (RD) for a range of common excipients.

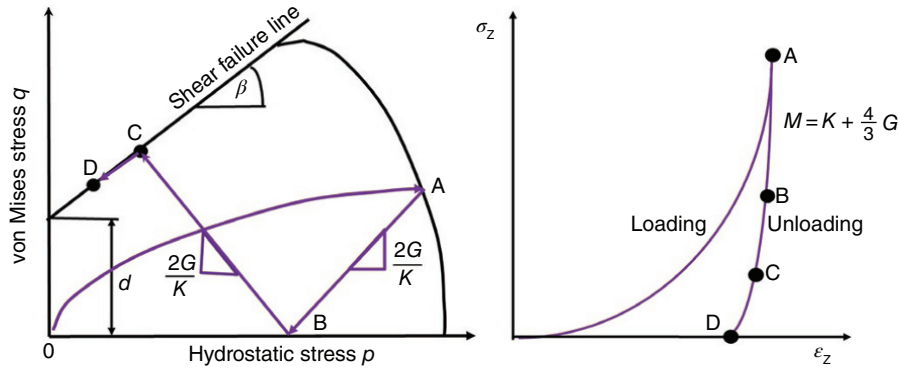


FIGURE 9.10 (Left) loading path plotted on the p - q space; (right) axial stress–axial strain curve, which are used to calculate the elasticity of the compact.

data of die compaction studies in which the axial and radial stresses and axial strain are measured (see Han et al. [41]) (Eqs. 9.22 and 9.23):

$$\nu = \frac{d\sigma_r}{d\sigma_r - d\sigma_z} \tag{9.22}$$

$$E = \frac{d\sigma_z(1 - 2\nu^2)}{d\epsilon_z(1 - \nu)} \tag{9.23}$$

Thus, using Eq. (9.22), we can obtain the Poisson’s ratio ν as a function of the relative density of compact, ρ , from the axial stress/radial stress unloading curve (Figure 9.10 right). Consequently, using Eq. (9.23) and Poisson’s ratio, we can obtain the Young’s modulus E as a function of the axial stress/strain, $(d\sigma_z/d\epsilon_z)$ at the same relative density of compact, ρ (Figure 9.11).

Based on determination of the DPC parameters against relative density, LaMarche et al. [44] considered data for five commonly used excipients and found that MCC Young’s

modulus is on the lower side of the range observed (Figure 9.11, right), while the Poisson’s ratio is higher than most other materials (Figure 9.11, left). The internal friction angle, hydrostatic yield stress, and cap eccentricity were all about mid-range in comparison with the other materials. The cohesion of MCC is much higher than any other material over much of the range of relative density (Figure 9.9 right). LaMarche et al. [44] suggested that these parameters probably account for the superior compaction behavior of MCC. In addition they found that the two lactose grades had much smaller values of cohesion than MCC (Figure 9.9, left). Lactose is also commonly used in formulations for its compaction properties and they suggested that the high Young’s modulus and low Poisson’s ratio (Figure 9.11) show lactose is stiff and does not expand during decompression/ejection; this behavior might be important for formulations. LaMarche et al. [44] commented that dicalcium phosphate had a very large cohesion at low relative densities (Figure 9.9) and like lactose has a high Young’s modulus and low Poisson’s ratio (Figure 9.11). Therefore, dicalcium phosphate is able to

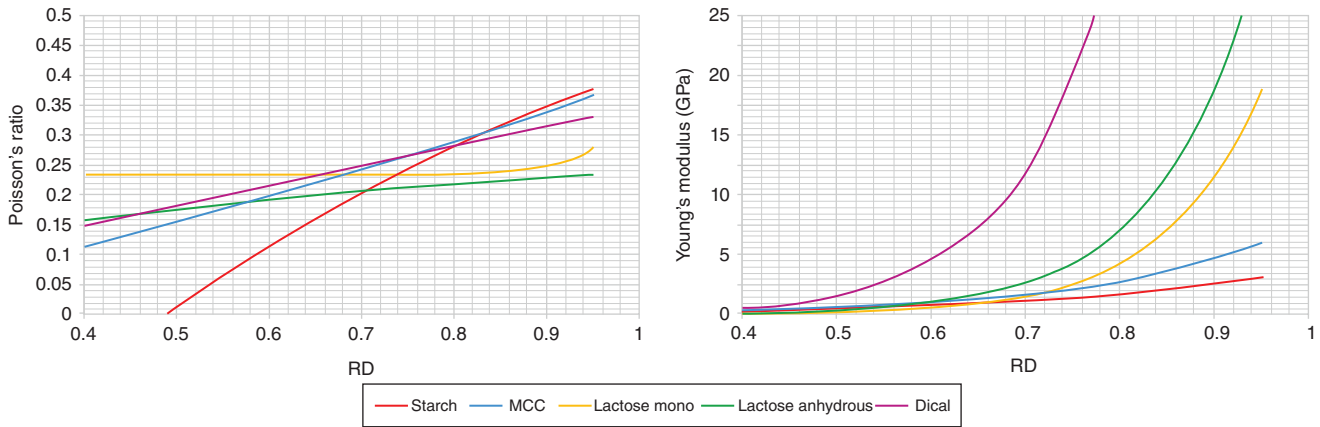


FIGURE 9.11 Excipient elastic properties: (left) Poisson’s ratio versus relative density (RD) and (right) Young’s modulus versus relative density (RD).

produce tablets that are strong but stiff, showing little relaxation during ejection. Finally, starch has the lowest Young's modulus (Figure 9.11, right) and low cohesion (Figure 9.9, left), resulting in tablets that expand a great deal during unloading and ejection [44]. Therefore bonds that are formed during compaction might not survive, resulting in lamination/capping failure.

9.11.2 FEM Models

Sinka et al. [38] investigated the effect of powder die wall friction and punch shape on the compaction of convex MCC tablets. In this investigation, the coefficient of friction at the powder–die wall interface varied inversely over the tableting cycle as a function of the compression pressure, and compaction in lubricated and unlubricated dies were simulated using FEM. The results indicated conclusively that powder–die wall friction has a significant effect on the density distribution within a tablet. Tablets produced in unlubricated dies, having a higher coefficient of powder–die wall friction, were shown to densify preferentially at the edge of the tablet, while those produced in lubricated dies showed the opposite pattern and densified at the center of the tablet. This type of predictive capability could potentially be utilized to optimize the formulation process and aid process design to produce tablets with known or desired properties.

Using the modified DPC model, Han et al. [41] assessed the predictive capability of the model, different tablet sizes and shapes were compared using three different tablet shapes – flat face (FF), single radius concave (SRC), and double radius concave (DRC) – and two different tablet diameters: 8 and 11 mm. FEMs were used to predict the compaction behavior of powders during the manufacture of these tablets. Pressure versus density plots for the different shapes and sizes determined by experiment agreed very well with those predicted using FEM method. The detailed stress and density distributions of FF and SRC round tablets during compression, decompression, and ejection phases were analyzed. FEM results showed that the density and stress distributions could be used to analyze and explain tablet defects such as edge chipping, capping, and lamination. For example, nonuniform density distributions due to wall friction or the punch geometry could be used to explain the phenomena in the practical manufacture of pharmaceutical tablets that concave tablets had more resistance against edge chipping and less resistance against capping, compared with FF tablets. The local stress concentration and large shear stress due to the radial elastic recovery during the decompression and ejection phases could be used to analyze capping and laminar cracks. The results obtained so far indicate the model has a great potential for the tableting process modeling of pharmaceutical powders and the failure investigation of tablets.

Wu et al. [37] showed that there was a strong correlation between the stress distributions from FEM with the potential

fracture patterns during powder compaction from internal cracks identified using X-ray computed tomography. From the distribution of shear stress (FEM) at the early stage of the unloading/decompression, there was a localized shear band induced, running from the top edge toward the mid-center. Shear deformation in powders generally leads to powder dilation. Consequently, the strength in this zone is relatively weak and the von Mises stress in this region is low [45]. Wu et al. [37] found that cracks were developed from the top edge toward the bottom center, similar to the pattern of shear band obtained from FEM analysis. This suggests that capping can be attributed to the formation of the shear band or the highly localized stress distributions.

Krok et al. [46] examined the influence of friction on the compaction process for the FF, FF with radial edge, and standard convex tablets using FEM and examined how changes in friction affected the density distribution, von Mises stress, and shear stress. They found that reducing wall friction should help the tablet retain its strength and density while distributing these parameters homogeneously through the tablet. However, to obtain the desired properties, the correct values for lubrication must be chosen, and this depends on the shape of the tablet. In order to produce a homogeneous tablet, they found that there was an optimal value for the friction coefficient of 0.2 for both flat-faced radial edge and shallow concave tablets since at this value the shear stress banding was reduced and hence damage would be reduced.

The temperature distribution within the tablet and on the surface during the compression, decompression, and ejection can be predicted using FEM [47]. It was shown at high compression speed, part of the generated heat near the die wall is transferred by conduction into the die tools and punches and that the highest temperature was induced inside the tablet in a ring close to the wall, which extended about half way to the tablet center. After unloading, the temperatures inside the tablet decreased slightly and the upper surface has a slightly higher temperature than the lower surface or the wall of the tablet, because some heat in these areas is dissipated by conduction. Krok et al. [47] measured temperature distributions using an infrared camera on the surface of FF and the shallow convex tablets, immediately after ejection. For both cases the highest temperatures were obtained on the centerline of the upper surface, which agreed with the FEM model.

9.11.3 Discrete Element Method (DEM)

This method used to investigate the behavior of powder systems, where the motion and forces associated with each particle are tracked individually, is known as the DEM. This approach differs significantly from the continuum model (FEM) described previously. It is primarily used as a research tool with customized codes utilized mainly in academic investigations, and, as such, the software necessary to perform DEM simulations has only become commercially

available in the last 10 years. DEM simulations also require a great deal of computational power and are currently limited in scope to simulations involving approximately 10^7 [7] particles [48, 49]. To some extent, this limitation is inherent in the formulation of these models, and thus the limitation in simulation size is not expected to change dramatically in the near future. In general, all DEM codes utilize Newton's second law to track the motion of individual particles, and the forces resulting from particle–particle interaction are handled through two distinct approaches. Hard-sphere DEM models are typically employed in the simulation of rapid powder flows, and in this approach particle–particle collisions are considered to be binary and instantaneous. Soft-sphere DEM models are utilized in simulations where the particles remain in contact for longer periods of time, such as low-speed, frictionally dominated flows where particle–particle collisions are more complex. Researchers have used both approaches to investigate situations of interest to the pharmaceutical industry.

The particle interaction model is used in DEM to calculate the forces acting in either particle–particle or particle–wall contacts. Each of these two contact types can be resolved using the same contact model. Furthermore, the material properties (e.g. coefficient of restitution, friction coefficient, etc.) for each contact type can differ. Consequently dissimilar materials can be simulated. The forces are used to calculate accelerations, which are then integrated in time for updated particle velocities and positions. The modeled forces in a soft-particle system are subdivided into normal and tangential components. A number of different normal force–displacement models have been proposed, and these have been reviewed [50–52]. One of the simplest and most commonly used normal force model is the linear spring and dashpot (LSD) model [50, 52, 53]. Here, the repulsive portion of the contact is modeled with a linear spring, and the dissipative portion of the contact is in parallel with a viscous dashpot. Together, these elements represent the normal portion of a contact fairly well. This commonly used model is advantageous in that the coefficient of restitution, contact time, and maximum overlap can be analytically determined. However, the resulting coefficient of restitution was not dependent on the impact velocity as expected.

Other work by Kuwabara and Kon [54] and Brilliantov et al. [55] assumed the material to be viscoelastic instead of elastic. With this modified dissipation term, the coefficient of restitution decreases with increasing impact velocity, in accordance with experimental results. The simulation time step is selected by using an estimate of the maximum expected impact velocity.

A number of tangential force–displacement models of varying complexity have also been proposed. One of the simplest models was first introduced by Cundall and Strack [56] and subsequently used by several others [57, 58], where the tangential force is modeled with a linear spring and the

associated displacement is integrated from the relative velocity at the contact point. This tangential force is determined by the spring and is limited by the Coulomb criterion. This tangential force model is relatively easy to implement and reproduces experimental observations fairly well. Other tangential force models have been based on the detailed analysis of frictional elastic spheres, which specifies that the tangential force is dependent on the loading and unloading history of the contact. The full model considers eleven different loading/unloading possibilities, but generally, only two are considered in DEM implementations. The implementation of this model is significantly more complex than the Cundall and Strack [56] model discussed previously but compares only slightly better with experimental results, as reported by Schafer et al. [59]. Despite the additional complexity of this tangential model, it has been used by Vu-Quoc et al. [60], in 3D granular flow.

Noncontact forces due to capillary cohesion, electrostatics, or van der Waals interactions may significantly affect the behavior of bulk solids, especially for systems with small particle sizes or where moisture is present. The forces become quite significant relative to particle weight as the particle diameter decreases. The importance of these forces has been reviewed by Seville et al. [61] and the application of such forces in DEM models has been reviewed by Zhu et al. [62]. Cohesive forces strongly affect the flowability of powders [63] and can lead to rat-holing and arching during hopper discharge, blockage of pneumatic conveying lines, and agglomeration and caking of powders [64]. Other attractive forces such as van der Waals forces and surface adhesion have also been incorporated into DEM models. Van der Waals attractive forces have been added to study packing and rearrangement of fine particles [65] as well as fluidization [66]. Thornton and Yin [67] developed a theory for contacts in the presence of adhesion and later applied this model in the study of agglomerate fracture [68].

The verification and validation of DEM are very important; verification is the process of ensuring that the numerical solutions are an accurate solution to the model. Validation describes the process by which the accuracy of a model's predictions is compared with experimental data. Verification and validation of DEM models do not appear to be that rigorous and the procedures are not well defined. This is probably due to the difficulty in obtaining experimental data to compare with the DEM results. The extent and methods of verification and validation used by researchers vary widely; an example is the work of Asmar et al. [69], who outlined several fundamental, two-body contacts to verify their DEM model and is used as the benchmark for best practice.

Various aspects relating to tablet compression have been studied using DEM techniques. The process of die filling is one aspect that has been studied in both the pharmaceutical and powder metallurgy industries [70–72]. These studies typically describe the fill rate and particle flow field into a die

and make comparisons with experimental results. Other results show the effects of powder characteristics such as friction, particle-to-die size ratio, and the effect of air pressure [70, 71]. The DEM method was used [73] to measure the residence time distribution (RTD) of particles by size ranges and the mean residence time (MRT) inside the feed frame. The experimental and DEM simulation design of experiments were used to conduct a structured study of the variables of the process. Operating conditions, feed frame exit contributions, and outlet mass flow rate affect the MRT. The RTD analysis by particle size shows that particle size segregation occurs where the smaller particles are leaving the feed frame faster. The level of powder confinement inside the feed frames can have a significant effect on powder attrition. As particle size increases, the compressive and tangential forces applied to it increase. The tangential force data show that the level of attrition was also affected by the paddle wheel speed. Segregation of particles during die filling was investigated using DEM by Furukawa et al. [74], which composed of a horizontally moving die equipped with a vertically moving bottom punch. Accurate particle flow simulations and parameters of the pharmaceutical spherical granules were calibrated using experimental bounced height, rolling distance, and angle of repose. Simulations and experimental results showed reasonable qualitative and quantitative agreement, validating the simulation model.

DEM combined with RSM was carried out for investigating the segregation behavior of binary particles during die filling of the pharmaceutical tableting machine. These parameters were implemented in the DEM simulations and compared with experimental high-speed camera observations and measurements of the filling ratio. Simulations and experimental results showed reasonable qualitative and quantitative agreement, validating the simulation model. The validated model was employed to evaluate the segregation of binary particle mixtures and RSM was used for analyzing the DEM results of the segregation. DEM combined with RSM approach can offer an efficient method to understand the compaction process. In turn this allows the implementation of a quality by design approach. The RSM revealed that, unlike the die velocity, small particle size, horizontal position, and vertical position affected the segregation index significantly. The segregation resulted from percolation of small particles between large particles in the feed shoe and bottom back corner of the die.

Lewis et al. [75] have studied the compression of materials within a die using a finite discrete element method (FDEM). In this approach they modeled a small number of particles of different sizes that were meshed and were compacted at low pressures less than 100 MPa. For ductile particles Gethin et al. [76] were able to compress to high relative densities since the FDEM particles were able to fill the spaces between particles in a die; however at high

densities the agreement between theory and modeling was poor. In another FDEM study, Ransing et al. [77] investigated the compaction of assemblies comprising both ductile and brittle particles and found that the inclusion of ductile particles (25% by volume) suppressed brittle failure of the particles.

A disadvantage of DEM is that the interparticle contact behavior is simplified. This limits the applicability of DEM to low relative densities. Zavaliangos [78] discretized the particles at a sufficient level to provide adequate detail of the interparticle interaction. Simulations show that there is an effect of interparticle friction in the earlier stages of compaction, there is significant rearrangement even in highly constrained compaction, and the absence of friction promotes inhomogeneous deformation and conditions for fragmentation develop in particles with loose lateral constraints. Martin et al. [39, 40] carried out DEM simulations of the cold isostatic and close die compaction of mixtures of mono-size soft and hard powders starting at 0.64 relative density and compressed to 0.90 relative density. They studied the conditions under which the hard particles form clusters that hinders homogeneous distribution of the load and found that friction between particles affects the formation of clusters. Soft particles require additional deformation in the presence of hard non-deformable particles. In addition, in the constitutive equation of the soft phase, hardening of soft particle needs to be accounted for. In another similar study [79], the cold compaction of composite powders having eight different combinations of materials (hard or soft), size ratios, and loading conditions was investigated in 3D. Both isostatic and die compaction were investigated and the theoretical models were accurate in most of the cases for isostatic loading. An important exception is the case when the volume fraction of large and hard particles is high as the numerical predictions suggest a much stiffer response of the powder. For die compaction, theoretical results were in worse agreement than the isostatic compaction DEM calculations. One reason for this was that local rearrangements of particles strongly influences the results during die compaction and the assumption that hard and soft particles have the same contact force for the same orientation is clearly in error. Sheng et al. [80] used a 3D DEM program TRUBAL, which is capable of calculating the contact between particles considering friction and local plastic deformation. The uniaxial powder compaction process was simulated in a cubic periodic unit cell by applying the strain rate to the individual particles. Results on the particle scale (microscopic) were obtained and correlated to the statistical bulk response of the assembly. They found that it was possible to fit DEM results at a macroscopic scale to the experimental measurements by adjusting the particle properties in DEM calculation.

As described above DEM simulations are in general limited to a relative density of about 0.8, because the existing contact laws do not reproduce all the physical phenomena

involved in the densification of granular media. Harthong et al. [81] and Jerier et al. [82] used the meshed distinct element method (MDEM) to formulate a new contact law for higher relative densities, which was used to validate the DEM model. Thus, it was possible with DEM modeling to reproduce high-density compaction of random packings up to a relative density of about 0.95. Yohannes et al. [83] studied bilayer tablet compaction since the bonding formed between the first layer and the second layer particles is crucial for the mechanical strength of the bilayer tablet. The bonding and the contact forces between particles of the first layer and second layer are affected by the deformation and rearrangement of particles due to the compaction pressures. The model used takes into consideration the elastic and plastic deformations of the first layer particles due to the first layer compaction pressure, in addition to the mechanical and physical properties of the particles. The simulations show that the strength of the layer interface becomes weaker than the strength of the two layers as the first layer compaction pressure is increased. The reduction of strength at the layer interface is related to reduction of the first layer surface roughness. These results are in agreement with findings of experimental studies.

A few other aspects related to compression have been studied using DEM modeling; Couroyer et al. [84] used a particle fracture model [85] to examine crushing during compression for various material properties. Martin [39] investigated the springback during the unloading process. Springback was shown to be affected by the material properties, compaction history, and the die wall. In addition to elasticity, the separation of particles previously in contact was also shown to be an important factor during unloading. DEM simulations were conducted to study the effects of particle–particle and particle–wall frictions on the compaction behavior of particles containing a liquid bridge [86]. The results showed that in the die compaction the particle–wall friction had a minimum effect on the initial particle rearrangement stage and mainly affected the particle deformation stage. The interparticle friction reduced the particle rearrangement in the initial stage but had no significant effect on the deformation stage and was attributed to force transmission. This affected the unconfined axial compression: the compressive strength of compacts decreased with increasing particle–wall friction but increased with increasing interparticle friction.

Hassanpour and Ghadiri [87] examined the Heckel analysis used for powder compression and determined that it may not be valid for all conditions. Rather, results show it is valid only in a limited range of Young's moduli and yield stresses.

Although these examples are limited to compaction, DEM has been used to study other pharmaceutical and metal powder processing, and reviews available [62, 88–90] are excellent sources on the subject.

9.12 COMPRESSED TABLET ANALYSIS

As discussed previously, the given physical criteria that a pharmaceutical tablet must reach are dependent on a number of factors, such as clinical stage, API dosage, and API percentage composition of the tablet. This section details the likely day-to-day physical testing that will be performed upon a pharmaceutically active tablet and seeks to suggest route causes and corrective actions that may be undertaken if a given physical parameter is deemed to be insufficient.

9.12.1 Uniformity of Weight and Content (Uniformity of Dosage Unit)

Generally speaking tablet weight uniformity is a function of the flowability of the material to be compressed (poor granule or powder flow is often manifested in poor weight uniformity). The flowability of input material can often be improved by increasing the particle size of the granules [91] or powder or by reducing the compaction production speed (to allow a poorly flowing powder to adequately fill tablet dies).

Uniformity of content (either API uniformity across a number of tablets or indeed the uniformity of a performance critical excipient) is a more complex behavior and can be seen as a result of all previous mixing and segregation events. If poor uniformity of content is observed during a tablet process, analysis should be undertaken to understand if the tabletting operation is inducing segregation or indeed if the problem lies with a nonuniform blend of material being inputted into the tablet press. Uniformity of content is generally measured by taking a number of tablets and analyzing content via an appropriate analytical method (such as high-performance liquid chromatography [HPLC]).

9.12.2 Tablet Porosity

Tablet porosity or solid fraction is the volume occupied by any void space within the *envelop volume* of the tablet core. It is a highly important tablet property as it often has direct consequences on subsequent tablet performance. It is often inversely proportional to tablet robustness measures such as tablet tensile strength [92] and often directly proportional to tablet disintegration times and hence potentially dissolution performance [93]. Tablet porosity is also often directly proportional to compression force (Figure 9.12).

Tablet porosity can be quickly and easily calculated from knowledge of tablet apparent density and tablet true density (see Eq. 9.24). Tablet true density is routinely measured using helium pycnometer or calculated from a weighted average of the true density of its components, which in turn can either be measured or calculated from crystal structures if available. Tablet apparent density can then be measured using either volume displacement approaches, or calculated from tablet dimensions (automated tablet testers, which measure

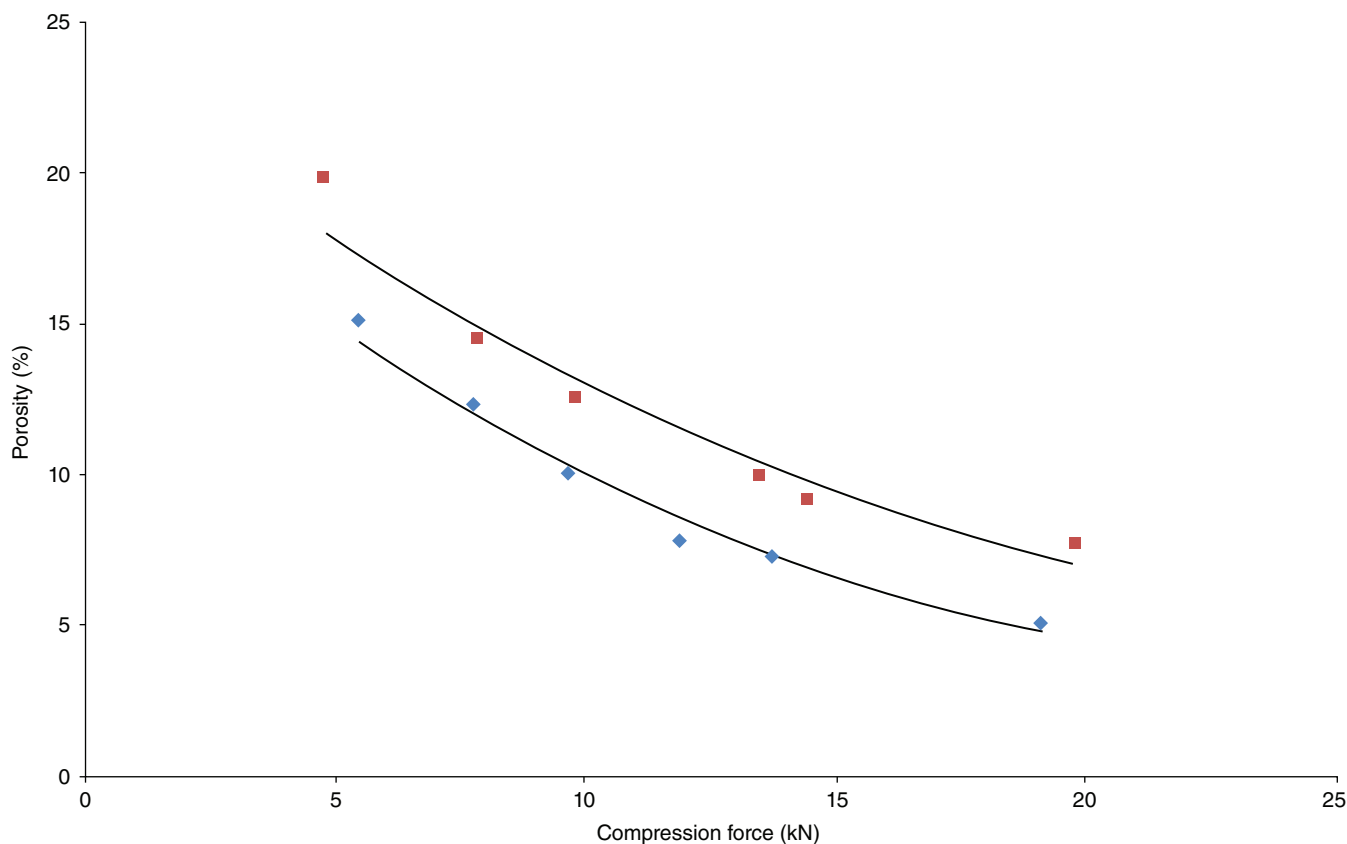


FIGURE 9.12 Response of tablet porosity to compression force.

diameter, weight, and thickness of hundreds of tablets in a matter of minutes, are ideal for such an operation).

$$\varepsilon = 100\% \times \left(1 - \frac{\rho_A}{\rho_T}\right) \quad (9.24)$$

Although tablet porosity would superficially appear to be a comprehensive and simple measurement, it is important to realize that tablet porosity and pore size distribution are two distinctly different phenomena. Porosity is the cumulative sum of all pores with a solid, whereas pore size distribution describes the distribution of pore size within a solid (it is quite possible for two solids to have the same porosity, but radically different pore size distributions). As mentioned, it is well understood that tablet envelop porosity has an effect upon the subsequent tablet properties; however the impact of tablet pore size distribution upon performance is much more poorly understood.

Pore size distribution is typically measured using mercury porosimetry, NMR, or X-ray microtomography. Experiments can be time consuming and difficult, and as such, research into the importance of pore size distribution is scant, although recent research has shown that differing pore size distributions are correlated with “loss of compressibility” phenomena and may indeed influence tablet dissolution performance.

9.12.2.1 Tablet Disintegration Rate The disintegration rate of a tablet core can be of fundamental importance to the dissolution performance of the drug product. Depending on the particle size, solubility, and dissolution kinetics of the drug substance particles, disintegration behavior may be a rate-limiting step. The compaction process can be influential in this regard as it is often the deciding factor in tablet core porosity, which in turn has been demonstrated previously to be influential for disintegration (Figure 9.13).

Disintegration can be measured using either the standard USP disintegration tester or through more advanced approaches, which seek to put disintegration into context with USP dissolution methods [93].

9.12.2.2 Tablet Tensile Strength, Crushing Force, and Hardness It is unfortunate that in the pharmaceutical industry, the term “hardness” is often used to describe what is actually crushing force. Hardness is in fact something distinctly different from crushing force and as a consequence of using the terms interchangeably, hardness is rarely measured.

Hardness is defined as the resistance of a material to a permanent shape change, such as resistance to scratching or indentation. Typical classification systems are the Mohs or Vickers hardness scale and experiments can be performed using nano-indentation approaches [94, 95].

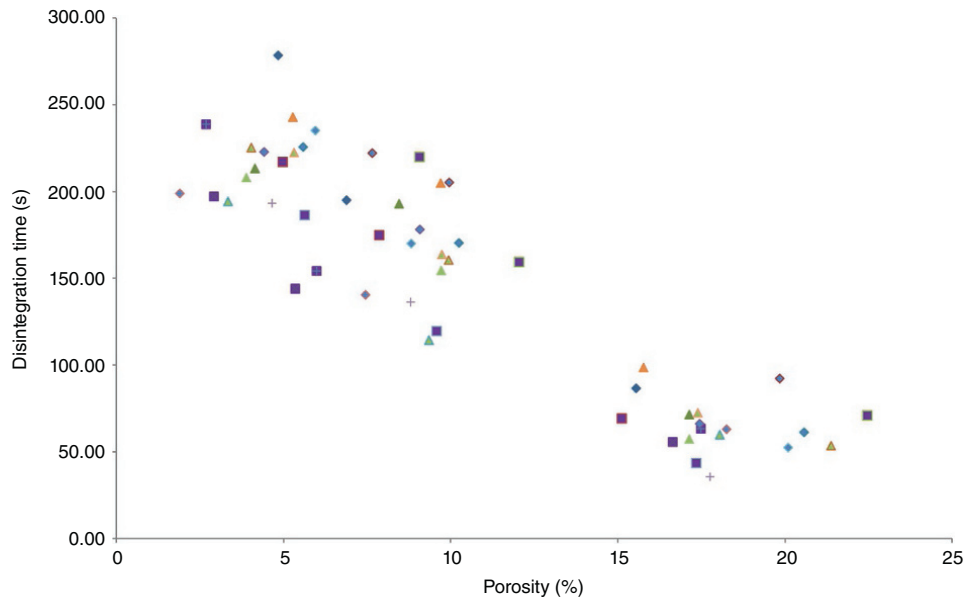


FIGURE 9.13 Tablet disintegration times as a function of tablet porosity for a typical pharmaceutical formulation.

The crushing force of a tablet compact is the force required to break the tablet along a given axis; it is directly related to the tensile strength of a material. Units of crushing force are Newtons or, if one wishes to be particularly old fashioned, kiloponds (1 kP = 9.81 N). It is not correct to compare crushing forces across different tablet sizes and shapes as crushing force is an extensive property (it is dependent on the size of a sample). Crushing force is typically dependent on the porosity of the tablet core, with (in general) a more porous tablet giving a lower crushing force.

The tensile strength of a material is defined as the stress required to induce plastic “flow” in the material. Imagine a metal bar being stretched: the material will resist deformation until a load sufficient to induce “necking” of the material is encountered and the bar lengthens and thins considerably (value of tensile strength) and ultimately (when enough stress is applied) fractures (fracture strength). In tablets necking and fracture are virtually indistinguishable from one another and are usually considered the same thing. The tensile strength of a tablet compact is calculated from values of crushing force and tablet dimensions. It is an intensive property and as such is not dependent on the size of the sample; thus it is possible to compare tensile strengths across a range of tablet sizes and shapes. As with crushing force and as discussed earlier, tensile strength is typically dependent on the porosity of the tablet core.

Calculation of tablet tensile strength (diametrical compression) can be performed by utilizing Eqs. (9.25)–(9.27) for typical tablet shapes [96]:

For a round flat faced tablet,

$$\sigma_d = \frac{2F}{\pi Dt} \quad (9.25)$$

For bi-convex faced tablets,

$$\sigma_d = \left(\frac{10F}{\pi D^2 \left(2.84 \frac{t}{D} - 0.126 \frac{t}{W} + 3.15 \frac{W}{D} + 0.01 \right)} \right) \quad (9.26)$$

For oval shaped tablet,

$$\sigma_d = \frac{2}{3} \left(\frac{10F}{\pi D^2 \left(2.84 \frac{t}{D} - 0.126 \frac{t}{W} + 3.15 \frac{W}{D} + 0.01 \right)} \right) \quad (9.27)$$

where

F is the crushing force of the tablet in Newtons.

D is the diameter of the tablet or shortest dimension in the case of an oval shaped tablet.

t is the tablet thickness.

W is the thickness of the central section of the tablet as shown in Figure 9.14.

The purpose of measuring tablet tensile strength/hardness/toughness is usually to predict the likelihood that tablets will withstand onward processing, such as coating, packaging, transportation, and handling before patient ingestion. This behavior is complex and no single number or measure is able to dictate satisfactory performance in all scenarios. Nevertheless, recent publications have sought to suggest that a tensile strength of greater than 1 MPa be as a minimum requirement for pharmaceutical compacts and greater than 1.7 MPa should be considered a desirable target [7].

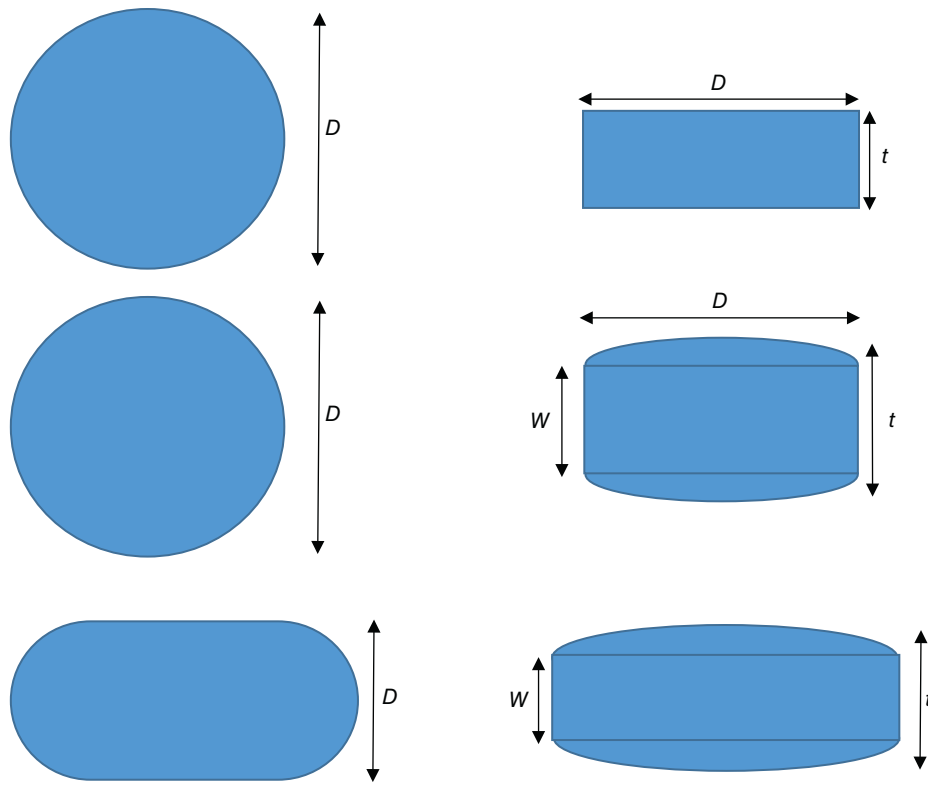


FIGURE 9.14 Dimensions of round flat-faced, biconvex, and oval-shaped tablets.

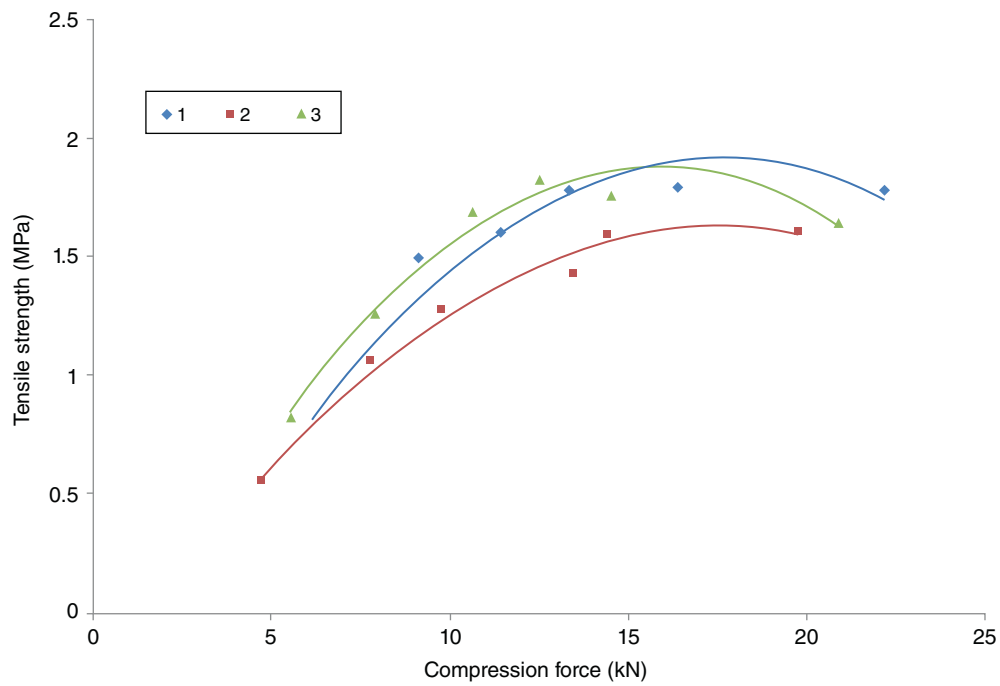


FIGURE 9.15 Response of Tensile strength to compression force.

9.12.3 Typical Compaction Curves and “Overcompression”

Figure 9.15 shows a model response of tablet tensile strength to compression force. It can be seen that an increase in compression force results in an increase in tablet tensile strength up until a maximum is reached after which tablet tensile strength begins to fall again. This is an example of the phenomena commonly referred to as “overcompression.” Subsequent to this maximum, tablets will become far more prone to capping and as such tablet appearance post coating may become problematic.

9.13 SUMMARY

The usage of tablets as an oral solid dosage form is likely to continue for a considerable amount of time to come. Their advantages include being a well-defined and discrete dosage of medicine with ubiquitous acceptance throughout the pharmaceutical industry, while the equipment to manufacture them is common across the world. Consequently it is highly advisable that process engineers involved in solid form design and production familiarize themselves thoroughly with this common unit operation.

LIST OF SYMBOLS

α	A transition parameter
α'	Quantity related to the pressure coefficient
α_a	The pressure coefficient of an agglomerate
β	Angle of friction
ε	Porosity
ε_n	The natural strain
ε_v^p	volumetric inelastic strain
ε_z	Axial strain
ν	Poisson's ratio
π	Mathematical Pi
ρ_r	Relative density
ρ_A	Apparent density
ρ_T	True density
σ_c	Uniaxial compression strength
σ_d	Diametrical breaking strength
σ_z^u	Axial stress on the upper punch
σ_z^L	Axial stress on the lower punch
σ_z	Axial stress
σ_r	Radial stress
τ'_0	Apparent granule agglomerate strength
τ_0	Actual granule agglomerate strength
A	Intercept of straight line in the Heckel equation
1/b	Kawakita parameter
C	Degree of reduction in volume
d	Cohesion
D	Compact diameter

E	Young's modulus
f	Constant for the hydrostatic yield pressure equation
F	Crushing force of a compact
F_s	Shear failure (linear segment)
F_c	Cap surface
F_t	Transition surface
F_z^u	Force on the upper punch
F_z^L	Force on the lower punch
G	Shear modulus
H	Height
ΔH	Change in height
H_0	Initial height H_p
H_p	Height of the powder at pressure P
K	Bulk modulus
k_p	Material related pressure constant
K_H	Material constant in the Heckel equation
p	Hydrostatic pressure
p_a	Hardening evolution parameter
p_b	Hydrostatic yield pressure
P	Pressure
P_k	The pressure required to reduce the powder volume by 50%
P_Y	Material yield pressure
q	Von Mises stress
R	Cap shape parameter
t	Compact thickness
V_p	Volume at given pressure
V_0	Volume at zero pressure/Initial volume of the powder
V_p	Volume of powder at pressure P
W	Thickness of the central section of the compact

REFERENCES

1. Ashford, M. (2002). Bioavailability – physicochemical and dosage form factors. In: *Pharmaceutics The Science of Dosage Form Design* (ed. M.E. Aulton), 234. Churchill Livingstone.
2. Sheth, B.B., Bandelin, F.J., and Shangraw, R.F. (1980). Compressed tablets. In: *Pharmaceutical Dosage Forms: Tablets*, vol. 1 (ed. H.A. Lieberman and L. Lachman), 109. New York: Marcel Dekker.
3. Burlinson, H. (1968). *Tablets and Tableting*. London: William Heinemann Medical Books Ltd.
4. Catellani, P.L., Santi, P., Gasperini, E. et al. (1992). *International Journal of Pharmaceutics* 1: 285–291.
5. Uchimoto, T., Iwao, Y., Yamamoto, T. et al. (2013). *International Journal of Pharmaceutics* 1: 128–134.
6. Roberts, M., Ford, J.L., MacLeod, G.S. et al. (2003). *Journal of Pharmacy and Pharmacology* 9: 1223.
7. Pitt, K.G., Webber, R.J., Hill, K.A. et al. (2015, Part B). *Powder Technology* 490–493.
8. Pavie, E. and Doremus, P. (1999). *Powder Metallurgy* 4: 345–352.
9. Rottmann, G., Coube, O., and Riedel, H. (2001). Comparison between triaxial results and models prediction with special

- consideration of the anisotropy. *Proceedings of the European Congress on Powder Metallurgy*, PM'2001, Nice, France, vol. 3, pp. 29–37.
10. Doelker, E. and Massuelle, D. (2004). *European Journal of Pharmaceutics and Biopharmaceutics* 2: 427–444.
 11. Nedderman, R.M. (1992). *Statics and Kinematics of Granular Materials*. New York: Cambridge University Press.
 12. Heckel, R.W. (1961). *Transactions of the Metallurgical Society of the AIME* 671–675.
 13. Hersey, J.A. and Rees, J.E. (1971). *Nature* 230.
 14. Roberts, R.J., Rowe, R., and Kendall, K. (1989). *Chemical Engineering Science* 1647–1651.
 15. Mahmoodi, F., Klevan, I., Nordström, J. et al. (2013). *International Journal of Pharmaceutics* 2: 295–299.
 16. Rowe, R.C. and Roberts, R.J. (1996). Mechanical properties. In: *Pharmaceutical Powder Compaction Technology* (ed. G. Alderborn and C. Nystrom), 283–322. Marcel Decker.
 17. Kawakita, K. and Lüdde, K. (1971). *Powder Technology* 2: 61–68.
 18. Adams, M.J., Mullier, M.A., and Seville, J.P.K. (1994). *Powder Technology* 1: 5–13.
 19. Fredrik Nicklasson, G.A. (2000). *Pharmaceutical Research* 8: 949.
 20. Samimi, A. (2003). *Characterisation of Deformation and Breakage of Agglomerates*. University of Surrey.
 21. Nordström, J., Welch, K., Frenning, G., and Alderborn, G. (2008). *Journal of Pharmaceutical Sciences* 11: 4807–4814.
 22. Long, W.M. and Alderton, J.R. (1960). *Powder Metallurgy* 6: 52–72.
 23. Fleck, N.A. (1995). *Journal of the Mechanics and Physics of Solids* 9: 1409–1431.
 24. Fleck, N.A., Kuhn, L.T., and McMeeking, R.M. (1992). *Journal of the Mechanics and Physics of Solids* 5: 1139–1162.
 25. Helle, A.S., Easterling, K.E., and Ashby, M.F. (1985). *Acta Metallurgica* 12: 2163–2174.
 26. DiMaggio, F.L. and Sandler, I.S. (1971). *Journal of the Engineering Mechanics Division* 935–950.
 27. Khoei, A.R. and Azizi, S. (2005). *Materials & Design* 2: 137–147.
 28. Schofield, A. and Wroth, C.P. (1968). *Critical State Soil Mechanics*. London: McGraw Hill.
 29. Aydin, İ., Briscoe, B.J., and Ekanlitürk, K.Y. (1996). *Powder Technology* 3: 239–254.
 30. Chtourou, H., Guillot, M., and Gakwaya, A. (2002). *International Journal of Solids and Structures* 4: 1059–1075.
 31. Coube, O. and Riedel, H. (2000). *Powder Metallurgy* 2: 123.
 32. Cunningham, J.C., Sinka, I.C., and Zavaliangos, A. (2004). *Journal of Pharmaceutical Sciences* 8: 2022–2039.
 33. Frenning, G. (2007). *Powder Technology* 2: 103–112.
 34. Michrafy, A., Ringenbacher, D., and Tchoreloff, P. (2002). *Powder Technology* 3: 257–266.
 35. Sinka, I.C., Cunningham, J.C., and Zavaliangos, A. (2003). *Powder Technology* 1: 33–43.
 36. Wu, C., Ruddy, O.M., Bentham, A.C. et al. (2005). *Powder Technology* 1: 107–117.
 37. Wu, C., Hancock, B.C., Mills, A. et al. (2008). *Powder Technology* 2: 121–129.
 38. Sinka, I.C., Cunningham, J.C., and Zavaliangos, A. (2004). *Journal of Pharmaceutical Sciences* 8: 2040–2053.
 39. Martin, C.L. (2003). *Acta Materialia* 15: 4589–4602.
 40. Martin, C.L., Bouvard, D., and Shima, S. (2003). *Journal of the Mechanics and Physics of Solids* 4: 667–693.
 41. Han, L.H., Elliott, J.A., Bentham, A.C. et al. (2008). *International Journal of Solids and Structures* 10: 3088–3106.
 42. Procopio, A., Zavaliangos, A., and Cunningham, J. (2003). *Journal of Materials Science* 17: 3629.
 43. Mazel, V., Diarra, H., Busignies, V., and Tchoreloff, P. (2014). *International Journal of Pharmaceutics* 1: 63–69.
 44. LaMarche, K.; Buckley, D.; Hartley, R.; Qian, F.; Badawy, S. *Powder Technology* 2014, Supplement C, 208–220.
 45. Frenning, G. (2008). *Computer Methods in Applied Mechanics and Engineering* 49: 4266–4272.
 46. Krok, A., Peciar, M., and Fekete, R. (2014). *Particuology* Supplement C: 116–131.
 47. Krok, A., García-Triñanes, P., Peciar, M., and Wu, C. (2016). *Chemical Engineering Research and Design* Supplement C: 141–151.
 48. Ottino, J.M. and Khakhar, D.V. (2001). *Powder Technology* 2: 117–122.
 49. Coetzee, C.J. (2017). Review: calibration of the discrete element method. *Powder Technology* 310: 104–142.
 50. Luding, S. (1998). Collisions and contacts between two particles. In: *Physics of Dry Granular Media* (ed. H.J. Herrmann, J.P. Hovi and S. Luding), 285–304. Dordrecht: Kluwer Academic Publishing.
 51. Herrmann, H.J. and Luding, S. (1998). *Continuum Mechanics and Thermodynamics* 4: 189–231.
 52. Stevens, A.B. and Hrenya, C.M. (2005). *Powder Technology* 2: 99–109.
 53. Haff, P.K. and Werner, B.T. (1986). *Powder Technology* 3: 239–245.
 54. Kuwabara, G. and Kono, K. (1987). *Japanese Journal of Applied Physics* 8R: 1230.
 55. Brilliantov, N.V., Spahn, F., Hertzsch, J., and Poschel, T. (1996). *Physical Review E* 5: 5382–5392.
 56. Cundall, P.A. and Strack, O.D.L. (1979). *Géotechnique* 1: 47–65.
 57. Ristow, G.H. and Herrmann, H.J. (1995). *Physica A: Statistical Mechanics and Its Applications* 4: 474–481.
 58. Ristow, G.H. (1997). *Physica A: Statistical Mechanics and Its Applications* 3: 319–326.
 59. Shafer, J., Dippel, S., and Wolf, D.E. (1996). *Journal de Physique I* 5–20.
 60. Vu-Quoc, L., Zhang, X., and Walton, O.R. (2000). *Computer Methods in Applied Mechanics and Engineering* 3: 483–528.
 61. Seville, J.P.K., Willett, C.D., and Knight, P.C. (2000). *Powder Technology* 3: 261–268.
 62. Zhu, H.P., Zhou, Z.Y., Yang, R.Y., and Yu, A.B. (2007). *Chemical Engineering Science* 13: 3378–3396.

63. Baxter, J., Abou-Chakra, H., Tüzün, U., and Mills Lamptey, B. (2000). *Chemical Engineering Research and Design* 7: 1019–1025.
64. Gröger, T., Tüzün, U., and Heyes, D.M. (2003). *Powder Technology* 1: 203–215.
65. Yen, K.Z.Y. and Chaki, T.K. (1992). *Journal of Applied Physics* 7: 3164–3173.
66. Tatemoto, Y., Mawatari, Y., and Noda, K. (2005). *Chemical Engineering Science* 18: 5010–5021.
67. Thornton, C. and Yin, K.K. (1991). *Powder Technology* 1: 153–166.
68. Thornton, C., Yin, K.K., and Adams, M.J. (1996). *Journal of Physics D* 2: 424.
69. Asmar, B.N., Langston, P.A., Matchett, A.J., and Walters, J.K. (2002). *Computers & Chemical Engineering* 6: 785–802.
70. Wu, C., Cocks, A.C.F., Gillia, O.T., and Thompson, D.A. (2003). *Powder Technology* 2: 216–228.
71. Wu, C. and Cocks, A.C.F. (2006). *Mechanics of Materials* 4: 304–324.
72. Ozaki, Y., Uenosono, S., Tagami, N. et al. (2006). *Advances in Powder Metallurgy and Particulate Materials* 3: 35–46.
73. Mateo-Ortiz, D. and Méndez, R. (2015). *Powder Technology Supplement C*: 111–117.
74. Furukawa, R., Shiosaka, Y., Kadota, K. et al. (2016). *Journal of Drug Delivery Science and Technology Supplement C*: 284–293.
75. Lewis, R.W., Gethin, D.T., Yang, X.S., and Rowe, R.C. (2005). *International Journal for Numerical Methods in Engineering* 7: 853–869.
76. Gethin, D.T., Ransing, R.S., Lewis, R.W. et al. (2001). *Computers & Structures* 13: 1287–1294.
77. Ransing, R.S., Lewis, R.W., and Gethin, D.T. (2004). *Philosophical Transactions of the Royal Society of London A: Mathematical, Physical and Engineering Sciences* 1822: 1867–1884.
78. Zavaliangos, A. (2002). *MRS Proceedings*. San Francisco Marriott and Argent Hotels. San Francisco, CA.
79. Skrinjar, O. and Larsson, P. (2004). *Acta Materialia* 7: 1871–1884.
80. Sheng, Y., Lawrence, C.J., Briscoe, B.J., and Thornton, C. (2004). *Engineering with Computers* 2: 304–317.
81. Harthong, B., Jérier, J., Dorémus, P. et al. (2009). *International Journal of Solids and Structures* 18: 3357–3364.
82. Jerier, J., Hathong, B., Richefeu, V. et al. (2011). *Powder Technology* 2: 537–541.
83. Yohannes, B., Gonzalez, M., Abebe, A. et al. (2017). *International Journal of Pharmaceutics* 1: 597–607.
84. Couroyer, C., Ning, Z., and Ghadiri, M. (2000). *Powder Technology* 1: 241–254.
85. Ghadiri, M. and Ning, Z. (1997). *In Effect of Shear Strain Rate on Attrition of Particulate Solids in a Shear Cell*, vol. 97 (ed. R. P. Behringer and J.T. Jenkins), 127–130. *Powders and Grains*.
86. He, Y., Evans, T.J., Yu, A.B., and Yang, R.Y. (2017). *Powder Technology Supplement C*: 183–190.
87. Hassanpour, A. and Ghadiri, M. (2004). *Powder Technology* 3: 251–261.
88. Ketterhagen, W.R., Curtis, J.S., Wassgren, C.R., and Hancock, B.C. (2009). *Powder Technology* 1: 1–10.
89. Zhu, H.P., Zhou, Z.Y., Yang, R.Y., and Yu, A.B. (2008). *Chemical Engineering Science* 23: 5728–5770.
90. Rogers, A., Hashemi, A., and Ierapetritou, M. (2013). *Processes* 67–127.
91. Sandler, N. and Wilson, D. (2010). *Journal of Pharmaceutical Sciences* 2: 958–968.
92. Pitt, K.G., Newton, J.M., and Stanley, P. (1991). *Journal of Pharmacy and Pharmacology* 4: 219–225.
93. Wilson, D., Reynolds, G., and Wren, S. (2012). *Pharmaceutical Research* 1: 198.
94. Meier, M., John, E., Wieckhusen, D. et al. (2009). *Powder Technology* 3: 301–313.
95. Masterson, V.M. and Cao, X. (2008). *International Journal of Pharmaceutics* 1: 163–171.
96. Pitt, K.G. and Heasley, M.G. (2013). *Powder Technology Supplement C*: 169–175.

10

PUNCH STICKING: FACTORS AND SOLUTIONS

DARYL M. SIMMONS*

Eurofins Lancaster Laboratories PSS, Lancaster, PA, USA

10.1 INTRODUCTION

During tablet manufacturing, as the powder is compressed, it sometimes adheres to the tablet punches. This is punch sticking, which has been defined as adherence of powder material to tablet punches as indicated by visual inspection of tablets and punch faces [1, 2]. It can be a significant obstacle to tablet manufacturing efficiency, product quality, and result in financial impact with the loss of active pharmaceutical ingredient (API), drug product, and production time [3]. Lost production time includes stopping the manufacturing process, disassembly of the tableting machine to remove the punches, cleaning and polishing tooling, and reassembly of the equipment. It is evident that a substantial amount of production time is lost to remove adhered API on punches to assure product quality. Punch picking is a type of adherence that is observed around the punch debossing that leads to a pitted tablet surface [4]. The precise definitions of punch sticking and picking have some differences among various sources in the literature, but it is generally agreed that they are both adhesion to punches and are a significant problem in the pharmaceutical industry with no single scientific cause or remedy [5, 6]. Regardless of other factors, it is undeniable that punch sticking occurs because an active ingredient or excipient has a greater affinity for tablet punches (adhesion), than for the formulation components (cohesion). In this chapter, picking and sticking are discussed under the overarching topic of punch sticking during tablet compression. Compression refers to the reduction in

powder volume during formation of a compact or tablet of certain tensile strength. Tablet punches are the tools compressing the powder in a die, and the punches and the die together are referred to as tooling.

Punch sticking becomes apparent when a powder is compacted and begins to adhere to the punches. The phenomenon is a result of three steps: initially (i) a film forms on the punch face and, with subsequent compressions, (ii) additional particles adhere to the punch surface rather than remaining intact on the tablet and powder continues to build up with successive compressions until (iii) defective tablets are observed. At this point, production stops and the punches are removed to be cleaned then reinstalled. The downtime can be significant, as can product loss.

Some formulations show a potential for punch sticking during earlier manufacturing steps. For example, in blending, API may adhere to the sides of the blender. For this reason, API is often sandwiched between excipients when added to the blender. In the granulation step, roller-compacted formulations may stick to the rollers, which by itself is not problematic since the amount of blend on the rollers does not increase once they are coated with it, but it does affect ribbon quality. The loss of gram quantities relative to hundreds of grams or kilograms in a typical batch will not likely affect tablet potency. However, it is indicative that punch sticking may occur in the tableting step. Tableting the granulation is a critical step to forming a unit dose. If the granulation adheres to the punches and accumulates with successive compressions, the buildup of added formulation can cause higher compression force which, if high enough, could be damaging to the punches and the tablet press.

Punch sticking has no known single cause or remedy. However, there are various testing procedures that can be

*Current address: Three Rivers Community College, Norwich, CT USA

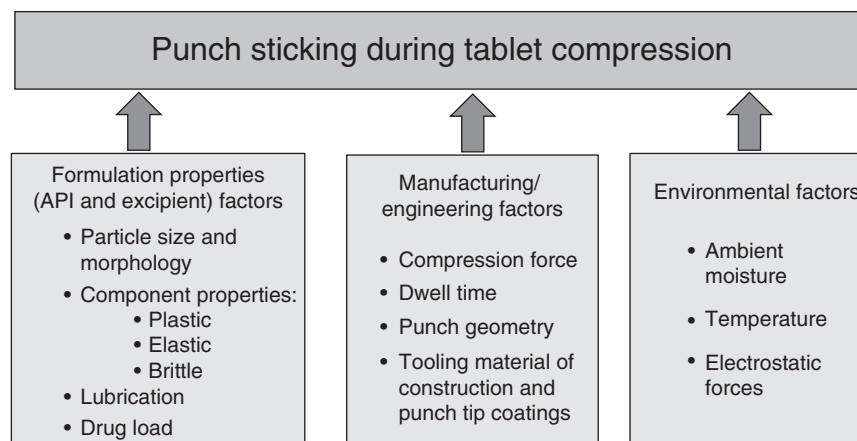


FIGURE 10.1 The punch sticking factors within the three domains of Formulation properties, Engineering, and Environment conditions. *Source:* Courtesy of Avik Sarkar.

used during development, and there are numerous solutions to help remediate the problem. Predictive punch sticking tests are explained as is the value of conducting them. To begin with, contributing factors to punch sticking are discussed. These lie within the three domains of Formulation, Engineering, and Environmental or ambient conditions (Figure 10.1).

10.2 PUNCH STICKING FACTORS

Punch sticking results from variables within three main factors: formulation, engineering/manufacturing equipment, and environmental conditions.

10.2.1 Formulation Factors

An adhesive formulation can be optimized to reduce its punch-sticking potential. The adjustable formulation factors include (i) particle sizes and morphologies of the API and excipients, (ii) the use of brittle, plastic, and elastic components, (iii) lubricant selection, amount, and mixing time, (iv) granulating, and (v) drug load. Much of the strategic design to reduce punch sticking is based on reducing the exposure of an adhesive API to tablet punches and dies. Minimizing contact between the API and the tooling lowers the probability for adhesive binding.

10.2.1.1 Particle Size and Morphology Exposed particle surfaces bond to other particles by cohesion (B_C) and to punches by adhesion (B_A). When particles have greater affinity for the punches than other particles, i.e. $B_A > B_C$, then punch sticking results. Larger particles of API, by having less surface area than smaller ones, are expected to be less adhesive. Smaller-sized excipients, having more surface area, will have more interparticulate bonding with API thereby reducing B_A while interfering with the API contacting the punches. Cohesive bonding and less contact with punches

reduces the adhesive bonding [7]. The particle size effect has also been demonstrated with different sizes of API in powder blends [8, 9]. Because a larger particle has less available surface area to contact the punch, the particle would be functionally less adhesive than would smaller particles, i.e. decreased B_A . This was shown when the use of smaller-sized ibuprofen particles resulted in greater punch sticking [7]. Larger particles also result in tablets of reduced tensile strength due to lower contact area for interparticulate bonding when compressed [10]. The conclusion of an investigation using two different-sized particles of ibuprofen with d_{50} of 90.3 μm (+0.7) for IBU90 and 32.1 μm (± 0.2) for S250 was that smaller-sized ibuprofen had pronounced sticking and poorer flowability [11]. Generally, the larger particle API should be formulated with smaller-sized excipients to reduce or eliminate punch sticking. The effects of excipient size and segregation also need to be considered in powder flow through the equipment. Segregation results when there are large differences in particle sizes of components. This is due to smaller particles not flowing as well as larger ones. The greater surface area of smaller-sized particles generates more flow resistance (friction) and interparticulate bonding [12].

In general, to prevent punch sticking when formulating an adhesive API, the formulator should consider using the largest acceptable particle size of the API and the smallest particle sizes of the excipients that do not detrimentally affect flow.

Particle shape affects the movement and rearrangement of the particles and the bonding as a consequence [13]. Both ibuprofen and acetaminophen have more than a single morphology and are problematically adhesive [14]. However, the more spherical morphologies are less sticky probably due in large part to the decreased surface area compared to the plate or needle shapes. The less round and more irregular the particle shape, the higher the resistance to movement and rearrangement of the particles during compression. Even

isomorphous APIs, where there are only differences in crystal habit, were shown to affect tableting and punch sticking. More energy is required to move a plate-shaped structure than a sphere.

Surface roughness of the API is sometimes a factor in punch sticking [15]. Particles with smoother surfaces also have less resistance to rearrangement during compression. Rougher surfaces resist movement and rearrangement during compression due to added friction. They also have more surface area for bonding, whether cohesively or adhesively sticking to punches. Karehill attributed the added sticking of his formulation to increased powder interlocking making bond formation easier [13]. The same logic applies to tooling surfaces whereby smooth, polished punches are expected to have less punch sticking than worn ones.

10.2.1.2 Excipient Properties: Elastic, Plastic, and Brittle

Powder excipients are used as binders for cohesiveness, as fillers for tablet tensile strength, as lubricants for successful processing of a blend or granulation into a tablet dosage, as disintegrants to facilitate the disintegration of the tablet *in vivo*, and sometimes as glidants to aid poorly flowing mixtures. Each component is predominantly elastic, plastic, or brittle. Elastic particles deform during compression up to their elastic limit after which they will fracture (Figure 10.4). Prior to reaching that stretch limit, the deformation of elastic materials is reversible just as any elastic substance retracts when stretching forces are removed. Elastic powders expand after compression forces are removed. This is known as relaxation. Plastic deformation maintains shape after compressive forces are removed, and therefore deforms without recovery to provide good particulate binding. A brittle component will fracture under compression stress when the force goes beyond its deformation limit. While more surface area forms after brittle fracture, it also results in more friction and resistance to movement within the die during compression. Plastic and brittle excipients form physical shapes and add stability to the shape after compression.

The particle differences are seen by the deformation and fragmentation during compaction. Examples of common binders are microcrystalline cellulose that is ductile and consolidates by plastic deformation [16]; lactose, which is plastic/brittle; and dibasic calcium phosphate (DCP), which is classified as brittle [9]. The brittle nature of lactose and DCP contribute to the tensile strength of the compact. Fracturing of brittle excipients during compression brings about increased B_C and a reduction of B_A while forming harder tablets [17]. This is due to the increased excipient surface area that prevents the adhesive API from contacting the punches.

10.2.1.3 Lubrication The primary function of lubricants is to reduce friction. Studies of lubrication and friction are known as tribology [18]. In solid dosage manufacturing, lubricants reduce interparticulate friction as well as friction

between the powder blend and the equipment used in blending, roller compacting, milling, tableting, and capsule filling [19]. Lubricants also improve flowability and decrease punch sticking through interference with the adhesive bonding. Friction and punch adhesion are related. As punch adhesion increases, friction also increases. Li and Wu [20] proposed the following equation that includes the adhesive forces in the second term:

$$F_{\parallel} = \mu F_{\perp} + 2\epsilon A \left(\frac{\Delta\gamma}{\delta} \right) \quad (10.1)$$

Where in the first term, F_{\parallel} is normal friction, μ is the coefficient of friction, F_{\perp} is friction from the load applied, ϵ represents the transferred coefficient, A is the area of contact, $\Delta\gamma$ is the difference in surface energy, and δ is the elemental distance. Factors such as surface roughness of equipment, especially the punches and dies, contribute to friction and add surface area for more B_A and therefore increase the potential for punch sticking.

Pharmaceutical lubricants are available in four types: (i) metallic salts of fatty acids, such as magnesium stearate (MgSt), calcium stearate, and zinc stearate; (ii) fatty acids, including stearic acid, palmitic acid, and myristic acid; (iii) fatty acid esters, some examples of which are glyceryl monostearate, glyceryl behenate, and sucrose monopalmitate; and (iv) inorganic materials and polymers that include magnesium silicate (talc), and colloidal silicon dioxide (CS) which is also a glidant. Lubricant selection is based on chemical compatibility with the API as well as functionality as a lubricant. While there are numerous lubricants used in the pharmaceutical industry, MgSt is the most common. MgSt also has glidant properties and provides multiple functions to the formulation. It easily forms a hydrophobic film over the surfaces of particles and reduces flow resistance by reducing interparticulate cohesion. Excess MgSt interferes with cohesion and can result in weaker tablets. Lubricants can also reduce punch sticking by reducing the friction between compacts and tooling during compression.

A property of MgSt is its ability to form what are known as pseudo-polymorphs [21]. Four hydrates have been reported: anhydrous, monohydrate, dihydrate, and trihydrate. They are relatively stable but can change hydrate form, depending on ambient temperature and humidity [22]. The dihydrate form is the most efficient at lubricating pharmaceutical powders.

10.2.1.4 Drug Load Compression enables particles to bond to each other (B_C) and, when an adhesive API is used, to the punches (B_A). When increasing the drug load of an adhesive API, it would be expected that punch sticking would occur with fewer compressions, just as decreasing the amount of API would reduce, and possibly eliminate, the problem.

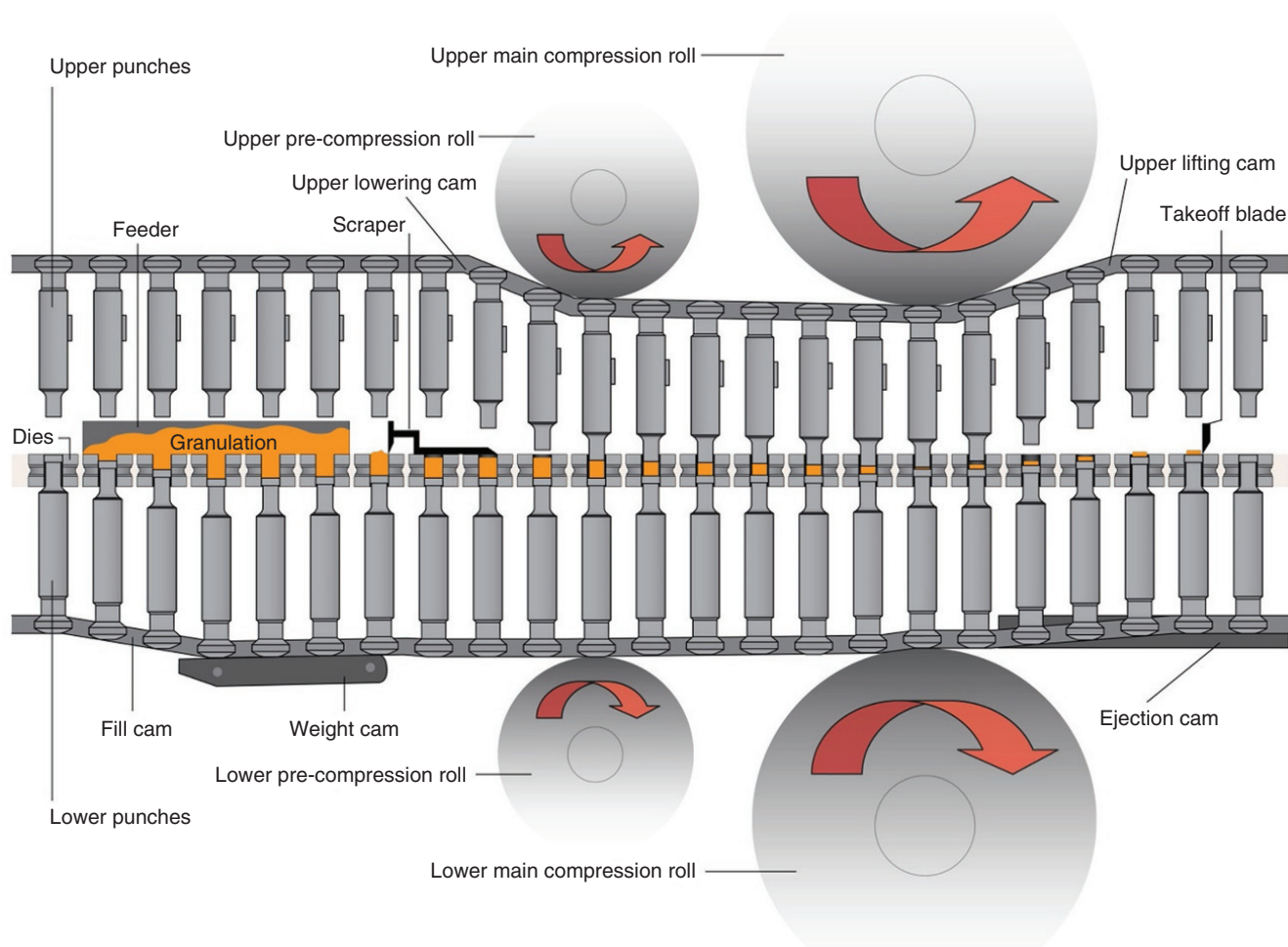


FIGURE 10.2 Typical rotary tablet press and the process of compressing powder into compacts. *Source:* Courtesy of Natoli Engineering Company, Inc., Saint Charles, MO.

Higher drug load may require more lubricant or possibly a combination of lubricants. This aspect of formulating is addressed in Section 10.4.

10.2.2 Engineering Factors

The engineering factors include compression force, tableting speed (punch dwell time in the die), punch geometry, and tooling material (punch coatings). If punch sticking is a problem, compression force and dwell time would be investigated first. Next would be punch geometry and punch coatings.

10.2.2.1 Compression The tableting process in a typical rotary tablet press is illustrated in Figure 10.2. Tablet punches are used to compress a blend or granulation into a compacted unit dose. First, the bottom punch is lowered so that the feeder can dispense powder into a die. A scraper removes excess powder so that the formulation is level with the top of the die. The amount of powder in the die is the fill. The upper punch is then lowered into the die. The

pre-compression rolls apply pressure to deaerate the fill in the die by forcing the particles closer together. This is followed by the main compression rollers applying force to compress the powder into a compact. The upper punch is then lifted from the die by the upper-lifting cam. The lower punch is then raised by the ejection cam to eject the compact out of the die so that a scraper can remove it.

Compression enables particle-to-particle bonding for the purpose of making a tablet of sufficient cohesive strength or hardness, to maintain its shape and appearance. Compression includes the compression force (F_D) and resistance in the die. In Figure 10.3, compression is represented by the two large vertical black arrows. A portion of that uniaxial force is distributed as a lateral radial force on the formulation that pushes against the die wall [23].

Most of the force consolidating the powder is at first generated by the upper punch, delivering the main compression force (F_{UF}). The lower punch applies both resistance to the upper punch and compression force (F_{LF}). As these forces are applied, the powder moves within the die, and applies

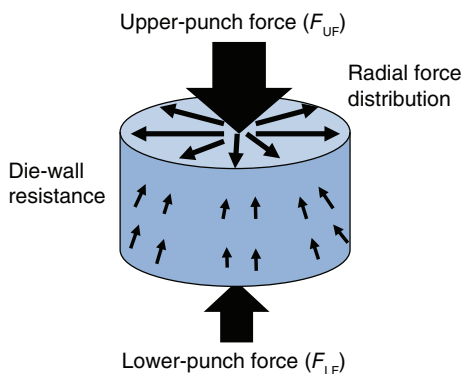


FIGURE 10.3 Forces on a cylindrical flat-faced tablet generated during compression. *Source:* Courtesy of Avik Sarkar.

radial force against the die wall. The die-wall pressure is the resistance applied back to the compact as indicated by the smallest black arrows. If the compact adheres to the die wall, more force is required to eject it due to the die-wall friction.

When powder is added to a die, there is space between the particles (Figure 10.4a). As the powder undergoes compression, the particles are shown to consolidate and the powder fill deaerates (Figure 10.4b). The die constrains powder movement and provides time for adhesive particles to bond to the die wall as well as for cohesive bonding of particles to each other. This is followed by particle rearrangement, plastic and elastic deformation, and then fragmenting of the brittle components once they have exceeded their deformation limit. The pressure enables formation of interparticulate bonds that provide the compact's tensile strength. After ejection from the die, the compact undergoes a type of elastic recovery that is sometimes referred to as relaxation, whereby the elastic materials partially return to their pre-compressed forms (Figure 10.4c).

The cohesiveness of compressed particles holds the shape of the compact after compression. If the adhesive attraction is greater than the cohesion of the particles, some of the powder sticks to the punches and is known as punch sticking (Figure 10.4c). Brittle excipients would fracture first and generate more surface area. With increased surface area, more excipient contacts the die and punches, and less API can bind to them, thereby reducing punch sticking.

When a formulation sticks to the punches, the adhesive bond (B_A) with the tooling is stronger than the interparticulate cohesive bonds (B_C) that hold the compact together: i.e. $B_A > B_C$. During production runs, repetitive compressions of an adhesive formulation result in a buildup of material on the punch surface. As this layering effect continues, the buildup added to the die fill causes higher compression force and a potential for damage to the punches and or the tablet press [24]. It has been shown that the mass of formulation adhering to the punch can also decrease with subsequent compressions as has happened with ibuprofen and mannitol formulations [25]. In those experiments, the formulations continued to

build up on the punch tip until a critical accumulation was achieved and some of the adhered material transfers to a compact surface, creating a defect. The nature of the punch-sticking phenomena has been found to involve numerous factors and the unique properties of each adhesive API [6, 26].

Ibuprofen has been used in punch-sticking investigations because of its propensity to adhere and layer on tablet punches [1, 27–29]. It is a low-melting solid and Bechard et al. [30] stated that the localized high-temperature zones, which are generated by compaction forces and friction during particulate movement, could exceed the melting temperature of low-melting compounds, thereby facilitating powder adhesion to punches. By this reasoning, low-melting solids would be more likely candidates for punch sticking.

10.2.2.2 Dwell Time Interparticulate and adhesive bonding can both be affected by not only the force of the compression but also by the punch dwell time during which the powder is compressed in the die. Dwell time refers to the length of time that a punch is compressing powder in the die, that is, the time the dies are between the rollers as shown in Figure 10.2. It is a function of tableting speed, whereby faster tableting leaves less time for the punch to compress the powder in the die. Lengthening the dwell time provides more time for particle deformation, fragmentation for increased surface area, and for interparticulate bonds to form. It also provides more time for adhesive bonding to the punches and die [1, 31]. A kind of bonding strength contest exists between B_A and B_C and if B_A becomes greater than B_C by compression force and/or dwell time, then the formulation becomes more susceptible to punch sticking.

Altering the compression force and/or dwell time can either improve or negatively impact punch sticking [32, 33]. If higher compression fractures API particles, the increased particle surface area permits more contact with the punches. Therefore, increasing the compression force could increase punch sticking if the API's $B_A > B_C$. This result has been reported with ibuprofen [34]. Changing the tableting speed, or punch dwell time, also has different effects on punch sticking. During compressions, if B_A requires more time to become stronger than B_C , then shortening the dwell time can help reduce punch sticking. Conversely, if the particles require more time to form interparticulate bonds, then increasing the dwell time could reduce punch sticking. This means that contact time with a punch is specific for B_A to be stronger than B_C . A shorter dwell time also provides less time for an adhesive API to make contact with the punch surface. Both compression force and dwell time should be investigated with predictive punch-sticking tests since formulation changes might not be a possible method for solving the issue. At the very least, testing these two parameters would show the best conditions to manufacture tablets that contain an adhesive API. It should also be noted that while a change

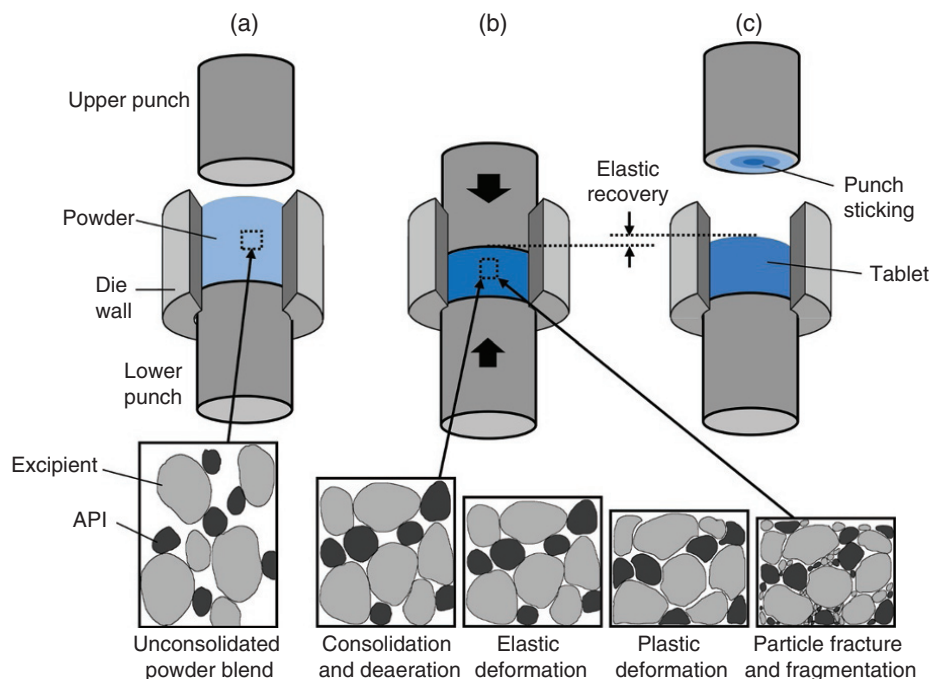


FIGURE 10.4 Stages of powder compression within the tableting die: (a) powder blend with spaces among the irregularly shaped particles of API and excipients during fill and pre-compression phase; (b) powder compression showing particle consolidation and deaeration, followed by elastic and plastic deformation, with particle fracture when compression force exceeds the elastic limits; (c) and finally, elastic recovery after compression force is removed when powder sticking to the upper punch becomes evident (punch sticking). *Source:* Courtesy of Avik Sarkar.

in the compression force is the simplest variable to adjust during development, doing so affects tablet hardness and subsequently, harder tablets may have longer disintegration times and dissolution rates [24]. When finding the balance between compression force and dwell time, tablet attributes must be considered.

A target solid fraction (SF) is a useful physical characteristic to guide tableting process development balancing the optimal compression force for the preferred tablet size, disintegration time, and friability. SF is a reduction in volume due to the compression events of deaeration and consolidation, and is a convenient factor to monitor. It is calculated using the true density (ρ) of the powder, and the volume (v) and mass (m) of the compact. The SF of a cylindrical compact can be calculated from Eq. (10.2).

$$SF = \frac{m}{\rho v} \quad (10.2)$$

Tablets with curved surfaces require a modified equation to account for the correct volume. The punch drawings from the manufacturer supply the punch cup depth necessary to calculate the volume for a single-radius tablet, e.g. standard round convex tablet that has two convex sides. A biconvex tablet shape is represented in Figure 10.3. The distance between the top of the band and the top of the tablet corresponds to the punch cup volume (C_v).

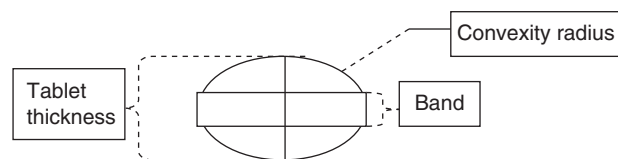


FIGURE 10.5 Biconvex tablet shape showing the radius of the convex surface and the tablet band used in calculating tablet volume.

The calculation for volume of a biconvex tablet with a single radius includes the cup volume (C_v) of the punches and the band volume (B_v) of the tablet, i.e. tablet thickness minus twice the cup depth (Eq. 10.3) (Figure 10.5).

$$v = 2(C_v) + B_v \quad (10.3)$$

The same tablet can be used to measure and monitor tensile strength, a second convenient factor, which is based on a tablet's diametral crushing strength, and is known as tablet hardness. This test can also be used when optimizing tableting conditions for each adhesive API. An example of a tablet tested for hardness is shown in Figure 10.6.

The tensile strength calculation for cylindrical tablets is shown in Eq. (10.4), where σ_x is tensile strength, F is the force applied to fracture the tablet along its diameter, D is the tablet diameter, and H is the tablet thickness [35].

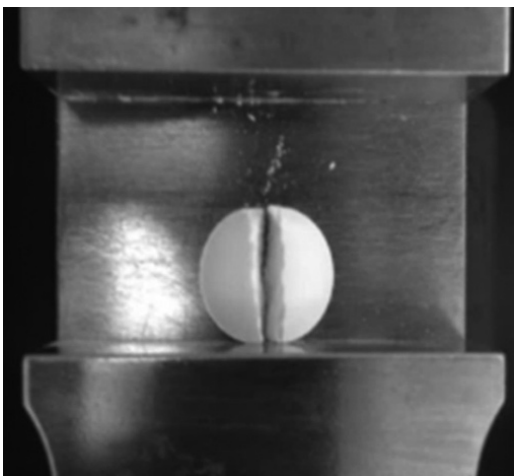


FIGURE 10.6 Example of a tablet tested for a hardness value to be used in calculating tensile strength.

$$\sigma_x = \frac{2F}{\pi DH} \quad (10.4)$$

10.2.2.3 Tooling: Material and Punch Tip Coatings

The punch surface material of construction (type of metal), physical condition, along with applied coatings are factors to be considered during development [1, 36].

Punch surfaces that are worn are more likely to have surface imperfections and to show punch sticking with fewer compressions than newer and smoother surfaces. Polishing the punches prior to use is a common manufacturing practice to produce a better tablet surface appearance. Smooth, polished surfaces can delay the onset of punch sticking. With new punches, there is less sticking than with older punches that have rougher, abraded surfaces from wear and/or corrosion [32]. However, a contradictory result was reported by Schuman and Searle [37], whereby a more pitted punch surface provided less sticking than the smooth one. They suggested that the imperfections “broke the suction” between the tablets and the punches.

During late-stage development, it is not always possible to make formulation changes once ICH stability batches have been prepared, without significant delay to the registration of the product. If punch sticking then becomes evident, testing various punch tip coatings at the development scale is suggested, prior to attempting a manufacturing run. Certain coatings provide a surface that is less attractive for bonding and less prone to wear or corrosion [38, 39], but punch sticking on those coated surfaces will be API or formulation dependent. The exact processes used to apply the coatings are proprietary; however, the coatings themselves are published in the literature. The application processes include electroplating for nickel and chromium, to which additions can include either boron for hardness or Teflon™ (PTFE)

for lubricity. Other processes are physical vapor deposition (PVD) for diamond-like carbon, titanium nitride (TiN), and chromium nitride (CrN) during which the punches are treated to temperatures in excess of 900 °F, which can cause them to be out of tolerance. Both TiN and chromium nitride (Cr₂N) coatings can be applied by ion beam-enhanced deposition (IBED), which is a lower temperature procedure that does not risk alteration to the tooling dimensions [40]. Other successful coatings are TiN and chromium nitride which provide very hard and very smooth surfaces that are durable and able to resist abrasion and wear [41–44].

While chromium coatings are used most predominantly, it has also been shown to be ineffective in reducing punch sticking for certain APIs [32, 37]. Tsiftsoglu and Mendes used a boron alloy coating that reduced punch tip adhesion and capping tendencies while lowering ejection force [45]. However, Shah et al. reported higher punch tip adhesion and increased ejection force with a boron coating [46]. As with each solution to punch sticking there is no universally successful coating. Therefore, having a toolbox of punches with a variety of coatings to use during screening runs and during development would help to determine if a punch tip coating could reduce or eliminate punch sticking for each API. Reed et al. reported the simultaneous use, in a single compression run on a rotary tablet press, of several punches with different proprietary coatings, each having a unique identifier that made tablets from each punch identifiable [47]. This enabled determination of the best punch coating to eliminate sticking for each API, and determination of the drug load in the tablets that results in no punch sticking.

10.2.2.4 Punch Geometry

Since compression force, whether relatively high or low, is a factor in punch sticking, it is reasonable to expect punch geometry to also be a factor in punch sticking due to its effect on the distribution of compression force and resulting tablet hardness. In fact, punch geometries have been shown to affect both compression distribution and punch sticking [48, 49]. Roberts et al. showed more punch sticking with flat-faced punches (FF) than with concave [28]. Aoki and Danjo similarly found less sticking with standard round concave (SRC) punches vs beveledged or FF [34]. Since the shape of the punch face affects distribution of the tablet’s surface hardness, and compression force affects punch sticking, it follows that punch sticking will differ among punches having different cup curvatures. The distribution of tablet surface hardness from compression force is due to the punch face curvature. It is known that tablets with standard convex shapes have greater hardness near the tablet band than do FF tablets [50, 51]. Hardness is greatest at the curved tablet’s periphery but flat tablets have greater hardness in the center. Tablet tooling has been evaluated for geometric shape effects on hardness: FF, shallow convex, SRC, deep convex (DC), and ball shaped. The tablets showed that indentation hardness decreases as tablets’ convexity

increases [50]. This was the basis for a predictive punch sticking test that was developed using two proprietary blends, whereby FF punches were shown to result in punch sticking with less than 50 g of formulation, even though punch sticking did not become evident with SRC until large batches of formulation were used in a manufacturing run [4].

Because of the multiple mechanisms that cause punch sticking, one or more predictive tests should be utilized as part of the development process and engineering factors should be tested for potential solutions. Those factors include compression force, punch dwell time, punch geometry, and punch coatings. Combining the partial solutions instead of searching for a reliable single one decreases risk, saves development time, and therefore has positive financial impact.

Each engineering factor is independent of the others with regard to punch sticking, and must be tested independent of the others, preferably in a designed experiment to select the best condition of each.

10.2.3 Environmental Factors

An environmentally controlled production area is ideal for a commercial drug product-manufacturing facility. Both temperature and moisture have been reported to affect punch sticking during manufacturing, as can electrostatic forces generated from powder flow over surfaces. Therefore, this section discusses the three main environmental factors that influence punch sticking.

10.2.3.1 Moisture To test the effects of moisture on punch sticking, scraper pressure (SP, i.e. force/area) was monitored on a rotary tableting machine with SRC tooling. When the scraper removed a tablet from the lower punch, the takeoff force against the scraper is measured. Danjo et al. showed that the SP increased with water content up to 3%, after which SP declined with higher water content [52]. They found a similar relationship with tablet surface roughness (Ra) measured by scanning laser microscopy. The simultaneous increases in SP and Ra suggested a relationship between punch sticking and water content, up to 3% moisture. Moisture values above 3% did not contribute to higher SP. It was demonstrated that powders adsorb moisture from ambient vapor. Ahlneck and Alderborn studied the impact of relative humidities (RH) from 33 to 100% on sticking [53]. When adsorbed to the particulate surfaces, the water is a condensate. In the case of sodium citrate, a large amount of moisture was absorbed as crystal hydrate. Water content due to humidity can increase or decrease tablet tensile strength. Since tablet strength has a role in punch sticking, it follows that moisture could also affect punch sticking for some materials. Ibuprofen tablets were found to have increased tensile strength with rising moisture levels from 1 to 3.5% [54]. This was explained as being due to the lubricating effects of moisture more evenly distributing the

compaction forces and improving plastic deformation. Higher moisture content resulted in weaker tablets and it was concluded that ibuprofen, as an adhesive agent, would therefore be susceptible to punch sticking due in part to its moisture content. Environmental control of humidity in the manufacturing facility is beneficial not only to controlling punch sticking but also to powder flow, adhesion to other manufacturing surfaces, cohesion, and tablet tensile strength.

10.2.3.2 Temperature Temperature effects on punch sticking are more noticeable with low-melting solids because they are more susceptible to heat effects during compression. The heat generated by mechanical energy and its conversion to thermal energy during compression is due to frictional forces of compressing powders and ejection from the tablet die [55]. Danjo et al. conducted compression experiments with butyl *p*-benzoate, which has a melting point of 68 °C. Punch sticking was indicated by increasing SP and by examination of the tablet surfaces. The results indicated that punch sticking occurred under constant conditions of temperature and humidity, due to the temperature of the tablet press and particle diameter of the API [56]. The tablet press became warmer with continued operation due to the cumulative effect of frictional forces, most notably the compression force on the powder blend as thermal energy is transferred from the punches to the powder. A relevant factor to equipment temperature is the glass transition experienced by some materials such as Eudragit™ [57]. During compression, an excipient with a low glass transition temperature could induce transition to the glassy state, which may be sticky or adhesive. Cespi et al. concluded that determining the temperature sensitivity of materials, particularly those characterized by glass transition at low temperature, could provide useful predictive information about punch sticking potential [58].

10.2.3.3 Electrostatic Forces Manufacturing processes involve powder flow and sliding friction which can impart electrical charge to the material. This is known as triboelectrification [59]. Combined with van der Waals forces, charging particles can result in powder adhering to equipment surfaces, including tablet punches. As charge builds in a blend, the adhesion potential increases. Samiei et al. demonstrated that the charge in a blend has a decay time, and that punch adhesion had a linear relationship with the decay [60]. They found that the API alone could be electrostatically charged more than could an active blend. Yet, API alone had a faster charge decay than did the blend, even when drug loading was as low as 10%. The excipients provided an insulating effect on loss of charge for the API. They theorized that a formulation with a fast charge decay rate would have reduced punch sticking potential. Negating electrostatic forces could require grounding for large equipment if punch sticking potential exists.

10.3 PUNCH STICKING TESTS

There are numerous techniques for determining punch sticking. The final proof of whether there is a punch sticking problem is at the tableting process stage when increased ejection forces, dulling, or defects on tablet surfaces are observed as detrimental impact to the product quality. However, the obvious advantages of being able to predict the problem makes it desirable, if not necessary, to test the probability of punch sticking before production. The earlier that potential for punch sticking can be raised as a concern, the more options there are to address or eliminate the issue. At the same time, the drug load of a formulation could be optimized, comparisons of granulations and direct compression blends could be made, engineering and environmental conditions could be evaluated and optimized, and the lubricant selected.

There are several published testing methods for identifying the potential for punch sticking. Testing for punch sticking should be a part of the development process. Some test methods are complex and measurements involved may require formulation modifications. The formulator must consider whether quantitative or qualitative methods are preferred, whether a test would be for screening or manufacturing or both, and the time and resources required. Each test should be coupled with visual inspection of both the tablets and the tooling. Punch sticking tests should provide information related to the best engineering conditions and the preferred environmental conditions. If formulation adjustments are an option, then testing should show the best formulation to use for that API.

10.3.1 Compression Tests with Challenging Punches

Simply producing tablets to determine punch sticking propensity could be expensive since a significant amount of API might be used. Yet, this would be the most reliable way to uncover the problem, i.e. predicting whether there would be a punch sticking problem. It has been shown that punch geometry is a factor in tablet properties [28, 49–51]. Much of the experimentation has been done with FF punches because studies have shown that FF punches are likely to result in sticking much earlier than SRC punches. An investigation using FF to develop a quick, qualitative, predictive visual test for punch sticking was successfully concluded [4]. The approach to this study was simply to observe whether defective tablets would result due to punch sticking and to observe whether any of the formulation was adhering to the punches. The test method showed punch sticking on FF punches with 200 g of formulation using a compaction simulator compared to 2.5 kg with SRC punches in production scale. This test was later used in the development of six formulations that were manufactured successfully at clinical scale. Only 200 g of formulation were required with a compaction simulator. The test was later refined to use 50 g of

blend with 0.25" FF punches. The extent of punch sticking was not found to differ between 0.50" and 0.25" round FF. This is in agreement with the findings of Roberts et al., who demonstrated reduced punch sticking with increased punch tip curvature, but found no difference related to punch tip diameter [28]. Therefore, this material sparing visual test could be performed with small-sized FF punches with the advantage of visually confirming both the defective tablets and buildup on the punches. A numerical ranking of the punch sticking amount was used in this method whereby (i) indicated some film on the punch surface but no tablet defects were visible; (ii) indicated some accumulation of material on the punches; however, the tablets showed acceptable appearance; and a rating of (iii) was assigned when formulation accumulated on the punches and defective tablets resulted.

10.3.2 Quantitative Accumulation

A quantitative test was developed using a removable FF punch tip that was weighed between a predetermined number of compressions to show accumulating mass of formulation adhering to the punch tip over time [25]. The advantages of the removable tip include (i) the ability to remove and weigh the tip on an analytical balance since the tip mass is less than 5 g compared to removing the entire punch, which would have a mass greater than 100 g, (ii) eliminating a need to recalibrate the press as is required after removing and reinstalling a punch, and (iii) being able to do a visual assessment of the tip. This test was successfully used with ibuprofen formulations, and in a comparison of mannitol grades. It has been used subsequently in the development of problematic APIs to produce successful formulations. An alternative to weighing the punch tip would be to assay the adhering material after some number of compressions and determine whether the API content is higher in that material than in the formulation. This would demonstrate that the API is the component that caused the issue.

10.3.3 Use a Toolbox of Various Punch Coatings

The punch roughness and surface properties can be altered through polishing or by altering the surface with a new material that minimizes adhesive interactions (coatings). As mentioned previously, no one punch coating solved punch sticking for all formulations. Contrasting findings bear this out. Pedersen's chrome plated tip was sometimes effective against punch sticking [61]. Other investigators found chromium electroplated tooling to be no better than plain tooling, yet chromium nitride prepared by ion-bombardment eliminated punch sticking for other investigators [38]. Boron-alloy coating increased ejection and adhesion forces in one report, yet it lowered ejection forces in another study and resulted in tablets with better surface quality [45, 46]. These contrary

findings support the fact that punch sticking occurs by multiple mechanisms.

In an investigation, a “compression toolbox” of punches with different coatings was used to test the sticking potential of five different APIs in formulations [47]. The coatings were not named due to confidentiality agreements with each of the tooling manufacturers, but the authors claim that the 24 coatings are freely available from manufacturers. Each of the coated punches had a unique identifying notch that allowed tablets to be distinguished. The tablet press was set up using nine different coated punches with their distinctive markings so that a variety of punch coatings could be evaluated in a single tableting run. Tablets from each punch were subsequently segregated for later analysis. This process was repeated with another set of nine coated punches until all punches had been tested. The advantage of testing multiple tooling coatings in one run is that it eliminates batch-to-batch variables for the formulations and the press operations. It also saves time and API. When a formulator is unable to change the components or levels in the formulation, such an efficient test of multiple coatings reveals whether there is sticking potential, and what might be the best punch coating to solve it.

10.3.4 Takeoff and Ejection Forces

Takeoff force is the amount of energy necessary to remove a tablet from the bottom punch after compression. It has been used as a quantitative test to indicate punch-sticking potential [5, 62]. A limitation of this technique on a rotary tablet press is that the relatively low amount of energy for tablet removal is coupled with the large rotational energy of the press as tablets make contact with the takeoff bar. Given the fluctuations for takeoff during normal tablet production, the accuracy of the takeoff measurement changes due to sticking has a large error. The variability of normal takeoff forces could exceed the sensitivity to takeoff values that would be attributable to punch sticking. Compaction simulators are more suitable in measuring takeoff force since they have no rotary force in the measurement [63, 64].

Ibuprofen has been studied extensively as a model API for punch sticking. Saniocki et al. used ibuprofen in studies of both takeoff and ejection forces [64]. It was found that the takeoff forces with all ibuprofen formulations were low despite the obvious sticking [27]. Detachment of a tablet from the punch occurs as a result of either weak adhesive bonding (B_A) to the punch or weak cohesive bonding (B_C) in the compact. When B_C is weaker than B_A , there is punch sticking. Since some compacts broke upon contact with a takeoff bar on the tablet press, any obvious sticking would not relate to detachment from the lower punch. Due to the complexity of sorting these forces and determining any other factors involved, takeoff force measurement may be a less-reliable test method to predict punch sticking.

Ejection force, i.e. the force required to remove a compact from the die, varies with compression force, the formulation excipients, tablet hardness, and additional factors. An instrumented tablet press or a compaction simulator could measure the forces. Wang et al. concluded that ejection and takeoff forces were related to punch sticking in their experiments with ibuprofen, ketoprofen, and flurbiprofen [5]. However, they also remarked that due to the numerous factors involved in ejection force, it is difficult to distinguish the contributions from those factors. It was determined that ejection force could not be used to rank-order punch sticking. Conversely, it was reported that punch sticking with plastic/brittle formulations was confirmed by high maximum die wall pressure [65].

10.3.5 Surface Profilometry

The expectation in measuring surface roughness (R_A) is that as punch sticking develops during a tableting run, more formulation would accumulate on the punch and the R_A of the tablets would increase over the run duration. R_A can be measured by atomic force microscopy and an image of the tablet and/or punch surface can be generated with nanometer resolution. By this method, a probe on a cantilever is drawn over the surface of a tablet or punch and moves along the z -axis in correlation to the surface quality. From the cantilever deflections, an image can be generated along with the measurements taken during the x - y scan. The theory holds that as material increasingly adheres to the punches, the tablet surface would become rougher and therefore R_A values increase. AFM limitations include the small imaging area and shallow height that it is capable of measuring. The test often requires several minutes but investigators have claimed success with this technique [2]. AFM testing requires method development because some API formulations require specific conditions for this type of assessment. One such study was conducted with silicon nitride probes that were modified by sputter coating with iron and chromium. The cantilever tips were then examined by SEM with dispersive X-ray. At first, the measurement of force–displacement was done in ambient conditions, but to improve the resolution, it was also performed in water saturated with the API of the compact being examined. AFM was shown to be a valuable tool in studying punch sticking; however, it is not quick and not always convenient for screening.

10.4 PUNCH STICKING SOLUTIONS

There is no single way that is guaranteed to solve all punch sticking, but there are numerous solutions. Once the problem is known, testing the utility of individual remedies allows a formulator to select the best means of eliminating it, and if necessary, combining strategies to overcome it.

10.4.1 Predictive Test During Development to Identify the Problem Before Manufacturing

This is a preemptive measure of determining whether or not punch sticking could be a problem. Small batches, tableted under challenging conditions, serve as a conservative prediction as mentioned in the testing section. The test selection criteria should ensure rapid turnaround to allow screening several formulations and to provide tablets that show defects if the formulation is problematic. Preparing small formulation batches and tableting with challenging tooling such as FF enables the formulator to test the variables of lubricant levels, excipient selection and amounts, and drug load.

10.4.2 Formulation

10.4.2.1 Granulation Granulation is the process of causing primary powder particles, such as API and excipients, to form larger, denser granules that have improved flow and reduced API exposure and therefore, reduced punch sticking.

Granulating an adhesive API encases it, reducing exposed surface for binding to equipment. The two main granulation methods are dry granulation and wet granulation. The dry granulation process requires compressing the initial blend into slugs or a ribbon, which is then milled. Lubricant is mixed into the blend prior to compacting ribbons as intragranular (IG) lubrication, and then again mixed into the granulation as extragranular (EG) process. With dry granulation, some of the achievable tensile strength of a tablet is lost due to compaction of the initial blend. After the first densification there is less compressibility for a second compaction during tableting. However, the achievable goals of granulation are improved flow; larger, denser particles; and improved uniformity which all help to reduce or eliminate punch sticking. Granulating the adhesive API reduces available surface area for bonding to equipment surfaces and insulates the API against contacting the punches. The lubricant content of a formulation can be portioned between IG and EG fractions to reduce or eliminate punch sticking.

In wet granulation, a liquid, usually water, is added in very low amounts to the powder blend in a closed vessel with agitation. The result is the powder particles forming wet agglomerates; this part of the process being referred to as wet massing [66]. Mixing the liquid into the powder blend aids in formation of interparticulate bonds. To facilitate that bonding, a binder is used, such as polyvinyl pyrrolidone. The resulting mass is denser than the original blend and after the drying step, is milled to form the final granules. Once lubricated, the granulation is ready for tableting. The wet granulation process takes more time and energy than dry granulation. However, both granulation processes can result in reduction or elimination of punch sticking.

In an alternative approach to granulating, Yoshikawa et al. coated ibuprofen with an aqueous mixture of hydroxypropyl

methylcellulose (HPMC) and Plasdone S360 prior to successful tableting [67]. The process formed granules with reduced excipient load and elimination of punch sticking.

To avoid or reduce punch sticking, the same strategy that is used in formulating blends for direct compression applies to granulations. Larger API particles and smaller excipient particles reduce available API surface area for adhesive interaction with the tablet punches.

10.4.2.2 Lubrication If punch sticking becomes apparent, one logical approach to solving the problem would at first seem to be to add more lubricant or to lubricate the blend with a longer mixing time. This would require development and testing to determine the appropriate amount of lubricant and the optimal mixing time. It has been shown that the reasons for excessive lubrication include higher tablet friability, lower compact strength, and due to the hydrophobicity of MgSt, a longer dissolution and extended disintegration time [68]. Solutions to eliminate punch sticking must be balanced against these possibilities. It should be noted that variations may exist in excipients from different suppliers and among batches from the same supplier. It is also known that glidants such as silicon dioxide and CS reduce the lubricating efficiency of MgSt [69]. The use of colloidal silica can improve tablet strength, probably by interfering with MgSt lubrication efficiency. Pingali et al. demonstrated that lubrication effects are impacted by the mixing order of CS and MgSt [22]. They explained that due to the lubricant forming a film on the outer layers of the CS particles, MgSt becomes less effective as a formulation lubricant. This was supported by the higher hydrophobicity of the particles in the ordered mixing as compared to adding both CS and MgSt at the same time.

Because lubricants like MgSt form a film over particle surfaces, and filming is affected by the particle size of the lubricant, it is recommended to de-lump the lubricant by sieving before blending, in order to maximize lubrication efficiency.

10.4.2.2.1 Internal Lubrication Internal lubrication refers to lubricant in the granules portion of the formulation. This includes lubricant added to a blend for direct compression or for granulating, and lubricant added to a granulation.

10.4.2.2.1.1 Intragranular and Extragranular Blending API and excipients, and then compacting and milling ribbons require addition of an IG lubricant. The EG lubricant is blended with the granules to permit tableting. Where adhesion events remain problematic, the formulator can reappportion the amounts of IG and EG lubricants. Combinations of different lubricants might also be used, such as glycerol behenate IG and MgSt EG. In some cases, two different lubricants may both be added, IG or EG. Punch sticking tests would confirm the value of the lubricants.

10.4.2.2 External Lubrication External lubrication (EL) is an option that does not blend a lubricant into the formulation. Instead, the lubricant is sprayed onto the tablet tooling prior to blend addition into the die. The lubricant may still be in the blend to aid powder flow. However, there can be a substantial reduction in the amount of blended lubricant when it is applied directly to the tablet tooling. It should be noted that since application of the external lubricant involves spraying, the lubricant will be found on more than the tooling. Even though this technique requires optimizing the spray rate, there is still a more extensive cleanup than with internal lubrication alone. Yamamura et al. found that in an eprazinone hydrochloride formulation, only 1/13th of the lubricant was required with EL compared to internal lubricant in the formulation content [70]. By using EL, they achieved a 40% higher tablet crushing strength. The lubricant that adhered to the tablets formed a thin layer on the surface that was thinnest at the tablet center. Another investigation utilized ejection force to test EL effects on punch sticking [71]. By using ejection force test measurements, it was possible to obtain stronger tablets and eliminate punch sticking.

Other than the favorable effects on punch sticking, there are added benefits of this method that include the use of less lubricant and the ability to make harder tablets. EL has great success and may turn out to be the one universal solution to punch sticking.

10.4.3 Engineering

10.4.3.1 Compression Force and Dwell Time Compression force and dwell time are discussed in the Engineering Factors of punch sticking. Optimizing these conditions in small-scale development batches prior to large-scale manufacturing is an important measure against adhesion events. Whether to increase or decrease compression force and/or dwell time during compression needs to be determined by small-scale batches or science of scale tools to provide an avenue to solve punch sticking. Optimized compression force and dwell time could be coupled with changes in lubricant levels or selection of punch tip coating.

10.4.3.2 Punch Coatings The compression toolbox consisting of challenging punch shapes with different coatings requires a larger amount of blend if run on pilot or commercial scale than does the predictive test. However, this permits the formulator to find out which punch coating type is the least adhesive to the API. Coupling that knowledge with the right compression force and dwell time helps in selecting the formulations and the preferred conditions under which to manufacture tablets.

10.4.4 Environmental

The three environmental factors of moisture, temperature, and electrostatic forces are external to formulations and

equipment. In a development laboratory and in a manufacturing facility, they should be under control. This is especially important if they are found to contribute to punch sticking in order to make adjustments.

10.4.4.1 Temperature and Humidity During development, temperature and RH (moisture) should be noted so that if punch sticking occurs, the formulator could investigate the effects of alternative conditions. A climate-controlled system maintains relatively stable conditions, usually within a range set for comfort of the manufacturing staff. When a batch is manufactured, climate factors are required to be monitored and recorded for the duration of production. If punch sticking occurs, adjusting the temperature and/or humidity may help reduce it and the preferred manufacturing conditions would be known for a specific product.

10.4.4.2 Electrostatic Forces Electrostatic charge builds with continuous movement of powder across surfaces, and with metal machine parts rapid movement, such as that occurs with a rotary tablet press. The static charge from this movement makes the machine parts attractive to the powder. If punch sticking occurs with a tablet press, electrical grounding is a factor that could be quickly checked, as it can be a contributing factor to punch sticking. Fortunately, modern manufacturing equipment is grounded to prevent static buildup.

EXAMPLE PROBLEMS

Problem 10.1

A series of tablets were made at seven different compression forces, using FF punches with 8.00 mm diameter (D), for a hardness/compression profile. The compression and tablet data were used to calculate and plot a tensile strength/solid fraction (TS/SF) profile. The true density (ρ) of the blend is 1.48 g/cc, measured with a helium pycnometer. The tablet target weight is 200 mg.

- Using the data given above and in Table 1-1, and Eq. (10.2) for SF and Eq. (10.4) for tensile strength (σ_x), complete the table. Convert σ_x units from kilopond (kp) to megapascal (MPa) (1 kp = 9.807 MPa) and plot a hardness/compression profile, and then plot a TS/SF curve.
- Use the formula for volume of a cylinder (FF tablet) to calculate the tablet volume [$V = (\pi r^2) \times H$]. The first row is done for you.
- The ideal TS lies between 1.25 and 2.25 MPa. It was found that punch sticking did not occur at higher compression. Based on this information, which SF should the formulator use as a target value to achieve a TS near the midpoint?

TABLE 1-1 Compression Data of an Active Blend and the Resulting Solid Fractions and Tensile Strength

Compression Force (kN)	Tablet Weight (m) (mg)	Tablet Thickness (H) (mm)	Tablet Hardness (F) (kp)	Tablet Volume (V) (mm ³)	Solid Fraction (SF)	Tensile Strength (σ_x) (kp)	Tensile Strength (TS) (MPa)
2.22	201.2	3.55	2.1	178.44	0.762	0.047	0.46
3.13	200.1	3.34	3.3				
4.97	200.9	3.21	6.3				
8.00	202.1	3.01	10.1				
14.22	201.9	2.96	12.2				
23.57	201.4	2.94	14.6				
30.10	202.2	2.89	14.8				

Problem 10.2

An active formulation made with lactose exhibited punch sticking. A subsequent formulation was made, substituting dicalcium phosphate (DCP) for lactose. A quantitative punch sticking test was conducted, using a removable punch tip that was weighed after each five compressions. For both formulations the mass of formulation plus the tare weight of the punch tip at each compression interval are given in Table 2-1.

TABLE 2-1 Punch Sticking Test Data from an API Formulated with (a) Lactose and (b) DCP

No. of Compressions	Mass of Punch Tip + Adhering Lactose Blend (mg)	Mass of Blend with Lactose (mg)	Mass of Punch Tip + Adhering DCP Blend (mg)	Mass of Blend with DCP (mg)
0	1682	0	1682	0
5	1685.2		1683.1	
10	1691.4		1683.3	
15	1699.3		1683.2	
20	1694.2		1683.1	
25	1700.6		1683.2	
30	1700.4		1683.1	
35	1705.3		1683.1	
40	1699.6		1683.2	
45	1700.4		1683.4	
50	1691.4		1683.1	
55	1702.6		1683.3	
60	1695.4		1683.2	
65	1688.3		1683.3	
70	1699.2		1683.2	
75	1696.2		1683.1	
80	1703.7		1683.4	
85	1687.5		1683.2	
90	1703.2		1683.3	
95	1706.5		1683.1	
100	1693.3		1683.2	

- A. Calculate and plot the mass of the material adhering to the punch tip for each formulation, at each compression interval.
- B. Using the two particle factors of cohesion and adhesion, explain what happened to cause decreases in powder mass on the punch tip for the lactose formulation.
- C. Does the plot of the mass of the blend containing DCP indicate punch sticking (observe the y-axis scale)? Explain the results of the plot. By what other means would a supporting determination for punch sticking be made for these results?

10.5 SUMMARY

Punch sticking is a challenging and significant problem throughout the pharmaceutical industry because there is no single known cause for it and there is no single solution at this time. It results as a characteristic of the API having greater bonding affinity to the equipment than to the formulation particles. Known contributing factors are classified within the three domains: formulation, engineering, and environmental conditions. To further complicate the issue, solutions to punch sticking can have negative effects on other aspects to production, such as flow, stability, or tableting.

Finding a solution first requires selecting a predictive punch sticking test method, whether qualitative or quantitative. The criteria for selecting a test should include the ability to predict punch sticking, rapid turnaround, use of small batches, and the ability to observe tablet defects should they occur. As mentioned in this chapter, not all tests work with all formulations. Therefore, it is suggested that more than one method be available. A predictive test example would be the use of a removable punch tip to weigh accumulating formulation as evidence of adhesion with successive compressions. Qualitatively, this also allows examination of the tablets from those compressions for defect confirmation. Predictive tests should be part of the development paradigm.

They are cost-effective measures that help identify successful formulations and maximum drug load, and manufacturing conditions.

Formulators may choose the larger particle-sized API and smaller-sized excipients to facilitate interparticulate bonding, but must also consider powder flow in production. A glidant can be added for improved flow but it interferes with the lubricant efficiency. Lubricants used in excess can result in low tensile strength tablets. Therefore, several small-scale batches would be made in testing various formulation factors' effects on punch sticking. These would include drug load and tablet size, particle size selection of API and excipients, chemically compatible excipients for stable dosages, lubricant type and amount, and blending time. The use of more plastic or brittle fillers affects interparticulate bonding by either deforming around the API to bond with it or fracturing into smaller particles that form more interparticulate bonds allowing for fewer adhesive bonds.

Engineering factors selected during development include punch geometry and type of material, punch coatings, and compression force and dwell time. Punches with concave surfaces tend to have less sticking than flat ones, making flat punches useful in punch sticking tests. Some types of steel have more affinity for adhesive APIs than do chrome punches. Punch coatings have been successful in overcoming punch sticking, but no one coating or type of metal works with all adhesive APIs. Therefore, a toolbox of punches with various coatings and types of metal is useful in testing punch adhesion and selecting the punch types. Using high or low compression forces, and long or short dwell times, has been shown to reduce punch sticking as well as to exacerbate the problem. These factors, coupled with preferred formulation factors can provide added assurance against punch sticking.

Environmental conditions include temperature, humidity, and static. Manufacturing is expected to be maintained at constant temperature and humidity for worker comfort and modern equipment is grounded. Deviations due to equipment failure are possible. Fortunately, environmental impact on punch sticking is minimal. Considering that development laboratory conditions would not have the same environmental requirements as a manufacturing facility, it is expected that punch sticking solutions, determined in a development lab, are generally sufficient for production.

ACKNOWLEDGEMENTS

Thanks to Avik Sarkar for his time and effort in constructing the figures for this chapter, and to Bruce MacDonald for valuable critique.

Answers and Explanations

Problem 1A FF tablets are cylindrical. Therefore, the volume is determined using the formula for a cylinder, i.e. area of a circle times the height (thickness). $V = H(\pi r^2)$.

The diameter (D) is given as 8 mm, therefore, $r = 4$ mm. The first line of the table is completed this way:

$$\text{Volume} = H(\pi r^2)$$

$$V = 3.55(3.14 \times 4^2)$$

$$V = 3.55(50.24) = 178.4$$

$$\text{Solid Fraction} = m/(\rho v)$$

$$\text{SF} = 201.2/(1.48 \times 178.4) = 0.762$$

$$\text{Tensile Strength}$$

$$\sigma_x = (2F)/(\pi DH)$$

$$\sigma_x = (2 \times 2.1)/(3.14 \times 8 \times 3.55) = 0.047$$

$$\text{Convert TS to MPa:}$$

$$\text{TS} = 0.047 \times 9.807 = 0.46$$

TABLE 1-1 Compression Data of an Active Blend and the Resulting Solid Fractions and Tensile Strength

Compression Force (kN)	Tablet Weight (m) (mg)	Tablet Thickness (H) (mm)	Tablet Hardness (F) (kp)	Tablet Volume (V) (mm ³)	Solid Fraction (SF)	Tensile Strength (σ_x) (kp)	Tensile strength (TS) (MPa)
2.22	201.2	3.55	2.1	178.44	0.762	0.047	0.46
3.13	200.1	3.34	3.3	167.89	0.805	0.079	0.77
4.97	200.9	3.21	6.3	161.35	0.841	0.156	1.53
8.00	202.1	3.01	10.1	151.30	0.903	0.267	2.62
14.22	201.9	2.96	12.2	148.79	0.917	0.328	3.22
23.57	201.4	2.94	14.6	147.78	0.921	0.395	3.88
30.10	202.2	2.89	14.8	145.27	0.940	0.408	4.00

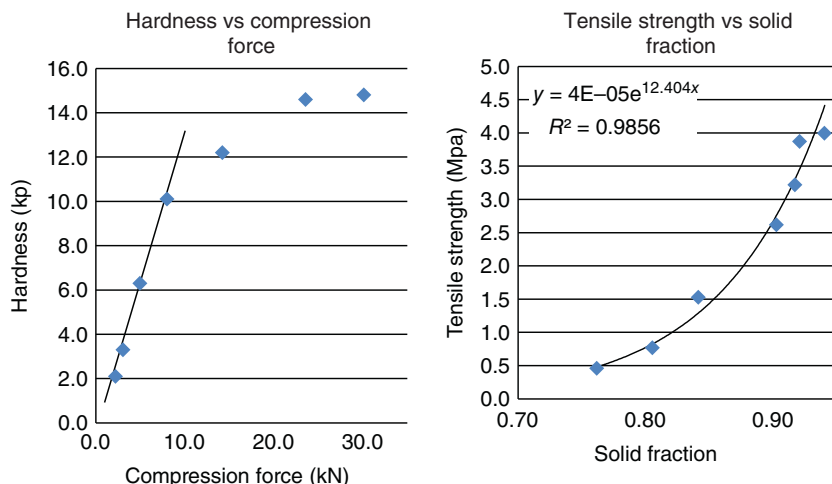


FIGURE 1-1 Hardness/compression and tensile strength/solid fraction curves showing the relationships.

Problem 1B The curve for hardness/compression begins to show a plateau because the formulation is reaching the limits of its compressibility. The line equation is shown to illustrate the exponential relationship between solid fraction and tensile strength. R^2 is a statistical correlation that provides a reliability factor in the data.

Problem 1C From the calculated values in Table 1-1, a solid fraction of 0.84 would be a targeted value.

Problem 2A

TABLE 2-1 Punch Sticking Test Data from an API Formulated with (a) Lactose and (b) DCP

No. of Compressions	Mass of Punch Tip + Adhering Lactose Blend (mg)	Mass of Blend with Lactose (mg)	Mass of Punch Tip + Adhering DCP Blend (mg)	Mass of Blend with DCP (mg)
0	1682	0	1682	0
5	1685.2	3.2	1683.1	1.1
10	1691.4	9.4	1683.3	1.3
15	1699.3	17.3	1683.2	1.2
20	1694.2	12.2	1683.1	1.1
25	1700.6	18.6	1683.2	1.2
30	1700.4	18.4	1683.1	1.1
35	1705.3	23.3	1683.1	1.1
40	1699.6	17.6	1683.2	1.2
45	1700.4	18.4	1683.4	1.4
50	1691.4	9.4	1683.1	1.1
55	1702.6	20.6	1683.3	1.3
60	1695.4	13.4	1683.2	1.2
65	1688.3	6.3	1683.3	1.3
70	1699.2	17.2	1683.2	1.2
75	1696.2	14.2	1683.1	1.1
80	1703.7	21.7	1683.4	1.4
85	1687.5	5.5	1683.2	1.2
90	1703.2	21.2	1683.3	1.3
95	1706.5	24.5	1683.1	1.1
100	1693.3	11.3	1683.2	1.2

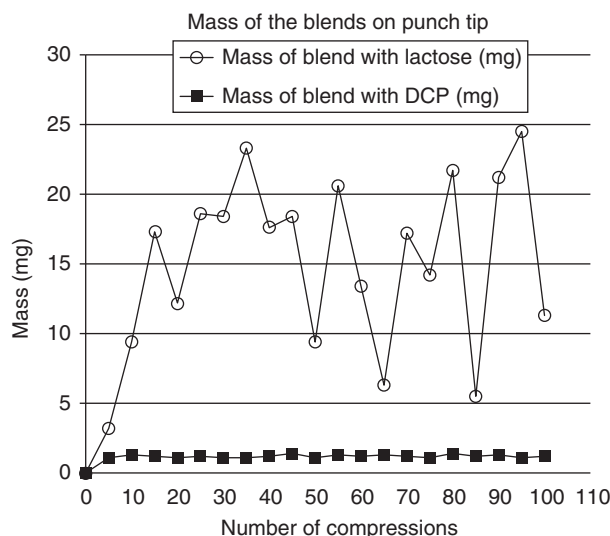


FIGURE 2-1 Blend mass adhering to a removable punch tip after each five compressions (total of 100 compressions).

Problem 2B The adhesive interparticulate bonding is demonstrated by increasing mass of the formulation on the punch tip. Where the mass decreased, the cohesive bonding was stronger than the adhesive bonding and caused displacement of the accumulated formulation.

Problem 2C The DCP formulation does not show punch sticking beyond the initial amount of formulation on the punch tip. The lactose formulation has measurable amounts of material adhering to the tip which indicates its punch sticking potential. Tablet quality by visual inspection is a supporting indication of punch sticking.

REFERENCES

1. Roberts, M., Ford, J.L., MacLeod, G.S. et al. (2003). Effects of surface roughness and chrome plating of punch tips on the sticking tendencies of model ibuprofen formulations. *J. Pharm. Pharmacol.* 55: 1223–1228.

2. Wang, J.J., Tonglei, L., Bateman, S.D. et al. (2003). Modeling adhesion in tablet compression-I. Atomic force microscopy and molecular simulation. *J. Pharm. Sci.* 2: 798–814.
3. Mazel, V., Busignies, V., Diarra, H. et al. (2013). The surface layer of pharmaceutical compacts: the role of the punch surface and its impact on the mechanical properties of the compacts. *Int. J. Pharm.* 442: 42–48.
4. Simmons, D.M. and Gierer, D.S. (2012). A material sparing test to predict punch sticking during formulation development. *Drug Dev. Ind. Pharm.* 38 (9): 1054–1060.
5. Wang, J.J., Guillot, M.A., Bateman, S.D., and Morris, K.R. (2004). Modeling of adhesion in tablet compression –II. Compaction studies using a compaction simulator and an instrumented tablet press. *J. Pharm. Sci.* 93: 407–417.
6. Markarian, J. (2013). Tablet-sticking solutions. *Pharm. Technol.* 37 (10): 52–59.
7. Saniocki, I., Sakmann, A., and Leopold, C.S. (2012). Direct compression of ibuprofen containing powder blends. *Pharm. Ind.* 11: 1842–1852.
8. Çelik, M., Ong, J.T., Chowhan, Z.T., and Samuel, G.J. (1996). Compaction simulator studies of a new drug substance: effect of particle size and shape, and its binary mixtures with microcrystalline cellulose. *Pharm. Dev. Technol.* 1 (2): 119–126.
9. Abdel-Hamid, S., Alshihabi, F., and Betz, G. (2011). Investigating the effect of particle size and shape on high speed tableting through radial die-wall pressure monitoring. *Int. J. Pharm.* 413: 29–35.
10. Caramella, C., Colombo, P., Conte, U. et al. (1988). A physical analysis of the phenomenon of tablet disintegration. *Int. J. Pharm.* 44: 177–186.
11. Saniocki, I. (2014). New insights into tablet sticking: characterization and quantification of sticking to punch surfaces during tablet manufacture by direct compaction. Doctoral dissertation, Universität Hamburg.
12. Huttenrauch, R., Fricke, S., and Zielke, P. (1985). Mechanical activation of pharmaceutical systems. *Pharm. Res.* 2: 302–306.
13. Karehill, P.G., Glazer, M., and Nystrom, C. (1990). Studies on direct compression of tablets. XXIII. The importance of surface roughness for compactability of some directly compressible materials with different bonding and volume reduction properties. *Int. J. Pharm.* 64: 35–43.
14. Rasenack, N. and Müller, B.W. (2002). Crystal habit and tableting behavior. *Int. J. Pharm.* 244 (1): 45–57.
15. Nystrom, C., Alderborn, G., Duberg, M., and Karhill, P.G. (1993). Bonding surface areas and bonding mechanism: two important factors for the understanding of powder comparability. *Drug Dev. Ind. Pharm.* 19: 2143–2196.
16. Roberts, R.J. and Rowe, R.C. (1987). Brittle/ductile behavior in pharmaceutical materials used in tableting. *Int. J. Pharm.* 36: 205–209.
17. Vromans, H. and Lerk, C.F. (1988). Densification properties and compactibility of mixtures of pharmaceutical excipients with and without magnesium stearate. *Int. J. Pharm.* 46: 183–192.
18. Bowden, F.P. and Tabor, D. (2001). *The Friction and Lubrication of Solids*. Oxford: Clarendon Press.
19. Wang, J., Wen, H., and Desai, D. (2010). Lubrication in tablet formulations. *Eur. J. Pharm. Biopharm.* 75: 1–15.
20. Li, J. and Wu, Y. (2014). Lubricants in pharmaceutical solid dosage forms. *Lubricants* 2: 21–43.
21. Braconi, P., Andres, C., and Ndiaye, A. (2003). Structural properties of magnesium stearate pseudo-polymorphs: effect of temperature. *Int. J. Pharm.* 262: 109–124.
22. Pingali, K., Mendez, R., Lewis, D. et al. (2011). Mixing order of glidants and lubricant: influence of powder and tablet properties. *Int. J. Pharm.* 409: 269–277.
23. Patel, S. and Bansal, A.K. (2011). Prediction of mechanical properties of compacted binary mixtures containing high-dose poorly compressible drug. *Int. J. Pharm.* 403: 109–114.
24. McDermott, T.S., Farrenkopf, J., Hlinak, A. et al. (2011). A material sparing method for quantitatively measuring tablet sticking. *Powder Technol.* 212: 240–252.
25. Mullarney, M.P., MacDonald, B.C., and Hutchins, A. (2012). Assessing tablet sticking propensity. *Pharm. Technol. Eur.* 36: 57–62.
26. Markarian, J. (2013). Researchers seek solutions to tablet sticking. *Pharm. Technol.* 37: 1–2.
27. Saniocki, I., Sakmann, A., and Leopold, C.S. (2013). How suitable is the measurement of take-off forces for detection of sticking during direct compression of various ibuprofen tablet formulations? *Pharm. Dev. Technol.* 18 (1): 257–265.
28. Roberts, M., For, J.L., MacLeod, G.S. et al. (2004b). Effect of punch tip geometry and embossment on the punch tip adherence of a model ibuprofen formulation. *J. Pharm. Pharmacol.* 56: 947–950.
29. Roberts, M., For, J.L., MacLeod, G.S. et al. (2004a). The effect of lubricant type and concentration on the punch tip adherence of model ibuprofen formulations. *J. Pharm. Pharmacol.* 56: 299–305.
30. Bechard, S.R. and Down, G.R. (1992). Infrared imaging of pharmaceutical materials undergoing compaction. *Pharm. Res.* 9: 521–528.
31. Malamartaris, S. and Rees, J.E. (1993). Viscoelastic properties of some pharmaceutical powders compared using creep compliance, extended Heckel analysis and tablet strength measurements. *Int. J. Pharm.* 92: 123–135.
32. Waimer, F., Krumme, M., Danze, P. et al. (1999). A novel method for the detection of sticking of tablets. *Pharm. Dev. Technol.* 4 (3): 359–367.
33. Cheng, Y., Hudacek, K., Kamali, A. et al. (2006). Small scale diagnostic tools for stickiness evaluation. World Congress on Particle Technology 5 Program, Orlando, FL (April 2006), pp. 23–27.
34. Aoki, S. and Danjo, K. (1998). Effect of tableting conditions on the sticking of tablets using ibuprofen. *Yakugaku Zasshi* 118: 511–518.
35. USP (2016). General chapter prospectus: tablet breaking force <1217>. <http://www.usp.org/node/46631> (accessed September 2017).
36. Kakimi, K., Niwa, T., and Danjo, K. (2010). Influence of compression pressure and velocity on tablet sticking. *Chem. Pharm. Bull.* 58 (12): 1565–1568.

37. Schumann, S. and Searle, G.D. (1992). The effects of chromium ion bombardment treatment of tablet tooling on tablet adherence. *Drug Dev. Ind. Pharm.* 18: 1037–1061.
38. Jehn, H.A. (2000). Multicomponent and multiphase hard coatings for tribological applications. *Surf. Coat. Technol.* 131: 433–440.
39. Konig, W., Fritsch, R., and Kammermeier, D. (1991). Physically vapor deposited coatings on tools: performance and wear phenomena. *Surf. Coat. Technol.* 49: 316–324.
40. Deutchman, A.H., and Natoli, D.C. (2002). Tablet punch performance can be improved with precision coatings. http://www.beamalloy.net/client_images/catalog19951/pages/files/Pharm-proc902B.pdf (accessed October 2017).
41. Universal Punch Corporation High Tech Coatings. http://www.universalpunch.com/high_tech_coatings.htm (accessed October 2017).
42. Cunha, L., Andritschky, M., Rebouta, L., and Silva, R. (1998). Corrosion of TiN and CrN hard coatings produced by magnetron sputtering. *Thin Solid Films* 317: 351–355.
43. Stoiber, M., Panzenbock, M., Mitterer, C., and Lugmair, C. (2001). Fatigue properties of Ti-based hard coatings deposited onto tool steels. *Surf. Coat. Technol.* 142–144: 117–124.
44. Dorfel, I., Osterle, W., Urban, I., and Bouzy, E. (1999). Microstructural characterization of binary and ternary hard coating systems for wear protection. Part I: PVD coatings. *Surf. Coat. Technol.* 111: 199–209.
45. Tsiftoglou, T.B. and Mendes, R.W. (1982). The effect of boron-alloy coating of tableting tools. *Pharm. Technol.* 6: 30–38.
46. Shah, K.B., Augsberger, L.L., and Shangraw, R.F. (1982). The effect of boron-alloy coating of tableting tools. *Pharm. Technol.* 6: 31–54.
47. Reed, K., Davies, C., and Kelly, K. (2015). Tablet sticking: using a ‘compression toolbox’ to assess multiple tooling coating options. *Powder Technol.* 285: 103–109.
48. Newton, J.M., Haririan, I., and Podzeck, F. (2000). The influence of punch curvature on the mechanical properties of compacted powders. *Powder Technol.* 107: 79–83.
49. Sixsmith, D. and McCluskey, D. (1981). The effect of punch tip geometry on powder movement during the tableting process. *J. Pharm. Pharmacol.* 33: 79–81.
50. Aulton, M.E. and Tebby, H.G. (1975). Proceedings: hardness distributions over tablet faces with different curvatures. *J. Pharm. Pharmacol.* 27 (Suppl): 4P.
51. Sixsmith, D. (1980). Punch tip geometry effects on powder compression. *J. Pharm. Pharmacol.* 32: 854–855.
52. Danjo, K., Sakiko, K., Cheng, Y.C. et al. (1997). Effect of water content on sticking during compression. *Chem. Pharm. Bull.* 45 (4): 706–709.
53. Ahlneck, C. and Alderborn, G. (1989). Moisture adsorption and tableting. I. Effect on volume reduction properties and tablet strength for some crystalline materials. *Int. J. Pharm.* 54: 131–141.
54. Nokhodchi, A., Rubinstein, M.H., Larhib, H., and Guyout, J.C. (1995). The effect of moisture on the properties of ibuprofen tablets. *Int. J. Pharm.* 118: 191–197.
55. Ketolainen, J., Ilkka, J., and Paronen, P. (1993). Temperature changes during tableting measured using infrared thermometer. *Int. J. Pharm.* 92: 157–166.
56. Danjo, K., Kamiya, K., and Otsuka, A. (1993). Effect of temperature on the sticking of low melting point materials. *Chem. Pharm. Bull.* 41 (8): 1423–1427.
57. Fujimori, J., Yonemochi, E., Fukuoka, E., and Terada, K. (2002). Application of eudragit RS to thermo-sensitive drug delivery systems. I. Thermo sensitive drug release from acetaminophen matrix tablets consisting of eudragit RS/PEG 400blend polymers. *Chem. Pharm. Bull.* 50: 408–412.
58. Cespi, M., Bonacucina, G., Casettari, L., and Ronchi, S. (2013). Effect of temperature increase during the tableting of pharmaceutical materials. *Int. J. Pharm.* 448: 320–326.
59. Bailey, A.G. (1984). Electrostatic phenomena during powder handling. *Powder Technol.* 37: 71–85.
60. Samiei, L., Kelly, K., Taylor, L. et al. (2017). The influence of electrostatic properties on the punch sticking propensity of pharmaceutical blends. *Powder Technol.* 305: 509–517.
61. Pedersen, M. (1999). Tablet tooling: design, maintenance, and troubleshooting. *Pharm. Technol. Eur.* 2: 22–28.
62. Mitrevej, K.T. and Augsbuger, L.L. (1982). Adhesion of tablets in a rotary tablet press II. Effects of blending time, running time, and lubricant concentration. *Drug Dev. Ind. Pharm.* 8: 237–282.
63. Mitrevej, K.T. and Augsbuger, L.L. (1980). Adhesion of tablets in a rotary tablet press I. Instrumentation and preliminary study of variables affecting adhesion. *Drug Dev. Ind. Pharm.* 6: 331–377.
64. Neuhaus, T. (2007). Investigation and optimization of the Presster: a linear compaction simulator for rotary tablet presses. PhD thesis, University of Bonn.
65. Abdel-Hamid, S. and Betz, G.A. (1012). Novel tool for the prediction of tablet sticking during high speed compaction. *Pharm. Dev. Technol.* 17 (6): 747–754.
66. Tousey, M. (2002). The granulation process 101. Basic technologies for tablet making. *Pharm. Technol.* 8–13.
67. Yoshikawa, T., Matsuoka, M., and Matoba, H. (2007) Pharmaceutical compositions containing coated granules of sticky drug and production thereof. Patent No. JP 2007169273 A 20070705.
68. Kushner, J. (2012). Incorporating Turbula mixers into a blending scale-up model for evaluating the effect of magnesium stearate on tablet tensile strength and bulk volume. *Int. J. Pharm.* 429: 1–11.
69. Ragnarsson, G., Holzer, A.W., and Sjogren, J. (1979). The influence of mixing time and colloidal silica on the lubricating properties of magnesium stearate. *Int. J. Pharm.* 3: 127–131.
70. Yamamura, T., Ohtaa, T., Tairaa, T. et al. (2009). Effects of automated external lubrication on tablet properties and the stability of eprazinone hydrochloride. *Int. J. Pharm.* 370 (1–2): 1–7.
71. Jahn, T. and Steffens, K.J. (2005). Press chamber coating as external lubrication for high speed rotary presses: lubricant spray rate optimization. *Drug Dev. Ind. Pharm.* 31 (10): 951–957.

11

SPRAY ATOMIZATION MODELING FOR TABLET FILM COATING PROCESSES

ALFRED BERCHIELLI

Drug Product Design, Pfizer Inc., Groton, CT, USA

PANKAJ DOSHI

Process Modeling & Engineering Technology Group, Pfizer, Inc., Groton, CT, USA

ALBERTO ALISEDA

Department of Mechanical Engineering, University of Washington, Seattle, WA, USA

JUAN C. LASHERAS

Applied Sciences and Mechanical & Aerospace Engineering and Bioengineering, Jacobs School of Engineering, University of California San Diego, La Jolla, CA, USA

11.1 INTRODUCTION

Film coatings are often used to enhance pharmaceutical tablet products. Tablet film coatings can provide many benefits including improved appearance, added functionality (e.g. sustained release, delayed release, or coatings with active ingredients), brand identity, dose strength identification, ease of swallowing, masking of bad taste, improved mechanical strength for improved handling (e.g. during production packaging), and reduced worker exposure (e.g. when dispensing in a pharmacy or hospital setting). There is also potential to improve chemical stability by separation of incompatible ingredients or through reduced oxygen or vapor transmission. Improvements in stability can be significant for maintaining potency, reducing or masking potential color changes in the tablet core, or reducing the development of odors due to chemical reactions. Coating ingredients may also cause instability, so selection of coating components and the coating process conditions are critical for successful product development.

The application of coating material to the tablets is carried out in a complex process that includes four key steps (spray atomization, droplet transport, droplet impact/spreading/drying, and tablet mixing). A schematic of the four elementary processes is shown in Figure 11.1. A photograph of the exterior and

close-up view of the corresponding interior of an actual chamber where tablet coating is being performed can be seen in Figure 11.2. First, the coating formulation is atomized into small droplets. This is done by using two-fluid coaxial atomizers in which the liquid formulation is injected through the inner nozzle at low speeds and a process gas, typically air, is injected through the outer nozzle at very high speeds. The atomizing gas exerts shear and pressure forces on the liquid jet and breaks it into droplets. This type of atomizer allows independent control of the liquid mass flow rate and droplet size. This is because, for all practical liquid flow rates of interest, the liquid speed at the nozzle is low and the breakup process that determines the droplet size is a function of the velocity differential between the two fluids, which can be modified by changing the gas speed (i.e. flow rate). The description and modeling of the physics behind this complex process is the focus of this chapter.

The second step in the process is the transport of the liquid droplets onto the tablets' surface. The droplets move under the velocity transferred to them by the atomizing gas and are dried by a secondary flow of hot dry gas that is introduced with negligible momentum into the process chamber, so as not to disturb the droplet motion. This secondary gas flow plays the role of providing the energy source needed to balance the latent heat of evaporation of the solvent used in the

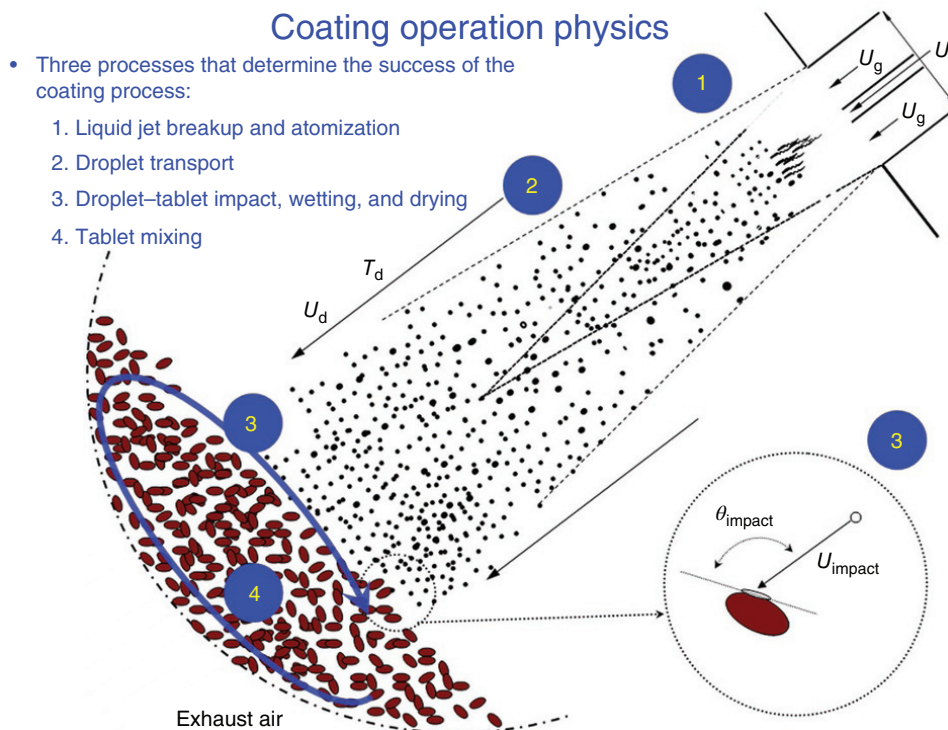


FIGURE 11.1 Schematic of the four elementary processes in the tablet coating operation.



FIGURE 11.2 Left – external view of Vector corp. laboratory development coating system (LDCS 20/30). Right – internal close-up view of cascading tablets during a development coating run. Note: In the image above on the right, a clear coating is added to white tablets. Numbers have been added in order to track individual tablets during the development coating run. The arrow indicates the direction of tablet movement.

coating formulation while keeping a sensitive balance between too little drying (droplets hitting the tablets with too much solvent) and too much drying (droplets drying out entirely during their transport). The third step is the impact and spreading of the droplets onto the tablets. The droplets spray hits the tablets at or near the surface of the bed. By rotating through the bed in a random manner,

the individual tablets form a continuous film coating by integrating the effects of many tiny droplets that hit it at different times during the process. This last aspect forms the fourth elementary process that determines the success of the operation. For the droplet spreading on the individual tablets to lead to a uniform film coating, the tablets need to be adequately mixed in their recirculating porous bed.

While there are extensive studies in the pharmaceutical literature on the granular flow mechanics that dominate mixing in the tablet bed, evaluation of coating quality, and properties, as well as on the thermodynamics that set droplet and tablet drying [1–8], much less attention has been dedicated to the atomization process [9–12]. Mathematical modeling of all four aspects of tablet film coating is important and has far-reaching impact on the economics of pharmaceutical commercialization, potentially reducing both cost and time to market. Generally, this is the last unit process where a batch may be rejected if process conditions are poorly selected or controlled, resulting in an unacceptable defect rate. Tablets with surface defects may be removed from the batch after a complete visual quality inspection; however this added step is time consuming and should be avoided.

Tablet coating has evolved throughout history from sugar-based coatings, to solvent-based polymeric film coatings, to today's aqueous-based polymeric film coatings. Equipment preferences have also changed from non-perforated pans to partially perforated and fully perforated pans. Coating application has also drastically improved from ladling of sugar coatings to films applied with modern pneumatic or coaxial air blast spray guns. This chapter describes advances in mathematical modeling of the atomization process that enable droplet size predictions, based on models of the physical processes that control liquid jet breakup in two-fluid coaxial atomizers, for pharmaceutically relevant aqueous and organic polymeric film coatings.

In the process control and experimentation that is involved in the design and scale-up of coating operations, atomization is one of the most difficult elements to understand and optimize. Direct measurements of droplet size are typically not available during coating operations and scale-up studies. Additionally, droplet size effects are often masked during development trials by the large variability in the results associated with the wide ranges of liquid and gas flow rates and velocities that need to be investigated. Predicting the size of the coating formulation droplets resulting from the atomization process is essential to the success of the coating process, as small droplets can dry and lead to reduced coating efficiency, or they may fail to coalesce on the tablet surface, not producing a smooth continuous film. Droplets that are too large, on the other hand, can have reduced solvent evaporation and lead to over-wetting. This causes coating defects (e.g. sticking marks where tablets contact each other and stick together or rough coatings due to erosion of the tablet core components). Similarities between the effects of droplets below or above the valid size range make the diagnostic of these defects difficult.

As we will show in the following sections, droplet size is a function of fluid properties (e.g. surface tension, viscosity), atomizer geometry (e.g. liquid and gas nozzle diameters), and process conditions (e.g. liquid and gas flow rates). The selection of a coating formulation for a specific pharmaceutical tablet is influenced by many factors that may be

unrelated to the physical processes involved. Once approved by regulatory agencies, however, changes to the formulation are not made unless necessary and often require justification and supporting regulatory data. The scale-up process, or changes in the industrial tablet manufacturing process that may occur during the lifetime of the product, may require modifications in atomizer geometry or process conditions. Under these circumstances, the atomization of the coating formulation will change. Unless a physics-based model is available to predict the direction and magnitude of those changes, a costly and time-consuming process is required to adapt the coating conditions to the new setup. These full-scale trials represent a major disruption to the manufacturing and commercialization, particularly if active ingredient costs are high or the delays lead to reductions in product supply.

The atomization process is a key part of the tablet coating process used in the pharmaceutical industry. It is therefore of great importance to develop physics-based atomization models that can predict droplet size and volume density as a function of fluid properties, atomizer configuration, and process conditions. These models can play an important role in the design of the process, including selection of equipment and process parameters, and in the optimization at all different levels: lab, pilot, and full scale. In the following sections of this chapter, we will discuss the rheology of coating formulations, characterization of coating suspensions and solutions, the detailed physics of atomization and the associated equipment, droplet size and velocity measurement, a mathematical model for mean droplet size, insight into the scale-up of the coating operation, and some conclusions on the current state and future directions of this technology.

11.2 COATING FORMULATIONS, PHYSICAL PROPERTIES, AND RHEOLOGY CHARACTERIZATION

Coating formulations typically contain significant amounts of polymers, plasticizers, surfactants, pigments/colorants, anti-foaming, anti-tack ingredients, film modifiers (i.e. sugars), and opacifiers. Because of this high content of large molecules in solution or of solids in colloidal dispersion (e.g. titanium dioxide, talc), coating fluids have complex rheology, exhibiting non-Newtonian behavior during the atomization process.

Tablet film coatings can be soluble (e.g. color coatings for immediate release formulations), insoluble (e.g. enteric coatings that are insoluble at gastric pH), or partially insoluble (e.g. porous or semipermeable membranes that may contain insoluble and soluble ingredients). Immediate release coating formulations are commonly based on polymers such as hypromellose (also known as hydroxypropyl methylcellulose or HPMC) and polyvinyl alcohol (PVA), but other polymers or natural products such as shellac have also been used.

Functional coatings for controlled release are often based on polymers such as cellulose acetate, ethyl cellulose, or methacrylates. In this chapter, we will present experimental results obtained from two characteristic coating formulations for immediate release tablets: HPMC (Colorcon Codes: Y-30-18037, OY-LS-28914) and PVA (Colorcon Code: 85F18422), and a characteristic coating for controlled release that is composed of a mixture of cellulose acetate/polyethylene glycol and forms a semipermeable membrane allowing for osmotic drug delivery [13, 14]. Table 11.1 lists these examples of coating formulations discussed in this chapter.

As we have pointed out in the previous sections and will show rigorously in the formulation of a mathematical model for the breakup of the liquid into individual droplets, the atomization process that constitutes the first stage of the pharmaceutical coating operation is strongly influenced by the physical properties of coating solutions. The two key nondimensional parameters that control this process are the Weber number ($W_e = \rho_g U_g^2 d / \sigma$) and the Ohnesorge number ($O_h = \mu_l / \sqrt{\rho_l \sigma d}$),

where

ρ_g is the gas density.

U_g is the gas velocity at the nozzle.

d is the liquid jet diameter.

μ_l is the liquid viscosity.

ρ_l is the liquid density.

σ is the value of the surface tension between the liquid and the gas.

The Weber number represents the relative importance of the inertia of the high-speed gas stream that disturbs the liquid jet and the surface tension stresses that minimize surface energy, bringing cohesion to the liquid. The Ohnesorge number characterizes the relative importance of viscous stresses and surface tension stresses. Clearly then, the characterization of the physical properties of the coating fluids is a key preliminary step in the understanding and modeling of the liquid atomization in the tablet coating processes.

11.2.1 Coating Solution Viscosity Measurement

A typical film coating formulation usually contains a mixture of solids comprising polymer, plasticizer, pigment/opacifier, and other film modifiers dissolved in water or a nonaqueous organic solvent. The coating formulations could either be a solution if solids are soluble in the solvent or else they could form a colloidal suspension. Figure 11.3 shows the viscosity for previously mentioned four coating fluids at 25 °C for different shear rates. These fluids show slight shear thinning behavior (i.e. viscosity decreases with increasing shear rate), up until shear rate reaches 200 s⁻¹. However, beyond this value of shear rate, viscosities are almost constant. In a typical atomization process the shear rate is well above 200 s⁻¹ [9], so coating solution viscosity can be assumed to be constant for a given coating formulation.

11.2.2 Effect of Solids Content on Viscosity

Figure 11.4 shows the variation of viscosity for coating 3 with increasing solids content. For this formulation the viscosity at low solids content, i.e. up to 5%, is quite close to that of the solvent. However, with further increase in solids content, the viscosity increases in an exponential fashion, typical of this type of colloidal dispersion.

11.2.3 Effect of Temperature on Coating Solution Viscosity

Figure 11.5 shows the variation of viscosity of three different coating fluids with temperature. The viscosity of all three fluids decreases linearly with increasing temperature. The viscosity reduces to almost half their values when the fluids are heated from 20 to 35 °C. This behavior can be taken advantage of, by heating very viscous coating formulation prior to injection in the atomizer, to facilitate tablet coatings with very high solids content, where the high viscosity makes atomization difficult. At the same time, this points out to the need to monitor thermal gelation point of coating formulations, since operating conditions that include temperatures above that threshold will lead to semisolid and almost unsprayable coating fluids [15].

TABLE 11.1 Example Coatings Discussed in This Chapter

Coating Number	Abbreviation	Vendor Code or Components	Composition
1	Opadry II White-HPMC	Colorcon code: Y-30-18 037	15% solids, 85% water
2	Opadry II White-PVA	Colorcon code: 85F18422	20% solids, 80% water
3	Opadry II White-HPMC	Colorcon code: OY-LS-28914	10, 12, 15% solids, 90, 88, 85% water
4	CA-PEG	Cellulose acetate 398–10	8%
		Polyethylene glycol 3350	2%
		Acetone	87%
		Water	3%

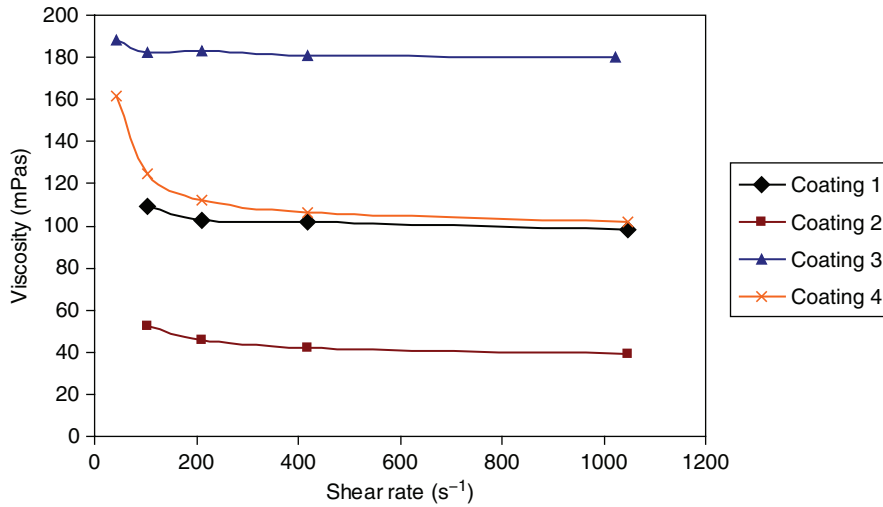


FIGURE 11.3 Viscosity versus shear rate for four coating formulations at 25 °C.

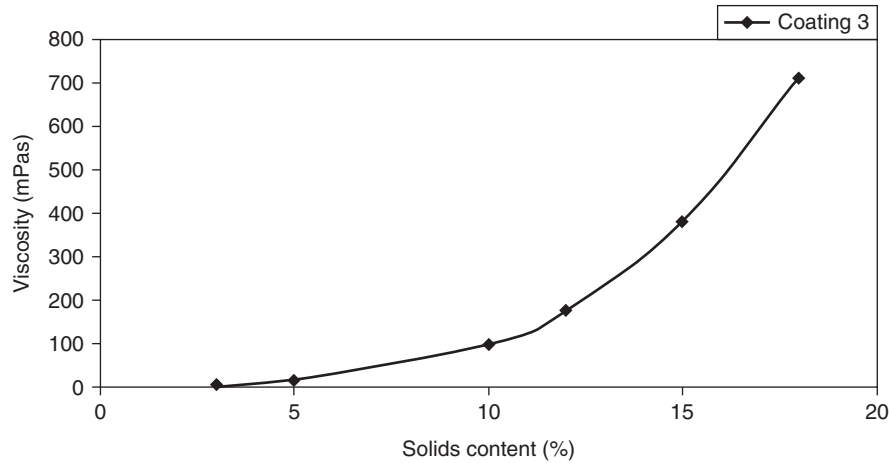


FIGURE 11.4 Solution viscosity versus solids content for coating 3 at 25 °C and a shear rate of 1000 1/s.

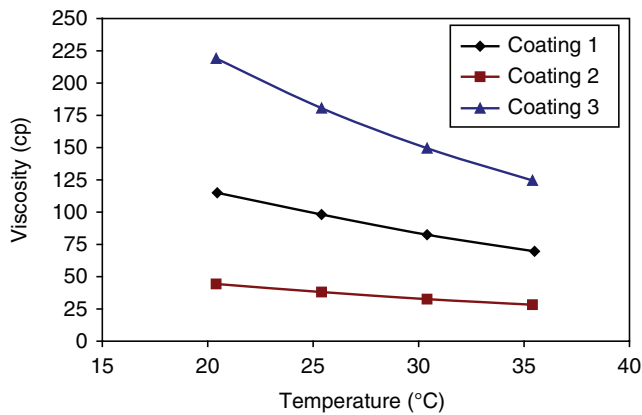


FIGURE 11.5 Effect of temperature on viscosity of coating solutions.

11.2.4 Coating Solution Surface Tension Measurement

Table 11.2 shows the surface tension data for different coating fluids and their respective solvents measured at approximately 23 °C. The values of surface tension for all the aqueous coating fluids (1–3) are quite close to each other, irrespective of the solids concentration. However, they are quite low compared to the surface tension of the solvent (water). This observation is not true, however, for the coating formulation with organic solvent (acetone). This data accords with the finding in the literature that HPMC acts as a surfactant and reduces the surface tension of water for very low concentrations ($\sim 2 \times 10^{-5}\%$ (w/w) [15]). After this point further increase in HPMC concentration has little effect on reducing the surface tension. This behavior is similar to the

TABLE 11.2 Effect of Solids Concentration on the Viscosity, Density, and Surface Tension of Coating Fluids Measured

Coating No.	Coating Type	% Solids	Viscosity of Coating Fluid at 25 °C, and 1048 (s ⁻¹) (kg/ms)	Density of Coating Fluid (g/cc)	Density of Solvent (g/cc)	Surface Tension of Coating Fluid (dynes/cm)	Surface Tension of Solvent (dynes/cm)
1	Opadry II White(Y-30-18 037) HPMC/lactose/TiO ₂ /triacetin/ water	15	0.098	1.05	1.00	46.98	72.8
2	Opadry II White (85F18422) PVA/PEG/water	20	0.039	1.07	1.00	43.93	72.8
3	Opadry II White (OY-LS-28914) HPMC/lactose/TiO ₂ /triacetin	10	0.095	1.02	1.00	48.22	72.8
3	Opadry II White(OY-LS-28914) HPMC/lactose/TiO ₂ /triacetin	12	0.181	1.03	1.00	47.66	72.8
3	Opadry II White (OY-LS-28914) HPMC/lactose/TiO ₂ /triacetin	15	0.419	1.04	1.00	46.67	72.8
4	CA/PEG/acetone/water	10	0.102	0.82	0.79/1.00	29.71	29.7

critical micelle concentration shown by surface-active materials. The typical composition of coating formulations lead to surface concentrations that are always above this threshold, so the saturation effect in the values of surface tension found in our experiments are characteristic of these types of coatings.

11.3 SPRAYING PROCESS AND EQUIPMENT

As highlighted in previous sections, the atomization of coating liquid formulations into small droplets is a key step in the process of pharmaceutical tablet coating, and the focus of this chapter. To contribute to success in the coating operation, atomization of the coating formulation must provide controllable droplet size distributions for a wide range of liquid flow rates. Additionally, the concentration of droplets across the spray should be spatially uniform. These two requirements make two-fluid coaxial atomizers the most common in this setting. Unlike pressure atomizers, commonly used in diesel fuel injection, where the size of the liquid droplets is dependent on liquid flow rate, two-fluid atomizers allow for control of the droplets size independent of liquid flow rates. In its simple configuration, they produce a very predictable spatial distribution of droplets within the spray, with a significant radial gradient of concentration, and high uniformity along concentric rings. To improve the spatial uniformity of the droplet flux, two auxiliary gas jets are used. In the atomizer design, these side jets are located a small distance from the nozzle, diametrically opposed to each other, breaking the cylindrical symmetry of the coaxial atomizer and creating two symmetry planes. Because they help shape the flow of gas and liquid droplets from the atomizer, these side jets



FIGURE 11.6 Front view of a laboratory-scale spray gun for tablet coating (Schlick fluid nozzle and air cap with spray gun body by Vector Inc.).

are commonly referred to as pattern air. The spray takes on an elliptical cross section, consistent with the two symmetry planes, and the droplet spatial distribution becomes very homogeneous.

Spray guns typically used in tablet spray coating include Schlick, Spraying Systems, Freund, Binks, and Walther Pilot. Typical spray gun designs can be seen in Figure 11.6 (front view of a Schlick fluid nozzle and air cap setup for creating an elliptical spray pattern), Figure 11.7 (side view of a Schlick fluid nozzle and air cap setup), Figure 11.8 (front view of a Spraying Systems fluid nozzle and air cap setup for creating an elliptical spray pattern), Figure 11.9 (side view of a Spraying Systems fluid nozzle and air cap setup), Figure 11.10 (front view of a Spraying Systems fluid nozzle

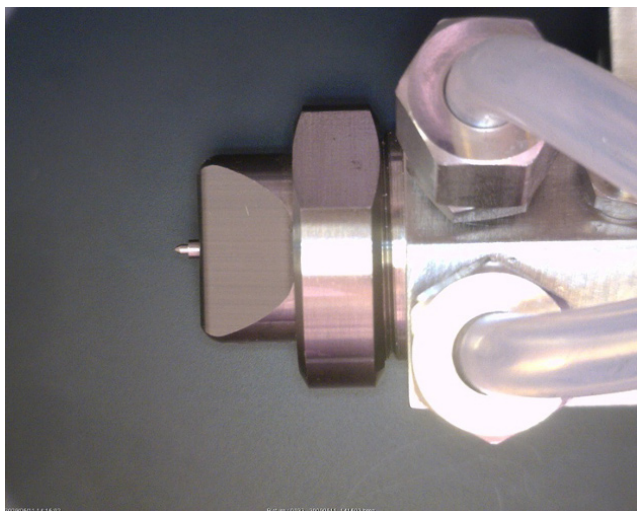


FIGURE 11.7 Side view of a laboratory-scale spray gun for tablet coating (Schlick fluid nozzle and air cap with spray gun body by Vector Inc.).

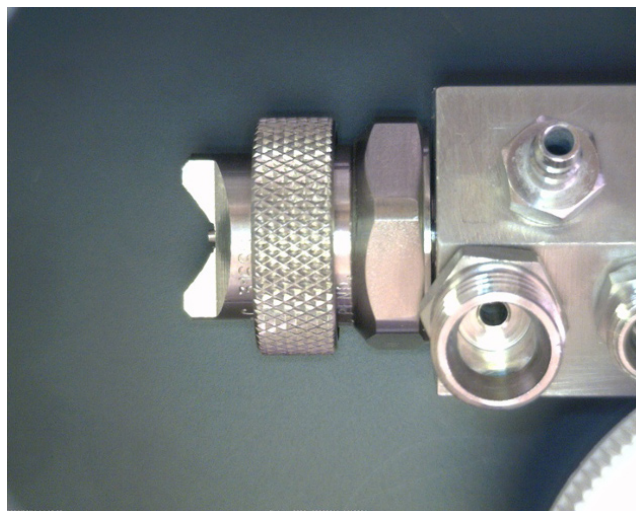


FIGURE 11.9 Side view of a laboratory-scale spray gun for tablet coating (Spraying Systems fluid nozzle and air cap with spray gun body by Vector Inc.).



FIGURE 11.8 Front view of a laboratory-scale spray gun for tablet coating (Spraying Systems fluid nozzle and air cap with spray gun body by Vector Inc.).



FIGURE 11.10 Side view of a laboratory-scale spray gun for tablet coating (Spraying Systems fluid nozzle and air cap with spray gun body by Vector Inc.).

and air cap setup for creating a round spray pattern), and Figure 11.11 (side view of a Spraying Systems fluid nozzle and air cap setup). Figure 11.12 provides a sketch of the spray nozzle showing key geometric parameters that must be provided to successfully model the atomization process.

The use of side gas jets in pharmaceutical coating atomizers also improves spray coverage. This is defined as the cross-sectional area of the spray where it intersects with the tablet bed. This area, which is obviously determined by the distance from the spray gun to the tablet surface and the spray shape, is important because it determines the probability of a tablet moving along the surface of the tablet bed to

collide with one or more coating formulation droplets. By stretching the spray in the direction perpendicular to the tablet motion (major axis of the spray elliptical cross section) and reducing the coverage in the perpendicular direction (minor axis of the ellipse), pattern air increases the number of tablets that will receive the impact of a coating droplets as they are distributed on the tablet bed's surface, since it increases the width covered by the spray. At the same time, the shorter coverage along the trajectory of the tablets reduces the probability of a tablet being hit by too many droplets over a short time, while it traverses the spray in its motion, thereby reducing the likelihood of over-wetting or tablets sticking to each other.

Coating operations requires atomization to work efficiently over a wide range of operating conditions. Poor selection or poor control of process conditions can lead to nonoptimal results (i.e. poor coating appearance, tablet core erosion, tablet breakage, potency loss, or dosage form functionality changes) or potentially batch failure. Successful tablet coating requires selecting the appropriate tablet formulation components, tablet properties, coating formulation, coating properties, and coating process conditions. The liquid flow rate (i.e. spray rate) needs to be determined in

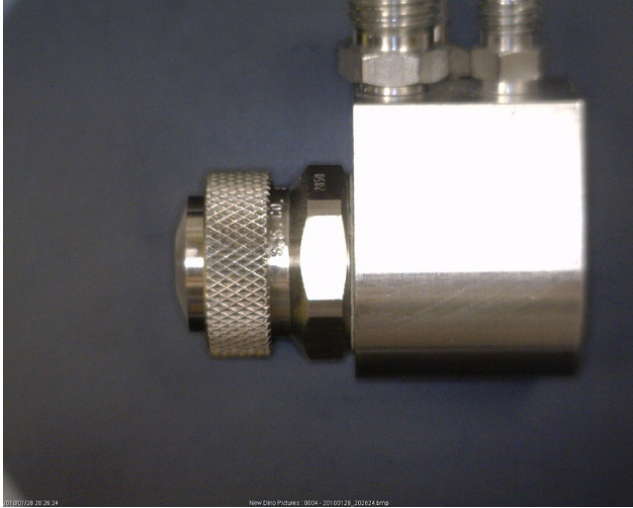


FIGURE 11.11 Side view of a laboratory-scale spray gun for tablet coating (Spraying Systems fluid nozzle and air cap with spray gun body by Vector Inc.).

combination with the appropriate thermodynamic [2] drying conditions (air flow, temperature, and humidity) so that evaporation of the solvent occurs at an adequate rate (too much evaporation leads to dry droplets not spreading over tablets, while too little leads to over-wetting). Spray rate and coating formulation solid content determine the total flux of coating material onto the tablets, and, therefore, these two parameters need to be selected in conjunction to control the overall coating time. There is a minimum number of passes through the spray that a tablet needs to go through in order to develop a uniform coating. Longer spray times allow for increased passes of tablets under the spray zone, and when combined with good mixing and appropriate tablet shape selection results in improved intra-tablet and inter-tablet coating uniformity; however an efficient process should not be longer than what is required to ensure product specifications are met. Changes in spray rate or coating fluid solids content will affect the droplet size of the atomized coating. Other atomization parameters need to be adjusted to maintain optimal droplet size under these changing conditions. Predictive models and accurate measurements of process conditions are essential to minimize the challenges presented by the wide range of needs within the coating operation. Manufacturers of atomization equipment typically provide values of the atomization and pattern air pressures as the key process settings, because they are easy to measure in the industrial environment. These are, however, surrogates for the gas flow rates, which play a key role in the atomization process and droplet redistribution within the spray. Different atomizer setups inside the coating equipment may translate to different pressure losses from the

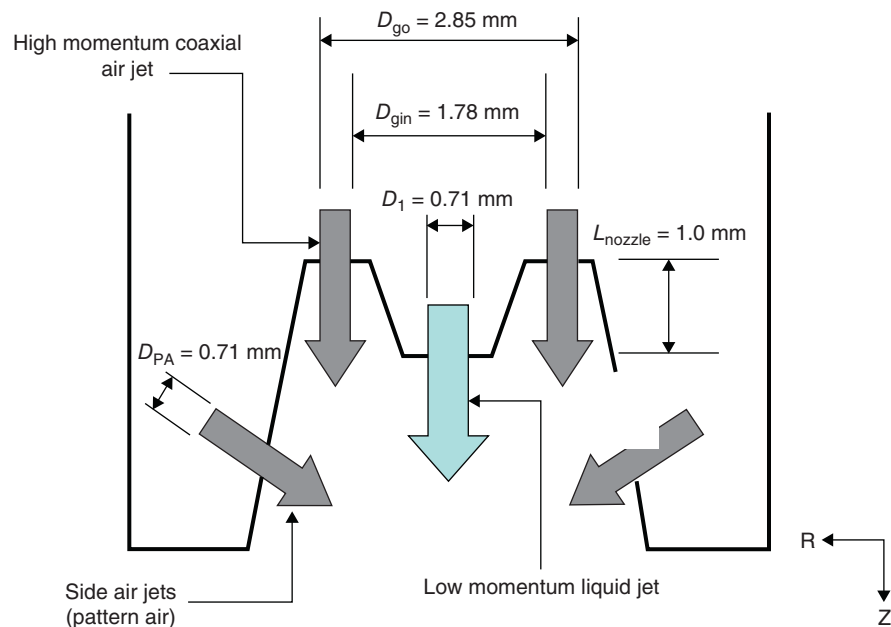


FIGURE 11.12 Side view – sketch of a laboratory-scale spray gun for tablet coating (Spraying Systems fluid nozzle and air cap with corresponding measurements used in modeling).

measurement to the atomizer head, introducing uncertainty about the actual flow rates in the process. It is, therefore, preferable to measure volumetric flow rates for both atomizing and pattern air. Air flow rates can economically be measured with rotameters (i.e. a graduated meter consisting of a tapered tube containing a free float for air flow measurement). The Key Instruments Flo-Rite Series is an example of a commercially available rotameter. Volumetric flow rates of air can also be measured with mass flowmeters, providing a digital output that is easy for the operator to read and allows coaters with data collection capabilities to store atomization and pattern air flow values electronically. The Sierra Top-Trak 824 Series is an example of a mass flowmeter. A combination of the appropriate monitoring equipment and accurate physics-based models greatly improves the effectiveness and efficiency of coating operations.

11.4 DROPLET SIZE AND VELOCITY MEASUREMENTS

The result of the atomization process described in previous sections, and modeled mathematically in the next section, is a distribution of droplets with a wide range of diameters and velocities. The combination of these two properties determines the droplet inertia and therefore how the droplets interact with the underlying high-speed turbulent air jet in which they are immersed. These interactions, together with the initial conditions for the motion of the droplets, will ultimately shape the spray coating process, setting the values of

the droplet number density, droplet velocity at the impact with the tablet bed, and mass flux per unit area. These fluid variables determine the fate of individual droplets impacting on the tablets and, when ensembled over the whole droplet distribution, the success of the coating operation.

The droplet size distribution is given by the balance between cohesive forces, such as surface tension and viscosity, and disruptive forces, namely, pressure and shear forces exerted on the liquid surface by the atomizing gas stream. Because of the turbulent character of the high-speed gas stream and the nature of the instability with a wide range of unstable wavelengths with similar growth rates, the resulting droplets have a distribution of sizes, not a unique diameter as expected in other types of liquid breakup such as Rayleigh–Plateau instability [16]. Because of this, measurements and modeling of the droplet size are focused on characterizing the full distribution of diameters. One way to do this is to describe the different statistical moments of the distribution of diameters. Table 11.3 gives the mathematical definition and physical interpretation of the most used moments of the size distribution.

Droplet size distributions can be measured by a number of instruments that rely on light scattering to determine the droplet size unintrusively. The most widely used measurement technique is based on light interferometry to evaluate the phase shift introduced by a scatterer spherical droplet on the fringe pattern created by the superposition of two laser beams with slightly different frequencies. This phase shift is proportional to the scatterer’s diameter. A much more detailed description of the measurement principle was

TABLE 11.3 Mathematical Definition and Physical Interpretation of the Most Used Moments of the Size Distribution

Statistical Moment	Mathematical Definition	Physical Meaning
d_{10}	$d_{10} = \frac{\sum_{i=1}^N n_i \cdot d_i}{\sum_{i=1}^N n_i}$	Diameter of the droplet that represents the arithmetic average of all droplet diameters. d_i is the diameter of droplets in size bin i , and n_i is the number of droplets in that bin of the distribution
d_{20}	$d_{20} = \sqrt{\frac{\sum_{i=1}^N n_i \cdot d_i^2}{\sum_{i=1}^N n_i}}$	Characteristic diameter that, if all droplets were this same size (monodisperse), they would contain the same surface area as the whole distribution under study
d_{30}	$d_{30} = \sqrt[3]{\frac{\sum_{i=1}^N n_i \cdot d_i^3}{\sum_{i=1}^N n_i}}$	Characteristic diameter that, if all droplets were this same size (monodisperse), they would contain the same volume (or mass) as the whole distribution under study
d_{32}	$d_{32} = \frac{\sum_{i=1}^N n_i \cdot d_i^3}{\sum_{i=1}^N n_i \cdot d_i^2}$	Characteristic diameter that, if all droplets were this same size (monodisperse), they would have the same volume to surface ratio as the whole distribution under study
D^{50}	$\frac{\sum_{i=1}^{M(D^{50})} n_i \cdot d_i^3}{\sum_{i=1}^N n_i \cdot d_i^3} = 0.5$	Characteristic diameter for which all droplets smaller than this contain 50% of the volume (or mass) of the whole distribution under study. M is the bin for which the cumulative volume summation reaches 50% of the total. The diameter corresponding to that bin is D^{50}
D^{90}	$\frac{\sum_{i=1}^{L(D^{90})} n_i \cdot d_i^3}{\sum_{i=1}^N n_i \cdot d_i^3} = 0.9$	Characteristic diameter for which all droplets smaller than this contain 90% of the volume (or mass) of the whole distribution under study. L is the bin for which the cumulative volume summation reaches 90% of the total. The diameter corresponding to that bin is D^{90}

published by Bachalo [17]. This measurement technique has the added advantage that, with some further processing of the information in the collected light scattered by the droplets, it also provides the droplet velocity distribution from the Doppler frequency shift. Although the physical principles behind these measurements are well established, the technological details involved in getting robust, accurate measurements in a wide variety of operating conditions are extremely complicated. There are currently only two instrument manufacturers that provide commercial solutions based on this technique: TSI Inc., Shoreview, MN (PDPA, phase Doppler particle analysis) and Dantec Dynamics A/S, Skovlunde, Denmark (PDA, phase Doppler anemometry).

The characteristics of the droplets in the spray vary spatially both radially and axially. The latter dependency is easy to describe and is universal for different spray types, with different phenomena having different relative weights in the resulting droplet distribution. An example of this axial evolution is shown in Figure 11.13. In the close proximity of the nozzle, droplets are very large or are growing to a large diameter. In this region, the liquid jet has been shattered by the atomizing air, and large liquid masses, not necessarily spherical in shape, are being stripped from the liquid core. As these liquid masses get deformed by the high-speed gas stream, they take on shapes with low curvature that corresponds to the particle sizing measurements as increasingly large droplets. It is not until the deformation stage ends and the breakup into smaller droplets begin that the shape

become truly spherical and accurate measurements are possible. This happens in the region between 5 and 10 diameters downstream. After that region, the breakup proceeds with the droplet diameter decreasing sharply before it plateaus in the region between 20 and 30 diameters. At this point the breakup process has finished. The breakup of the liquid into smaller droplets makes surface tension a stronger cohesive force. Together with the dilution of the high momentum jet by expansion of its cross-sectional area due to entrainment, the balance between disruptive and cohesive forces tilts very soon toward the cohesive side, thereby stopping the breakup. From a practical point of view, this process usually takes place within the first two inches of the jet, giving rise to a minimum standoff distance between the coating gun and the tablet bed, since the coating is designed to impact spherical droplets that are small and do not extend beyond the boundaries of a single tablet. In the region beyond where breakup takes place, the droplet size is dependent on secondary processes that determine the dynamics of small spherical droplets immersed in a turbulent gas jet. At this stage the evolution of the droplet size distribution can be evaluated with a population dynamics equation [18] where the sources on the right-hand side are turbulent transport, coalescence, and evaporation. Typically these phenomena alter the droplet diameter very slowly in the range of interest for the coating operation, as seen by the constant mean diameter measured in the second half of the plot in Figure 11.13 (between 25 and 50 diameters downstream).

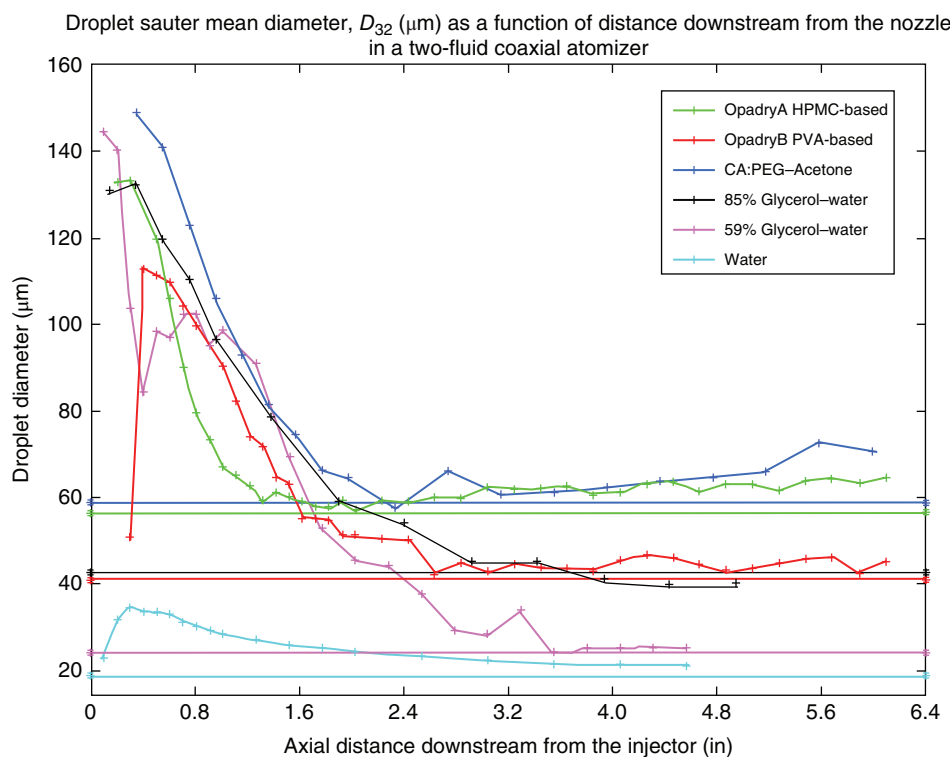


FIGURE 11.13 Droplet diameter versus axial distance downstream from the spray gun.

The evolution of the radial distribution of droplet sizes is determined by the different response of droplets to the high-speed gas flow due to their different inertia. The larger droplets are found along the centerline, where they maintain their strong axial momentum communicated by the high-speed jet near the nozzle. Smaller droplets have lower inertia and are therefore more prone to radial excursions induced by small velocity fluctuations in that direction. Finally, there is typically an increase in the droplet size near the outer boundary of the jet. In swirled sprays this is very acute and is caused by the centrifugal migration of large inertia droplets that sustain the initial azimuthal momentum longer and therefore travel farther distance radially. But even in non-swirled flows, the instability that is responsible for liquid breakup requires a radial velocity on the liquid that induces acceleration when it is exposed to the high-speed gas stream. Due to the natural stochastic distribution of this radial velocity, some liquid masses have high initial radial velocity and cross the gas jet quickly. Because of this, these droplets have reduced time of exposure to the high-speed gas stream, which decays exponentially in the radial direction, and have reduced breakup frequencies as a result. The tails of the radial velocity distribution are then responsible of those large droplets and their location in the outskirts of the spray.

The droplet sizes resulting from coaxial atomization of a low-speed liquid jet and a coaxial high-speed gas jet have been typically described in terms of log-normal or beta distributions. If a specific distribution is used to characterize the spray, two or sometimes only one moment can be used to quantify the droplet size over its entire range.

11.4.1 Velocity Measurements

The velocity of the two phases present in the spray can be measured by the same phase Doppler technique used to determine the droplet diameter distribution. The basis for velocity measurements is the same as Laser Doppler Velocimetry. Taking advantage of the size measurements, the velocity data can be segregated for tracer droplets that represent the velocity of the carrier gas and the velocity of the larger droplets that, because of their inertia, do not track the gas velocity.

For a typical axisymmetric coaxial injector, the gas velocity behaves as a classical round jet. After a short development length, during which the flattop profile injected at the nozzle evolves to adjust to the boundary conditions, the flow dynamics become self-similar and the velocity profile Gaussian. In a coaxial atomizer the gas jet flows through a ring-shaped nozzle that leaves inside space for the liquid nozzle. Given the small cross section necessary for the liquid injection, the gas coming out of the nozzle expands into the centerline so that, at the end of the development length, the gas in the spray behaves exactly as a round jet. Figure 11.14 shows velocity measurements in an axisymmetric coaxial injector. The gas velocity can be fitted very accurately by a Gaussian profile.

The velocity of the droplets tend to follow a very similar profile, but the magnitude of the velocity is slightly higher. Considering the inertia of the droplets and that the gas velocity is decreasing as it flows downstream from injection due to entrainment of ambient air, this is to be expected. Of particular interest is the strong deviation associated with the droplets near the edge of the jet. Those droplets, typically larger than the average across all radial positions, traverse the jet from the centerline to the edge with a significant radial velocity and therefore reach the region of very low gas velocity before they lose their forward momentum. This produces the largest velocity difference between the two phases.

11.4.2 Effect of the Pattern Air on Turbulent Droplet Transport within the Spray

The model for the breakup of the liquid jet into droplets does not take into account the existence of the two side jets that inject the pattern air into the spray.

The geometry of most pharmaceutical coating atomizers includes two auxiliary air jets that inject air from the edges of the liquid/air jet with a significant inward radial velocity component. Typical angles are between 30 and 45° with respect to the jet axial direction. Depending on the location and angle of the pattern air nozzles, these jets impinge on the droplet-laden jet after the liquid breakup process has finished or on the atomizing air as it is distorting the liquid jet. Pattern air plays an important role in the transport of the droplets of different sizes. Its effect on the liquid breakup is described in the next section. The goal of pattern air is to flatten the spray so that the spatial coverage of the atomizer in the coating pan is increased. To do this the pattern air induces an asymmetry in the velocity field of the spray that makes its cross section become elliptical rather than circular. The spray is narrower along the axis of the side jets, where the pattern air momentum forces the spray inward (minor axis of the ellipse), and broader in the perpendicular axis (major axis of the ellipse). This is shown in Figure 11.15, where the size distribution is plotted in the radial direction along the minor and major axes, at two distances downstream of the injector. The effect of the pattern air on the droplets size distribution is rather complicated. First, the size distribution is concave along the major axis, and larger droplets are found as we move away from the axis due to the larger droplets conserving their radial momentum longer than the smaller ones. The contrary is true along the minor axis, where the effect of the side jets is more noticeable for the larger droplets that are subject to higher aerodynamic focusing toward the spray axis due to their larger inertia and lower diffusivity. The smallest droplets quickly adjust to the local gas velocity, minimizing any local perturbation inflicted by the pattern air. This explains the convex shape of the diameter distribution found along the cross section minor axis. Furthermore, some very large droplets acquire a large radial velocity

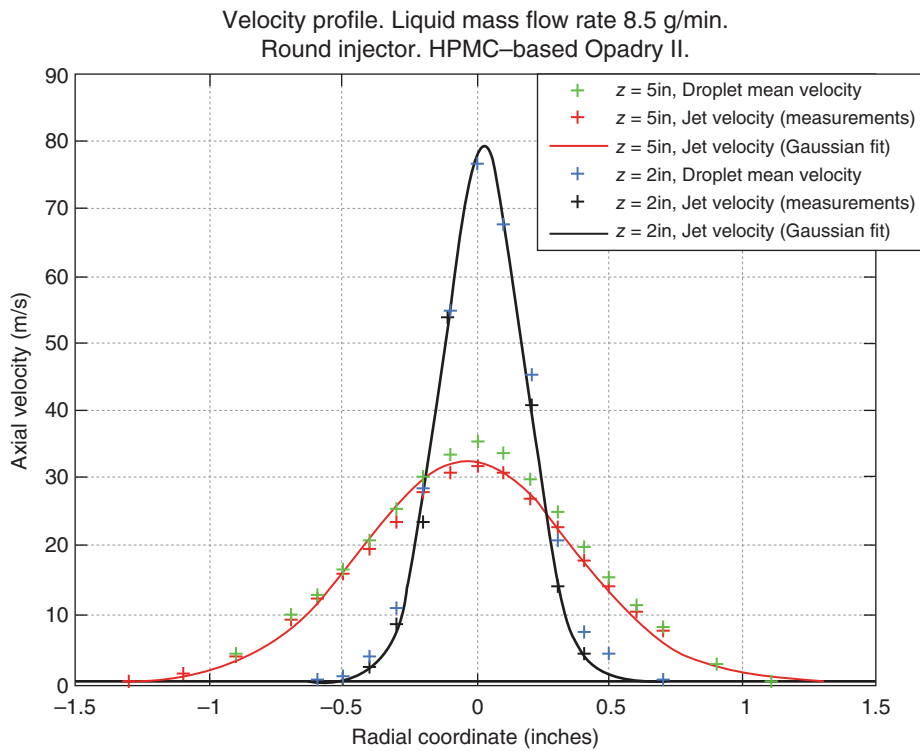


FIGURE 11.14 Droplet and gas jet velocity profile.

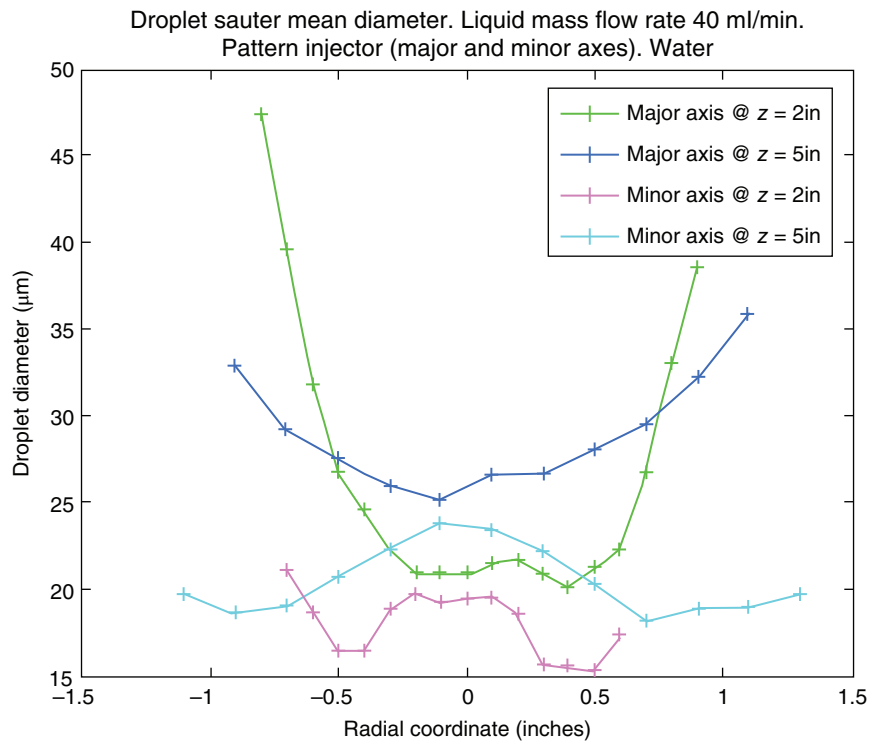


FIGURE 11.15 Droplet diameter versus radial distance from the center of the spray gun.

component as they interact with the side jets and, because of their large inertia, do not lose that momentum as they cross the sprays axis. These droplets are then found in relative high proportion in the outskirts of the jet, where the droplet number density is low and a small number of very large droplets have a significant impact on the average diameter of the distribution. This explains the change from convex to concave in the diameter radial distribution. Note that the inflection point associated with this change in character moves outward from the spray axis with distance downstream, as one would expect those large droplets to keep moving in a radial direction, occupying the outer region of the spray. This inhomogeneity in the diameter distribution along the radial coordinate in the spray is a typical effect of pattern air that needs to be monitored to avoid large droplets to accumulate at the edge of the spray.

11.5 MATHEMATICAL MODELING OF THE LIQUID BREAKUP: AVERAGE DROPLET SIZE

The atomization of a liquid jet by a co-flowing, high-speed gas stream occurs via a series of fluid instabilities that lead to the disruption of the liquid jet and results in the liquid mass being broken into small spherical inclusions transported within a gas stream. Although the nature of the instability depends on the specific conditions present in the jet

(Reynolds and Weber numbers, mass and dynamics pressure ratios between the air and liquid streams, etc.) and is the source of significant controversy, for the conditions typically present in pharmaceutical coating operations, it has been well documented [9, 19] that the breakup process follows a sequence of Kelvin–Helmholtz (K-H) instability disrupting the liquid jet, followed by a Rayleigh–Taylor (R-T) instability resulting in the breakup of the liquid cylinder into individual droplets. The K–H instability develops in the annular shear layer that exists between the low speed liquid injection and the high-speed coaxial gas jet. Once the liquid has been displaced from its axisymmetric position, it suffers a sudden acceleration as a result of the drag imposed by the gas flow and a R–T instability at the interface ensues. This process is shown in schematic form in Figure 11.16. The wavelength of the primary instability, λ_1 , depends on the gas boundary layer thickness, δ_g , at the gas nozzle, as described by [20], and can be computed by:

$$\lambda_1 \approx 2\delta_g \sqrt{\frac{\rho_l}{\rho_g}} \tag{11.1}$$

where

ρ_l and ρ_g are the liquid and gas densities, respectively.

Spray atomization nozzles are typically designed with a strong convergence upstream of the nozzle so that the flow acceleration will keep the boundary layer laminar, and

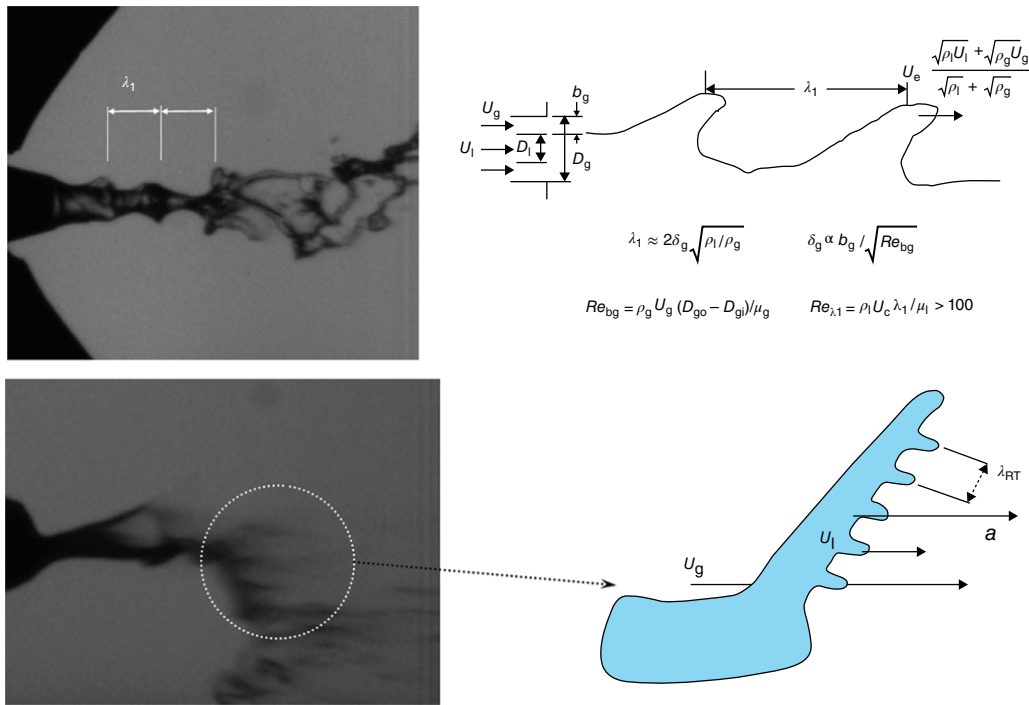


FIGURE 11.16 High-speed photograph (top left) and sketch (top right) of the primary Kelvin–Helmholtz instability. High-speed photograph (bottom left) and sketch (bottom right) of the secondary Rayleigh–Taylor (R-T) instability.

therefore as thin as possible. The boundary layer thickness can then be calculated as

$$\delta_g = \frac{C b_g}{\sqrt{\text{Re}_{bg}}} \quad (11.2)$$

where

$\text{Re}_{bg} \equiv U_{\text{Gas}} b_g / \nu_{\text{Gas}}$ and the coefficient of proportionality C depends on nozzle design.

Considering air as the working gas and typical flow rates used in coating operations (1–5 SCFM, 20–100 SLPM), the Reynolds number is of the order 10^4 . The disrupted liquid mass is accelerated by the surrounding gas and its resulting convective velocity is

$$U_c = \frac{\sqrt{\rho_1} U_{\text{Liquid}} + \sqrt{\rho_g} U_{\text{Gas}}}{\sqrt{\rho_1} + \sqrt{\rho_g}} \quad (11.3)$$

For the K-H instability to develop quickly, before the gas jet velocity decays and becomes close to the liquid injection velocity, the Reynolds number of the liquid shear layer must be large:

$$\text{Re}_{\lambda_1} = \frac{(U_c - U_{\text{Liquid}}) \lambda_1}{\nu_1} > 10. \quad (11.4)$$

This condition is necessary even though the instability is driven by the gas. In the case of coating solutions that may have non-Newtonian rheology, the applicable viscosity, ν_1 , is the effective shear viscosity at the injection shear rate.

The liquid excursion resulting from the K–H instability, of thickness b_1 , grows rapidly and is exposed to and accelerated by the high-speed gas stream. This liquid is subsequently subject to a R-T instability similar to water placed over oil. Because oil is lighter than water, the interface between the two liquids is accelerated and creates fingers by which oil flows up and water goes down. Joseph et al. [21] studied the growth of a R-T instability in an accelerated drop suddenly injected into a high-speed gas stream. The dispersion relation for the case $\rho_g \ll \rho_1$ is given by

$$-\left[1 + \frac{1}{n^2} \left(-ak + \frac{\sigma k^3}{\rho_1}\right)\right] + 4 \frac{k^2 \alpha_1}{n \rho_1} + 4 \frac{k^3}{n^2} \left(\frac{\alpha_1}{\rho_1}\right)^2 (q_1 - k) = 0 \quad (11.5)$$

where

k is the magnitude of the wave vector,

n the amplification rate,

a the acceleration of the liquid tongue,

σ the surface tension,

α_1 the effective shear viscosity of the liquid in $\tau_{ij} = 2\alpha_1 e_{ij}$,

where τ_{ij} and e_{ij} are, respectively, the stress and rate of strain tensors in the liquid, and q_1 is given by:

$$q_1 = \sqrt{k^2 + n\rho_1/\alpha_1} \quad (11.6)$$

For very low Ohnesorge numbers, when viscous effects are negligible, the wave number corresponding to maximum amplification is

$$k_\sigma = \sqrt{\frac{a\rho_1}{3\sigma}} \quad (11.7)$$

This is the dominant case in water atomization and in fuel injectors where linear chain hydrocarbons with very low viscosity values are used, but it is rarely the case in coating operations where the coating solutions are typically high solid content solutions or particle dispersions with complex rheology.

When viscous terms are important, as is the case for the tablet coating operations, α_1 is large and it can be assumed that $\frac{n\rho_1}{k^2\alpha_1} \ll 1$ such that $(q_1 - k) \approx \frac{n\rho_1}{2k\alpha_1}$ in Eq. (11.6). The simplified dispersion relation from Eq. (11.5) then reads

$$n = -\frac{k^2\alpha_1}{\rho_1} \pm \sqrt{\frac{k^4\alpha_1^2}{\rho_1^2} - \left(\frac{k^3\sigma}{\rho_1} - ka\right)} \quad (11.8)$$

Disturbances will grow when the second term in Eq. (11.8) is positive and larger than the first term. It is useful to rewrite Eq. (11.8) in the form

$$n = \frac{k^2\alpha_1}{\rho_1} \left[\left(1 + \frac{a\rho_1^2}{k^3\alpha_1^2} - \frac{\sigma\rho_1}{k\alpha_1^2}\right)^{1/2} - 1 \right] \quad (11.9)$$

From Eq. (11.9), the amplification rate is zero when $k = \sqrt{a\rho_1/\sigma}$, which is the capillary cutoff wave number, and when $k = 0$. The wave number of maximum amplification is given by the third-order equation:

$$4 \frac{\alpha_1^2}{\rho_1^2} k^3 - \frac{3\sigma}{\rho_1} k^2 + a = 0 \quad (11.10)$$

The exact solution of this equation is too complex to be of practical interest. However, for the high viscosity fluids typical of tablet coating, the Ohnesorge number (which controls the relative importance of liquid viscosity and surface tension, $\text{Oh} = \mu_1 / \sqrt{\rho_1 \sigma D_1}$) based on the wavelength is large and the second term in Eq. (11.10) is small compared with the first one so that the wave number of maximum amplification is

$$k_{\text{max}} \approx \sqrt[3]{\frac{a\rho_1^2}{\alpha_1^2}} \quad (11.11)$$

The R-T wavelength is $\lambda_{\text{RT}} = 2\pi/k_{\text{max}}$ and ultimately the droplet diameter is a fraction of λ_{RT} [19]. Therefore,

assuming viscous and surface tension effects are additive to the leading order according to the dispersion relation, we look for a correlation in the form

$$\lambda_{RT} = 2\pi \left[\sqrt{\frac{3\sigma}{a\rho_1}} + C_2 \sqrt[3]{\frac{\alpha_1^2}{a\rho_1^2}} \right] \quad (11.12)$$

The acceleration a in Eq. (11.12) is simply $a = F/m = F/\rho_1 V$, where the force F is the drag force exerted by the gas stream on a liquid element, here the liquid tongue of the primary instability,

$$F = \frac{1}{2} C_D \rho_g (U_{Gas} - U_c)^2 A_e \quad (11.13)$$

where

$C_D \approx 2$ is the drag coefficient.
 A_e the projected area.

The mass of the liquid to be accelerated is $m = \rho_1 b_1 A_e$ with $b_1 \propto \lambda_1$. The expression for a is therefore given by

$$a \approx \frac{\rho_g (U_{Gas} - U_c)^2}{\rho_1 b_1} \quad \text{Weber number :} \quad (11.14)$$

Substitution of Eq. (11.14) in Eq. (11.12) gives

$$\lambda_{RT} \propto \left(\frac{\sigma \lambda_1}{\rho_g (U_g - U_c)^2} \right)^{1/2} \left(1 + C_2' \left\{ \frac{\rho_g (U_g - U_c)^2}{\lambda_1 \sigma} \right\}^{1/6} \left\{ \frac{\alpha_1^2}{\rho_1 \sigma} \right\}^{1/3} \right) \quad (11.15)$$

Further substituting for λ_1 from Eq. (11.1), using Eq. (11.2) and taking the drop diameter, say, the Sauter mean diameter (SMD), proportional to λ_{RT} gives

$$\frac{SMD}{D_1} = C_1 (1 + m_r) \left(\frac{b_g}{D_1} \right)^{1/2} \left(\frac{\rho_1 / \rho_g}{Re_{bg}} \right)^{1/4} \frac{1}{\sqrt{We_{D_1}}} \left\{ 1 + C_2 \left(\frac{D_1}{b_g} \right)^{1/6} \left(\frac{Re_{bg}}{\rho_1 / \rho_g} \right)^{1/12} We_{D_1}^{1/6} Oh^{2/3} \right\} \quad (11.16)$$

In Eq. (11.16), the mass loading effect in the form $(1+m_r)$ is obtained from energy arguments previously outlined by [22],

where

$$m_r = \frac{m_1}{m_g} = \frac{\rho_1 U_{Liquid} A_1}{\rho_g U_{Gas} A_g}$$

A_1 and A_g are the areas of the liquid and gas nozzle exit sections, respectively. Furthermore, this equation indicates

a dependency of the SMD on $U_{Gas}^{-5/4}$ and $\sigma^{-1/2}$. The drop diameter increases with $b_g^{1/4}$ if the coefficient of proportionality C in Eq. (11.3) remains constant when b_g is changed. As will be shown below, this would only be the case if the length of the gas jet potential cone is much larger than the liquid jet's intact length, which is not typical of atomizers designs.

The SMD in Eq. (11.16) has been made dimensionless by the liquid orifice diameter D_1 and the Weber and Ohnesorge numbers are based on D_1 following the usual convention. However, it should be emphasized that the drop diameter does not depend on the liquid orifice diameter but rather on the gas boundary layer thickness at the nozzle exit. This has been clearly demonstrated by [19] where the liquid orifice diameter was changed by a factor of 3 and the drop diameter remained practically identical for the same gas flow conditions.

For completeness, the various nondimensional parameters in Eq. (11.16) are defined as follows:

$$\begin{aligned} \text{Weber number : } We_{D_1} &= \frac{\rho_g (U_{Gas} - U_c)^2 D_1}{\sigma} \\ \text{Ohnesorge number : } Oh &= \frac{\alpha_1}{\sqrt{\rho_1 \sigma D_1}} \\ \text{Reynolds number : } Re_{bg} &= \frac{U_{Gas} b_g}{\nu_g} \\ \text{Massflux ratio : } m_r &= \frac{\rho_1 U_{Liquid} A_1}{\rho_g U_{Gas} A_g} \end{aligned} \quad (11.17)$$

The coefficients C_1 and C_2 in Eq. (11.16) are order 1 and values for both coefficients are determined from experiments. The value of C_1 depends on the gas nozzle geometry in general, and on the contraction ratio in particular, because for a given nozzle size the gas boundary layer thickness at the liquid nozzle discharge depends strongly on the contraction ratio. C_2 characterizes the viscosity dependence of the critical wave number in the R-T instability, compared to the surface tension dependence. This value is associated with the additivity and linearity of both cohesive effects, surface tension, and viscosity, which determine the growth rate of the instability. The validity of the linear theory for R-T instability has been confirmed for a wide parameter range via qualitative observation of the jet breakup process.

Another important parameter, which does not appear explicitly in Eq. (11.16), is the dynamic pressure ratio M that determines the rate of atomization and hence the intact length of the liquid stream [23].

This ratio is defined as

$$M = \rho_g \frac{U_{Gas}^2}{\rho_1 U_{Liquid}^2} \quad (11.18)$$

The dimensionless intact length of the liquid stream can be defined as $\frac{L}{D_1} \approx \frac{6}{\sqrt{M}}$ and in coaxial injectors used under usual

tablet coating parameters, M is typically large (of the order 100). The gas potential cone length is approximately $6b_g$. For efficient atomization it is desirable that the gas potential cone length be equal to or larger than the liquid intact length so that the primary atomization is completed before the gas velocity starts to decrease. This requirement is expressed by

$$\frac{b_g \sqrt{M}}{D_1} > 1 \quad (11.19)$$

It is worth noting that for the flow rates and atomizers utilized in pharmaceutical tablet coating, Eq. (11.19), is satisfied easily, with values typically exceeding 10, strongly suggesting that atomization is typically quite rapid and efficient. Finally, the fluid jets under the conditions of interest here are laminar at the injector nozzle but would potentially become turbulent if the flow rates are significantly increased or the gas contraction ratio decreased. Turbulent conditions of the liquid stream at the nozzle discharge plane would have little effect on the atomization process, while turbulent conditions in the high-speed gas stream would require altering the exponent of Re_{bg} in Eq. (11.16).

11.5.1 Pattern Air Effect on the Liquid Breakup

The previous model is derived for two coaxial streams of liquid and air, without taking into account the effect of pattern air on the liquid breakup. The effect of pattern air can be included in the model in two different ways.

In spray guns where the pattern air nozzles are oriented in a way that the jets impinge on the spray axis at a distance from the atomization nozzle at which liquid breakup is already underway, the effect of pattern air on droplet size is minor. Because the droplet size is set by the wavelength of the secondary R-T instability, pattern air does not modify it when the instability is already growing. Therefore, for this type of atomizers, only the gas flow through the atomization nozzle affects the breakup. In atomizers where the pattern air comes from the same supply line as the atomizer air, the effect of pattern air is simply to reroute some of the available gas from the atomizing nozzle to the side jets, therefore reducing the atomizing air flow rate and the exit velocity. To account for that, in the case when pattern air does not have independent flow control and measurements, it is easiest to measure the cross-sectional area of both the atomizing nozzle and the pattern air and to prorate the total flow rate into the two air streams according to the pressure loss across the two nozzles (which is proportional to the diameter to the fourth power). The atomizing air is the value that is input into the breakup model, setting U_g and consequently the Weber number and the gas Reynolds number in the model.

In spray guns where the position and orientation of the pattern air nozzles results in the side jets impinging on the spray axis very close to the atomizing nozzle, in the region where

the instability of the liquid jet is still developing, the impact of the pattern air jets on the droplet size resulting from the atomization is very strong. This effect can be modeled by considering that the atomizing air and the pattern air mix, while the liquid jet is breaking up. The resulting air stream has axial momentum equal to the sum of all the jets, but no radial momentum as the two pattern air side jets has equal but opposite values that cancel when the two streams mix. One can implement this into the previously derived model by computing the gas jet as the mean air velocity of the two streams, $U_g = (m_{atom} U_{atom} + m_{pattern} U_{pattern} \cos \theta) / (m_{atom} + m_{pattern})$,

where

θ is the angle between the pattern air jet axis and the atomizing air jet axis.

Although there are a number of approximations underlying this model, most importantly that the air streams mix instantly and that the instability is not modified by the lack of axial symmetry induced by the pattern air, it has been proven to provide accurate predictions for the droplet size under typical pharmaceutical coating conditions.

11.6 APPLICATION OF ATOMIZATION MODEL

In this section, few examples are presented that demonstrate the practical application of atomization model.

EXAMPLE 11.1

Evaluate the Effect of Coating Type on Sauter Mean Drop Size

Using the atomization model described above, calculate the SMD size for a film coating process using a 15 wt% solids, HPMC-based coating Y-30-18037. The coating formulation properties are given in Table 11.2 and the process conditions and spray nozzle geometry are listed in Table 6.1. How would the drop size be affected if a 20 wt% solids, PVA-based coating 85F18422 was to be used instead? How could one adjust the process conditions to change the drop size of the PVA-based coating to match that of the HPMC-based coating?

Solution

Given the material properties in Table 11.2, coating 1, and the process parameters in Table 11.4, Example 1-1, the atomization model in Eq. (11.3) predicts a droplet size (SMD) of 49.9 μm . If the same process were used for a PVA-based formulation (material properties in Table 11.2, coating 2), the predicted droplet size would be approximately 34% smaller, as shown in Table 11.4, Example 1-2. This may not be intuitive since the solids content was increased in the PVA

TABLE 11.4 Impact of Coating Formulation Polymer Type (HPMC vs. PVA) on Droplet Size and Selection of Atomization Air Flow Rate to Achieve a Similar Droplet Size

	Example 1-1 HPMC-Based Y-30-18037 (15% solids)	Example 1-2 PVA-Based 85F18422 (20% solids)	Example 1-3 PVA-Based 85F18422 (20% solids)
Spray Formulation, Parameters, and Variables			
Process type	Process A	Process A	Process B
Coating type	Coating 1	Coating 2	Coating 2
Action taken	None Base process	Base process with new coating	Atomizing gas decreased to match droplet size of Process A Coating 1
Spray pattern type	Ellipse	Ellipse	Ellipse
Fluid nozzle part no.	PF28100	PF28100	PF28100
Air cap part no.	PA110228-45	PA110228-45	PA110228-45
Input Parameters			
Inner diameter of liquid nozzle (mm)	0.71	0.71	0.71
Outer diameter of the atomizing gas cap (mm)	2.79	2.79	2.79
Inner diameter of the atomizing gas cap (mm)	1.78	1.78	1.78
Diameter of the pattern air side orifices (mm)	0.71	0.71	0.71
Liquid surface tension (mN/m)	47.0	43.9	43.9
Liquid density (kg/m ³)	1050.0	1070.0	1070.0
Infinite shear rate viscosity (kg/ms)	0.098	0.039	0.039
Spray rate (grams/min/spray gun)	10	10	10
Atomizing and pattern gas volumetric flow rate (m ³ /s)	9.44E-04	9.44E-04	6.63E-04
Calculated Values Used in Model			
Atomizing gas outlet velocity (m/s)	213.7	213.7	150.1
Model Output			
Sauter mean diameter (μm)	49.9	33.1	49.9

example, but due to the lower viscosity of the PVA-based formulation, the droplet size will decrease. One approach to match the droplet size of the HPMC-based coating in Example 1-1 (HPMC Process A) is to reduce the atomization air, as shown in Table 11.4 for the lower viscosity PVA coating (Example 1-3 (PVA Process B)).

EXAMPLE 11.2

Evaluate the Effect of Atomization and Pattern Air on Sauter Mean Drop Size

The coating process using the HPMC-based coating Y-30-18037, described in Example 11.1, is modified by changing the spray gun to a Schlick ABC anti-bearding cap design with dimensions given in Table 11.5. If the same atomization air flow rate is used without pattern air, what will be the mean droplet size for this new process? If the atomization air is set to a volumetric rate equivalent to the pattern air, how would the mean droplet size change? How should the air atomization and pattern air be adjusted, maintaining a 1 : 1 ratio of the volumetric flow rates, to reach a mean drop size of 49.9 μm?

Solution

Given the material properties in Table 11.2, coating 1, and the process parameters in Table 11.5, Example 2-1, the atomization model in Eq. (11.3) predicts a droplet size (SMD) of 33.9 μm due to differences in spray gun geometry. The addition of independent pattern air of equal volume (a relatively high setting) further reduces the droplet size to 27.4 μm due to the pattern air contribution to atomization, which was discussed earlier in the chapter (Table 11.5, Example 2-2). In order to meet a target droplet size of 49.9 μm, the atomization model can be used to back calculate the necessary air flow rates, assuming a 1 : 1 volumetric ratio, with inputs given in Table 11.5, Example 2-3. With this spray nozzle, the air flow rates must be reduced to 5.28E-04 m³/s to reach the target drop size of 49.9 μm.

The tablet coating process is often developed and optimized at a small scale and subsequently scaled up to larger equipment and batch size. The coating and atomization process are developed and optimized in a small lab coater, typically at 1–10l capacity using a few kilograms of tablets, through detailed experimentation. The main goal of this

TABLE 11.5 Impact of Spray Gun Type and Pattern Air Effect on Droplet Size for the Schlick ABC Setup

	Example 2-1 Spray Gun 2 Schlick ABC Y-30-18 037 (15% Solids)	Example 2-2 Spray Gun 2 Schlick ABC Y-30-18 037 (15% Solids)	Example 2-3 Spray Gun 2 Schlick ABC Y-30-18 037 (15% Solids)
Spray Formulation, Parameters, and Variables			
Action taken	No pattern air	High volume of pattern air	Change in process to give 49.9 μm droplets to match Example 1-1
Spray pattern type	Circular	Ellipse	Ellipse
Fluid nozzle part no.	10 w44019	10 w44019	10 w44019
Air cap part no.	27 w44183	27 w44183	27 w44183
Input Parameters			
Inner diameter of liquid nozzle (mm)	1.00	1.00	1.00
Outer diameter of the atomizing gas cap (mm)	2.70	2.70	2.70
Inner diameter of the atomizing gas cap (mm)	1.85	1.85	1.85
Diameter of the pattern air side orifices (mm)	1.50	1.50	1.50
Liquid surface tension (mN/m)	47.0	47.0	47.0
Liquid density (kg/m^3)	1050.0	1050.0	1050.0
Infinite shear rate viscosity (kg/ms)	0.098	0.098	0.098
Spray rate (grams/min/spray gun)	10	10	10
Atomizing gas volumetric flow rate (m^3/s)	9.44E-04	9.44E-04	5.28E-04
Pattern gas volumetric flow rate (m^3/s)	0	9.44E-04	5.28E-04
Calculated Values Used in Model			
Atomizing gas outlet velocity (m/s)	310.8	310.8	173.8
Model Output			
Sauter mean diameter (μm)	33.9	27.4	49.9

experimentation is to explore the design space of the process parameters such as the spray rate, atomization air flow rate, and pattern air flow rate which will produce spray that will result in the desired film coat on the tablet surface. When this process is scaled up, process parameters should be selected such that coating spray will be similar to the one generated at the smaller scale.

EXAMPLE 11.3

Scale-up of Atomization

The tablet coating process is often developed and optimized at a small scale and subsequently scaled up to larger equipment and batch size. The coating and atomization process are developed and optimized in a small lab coater, typically at 1–10 l capacity using a few kilograms of tablets, through detailed experimentation. The main goal of this experimentation is to explore the design space of the process parameters such as the spray rate, atomization air flow rate, and pattern air flow rate, which will produce spray that will result in the desired film coat on the tablet surface. When this process is scaled up, process parameters should be selected such that coating spray will be similar to the one generated at the

smaller scale. An important spray characteristic that should be kept the same across different scale experiments is the SMD droplet size. In this example, Eq. (11.16) will be used to calculate the SMD for 1–2, 30, and 500 l coating pans based on the spray parameters given in Table 11.6. Further, this model will be used to set the atomization gas flow rate, which will produce spray with same SMD for the given parameters.

11.7 GLOBAL SENSITIVITY ANALYSIS OF ATOMIZATION MODEL

The atomization model described in previous sections is dependent on several input parameters. The reliability of this model's prediction could depend on the source of uncertainty in the input parameters. Therefore, it is important to carry out a sensitivity analysis (SA) of the model to both quantify the impact of this variability on the output as well as rank order the importance of various input parameters. Sensitivity analysis has been succinctly defined by Saltelli et al. [24] as “the study of how the uncertainty in the output of a model can be apportioned to different sources of uncertainty in the model input.” Various techniques to carry out SA can be broadly

TABLE 11.6 Process Variables and Parameters for Spray Atomization for Three Different Scale Coaters Used in Scale-Up Example

Spray Formulation, Parameters, and Variables	Example 3-1 HPMC-Based Y-30-18 037 (15% Solids)	Example 3-2 HPMC-Based Y-30-18 037 (15% Solids)	Example 3-3 HPMC-Based Y-30-18 037 (15% Solids)
Process type	Process A	Process B	Process C
Coating type	Coating 1	Coating 1	Coating 1
Action taken	None	Atomizing gas increased to match droplet size of base Process	Atomizing gas increased to match droplet size of base Process
Spray pattern type	Ellipse	Ellipse	Ellipse
Fluid nozzle part no.			
Air cap part no.			
Input Parameters			
Inner diameter of liquid nozzle (m)	1.00E-03	1.50E-03	1.50E-03
Outer diameter of atomizing gas cap (m)	3.00E-03	4.00E-03	4.00E-03
Inner diameter of atomizing gas cap (m)	2.00E-03	3.00E-03	3.00E-03
Diameter of the pattern air side orifices (m)	1.0E-03	2.0E-03	2.0E-03
Liquid surface tension (N/m)	0.047	0.047	0.047
Liquid density (kg/m ³)	1050.0	1050.0	1050.0
Infinite shear rate viscosity (kg/ms)	0.0981	0.0981	0.0981
Spray rate (grams/min/spray gun)	10	60	120
Atomizing and pattern gas volumetric flow rate (m ³ /s)	1.18E-03	2.92E-03	3.41E-03
Calculated Values Used in Model			
Atomizing gas outlet velocity (m/s)	214.61	247.97	289.63
Model Output			
Sauter mean diameter (µm)	48.4	48.4	48.4

Fluid properties for Opadry II White (Colorcon Code: Y-30-18037).

Spray gun parameters are for example purposes and are not selected based on a specific spray gun model. Pattern air is not independent in this example.

classified into local and global strategies. Local sensitivity analysis is concerned with individual input–output relationships. It typically involves the calculation of partial derivatives of output with respect to various input parameters. The term local refers to the fact that all derivatives are taken at a single point. This kind of analysis cannot evaluate the effect of multiple inputs acting in concert on the output. In global sensitivity analysis (GSA), all inputs are varied simultaneously to evaluate the effect of individual inputs as well as the cooperative effect of multiple inputs on the output. This rich information content from global SA typically requires a large Monte Carlo sampling study and could be numerically expensive depending on the model at hand [25, 26].

Since a single invocation of the spray atomization model in MATLAB takes less than one second, global sensitivity analysis is being carried out and presented here. The goal of this analysis is to determine the impact of two operating parameters, namely, atomization air flow rate and coating fluid spray rates, along with three material parameters, namely, viscosity, surface tension and density, on the mean drop size. There are several methods to carry out GSA, one which is chosen is the Sobol (2001) approach based

on ANalysis Of VAriance (ANOVA) decomposition. Cannavó [27] has developed a MATLAB based toolbox, GSAT, to carry out GSA. This toolbox was obtained from MATLAB central repository and appropriately modified to analyze the output of atomization model. It uses the algorithm developed by Homma and Saltelli [25] to compute first-order and higher-order sensitivity indices. The nominal values of five input parameters were chosen as shown in Table 11.4 for coating 1 and process A and GSAT software computed (25-1) sensitivity indices. Figure 11.17 shows the results of Monte Carlo sampling of 1000 runs when five input parameters were varied between 30% of their nominal values using a pseudo-random distribution. The scatterplot shows that the mean drop size varies between 10 to 120 µm, and it increases with the liquid viscosity and decreases with the atomization air flow rate. Figure 11.18 shows the histogram representing the distribution of mean drop size. It is clear from this figure that distribution is asymmetric about mean and is skewed toward larger drop size, emphasizing the non-linearity of the model. Figure 11.19 shows the first-order sensitivity indices for the five input parameters. The dominant effects are contributed by atomization air flow rate and liquid

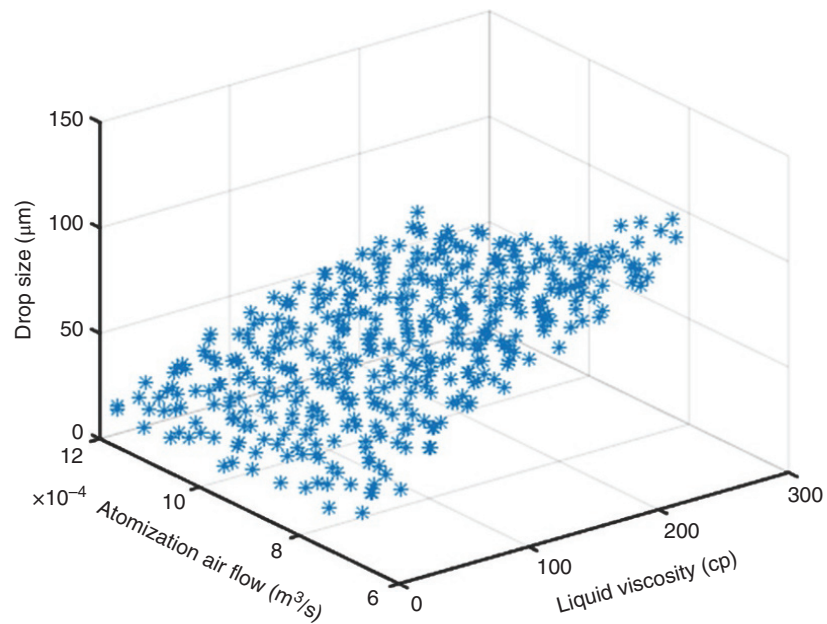


FIGURE 11.17 Mean drop diameter plotted against atomization air flow rate and liquid viscosity.

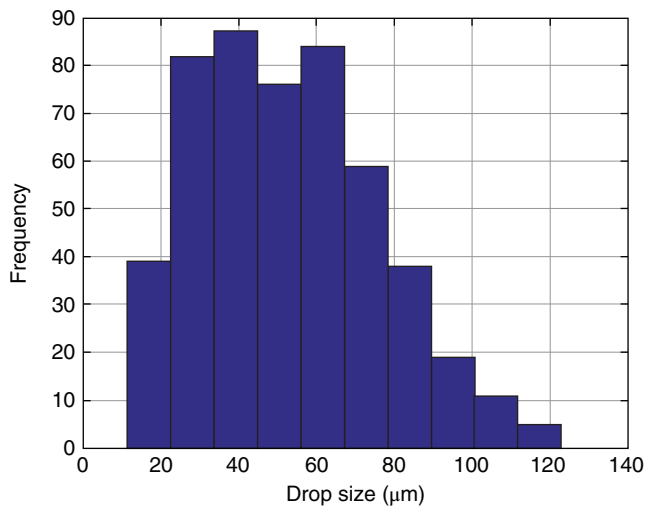


FIGURE 11.18 Histogram showing the distribution of mean drop size when five input parameters are varied within 30% of their nominal values. This distribution is obtained from a Monte Carlo sample of 1000 elements.

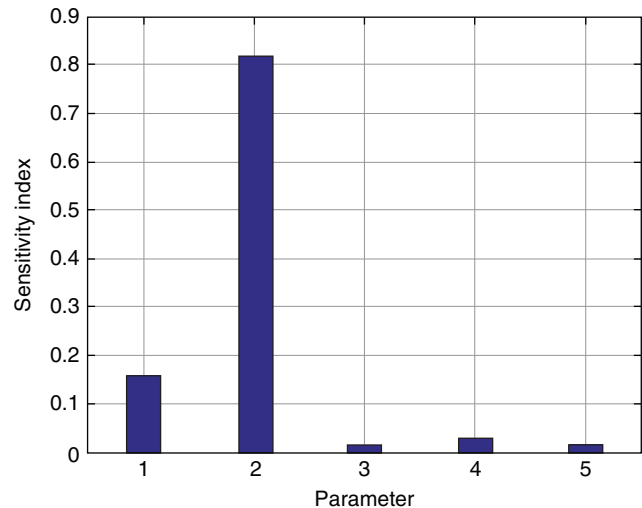


FIGURE 11.19 First-order sensitivity indices for five input parameters when they are allowed to vary within 30% of their nominal values. Parameters are defined as 1 – coating fluid viscosity, 2 – atomization air flow rate, 3 – spray rate, 4 – coating fluid surface tension, and 5 – coating fluid density.

viscosity, with sensitivity indices of 0.81 and 0.17, respectively. The remaining three other parameters along with all the higher-order indices have negligible impact on the output. SA clearly shows these two parameters should be closely monitored during atomization process design. As shown in Figure 11.4, the coating liquid viscosity exhibits large variation with increasing solids fraction. To account for this variability, SA is carried out where liquid viscosity was allowed to vary between 75% of its nominal value, whereas the

variability of other four parameters was kept at 30%. Figure 11.20 shows the first-order indices for the five input parameters. It is clear from this figure that liquid viscosity and atomization air flow rate still make the dominant contributions to the variability of the output. However, now the fluid viscosity assumes the role of most dominant input parameter. So, for the coating process that uses liquid with high solids fraction, fluid viscosity should be properly controlled and monitored.

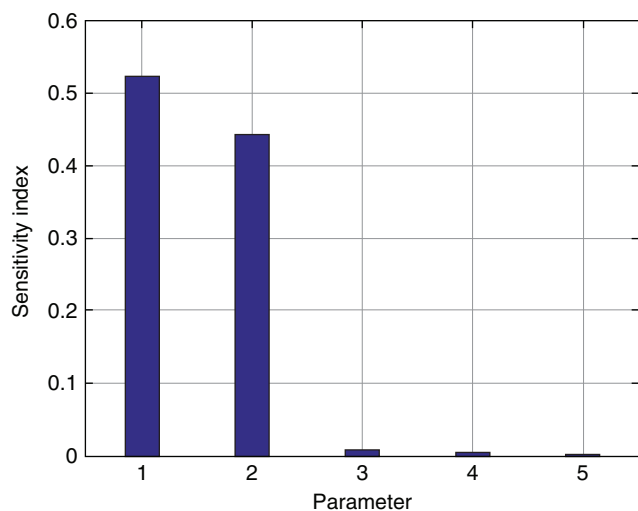


FIGURE 11.20 First-order sensitivity indices for five input parameters. Here, coating solution viscosity is allowed to vary between 75% of the nominal value, whereas the remaining four input parameter vary within 30% of their nominal values. Parameters are defined as 1 – coating fluid viscosity, 2 – atomization air flow rate, 3 – spray rate, 4 – coating fluid surface tension, and 5 – coating fluid density.

11.8 CONCLUSIONS

In this chapter, we have outlined the physical processes that are involved in the spray coating of pharmaceutical tablets. We have focused on the atomization of the complex rheology liquid formulations used in the coating because the size and spatial distribution of the droplets formed during the atomization process have a strong influence on the success or failure of the overall process. The essential physics of the problem have been described in detail, and a mathematical model for the prediction of the average droplet size has been presented. The hydrodynamic stability analysis of the two parallel streams of gas and liquid characteristic of two-fluid atomizers, commonly used in the coating operations, provides quantification of the functional dependency of the resulting droplet size with the coating formulation properties (surface tension, viscosity) and the operational conditions (liquid and gas flow rates). We have also investigated the spatial distribution of droplets within the spray and the effect of the symmetry-breaking side gas jets (commonly referred to in the industry as pattern air). We show the redistribution of droplets into an elliptical cross-sectional spray due to these side jets and a more uniform spray density. The conditions under which these side jets need to be taken into account in the calculation of the droplet size have been identified and a simple way to include them in the breakup model is presented.

The availability of quantitative physics-based models like the one presented in this chapter can change the way pharmaceutical manufacturing operations, like tablet

coating, is approached. The capability of predicting droplet size and distribution onto the coated material allows for the optimization of the process with reduced input from experiments, as opposed to the traditional empirical approach that requires costly and time-consuming tests at each step of the process design (lab, pilot, and full scale). The improved understanding of the relationship between coating liquid rheology and coating outcomes can also provide more flexibility in the use of advanced coating formulations, reducing the barriers to apply novel coatings to improve a product or process.

ACKNOWLEDGMENTS

The authors would like to acknowledge Pfizer Inc. for supporting the research reported in this chapter. Additionally, Douglas M. Kremer is thanked for his contributions to set up and develop this collaborative research effort. We also wish to thank Dr. Emil Hopfinger for his assistance in developing the atomization model described in Section 11.6. Katie Osterday is acknowledged for assisting in the viscosity measurements as well as in the measurements of the droplet size distributions using PDPA techniques. Daniel Bolleddula is acknowledged for helping setup the two-fluid coaxial atomizers for the independent pattern air experiments.

REFERENCES

1. Smith George, W., Macleod Graeme, S., and Fell John, T. (2003). Mixing efficiency in side-vented coating equipment. *AAPS PharmSciTech* 4 (3): E37.
2. am Ende, M.T. and Berchielli, A. (2005). A thermodynamic model for organic and aqueous tablet film coating. *Pharmaceutical Development and Technology* 10 (1): 47–58.
3. Kalbag, A., Wassgren, C., Penumetcha, S.S., and Pérez-Ramos, J.D. (2008). Inter-tablet coating variability: residence times in a horizontal pan coater. *Chemical Engineering Science* 63 (11): 2881–2894.
4. Ebey, G.C. (1987). A thermodynamic model for aqueous film-coating. *Pharmaceutical Technology* 11 (4): 40, 42–43, 46, 48, 50.
5. Porter, S.C. and Felton, L.A. Techniques to assess film coatings and evaluate film-coated products. *Drug Development and Industrial Pharmacy* 36 (2): 128–142.
6. McGinity, J.W., L.A. Felton, *Aqueous Polymeric Coatings for Pharmaceutical Dosage Forms*, 3rd. 2008. NewYork; Informa Healthcare, pp. 488.
7. Turton, R. and Cheng, X.X. (2005). The scale-up of spray coating processes for granular solids and tablets. *Powder Technology* 150 (2): 78–85.
8. Gibson, S.H.M., Rowe, R.C., and White, E.F.T. (1988). Mechanical properties of pigmented tablet coating formulations and their resistance to cracking. I. Static mechanical measurement. *International Journal of Pharmaceutics* 48 (1–3): 63–77.

9. Aliseda, A., Hopfinger, E.J., Lasheras, J.C. et al. (2008). Atomization of viscous and non-Newtonian liquids by a coaxial, high-speed gas jet. Experiments and droplet size modeling. *International Journal of Multiphase Flow* 34 (2): 161–175.
10. Mueller, R. and Kleinebudde, P. (2007). Comparison of a laboratory and a production coating spray gun with respect to scale-up. *AAPS PharmSciTech* 8 (1): 3.
11. Mueller, R. and Kleinebudde, P. (2006). Comparison study of laboratory and production spray guns in film coating: effect of pattern air and nozzle diameter. *Pharmaceutical Development and Technology* 11 (4): 425–433.
12. Tobiska, S. and Kleinebudde, P. (2003). Coating uniformity: influence of atomizing air pressure. *Pharmaceutical Development and Technology* 8 (1): 39–46.
13. Thombre, A.G., Appel, L.E., Chidlaw, M.B. et al. (2004). Osmotic drug delivery using swellable-core technology. *Journal of Controlled Release* 94 (1): 75–89.
14. Thombre, A.G., Cardinal, J.R., AR, D.N., and Gibbes, D.C. (1999). Asymmetric membrane capsules for osmotic drug delivery. II. in vitro and in vivo drug release performance. *Journal of Controlled Release* 57 (1): 65–73.
15. Cole, G., Hogan, J., and Aulton, M.E. (1995). *Pharmaceutical Coating Technology*. Informa Healthcare.
16. de Gennes, P.-G., Brochard-Wyart, F., and Quéré, D. (2002). *Capillary and Wetting Phenomena – Drops, Bubbles, Pearls, Waves*. Springer.
17. Bachalo, W.D. (1994). Experimental methods in multiphase flows. *International Journal of Multiphase Flow* 20 (Suppl., Annual Reviews in Multiphase Flow 1994): 233–259.
18. Williams, F.A., *Combustion Theory*. 2nd ed. 1985: Addison Wesley.
19. Varga, C.M., Lasheras, J.C., and Hopfinger, E.J. (2003). Initial breakup of a small-diameter liquid jet by a high speed gas stream. *Journal of Fluid Mechanics* 497: 405–434.
20. Marmottant, P. (2001). *Atomisation d'un courant liquide dans un courant gazeux*. Grenoble: Institut National Polytechnique de Grenoble.
21. Joseph, D.D., Beaver, G.S., and Funada, T. (2002). Rayleigh-Taylor instability of viscoelastic drops at high weber numbers. *Journal of Fluid Mechanics* 453: 109–132.
22. Mansour, A. and Chigier, N. (1995). Air-blast atomization of non-Newtonian liquids. *Journal of Non-Newtonian Fluid Mechanics* 58: 161–194.
23. Hopfinger, E.J. and Lasheras, J.C. (2000). Liquid jet instability and atomization in a coaxial gas stream. *Annual Review of Fluid Mechanics* 32: 275–308.
24. Saltelli, A., Tarantola, S., Campolongo, F., and Ratto, M. (2004). *Sensitivity Analysis in Practice: A Guide to Assessing Scientific Models*. Wiley.
25. Homma, T. and Saltelli, A. (1996). Importance measures in global sensitivity analysis of nonlinear models. *Reliability Engineering & System Safety* 52 (1): 1–17.
26. Sobol, I.M. (2001). Global sensitivity indices for nonlinear mathematical models and their Monte Carlo estimates. *Mathematics and Computers in Simulation* 55 (1): 271–280.
27. Cannavó, F. (2012). Sensitivity analysis for volcanic source modeling quality assessment and model selection. *Computers & Geosciences* 44: 52–59.

12

SPRAY DRYING AND AMORPHOUS DISPERSIONS

KRISTIN J.M. PLOEGER AND PAIGE ADACK

Pharmaceutical Commercialization Technology, Merck & Co., Inc., West Point, PA, USA

PAVITHRA SUNDARARAJAN

Formulation Sciences, Merck & Co., Inc., West Point, PA, USA

PEDRO C. VALENTE AND JOÃO G. HENRIQUES

R&D Drug Product Development, Hovione Farmaciência SA, Loures, Portugal

KENNETH J. ROSENBERG

Center for Materials Science and Engineering, Merck & Co., Inc., West Point, PA, USA

12.1 INTRODUCTION

Spray drying is used by numerous industries, with applications ranging from dairy and flavorings to detergents and ceramics. The earliest spray dryers were wood and copper, used for dehydrating dairy and eggs nearly a century ago. Recently, spray drying is increasingly used in pharmaceuticals, as it allows for the generation of materials and design of particles not previously possible [1, 2].

Spray drying has numerous applications within the pharmaceutical industry. Several common excipients used in oral solid dosage formulations are spray-dried, such as lactose (e.g. “FastFlo Lactose^(R)” by Foremost) and mannitol (e.g. “Pearlitol SD200^(R)” by Roquette). Nutraceuticals use the technology to generate colloidal dispersions and granular forms of waxy materials, such as vitamin D₃. The isolation of some drug substances that cannot be crystallized or are thermolabile utilize spray drying. There is significant interest in spray drying for large molecules (peptides and proteins), including stabilizing them for injection and as an alternative to lyophilization. Inhalation products benefit from the small particle size and low density of spray-dried particles [3].

Of recent interest in the pharmaceutical industry is the use of spray drying to produce amorphous solid dispersions (ASDs), which allows the formulation of drugs into solid dosage forms, which previously had an aqueous solubility

too low to be dosed orally [4–6]. This chapter will focus on this application, which is typically used by small molecule active ingredients in a polymer matrix and spray-dried out of an organic solvent (due to the low aqueous solubility). However, many of the concepts presented would apply to other applications, both within and outside of pharmaceuticals.

The chapter begins with an overview of the role of solid dispersions, in particular spray-dried solid dispersions, within the pharmaceutical industry. It will then consider these spray-dried particles from a pharmaceutical materials perspective, both in consideration of materials selected during formulation and in understanding of the stability of ASDs. It then considers spray drying in context of individual unit operations used to convert active ingredients and raw materials into a final dosage form. The spray drying unit operation is then considered from a thermodynamics perspective, where overall heat and mass energy balances yield information about the interrelationship of process parameters. Conservation equations, including fluid dynamics and mass and energy transfer, are then applied to spray drying on both the level of a single droplet drying and computational fluid dynamics (CFD) simulations of the entire spray chamber. The section ends with discussion of designs of experiments used to develop spray drying processes and how the engineering concepts in the preceding sections can inform the design of those studies.

12.2 ROLE OF SOLID DISPERSIONS IN THE PHARMACEUTICAL INDUSTRY

Developing and manufacturing small molecules that are highly soluble and highly permeable into oral drug products has become straightforward, reliable, and routine. However, 75% of the pharmaceutical pipeline [7] and 40–45% of the oral marketed products are classified as poorly soluble drugs [8]. *In vivo* absorption of these drugs is limited by solubility of the drug. In order to improve absorption, different delivery techniques for enhancing solubility of the drug, such as ASD formulations, nanoformulations, lipid-based formulations, and co-crystals, are being used. Significant progress has been made in understanding and designing these delivery techniques, and they are changing the landscape of drug product development [9].

ASDs are single-phase solid solutions of the poorly soluble drug, polymer, and sometimes surfactants. These ASDs maintain the drug in the metastable amorphous phase with higher apparent solubility, thereby improving absorption. Several researchers have shown enhancement in the *in vitro* dissolution performance, the *in vivo* pharmacokinetic (PK) performance in preclinical species, and ultimately the PK performance in human clinical studies with the use of ASD-based formulation relative to the crystalline drug formulation [10]. In addition to improvement in bioperformance, solid dispersions also provide robustness and insensitivity in PK performance when subject to variations in gastric pH conditions and food intake [11]. Figure 12.1 provides an example of enhanced bioavailability using ASD formulation relative to conventional crystalline formulation. There are several ways of generating ASDs, such as coprecipitation, extrusion, freeze drying, and spray drying. Of these technologies, spray drying and extrusion have been most commonly used, and the number of commercially marketed products using these two techniques has dramatically increased in the last decade [1].

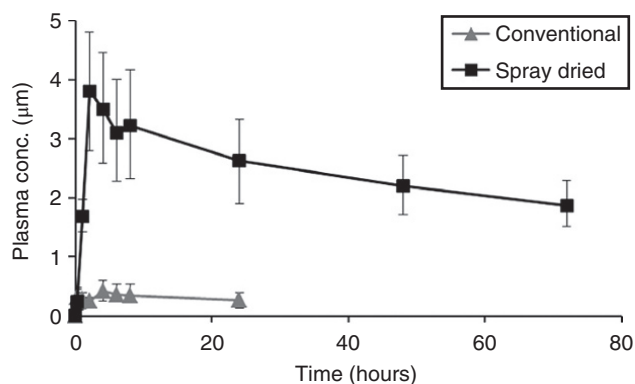


FIGURE 12.1 Comparison of the pharmacokinetic performance (plasma concentration as a function of time) between spray-dried dispersion formulation and conventional formulations.

Spray drying involves dissolving the solid components, namely, the drug, polymer, and surfactants, into solvents and drying the resulting solution by spraying the solution in a drying gas. Spray drying kinetically traps drug in the metastable amorphous phase through fast-drying kinetics. The ASD generated by spray drying will be referred to in this chapter as spray-dried dispersion (SDD). Extrusion, on the other hand, involves mechanically mixing and melting the drug with polymer and surfactants, creating an amorphous solid solution. The differences in the processes and their impacts on the attributes of resulting ASD determines their respective utility: drugs or excipients that exhibit heat sensitivity, high melting point, or low miscibility cannot be extruded easily. Since spray drying involves dissolving the components into solvents, drugs or excipients with high melting point or miscibility can be handled by selection of appropriate solvents. By employing organic solvents with lower boiling points and thermal capacity, rapid drying kinetics of the small atomized droplets can be achieved without going to high temperatures. Hence, spray drying is feasible for thermally sensitive materials. However, removal of some solvents to below levels acceptable by ICH guidelines can be difficult, often requiring an additional drying step, potentially restricting the choice of solvent–excipient combinations. In terms of ASD properties, SDDs are typically low-density, highly compactable particles, while hot melt extruded dispersions are dense and have limited compactibility. In terms of equipment size and cost relative to throughput, spray dryers have a significant disadvantage. The spray drying equipment footprint for a commercial-scale supply of spray-dried drug product is much larger than extruded product. These considerations, both the suitability of a given technique for a new drug in the pipeline and the cost of commercialization, play a key role in the selection of the appropriate formulation and process.

12.3 SOLID DISPERSIONS

12.3.1 Formulation Development of Spray-Dried Dispersions

The formulations used in SDDs have evolved since its inception in the 1960s. While early formulations combined drugs with low molecular weight matrices, the more recent ones use higher molecular weight matrices. The most recent formulations have also become complex including surfactants and mixed solvent systems in generating solid dispersions [12]. The key consideration for selection of SDD composition as highlighted before is the promotion of the apparent amorphous solubility to provide supersaturation and improve oral absorption. However, given that these SDDs are primarily metastable glasses, physical stability of the single-phase amorphous SDD in the final drug product throughout the

shelf life of the product is another important consideration for the selection of SDD composition. In recent times, given the inclination to develop fixed dose combinations of drugs in the pharmaceutical industry, the interaction of components of SDD with other drugs in the oral dosage form and within the body is another important consideration.

Drugs that are poorly soluble in gastric fluids may be formulated as ASDs. While most poorly soluble drugs that are spray-dried are crystalline and are made amorphous by spray drying, amorphous drugs that have a pH-dependent solubility profile may benefit by spray drying with polymers/surfactants to create insensitivity of PK performance to food and gastric conditions. Generally, for a poorly soluble new drug in the early stages of development, high throughput solvent cast screens of different combinations of the drug with polymers and surfactants are used to evaluate the suitability of generating SDD. Based on stability of the drug in these solvent screens, prototype SDD is selected for spray drying at the pilot scale and further evaluated as a drug product formulation. The guiding principles of composition selection are highlighted below.

The solvent(s) used for spray solution is selected based on the solubility of the drug. Since poorly soluble drugs are usually lipophilic, they are often soluble in organic solvents. Additionally, based on the solubility of the polymer and surfactants, a mixture of solvents may also be considered. Other key factors that determine solvent selection are its thermal properties (e.g. latent heat for evaporation, specific heat, thermal conductivity, and the boiling point) and physical properties (e.g. surface tension and viscosity). These properties play an important role in ensuring that spray solution atomizes into droplets and dries quickly. In the case of mixture of solvents, in addition to relative values for the abovementioned properties, other thermodynamic considerations are the miscibility of the solvents and the mutual diffusion coefficient of the solvent mixture. Other solvents may be considered but must be reduced to sufficient levels at the end of process, such as those described in ICH guidance Q3C.

The selection of polymers is central to generating an ASD that is both physically stable and achieves the desired PK performance. Polymers serve as diluents in the SDD and help

stabilize the drug in its amorphous state. The key physical properties of the polymers are its glass transition temperature (T_g , described in Section 12.3.2), the solubility in the spray solution, and the viscosity of the resulting spray solution. When SDDs were generated in the 1960s, they used low molecular weight matrices such as urea- and succinic acid-based formulations. Modern SDDs use higher molecular weight polymers, mostly commonly vinyl or cellulosic derivatives (Table 12.1).

Surfactants are another component of an SDD formulation whose function is usually to improve the solubility of the drug in the gastric fluids. The formation of micelles, colloids [13], and nanoparticles during dissolution [14] in the gastric fluids are some of the common mechanisms by which surfactants improve the drug solubility and stability. The key physical property of a surfactant that impacts the spray drying kinetics is its influence on the surface tension of the droplet.

In addition to the selection of the components of the spray solution, the drug loading (weight fraction of solids % w/w) in SDD and the solids loading in the spray solution are important factors that can impact the ability to improve dissolution, physical stability, and throughput. Some of the common challenges experienced during formulation development are precipitation of the supersaturated drug during dissolution, recrystallization or phase separation of the solid dispersion during shelf life, long disintegration times for the SDD containing tablet, and interaction of the polymer or surfactant with the drug preventing complete dissolution. Thus, while increased drug loading and solids loading allow for better packing and small footprint of the final drug product, the dissolution performance and physical stability determine the highest allowed loading in the SDD and the drug product. Note that processing conditions of the spray drying process also play an important role in determining the stability and dissolution performance of the SDD.

Since dissolution testing of the drug is commonly used to evaluate the performance of the SDD, the selection of an appropriate dissolution method is critical. As highlighted elsewhere [10], developing a dissolution method with ability to predict or correlate with *in vivo* bioperformance is challenging for drugs whose absorption may be dissolution

TABLE 12.1 Summary of Common Solvents, Polymers, and Surfactants Used in SDD Formulations

Solvents	Polymer	Surfactants
Acetone	Hydroxypropyl methylcellulose	Polysorbate
Ethanol	Hydroxypropyl methylcellulose acetate succinate	Vitamin E polyethylene glycol succinate
Methanol	Hydroxypropyl methylcellulose phthalate	Poloxamer
Isopropanol	Polyvinylpyrrolidone	Cremophor
Water	Polyvinyl acetate phthalate	Sodium lauryl sulfate
Tetrahydrofuran	Polyvinylpyrrolidone–polyvinyl alcohol copolymers	
Dichloromethane	Eudragit [®] Soluplus [®]	

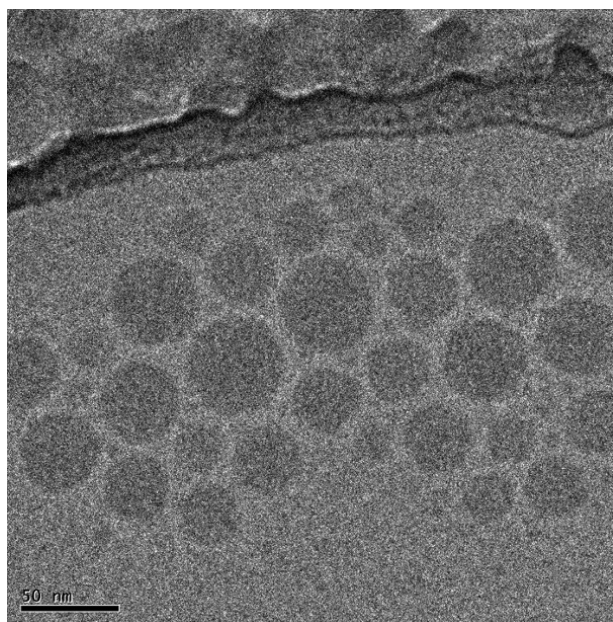


FIGURE 12.2 Cryo-TEM image of nanoparticle suspension obtained during dissolution of an amorphous solid dispersion containing surfactant. The edge of the carbon support film appears in the upper portion of the figure. Nanoparticles (darker gray circles) collect in the water film near the carbon film edge, packing closely and appearing approximately 20 minutes after start of dissolution [14]. *Source:* Reprinted with permission from Harmon et al. [14]. Copyright 2016 American Chemical Society.

limited. Often, dissolution methods are developed by building correlations with bioperformance in animals, but have not always been successful. The general guidance for assessing the performance of solid dispersions is discussed in Newman et al. [15]. Additionally, mechanistic understanding of dissolution of ASDs is still an active area of research [14, 16, 17]. Harmon et al. [14] through cryo-transmission electron microscopy imaging and microscopic imaging of the dissolution front built a holistic mechanistic rationalization of nanoparticle formation and release. They found that nanoparticles (Figure 12.2) are formed due to rapid dissolution of the polymer from the ASD (in their case, a polyvinylpyrrolidone–polyvinyl alcohol and vitamin E polyethylene glycol succinate-based hot melt extrudate). They also highlight the role of surfactants and drug loading on the nanoparticle yields and stabilization.

12.3.2 Impact of Formulation on Physical Stability

A schematic detailing the evolution of key thermodynamic properties during the formation of the amorphous glassy state by spray drying is illustrated in Figure 12.3. Once the initial droplet forms and the solvent begins to rapidly evaporate, there is an immediate decrease in the overall enthalpy, entropy, and specific volume of the dissolved solids in

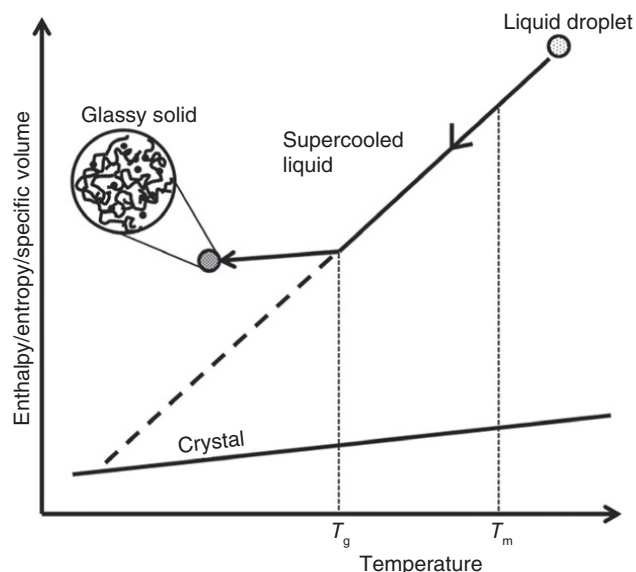


FIGURE 12.3 Thermodynamic schematic of spray-dried particle formation. The time scale for drying of the liquid droplet is fast enough to avoid the onset of crystallization (denoted by the melting point temperature T_m) and instead becomes a supercooled liquid. With continued evaporation, the viscosity of the remaining droplet increases and passes through the glass transition to reach the final glassy phase. *Source:* Adapted from Hancock and Zografis [18].

solution. By quickly drying the droplet, the thermodynamically favored crystalline state is bypassed, and the droplet enters a “supercooled” liquid state. Upon further evaporation, the viscosity of the drying droplet undergoes a significant increase due to diminished molecular mobility, causing the supercooled liquid state to deviate from the equilibrium liquid line. Eventually, the particle is effectively “frozen” in a nonequilibrium glassy solid where the final state is largely controlled by the drying rate. The temperature corresponding to this sudden change in the excess properties of enthalpy, entropy, and specific volume is termed the glass transition temperature (T_g) [19, 20]. The T_g is a function of both SDD composition and process parameters. For example, for a particular composition, spray drying conditions that lead to more rapid drying tend to result in the nonequilibrium solid state reached at higher temperatures, leading to a higher T_g , while a slower drying delays the conversion into the glassy state, often resulting in a lower T_g .

The elevated free energy of the amorphous state compared with the crystalline phase is one of the driving forces for both higher solubility and physical form instability [21]. Even after manufacture, the higher enthalpic level of the amorphous glassy state will continuously evolve to a lower enthalpic state, in a process known as physical aging or annealing. The rate of this relaxation process depends on a number of factors, especially both the time scales of interest and storage

temperature, any exposure to plasticizing agents like water, and the specific composition of the formulation. There are analytical methodologies to quantify the rate of structural relaxation [22, 23], but these can be challenging to interpret, particularly for ASD, where multiple relaxation times or additional sub- T_g relaxation modes may exist that are ultimately responsible for crystallization kinetics. With sufficient molecular mobility, the initially homogeneous ASD can phase separate into drug-rich and polymer-rich regions [24]. Ultimately, this phase separation can lead to crystallization and present a risk to product performance.

For multicomponent SDD systems consisting of a polymer and drug, it is possible to estimate the glass transition temperature of the composite ($T_{g,SD}$) from the known T_g 's of the individual components. One approach based on polymer free volume theory was developed by Gordon and Taylor and summarized by the Gordon–Taylor (GT) equation (shown here for two components but can be generalized to more):

$$T_{g,SD} = \frac{w_1 T_{g1} + K w_2 T_{g2}}{w_1 + K w_2} \quad (12.1)$$

where

w_1 and w_2 are the weight fractions of each component with the subscripts 1 and 2 representing the components with the lowest and highest T_g , respectively.

K is a constant approximated as

$$K = \frac{\rho_1 T_{g1}}{\rho_2 T_{g2}} \quad (12.2)$$

where

ρ_1 and ρ_2 are the densities of the amorphous components [25].

The GT equations (12.1) and (12.2) generally exhibits good agreement with the experimental composite T_g value for many SDD systems and is also equally applicable to polymer–solvent systems [26–28]. As the relationship was derived assuming ideal mixing between the components, experimental deviations from predicted values due to drug–polymer interactions that affect miscibility are not unexpected [29]. However, not all SDD systems behave similarly with respect to the GT equation, regardless of the extent of phase homogeneity. Specifically, excellent agreement to experimental results does not necessarily imply that the two components are completely miscible nor is the magnitude of deviation between data and prediction a reliable indicator of the strength of drug–polymer interactions [30]. Other similar relationships to approximate the composite T_g were also developed soon after, including one from Couchman

and Karasz [31] based on a thermodynamic approach or one from Fox that focused on components having the same density [32]. Despite the varying analytical complexity of some of these approaches, none of these relationships are universally successful in predicting the composite glass transition.

The desired goal of ensuring SDD physical stability is to maintain the system in an amorphous state, which ensures negligible diffusion and minimizes the risk of crystallization nucleation or growth. In effect, this translates to maximizing the temperature difference between the SDD T_g and the storage temperature [33]. However, this is not the sole factor in governing the stability, as the inherent crystallization tendency dictated by the particular polymer and drug is also an important property. For example, in an acetaminophen SDD, the crystallization rate depended on whether polyacrylic acid (PAA) or PVP was the polymer despite both formulations having similar T_g 's, indicating the polymers have different plasticizing behaviors [34]. Furthermore, although molecular mobility is decreased below the T_g , it is not completely absent as other types of secondary relaxation (i.e. β relaxations) can occur locally on small length scales, such as rotations of side chains [35, 36]. In some glassy materials, these secondary relaxations have been shown to affect the crystal growth rate [37]. In general, the ability to predict the kinetics underlying SDD physical stability is quite challenging and remains an area of active research. To emphasize the point that the T_g is not necessarily the sole determinant of physical stability, one study of an amorphous material indicated that crystal nucleation occurred 55 °C below T_g [38].

For clinical and commercial manufacture, proper manufacturing and packaging controls rely on determining the physical stability envelope for the SDD. Because most SDDs are hygroscopic due to the presence of a hydrophilic polymer, relative humidity (%RH) is an important factor since water acts as a plasticizer, decreasing the T_g , hence enhancing structural mobility [26]. The relationship between T_g and %RH is useful to predict SDD behavior at various storage conditions, with a typical correlation shown in Figure 12.4. Since predicting the physical stability envelope or the kinetics of crystallization remains a challenge, real-time stability mapping studies exposing SDD to accelerated stability conditions are critical to verify the environmental conditions that ensure physical stability. The physical stability of the SDD is not necessarily solely determined by environmental conditions as further processing, for example, by adding additional excipients or undergoing compression, can also cause subtle changes to either crystalline nucleation or growth [39, 40]. Moreover, in cases where the risk of crystallization is low, either due to the lack of a thermodynamically stable crystalline form or a weak crystallizer, other physical changes are likely to occur when the SDD is stored at temperatures near the T_g , including particle deformation or densification and increased particle “stickiness.” These changes may not

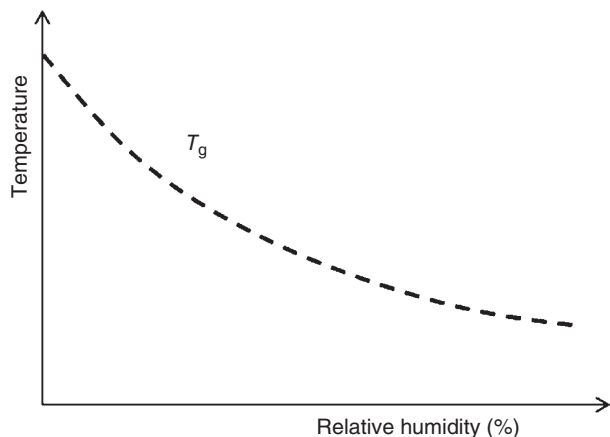


FIGURE 12.4 Schematic indicating the typical glass transition temperature (T_g) for an SDD as a function of storage %RH.

directly affect bioperformance, but can still negatively impact downstream processing due to difficulties arising from poor flow and compactibility.

12.3.3 Spray-Dried Solid Dispersions in Oral Solid Dosage

Spray drying and other technologies to create ASDs have great potential to increase the bioavailability of poorly soluble drugs and drug candidates, although the integration of spray-dried material into tablet formulations may pose a number of challenges. Due to the need for a large portion of stabilizing polymer and low active composition in most SDDs, as well as additional formulation excipients besides the SDD (which may often be more than 50% of the overall composition), the final oral solid dosage forms tend to be large. In most, but not all, cases, SDDs are poor flowing and cohesive, highly compressible, of low density, and sensitive to the effects of temperature and moisture in the manufacturing environment [41].

The poor flowing and extremely cohesive nature of SDD leaves downstream manufacturing processes vulnerable to a number of risks including assay loss due to material hang-up in equipment and poor control of weight and hardness in compression. Additionally, the hygroscopic nature of SDDs may require the implementation of enhanced temperature and humidity controls in the tableting facility to maintain the physical stability of the product.

Improvement in the flow of SDD and SDD blends is often bounded by the need to maintain sufficient compressibility to generate resilient tablets during the compression processes. Denser and less cohesive SDD can be generated by operating the spray drying process at conditions that decrease the rate of solvent evaporation. Similarly, SDDs are often blended with excipients and subjected to a dry granulation process such as roller compaction. Manipulating the SDD conditions or

creating SDD-based granulation will improve processability, but both would be expected to decrease tablet compressibility and potentially affect quality attributes such as dissolution and disintegration. Some SDD formulations may also show a decrease in physical stability following the application of compressive stress in either a dry granulation or compression process. Opportunities exist to mitigate most of the risks associated with the use of SDD in tablet formulations, although it is essential to optimize processability, physical stability, and tablet performance in parallel during the drug development process.

Given the decreased physical stability of SDDs upon exposure to moisture, wet processing unit operations (such as fluid bed granulation) are typically avoided. However, market conditions, patient needs, and product sensitivities often drive the need to apply a film coat to solid dosage forms. The intentional application of aqueous solutions and increased temperature during film coating pose a risk to the physical stability of SDDs. Film coating process development should seek to identify process conditions that produce elegant tablets without approaching the T_g of the SDD.

12.4 OVERVIEW OF SPRAY DRYING UNIT OPERATIONS

The spray drying process, while described as a single process step, is actually a sequence of several unit operations [3]. Here, it has been broken into five major steps; an overview of all steps has been provided in Figure 12.5.

12.4.1 Solvent and Solution Preparation

This chapter primarily focuses on ASDs. Under these circumstances the feed to the spray dryer is usually a solution and not a suspension. To begin the spray drying process, the solid components of the spray-dried product (active ingredient, polymers, surfactants, etc.) are introduced to the process solvents and mixed until a homogeneous solution is achieved. The solution can be entirely aqueous; although for most pharmaceutical spray drying applications, poor solubility of the active ingredient in water will drive the use of organic solvents or a combination of water and organic solvents as the medium for the spray drying solution. Numerous variations can exist on methods of agitation and solids addition, which are discussed in the drug substance chapters of this book and elsewhere [42]. The contents of the vessel are typically inerted. If the solution has significant viscosity that results in heating during agitation, or the solution requires an elevated temperature, the tank may be jacketed.

Additionally, a portion of solvents matching the composition of the spray drying solution are normally prepared in a separate tank to facilitate equipment warm-up and the start-up and shutdown of the spray drying operation.

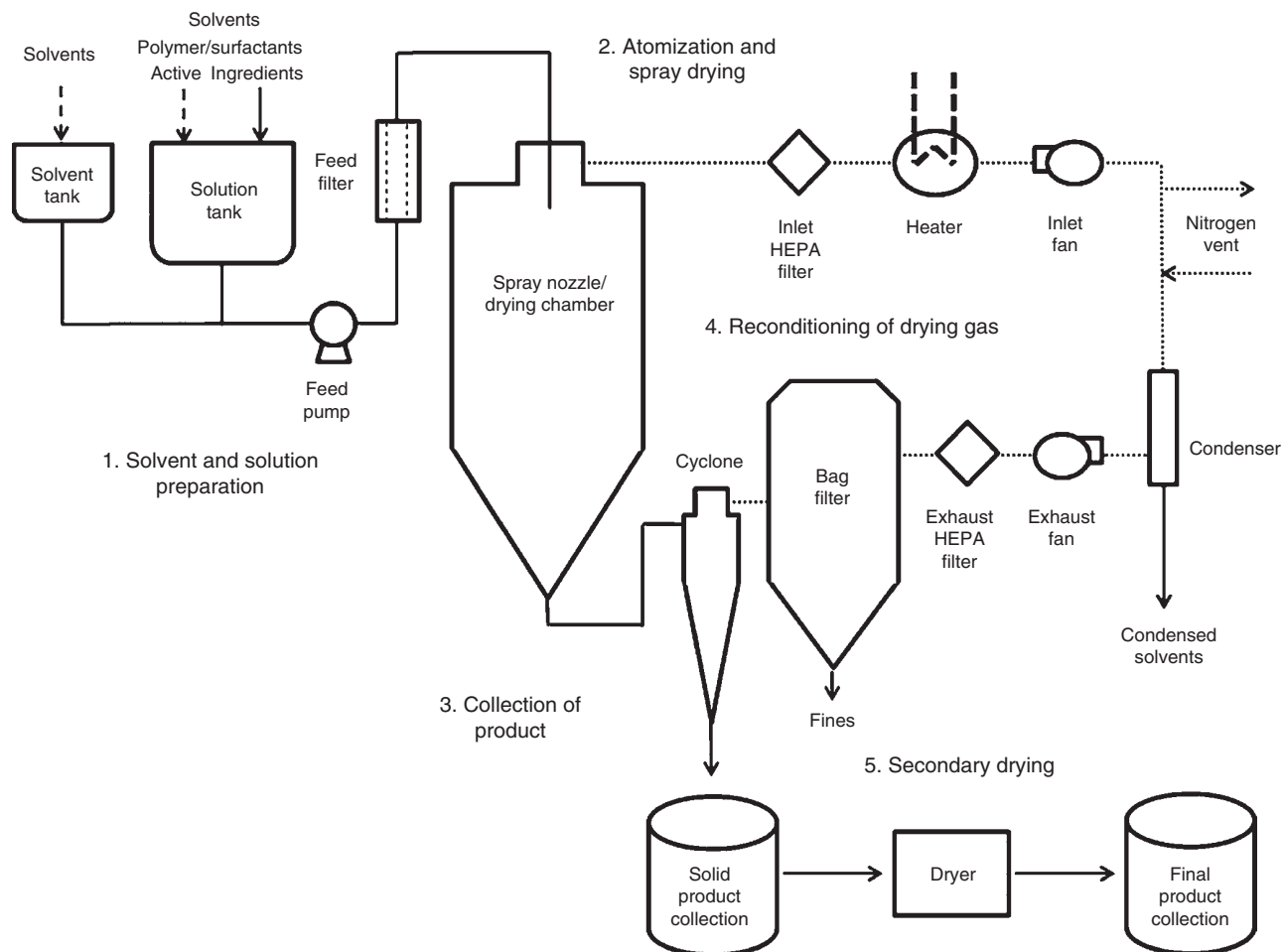


FIGURE 12.5 Spray drying process flow diagram.

12.4.2 Atomization and Spray Drying

The solution stream is pumped from the tank to a nozzle within the drying chamber, where it is atomized into droplets. The size of the droplets resulting from the atomization step will influence characteristics of the final spray-dried product in conjunction with the kinetics of the solvent evaporation process. Depending on the atomization method and product needs, the solvent and solution may be passed through a filter prior to reaching the spray dryer to prevent any particulates or agglomerates from causing clogging or other disruption to the atomization process.

Atomized particles are mixed with a cocurrent stream of hot drying gas. A gas disperser at the top of the drying chamber ensures that the drying gas mixes efficiently with the droplets efficiently. In a very short period of time, the solvents are evaporated from the droplets, and particles are swept out of the spray drying chamber. A detailed description of the interaction between droplets and the drying medium in the spray drying chamber is provided in Sections 12.6, 12.7, and 12.8.

Common methods for spray drying atomization include pressure nozzles, rotary atomizers, and two-fluid nozzles. Other atomization methods, such as ultrasonic nozzles, are also used in research applications [43].

12.4.2.1 Pressure-Swirl Nozzle Liquid is forced through a small orifice to break the solution stream into a hollow cone of droplets and typically requires the use of a high-pressure pump to deliver solution to the nozzle. The feed pressure to the nozzle is adjusted to influence droplet size, and this atomization method generates narrow particle size distributions (PSD). Use of pressure nozzles is usually limited to solutions with low to moderate solids content and viscosity. Solutions with high viscosity and solids content may fail to pass through a small orifice or atomize effectively.

12.4.2.2 Rotary Atomizer Liquid is atomized as it contacts a disk spinning at a high rate of speed. Droplet size can be manipulated by changing the size and velocity of the rotating disk. This atomization approach works best for

highly viscous solutions or slurries; however, it is less adaptable for a multiproduct facility.

12.4.2.3 Two-Fluid (Pneumatic) Nozzle The liquid is atomized by a cocurrent stream of compressed gas. Both the liquid and gas streams can be maintained at a relatively low pressure, and the particle size can be adjusted by changing the liquid to gas ratio in the nozzle. However, the flow rates that can be atomized by this methods are limited to small- and mid-sized spray dryers.

12.4.3 Collection of Product

Following spray drying, particles are normally separated from the gas stream using a cyclone. The individual cyclone design will dictate the separation efficiency and will affect the PSD of the final product as more or less fine particles created during the spray step are collected from the gas stream [44, 45].

The solid particles discharged from the cyclone are collected in bags or other intermediate containers suitable to protect the powder from the facility environment or are collected directly into a secondary dryer.

After separation in the cyclone, the drying gas is passed through a bag filter and a series of HEPA filters to reduce the presence of particles not captured by the cyclone. The material collected by the bag filter is usually treated as waste, and this additional solids filtration step helps to reduce the incidence of HEPA filter blockage or performance loss.

12.4.4 Reconditioning of Drying Gas

In most pharmaceutical ASD applications of spray drying, the use of organic solvents requires the use of an inert drying gas, like nitrogen, for safety reasons. Pilot- and commercial-scale pharmaceutical spray drying systems require large drying gas flow rates and are normally set up in a closed loop so that the drying gas may be recycled and operating costs minimized; smaller units or aqueous systems that use air as the drying gas may operate as an open system (e.g. single pass of the drying gas, without recycling).

Following separation of the solution from the drying gas, the recycled gas stream is passed through a condenser to reduce the level of residual solvent carried by the gas. The drying gas is then reheated to the targeted temperature before being recycled to the spray dryer. Small units may use only one fan to circulate the drying gas through the system, while larger systems may require the use of multiple fans in a “push–pull” design to circulate the drying gas effectively. A vent in the gas recycling system allows small quantities of nitrogen to be vented or fresh nitrogen reintroduced in order to maintain constant pressure and flow rate within the system.

12.4.5 Secondary Drying

If the residual solvent content of the product is above the prescribed limits following spray drying, the product will undergo a secondary drying step to decrease the residual solvent levels further. Multiple options exist to complete secondary drying for spray-dried products, but tray, rotary, and agitated dryers are common. All typically take place at low temperatures and under vacuum. Tray drying often requires extensive manual powder handling and is normally unsuitable for products with large production volumes. Rotary drying may result in agglomerated balls of SDD. Agitated drying is likely to damage fragile particles and cause product densification due to the mechanical shear imposed on the product. Drying temperature is the main process control parameter for all drying systems mentioned, although batch size, agitation rate, vacuum, and nitrogen sweep may impact drying performance depending on the mass transfer limitations of the drying process.

Following secondary drying, the final product is packaged in appropriate containers for storage and shipment to the point of tablet formulation. Care should be taken when choosing packaging and designing supply chains to adequately protect stored material from moisture absorption and extreme temperature to ensure the physical stability of the product.

12.4.6 Variations for Specific Applications

There are a number of specialized applications where the spray dryer design requires customization. For handling APIs with high potency (i.e. very low occupational exposure limits), the spray dryer installation is designed to ensure protection of the operators during operation, packaging, and cleaning (such as by operation under negative pressure relative to the processing room). For aseptic manufacturing, the design solutions allow a complete sterilization of the surfaces in contact with the drug product and ensure protection of the drug product to microbiological contamination during operation and packaging. There are also other specialized units that allow the combination of spray drying with fluid bed granulation, commonly known as fluidized spray dryers. In these dryers a fluidized bed chamber is attached to the outlet of the spray drying chamber, leading to a continuous granulation of the spray-dried material. As a last example, spray dryers have also been modified for use in spray congealing applications where the cocurrent nitrogen stream is cooled in order to congeal the melted feed flow that is atomized in the nozzle.

12.4.7 Process Analytical Technologies (PAT)

Continuous monitoring of the material attributes during a spray drying operation is typically carried out to assess if

the material properties are within the target ranges and assess the existence of trends or deviations from normal behavior during the process. A typical in-process control scheme for a spray drying operation includes particle size and density for an assessment of bulk properties of the material and gas chromatography for determination of residual solvents. For evaluation of the solid-state properties during the process, additional techniques such as X-ray powder diffraction (XRPD) or differential scanning calorimetry (DSC) may also be used.

The in-process control will, in many cases, translate into a significant analytical burden. On top of this, offline analysis may be of limited value in actual control or troubleshooting of the process. Analysis with long lead times, such as XRPD and HPLC, will have a delay of hours between the sample collection and the actual result. Results from analytical methods with shorter lead times, such as laser diffraction, will generally be offset by over one hour from the actual process time point due to sampling, transport, and analysis time, rendering any type of feedback control inefficient.

Taking into account the analytical burden, continuous nature of the spray drying process, and the often complex analytical methods required for characterization of ASDs, the use of in-line process analytical technologies (PAT) can bring a number of advantages for monitoring spray drying processes, including:

- Evaluation of process dynamics (stabilization times, process stability, shutdown procedures).
- Real-time measurements and feedback control.
- Process troubleshooting.
- Reduction of analytical burden.
- Elimination of sampling bias.

Three different PAT applications for spray drying processes are highlighted here: near-infrared spectroscopy (NIRS), laser diffraction, and multivariate analysis; however, the field is rapidly developing, and numerous other tools could also be applied.

12.4.7.1 Near-Infrared Spectroscopy The use of NIRS in the pharmaceutical industry is well established, particularly in secondary manufacturing [46]. The possibility of remote measurements (up to several meters), along with the physical and chemical information obtained with NIR, opens up interesting opportunities for monitoring spray drying processes [47]. Using appropriate chemometric modeling for correlation with reference analytical methods, an NIR spectrum may provide information on the bulk properties of the material (particle size and bulk density), as well as the level of residual solvents and the physical state of the solid dispersion [48]. Using an appropriate process interface, it is possible to generate real-time information on all these

attributes with a frequency that allows feedback control and immediate actions to be taken in case of deviation from the target material properties. Using multiple channels or a multiplexer, a single NIR spectrometer can also be used to monitor additional unit operations, including the solution preparation and secondary drying steps.

12.4.7.2 Laser Diffraction The PSD is one of the quality attributes typically monitored on pharmaceutical spray drying processes, as modest variations in the droplet size due to nozzle atomization issues may not be readily observable visually or from other process parameters. Traditional offline analysis relies on determining the distribution using lab-based laser diffraction equipment. With an appropriate process interface, the PSD of the material can be determined in real time without any sampling through the use of a process laser diffraction equipment [49]. The laser diffraction equipment can be interfaced with a spray drying process in an in-line or online configuration (Figure 12.6). The in-line configuration relies on the use of a process interface that diverts a fraction of the material being collected for measurement, which is then recirculated into the system. The online configuration will effectively measure the particle size of the material directly at the outlet of the spray drying chamber. This technology allows a continuous real-time evaluation of particle size, although there can be challenges in achieving a constant and adequate signal depending on solids throughput and adhesion of material to equipment walls.

12.4.7.3 Multivariate Analysis Monitoring of spray drying parameters in a routine way is complex, and a univariate approach provides limited information given the multivariate nature of the process (interrelation between all the variables) and the number of variables that should be registered. Multivariate data analysis allows the monitoring of the relationship between all the important process parameters [50]. The use of multivariate statistics for a set of variables, identified during development, as critical/relevant enhances the monitoring and fault detection capabilities for spray drying monitoring [51]. Additionally, the use of regression analysis allows establishing empirical relations between multiple process parameters with both process performance and material properties, creating a valuable tool for optimization and troubleshooting.

12.4.7.4 Other PAT Other PAT may also be used in addition to the ones mentioned above. Simple univariate sensors, including viscometers, turbidimeters, or UV spectrometers, can be used to monitor the feed solution preparation step and also assess the solution attributes over time. More advanced focused beam reflectance measurement (FBRM) units can be used in-line or online to evaluate suspension behavior and ensure uniform material attributes over time. For the post-drying step, both NIR for powder and LC-MS

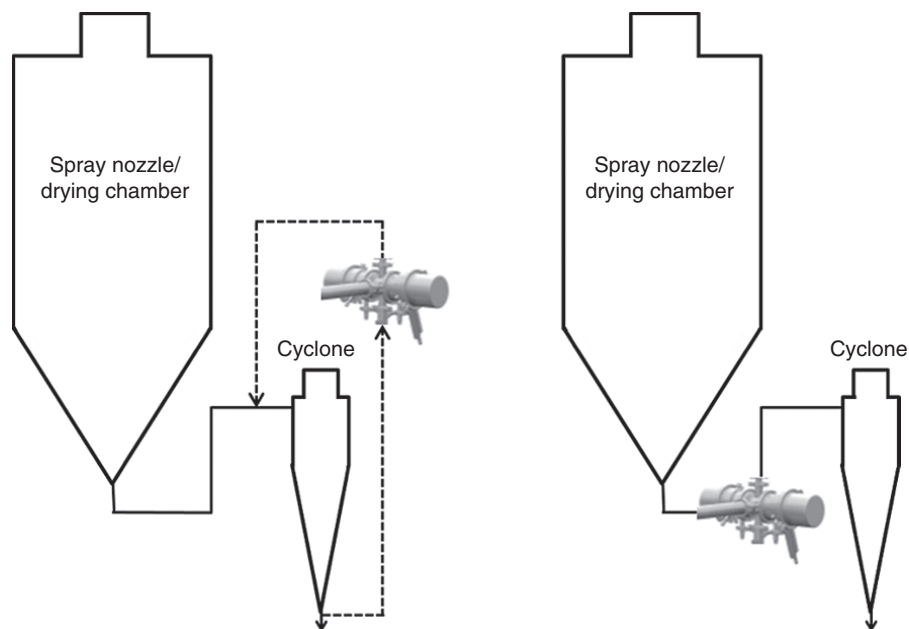


FIGURE 12.6 Representation of online (left) and in-line (right) interface of a process laser diffraction equipment with a spray drying unit.

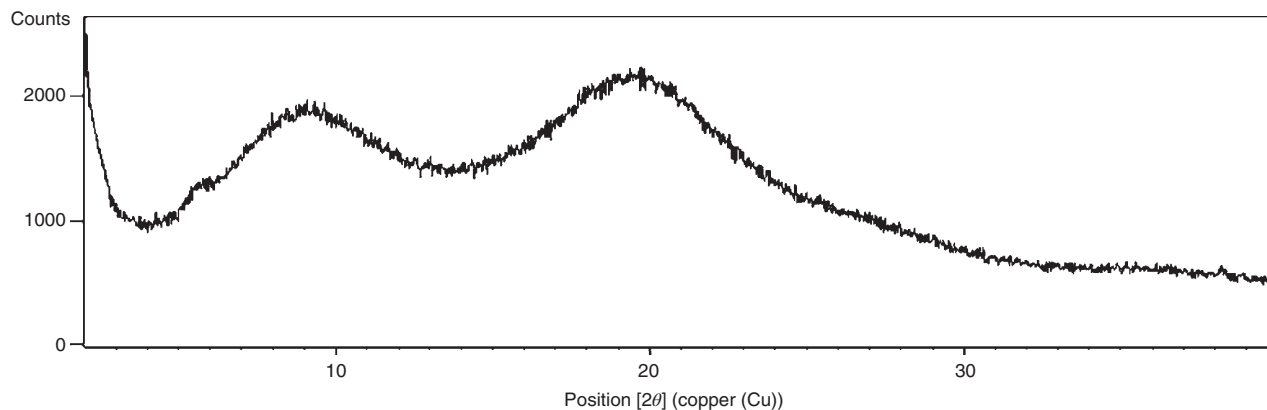


FIGURE 12.7 Example XRPD data of an ASD indicating the absence of sharp diffraction peaks.

for exhaust gas analysis can track the drying kinetics and determine drying end point in real time, reducing process cycle time, analytical burden, and heat stress on the material.

12.5 OVERVIEW OF SDD PARTICLE CHARACTERIZATION

12.5.1 Physical Form

The primary concern and focus with respect to monitoring physical stability of SDD is the inherent risk of recrystallization. An extensive array of solid-state methods exists for the detection of low levels of crystallinity in an amorphous matrix. Three commonly used techniques for this purpose are outlined in this section.

12.5.1.1 X-Ray Powder Diffraction (XRPD) The XRPD technique provides information regarding the presence of long-range order between molecular constituents consistent with crystalline material. Ideally, ASD systems lack this long-range order, giving rise to a diffraction pattern lacking well-defined peaks, but may contain short-range order owing to polymer structure and intermolecular interactions. A representative example of an XRPD pattern of an ASD is shown in Figure 12.7. In general, the technique is useful in providing quantitative information regarding the degree of crystallinity and distinguishing between crystal polymorphs. More recent work has utilized the scattering data collected by advanced X-ray sources to allow for atomic pair distribution function (PDF) analysis [52] to distinguish between a true amorphous dispersion and either an

amorphous systems with domains or one with nanocrystallinity [53].

12.5.1.2 Spectroscopy Vibrational spectroscopic techniques, such as FTIR and Raman, are frequently employed to monitor the presence of crystal polymorphs and can exhibit superior detection of low levels of crystalline content compared with XRPD or DSC [54]. Differences between amorphous phases generated by different approaches or formed by different cooling rates can be distinguished [55–57]. Another advantage of this technique is that it can be installed in a production environment and can be automated at-line or in-line more straightforwardly than many other analytical techniques [58]. The use of Raman spectral imaging or mapping provides detailed chemical images of the surface of the solid dispersion based on the sample's Raman spectrum and provides information on the spatial distribution and concentration of different components [59].

12.5.1.3 Thermal Analysis The thermal profile of an ASD is useful not only to reveal any potential melting events originating from a crystalline phase but is sensitive to subtle changes in the amorphous phase [30]. Although many instruments are applicable to use for this purpose, including dynamic mechanical analysis (DMA) or thermally stimulated depolarization current (TSDC), the most common is the DSC. These techniques are appropriate to monitor crystallization events, glass transitions, and relaxation events.

The DSC technique provides a way to probe the amorphous state via directly measuring the T_g . Because the T_g depends very much on method parameters like the heating rate and prior thermal history, both the temperature range associated with the glass transition and any single temperature are often used to define it (e.g. onset, midpoint, or end point) are not solely dependent on the formulation. The use of a non-monotonic heating rate, either using modulation, alternating heating and isothermal steps, or stochastically changing temperature pulses, can separate the heat flow into reversing and non-reversing components, thus more easily enabling the identification of glass transitions from thermal events associated with irreversible processes, like evaporation [60].

Other than recrystallization, phase separation into two or more amorphous phases, often a drug-rich and a drug-poor phase [61], is, in principle, detectable by thermal analysis. One cause of this amorphous–amorphous phase separation is due to the drug having low miscibility with the polymer carrier. Phase separation can exist in a range of metastability states, including a kinetically stable state due to the lack of possible crystalline forms or other kinetic barriers to achieving crystallization, or as a truly metastable state approaching the more thermodynamically stable crystalline phase. Each formulation will need to be assessed for this potential failure

mode based on an assessment of known crystal polymorphs and the kinetics of crystallization.

In practice, the existence of two (or more) separate glassy phases by thermal techniques relies on demonstrating the existence of two (or more) separate glass transitions. While the presence of two T_g 's is effectively incontrovertible evidence of phase separation, it is not considered a necessary condition as systems have shown to be phase separated (by microscopy) yet still indicate a single glass transition. A precursor of the splitting into multiple glass transitions is the broadening of the single glass transition as the initial phase becomes more heterogeneous although other factors can also explain this behavior. Barring any additional evidence, a small shift or narrowing of the glass temperature range is not by itself suggestive of a reliable signature of phase separation.

A few of the experimental limitations related to thermal analysis techniques are highlighted here. For example, if amorphous–amorphous phase separation exists but the two amorphous phases have similar or overlapping glass transitions, it will be very challenging to resolve them. Also, the T_g of an amorphous phase with very small domain sizes (<50 nm) will not necessarily be detected as the sample may change during the analysis. Partly as a result of these caveats, the appearance of a single glass transition is only a necessary, but not sufficient, condition to claim the SDD is a homogeneous amorphous phase at the molecular level. Furthermore, not all variability associated with the T_g is solely related to changes in the SDD phase. For example, as previously discussed, the moisture content of the SDD is a significant contributor to the T_g , so not only is it critical that all samples have the same water content prior to testing, but any modification in SDD properties that affect rates of moisture diffusion or evaporation, including a change in surface area or particle density, should be taken into account when interpreting the data.

12.5.2 Particle Morphology and Structure

Along with physical form, the spray drying process impacts physical particle properties like particle size and morphology directly, which in turn are related to derived properties like bulk density and compactibility. These particle attributes are often a useful gauge to assess and monitor performance of the spray dryer. Further development to map out the design space is required to determine which of these attributes, if any, affect downstream unit operations and ultimately any critical quality attributes of the drug product.

12.5.2.1 Particle Size The PSD can be a sensitive indicator of processing conditions, and an accurate, robust method may be helpful to optimize numerous process parameters. The size often directly reflects higher temperatures and faster drying times, resulting in “ballooned” spheres, while lower

temperatures and slower drying times are expected to increase the amount of collapsed “raisin-like” particles as the particles dry under wetter conditions. Typically, the PSD range of SDD is ideal for laser scattering, as values are often found between 10 and 100 μm . The nature of the dispersing mechanism is important to separate any adhering particles in order to measure the primary particle size. Excessive stress either from elevated air pressure using a dry dispersive technique or the use of sonication with a wet dispersive technique can result in particle attrition and breakage during the analysis. Frequently, at least two parameters of the PSD are required to fully characterize the distribution, including a central tendency (M_v , d_{50}) and a measure of the width (d_{10} , d_{90} , span, etc.). Choosing appropriate PSD parameters to track is often due to downstream concerns; for example, if the level of fines is perceived to affect subsequent handling, monitoring changes in d_{10} is recommended.

12.5.2.2 Morphology The particle morphology and surface texture is the result of complex interactions between spray drying process parameters and the composition. Some of the key factors that are known to impact particle shape include the inlet and outlet temperature, flow rate of drying gas, and the rheological properties of the feed solution [62]. Although particle morphology using microscopic techniques like scanning electron microscopy (SEM) is generally qualitative, in certain circumstances, the data may be used to diagnose the physical state of the SDD. For example, the appearance of crystalline domains, if present, are often readily observed in this manner. The particle shape is also fundamental in influencing properties like flow and bulk cohesivity. Images of particle populations enable an assessment of morphology attributes on the bulk powder, including the fraction of inflated, cracked, or shriveled particles and the relative surface smoothness. For example, the detection of features like fused or agglomerated particles may help understand challenges with flow or caking, while non-spherical, elongated particles could indicate incomplete atomization. Higher magnification images obtained on representative single particles often reveal additional information like wall thickness, surface texture (i.e. smooth versus wrinkled), and porosity. Example SEM images of spray-dried particles using different processing conditions are shown in Figure 12.8. Because of the relatively unique particle morphology compared with non-spray-dried components, SDD tends to be readily observable even after blending and mixing with other components.

12.5.2.3 Densitometry The bulk (random loose packing) and tapped (random dense packing) density are convoluted properties derived from both the PSD and morphology. The relative ease of generating this data coupled with its value in providing basic guidance on bulk storage and flowability concerns makes this a standard characterization

method for SDD, particularly since most SDDs exhibit very low bulk density and poor flow. The bulk density frequently exhibits good correlation with spray drying process parameters that are critical for controlling the drying rate, like the outlet temperature. The bulk and tapped density of the neat SDD usually trend with compactibility. In combination with the absolute or true density, the bulk density is used to calculate the overall porosity, P , or solids fraction of the SDD powder via the following equation:

$$P = 1 - \frac{\rho_b}{\rho_t} \quad (12.3)$$

where

ρ_b and ρ_t are the bulk density and true density, respectively.

For some SDD, a more rigorous porosity measurement can be obtained using mercury intrusion porosimetry, which can output the pore size distribution based on the change in intrusion volume of mercury as a function of pressure. This analysis is most beneficial for systems that exhibit more rigidity and often possess a complex microporous structure. However, for more fragile or brittle SDD, the interpretation of the data is fraught with difficulty due to the inability to differentiate between the interparticle and intraparticle void space [63] or the possibility that pore space is generated during the analysis.

Overall, the properties that govern the response of SDD particles to applied stress and whether the particles crack or plastically deform depend on a number of factors including particle shape, PSD, hardness, and surface friction or cohesivity. Since it is difficult to predict bulk behavior from single-particle attributes, bulk powder compaction of the neat SDD or after blending with excipients is a more specific test of the SDD mechanical properties and may be more correlated with downstream performance of the tablet, i.e. tablet hardness or disintegration.

12.5.2.4 Additional Techniques for SDD Physical Characterization Other conventional techniques for solid-state physical properties of powders remain potentially useful for certain circumstances. For example, the specific surface area is likely to be more discriminating than the PSD when the surface texture varies considerably or when there is an increase in the percentage of fines. To ascertain differences in cohesivity, a direct flowability measurement should be considered although the interparticle friction is known to be sensitive to factors like %RH and the presence of static charge in addition to particle attributes like size and shape. Furthermore, because flowability is not an additive property, the relevance of the SDD flowability to the final

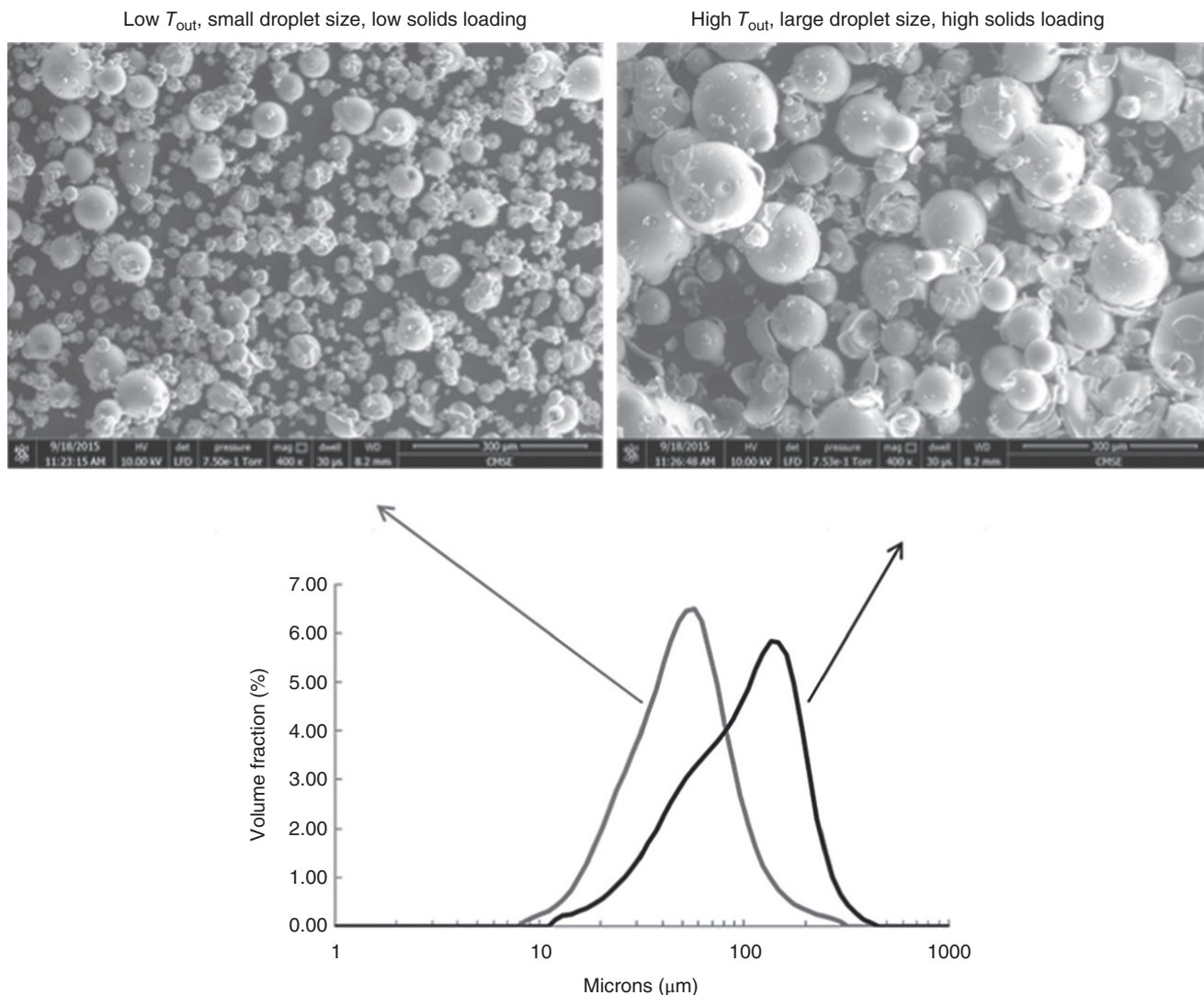


FIGURE 12.8 (Top) Representative ESEM micrographs of two designed experiments at the extremes of the processing range. (Bottom) PSD by laser diffraction of these conditions.

blend may be negligible, especially if the SDD loading is relatively low.

Certainly, a wide expanse of more sophisticated techniques are available to probe smaller length scales to interrogate single-particle properties. For example, X-ray photoelectron spectroscopy (XPS) can provide an elemental map of the surface that can be useful to investigate phase homogeneity especially for SDD containing surfactants. Another technique that provides structural and morphological information about the interior of a single particle is a focused ion beam coupled to a SEM (FIB-SEM) [64]. This instrument can section a particle to allow for imaging of the internal features, including the inner diameter and porosity. An example of a cross-sectioned particle following the FIB treatment is found in Figure 12.9. This technique, while powerful, tends to have very low throughput and usually

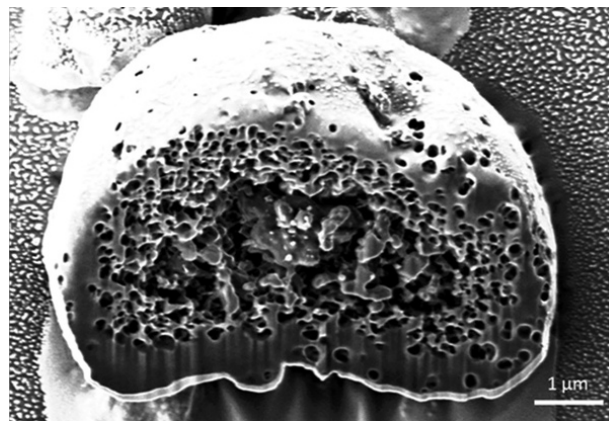


FIGURE 12.9 A cross section of a single SDD particle following FIB-SEM analysis.

requires many measurements to understand typical particle-to-particle variability. As is seen in the SEM images of Figure 12.8, there are a continuum of particle morphologies (fused, inflated, collapsed, etc.) produced by the same spray drying conditions, which may require the analysis of numerous particles to minimize errors due to unrepresentative subsampling.

12.6 THERMODYNAMICS AND OPERATIONAL PROCESS SPACES

In spray drying, after the selection of the solution composition, the drying of a droplet is primarily governed by the conditions in the spray chamber. The temperature and solvent content of the drying gas will determine the rate at which the droplet dries, as well as the temperature the resulting particle will experience. To a first approximation, the average drying conditions in the chamber can be determined from a thermodynamic model that applies mass and energy balances to the spray chamber and thermodynamic equilibrium in the condenser. This system-level model of the spray dryer allows for prediction of the dependent process parameters (outlet temperature, relative saturation, or dew point) from the independent process parameters (such as spray rate and gas inlet temperature). This proactive prediction of the process conditions identifies potential process failure models and can guide the process design for development, scale-up, and efficiency enhancement.

12.6.1 Thermodynamic Modeling

A simple flow sheet of a closed-loop spray drying system can be drawn, with the spray dryer as an ideal mixer, the cyclone as an ideal solids separator, and the condenser as a liquid separator at thermodynamic equilibrium. Conservation equations (mass and energy balances) can be applied to the spray chamber, with two incoming streams: the liquid solution stream and the vapor-phase drying gas stream. The liquid is characterized by the solvent spray rate and composition (and to a lesser extent its temperature). The drying gas stream is characterized by its temperature, flow rate, and residual solvent content (determined by the gas condenser temperature). Conservation equations for the nitrogen, solvent(s), solids, and enthalpy in the system define what the temperature and composition of the outlet stream from the spray chamber will be; these equations are explicitly shown in Ref. [65].

A thermodynamic model for a closed-loop spray dryer with a single solvent (or an open-loop spray dryer with multiple solvents) can be solved explicitly. For closed-loop systems with multiple solvents, the equations have to be solved implicitly, such as with a worksheet, numerical software, or a commercial flow sheet application (e.g. ASPEN^(R)).

Thermodynamic models of the system require several simplifying assumptions of what occurs within the spray chamber. While a later section of this work will discuss how the localized conditions within a spray drying may vary, thermodynamic models do not account for any spatial gradients within the spray chamber. The model also assumes that the material both in the spray chamber and in the condenser are full mixed and the phases are at equilibrium. It also assumes the system is running at a pseudo-steady state, which is a relatively valid assumption since spray dryers typically reach equilibrium in short times (much less than an hour) relative to the processing time (typically several hours). A few additional assumptions are typically made to simplify the calculations, such as neglecting the heat of formation (heat of dissolution) from solids formation. Heat losses from the spray chamber walls are typically not negligible (except in the largest equipment), and values must be experimentally determined for a given equipment installation.

12.6.2 Thermodynamic Models and Operational Spaces

The operating space of a spray drying process is constrained in a first stage by equipment ranges, including gas rate and heater capacity. This space is further constrained by the process itself to ensure smooth operation without product accumulation, condensation of process solvents, or the use of drying conditions that impact material properties. The process range will generally depend on the solvent system, solids concentration in the feed solution, and liquid to gas ratio. Understanding the limitations of the operating space requires an understanding of the thermodynamics of the spray drying process, and thermodynamic modeling of the spray drying operation is fundamental for a thorough characterization of the process, experimental design, scale-up, and troubleshooting. The thermodynamic model will provide the basis of the relation between the most important drying parameters such as feed rate, drying gas rate, inlet and outlet temperature, and also condenser temperature when operating in closed loop. This model will also allow estimating drying conditions such as relative saturation and other psychrometric parameters such as wet bulb temperature and dew point.

The most relevant factors determining the operating space of a spray drying process are discussed below.

12.6.2.1 Relative Saturation When drying water-based feed solutions, determination of relative humidity can easily be accomplished using in-line hygrometers. When drying organic solvent-based feed solutions however, which comprises the vast majority of SDD pharmaceutical applications, the measurement of relative saturation is not straightforward. In both cases, relative saturation can be determined using the partial pressure of the solvent, obtained from a mass balance

of the system, and the saturation pressure, determined from empirical correlations [66]:

$$\text{Relative saturation} = \frac{p_{\text{solvent}}}{p_{\text{sat}}} \quad (12.4)$$

The relative saturation in the drying chamber is an important parameter with a significant influence on material properties and process performance. This parameter will directly impact the drying rate and, therefore, particle density, morphology, residual solvent levels, and in some cases the particle size of the spray-dried material [67]. The level of residual solvents will, in turn, dictate the plasticization effect on the ASD and its respective glass transition temperature.

Low glass transition temperatures observed at high relative saturation conditions may impact process yield due to elevated plasticity and adhesion of material to the dryer walls and transfer lines. Operating under these conditions will also increase the risk of recrystallization in ASD formulations due to elevated molecular mobility.

Decreasing the relative saturation at a given condition may be accomplished by increasing the outlet temperature, reducing the condenser temperature, or reducing the liquid to gas ratio, the latter typically by reducing the feed rate, as the drying gas will generally be fixed at nominal conditions. This will have implications in the process throughput and material properties and will also be constrained by equipment limitations in the case of the drying gas flow and condenser temperature.

12.6.2.2 Dew Point Dew point is the temperature at which the partial pressure of the solvent equals the saturation pressure, thereby causing the onset of condensation. The dew point of the system defines one of operational limitations of a spray drying process; this is especially relevant in water-based processes where the dew point will often approach

room temperature. When the dew point is higher than room temperature, condensation may occur in cold spots in the process train, specifically material collectors, which compromises material quality, and filter bag housings, impacting the manufacturing process due to clogging of the filter. Decreasing the dew point can be accomplished by reducing the feed rate at the cost of throughput; therefore a balance between operationability and productivity must be met when drying at high dew point values.

12.6.2.3 Inlet Temperature The spray drying inlet temperature will dictate, along with the drying gas rate, the energy supplied for evaporation, which dictates the maximum throughput for a given feed composition. Working at low inlet temperatures will generally lead to low outlet temperatures for a given condition, resulting in an increase of the relative saturation. This will limit the amount of solvent that can be evaporated without compromising process performance and increase the drying time. Higher inlet temperatures within the equipment limits would generally be preferred in order to improve productivity and drying efficiency. However, most processes will likely be limited by the resulting outlet temperature, which is generally maintained below boiling point of the solvent system, as drying above this temperature will generate inflated or shattered particles. High inlet temperatures may also induce degradation of material that accumulates near the top of the chamber, thereby compromising material quality and adding an additional constraint. Figure 12.10 illustrates the difference in material properties when drying at high or low temperature.

12.6.2.4 Feed Rate Feed rate is directly linked to process throughput, productivity, and cycle time; therefore maximizing feed rate is often desired. The feed rate is limited by the drying capacity of the dryer and the amount of residual solvent in the gas stream that can be condensed before the

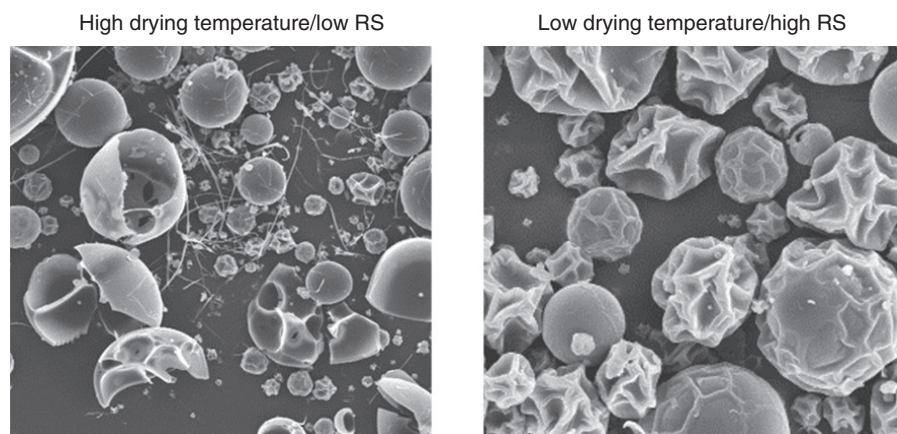


FIGURE 12.10 SEM pictures of spray-dried material dried obtained at fast drying, above boiling point (left), and slow drying (right) conditions.

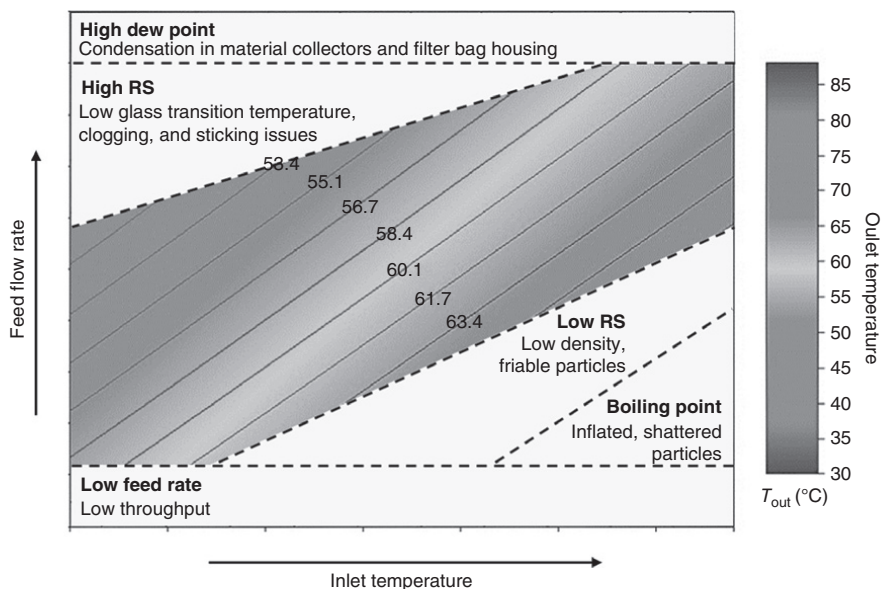


FIGURE 12.11 Spray drying operational space as a function of feed rate and inlet temperature.

relative saturation compromises process or material properties. Increasing feed rate at a constant outlet temperature will require an increase in inlet temperature and will result in higher relative saturations, and consequently the glass transition temperature of the wet SDD is reduced. Feed rate will thus be maximized based on the drying capacity of the unit, without compromising material properties or manufacturability resulting from operating at high relative saturation conditions. The higher feed rates used for large-scale dryers will also require increasing atomization energies; small particle size requirements for specific applications may also limit this parameter.

Figure 12.11 illustrates an operational space based on feed rate and inlet temperature. The failure modes associated with the different combinations of these variables are highlighted. In summary, the operational space of a spray drying operation can generally be described by a combination of feed rate and inlet temperature for a given feed composition. Feed rate may be maximized up to a point where (i) the amount of residual solvents can be properly removed, (ii) dew point is below room temperature, and (iii) the inlet and outlet temperatures are within the equipment capability and do not compromise material properties or process performance.

12.7 TRANSPORT EQUATIONS AND PARTICLE DRYING

In Section 12.6.2, the usefulness of using scale-independent parameters and mechanistic understanding for the support of process scale-up of transfer were illustrated. This allows a good estimate of process parameters and in some cases can

suffice for process scale-up design with at most confirmatory trials at scale. However, there are detailed physical phenomena that are scale and unit dependent that induce variability and require fine-tuning of process parameters to match the attributes of the SDD across scales. In this section the *in silico* approaches such as CFD and discrete element methods are discussed to understand the detailed physical phenomena during scale-up and process development. These approaches can support judicious selection of processing space and reduce the number of experiments required to achieve the desired SDD attributes. The typical challenges are illustrated with a case study of a process scale-up from a pilot unit to a larger unit with twice the chamber height and diameter.

12.7.1 Discussion of Single-Particle Drying

The modeling strategies for incorporating the various physical phenomena occurring during spray drying within the drying chamber are discussed in the following section from consideration of a single particle. Spray drying begins with the atomization of the spray solution into droplets. Following atomization, the resulting droplets are carried along by the flow and are dried due to the interaction with the flow. As the solvent leaves the droplets during drying, particles are formed whose morphology is dictated by the evaporation kinetics. Each of these steps is discussed below.

12.7.1.1 Atomization Scaling up the atomization system is a vital part of the process scale-up since it has a direct influence on the final particle size. The various technologies to atomize the feed solution into droplets have been discussed in Section 12.4. The focus here lies on discussing the physical

phenomena pertaining to pressure-swirl nozzles, which are commonly used for moderate to high throughput spray drying for SDDs. The atomization mechanism of pressure-swirl nozzles lies in the generation of a thin liquid film injected at high velocity into a gas stream. This injected film is geometrically shaped as a hollow cone with an aperture angle (or spray angle) determined by the swirl generated within the nozzle core and the diameter of the discharge orifice. Atomization occurs owing to the natural susceptibility of the liquid film to the Kelvin–Helmholtz instability and to the action of the surface tension, leading to the breakup of the sheet into ligaments and droplets [68, 69]. A recognized physical model to describe pressure-swirl atomization is the so-called Linearized Instability Sheet Atomization (LISA) model [69, 70], which is useful in predicting injection velocity, breakup length (distance for the liquid sheet to be atomized into droplets), and average droplet size. In this model, the velocity, thickness, and aperture angle of the hollow cone originating from the nozzle orifice are calculated using the nozzle geometry and measured pressure drop. A linearized perturbation analysis is then used to predict the most unstable instability wave numbers from which breakup distances and ligament or droplet diameters are calculated [69, 70]. These mechanistic models can guide the selection of the atomization system and elucidate the influence of process parameters and solution properties on the atomization process. In the case study presented below, it is shown that modifying the nozzle geometry (e.g. orifice diameter and core size) to cope with a larger solution feed rate may lead to a similar droplet size distribution and spray angle but different injection velocities and spray densities (i.e. droplet count per unit volume), which in turn influence the secondary breakup and droplet collision and coalescence.

12.7.1.2 Drying Kinetics Atomization creates a distribution of droplets with different initial velocities and sizes. The atomized droplets, depending on their initial conditions and imposed drag forces, follow different trajectories in the spray drying chamber. These differences contribute to differences in their drying kinetics and final physical and chemical characteristics when they are collected at the end of the cyclone as particles. Solvent evaporation from a droplet is driven by difference in the partial pressure of solvent at the droplet surface and that of the surrounding air and controlled by several flow characteristics and material properties. The kinetics of evaporation occurring during the life of the droplet within the dryer involves several time scales [71]:

- The deceleration time scale for the droplet ejected from the nozzle to become entrained in the flow.
- Time scale for temperature stabilization of the droplet when the heat loss due to evaporation is balanced by heat gained from the warmer gas to the droplet.

- Time scale of concentration gradient [72] being set up inside the droplet due to the relative differences in the solvent efflux and the diffusive flux of the dissolved solids within the droplet. This eventually leads to a transition of the droplet surface from a liquid interface to a concentrated gel-like phase called the shell.
- Time scale for morphological transition within the droplet with shell thickening, rehomogenization, pore formation, and structural heterogeneity – during which period the evaporation rate is drastically reduced and governed by the diffusion through the shell.
- Residence time of the droplet in the drying chamber.

Comparing the drying kinetics of a solvent droplet versus a solution droplet (Figure 12.12) informs us that the solution droplet drying largely follows the drying kinetics of the solvent droplets at early times until around shell formation. The early-stage drying kinetics of a solution droplet thus can be obtained by solving the governing equation for heat and mass transfer between the droplet and the surrounding drying gas. Assuming radial symmetry within the droplet and gas-phase processes are steady relative to the droplet, we find that evaporation kinetics are driven by the rate of convective transfer (governed by the Ranz–Marshall criterion) of solvent from the outer boundary of the droplet. Equation (12.5) describes the rate of change in droplet diameter (D) during the early stages of drying ([73]):

$$\frac{dD}{dt} = \frac{-4Md}{D\rho RT} \Delta p (1 + 0.276\text{Re}^{0.5}\text{Sc}^{0.33}) \quad (12.5)$$

where

M is the molecular weight of the solvent.

d is the diffusivity of the solvent in the droplet phase.

ρ is the density.

R is universal gas constant.

T is the temperature at the surface of the droplet.

Δp is the difference in the saturation vapor pressure of the solvent at the droplet surface and the bulk vapor pressure of the solvent.

Re is the Reynolds number of the droplet.

Sc is Schmidt number of the solvent in the gaseous phase.



FIGURE 12.12 Drying kinetics of a solvent droplet versus a solution droplet.

We note that the simple droplet drying model is commonly used in spray dryer scaled simulations. However, several detailed droplet-scale models incorporating effects such as droplet recirculations and surface tension effects have been studied numerically as discussed in George et al. [74].

When a mixture of solvents is used, the interplay between the mutual diffusion coefficient of the solvents and relative heat to mass transfer characteristics of the solvents determines the rate of evaporation. Additionally it also determines the relative concentration of the solvents within the drying droplets [75]. Thus, we find rich and complex transport phenomena that determine the course of evaporation of the droplets.

12.7.1.3 Particle Formation During the later stages of drying, the formation of shell on the droplets results in diffusion-limited evaporation kinetics of the droplet. In this case, diffusion refers to the mutual diffusion coefficient of solvent in the continuously evolution concentration of solution at the droplet surface. The resulting spray-dried particles as seen in Figure 12.12 may have different morphologies, such as blistered particle, shriveled particle, inflated hollow particle, puffed particle, and solid particle with uniform or varying porosity [76]. Several models have been proposed [76] to capture different morphologies but in principle follow two routes after the shell formation, namely, the route of evolving with a shrinking core and an outer shell versus the route of nucleation of a gaseous phase resulting in a hollow core.

In pharmaceutical spray drying processes considered here, the liquid sprayed is usually a solution without suspended solids. Considering the case of a single solute dissolved in a solvent, the evolution of the concentration gradient of the solute in the droplet is governed by convective–diffusive equations. The concentration gradient in the droplet is strongly dependent on the solute Péclet number, which is defined as the ratio of the instantaneous rate of change in droplet diameter and mutual diffusivity of the solute in the solution [72]. At this stage, different assumptions for surface concentration at which shell is formed and progression of crust thickness relative to the core [76, 77] or introduction of an internal bubble [78] are used to define the ultimate morphology of shell-like particles. These assumptions provide parameters that are used to fit to experimental data such as temporal variation of droplet diameter. Others have used simple empirical models transcending a wide temperature spectrum [79] or building characteristic drying curves using the reaction engineering approach [80] based on experimental data. Since the complexity of the model increases with the increase in the number of solutes and solvents present in the spray solution, the utility of empirical models for predicting final particle morphology is evident.

12.7.2 CFD Approaches for System-Wide Drying

In the previous paragraphs we summarized the physics underlying how the feed solution is atomized into droplets and how a droplet dries, forming first a shell that mechanically deforms until the final particle is obtained. In the spray dryer, each droplet will experience varying initial size, drying history, different temperature, and solvent concentration in the drying gas. The droplets can also interact with each other via their proximity, collisions, and coalescence. In order to simulate this additional complexity of interaction between drying gas and the spray, CFD is used. For the gas flow, the standard numerical codes integrate the ensemble-averaged governing equations (commonly known as Reynolds-averaged Navier–Stokes equations), which allow flow unsteadiness but average turbulent fluctuations. This requires modeling of turbulent momentum, mass, and energy transfer. In the example below, k – ϵ models for the continuous phase are used [81].

For the discrete phase, fluid parcels representing bundles of droplets are injected as a hollow cone with a predetermined size distribution and at a predetermined velocity, angle, and distance from the “virtual nozzle” (corresponding to the breakup length) given by the LISA model discussed in Section 12.7.1.1. Alternatively, experimental measurements of the droplet size and velocity could also be employed [82]. The discrete particle model (DPM) uses Lagrangian tracking of particles in the flow field with two-way coupling for exchanging mass, momentum, and energy. The trajectory of the fluid parcels is governed by their inertia and drag relative to the spatially varying gas flow (the droplets are assumed to be spherical) taking into account the solvent evaporation and drying kinetics as described by the Ranz–Marshall model supplemented with a standard phase change model. Turbulent dispersion and coalescence/breakup of the particles are neglected here for the purpose of simplicity.

For illustrative purposes, two spray drying scales are simulated, operating at nominal gas flows of 110 and 750 kg/h, respectively. The drying chambers also differ in diameter and height by a factor of two. The heat transfer coefficient of the walls is matched for the both spray drying scales. For the scale-up, the gas to feed flow ratios are kept at 10 : 1, and the inlet temperature was adjusted to obtain the same drying chamber outlet temperature. For atomization, a 0.40 mm diameter nozzle with a core size/geometry yielding a spray angle of 53° was used for the smaller-scale process (feed flow rate of 11 kg/h). For the larger unit operating with a feed flow rate of 75 kg/h, in order to maintain a similar average droplet size and spray angle, a nozzle with a diameter of 0.78 mm and matching core size was used. For inputs to the coupled CFD–DPM calculation, the LISA model described earlier is used to evaluate the spray characteristics, namely, the breakup length, injection velocity, and droplet size distribution, and whereas the droplet size distribution is matched, the injection

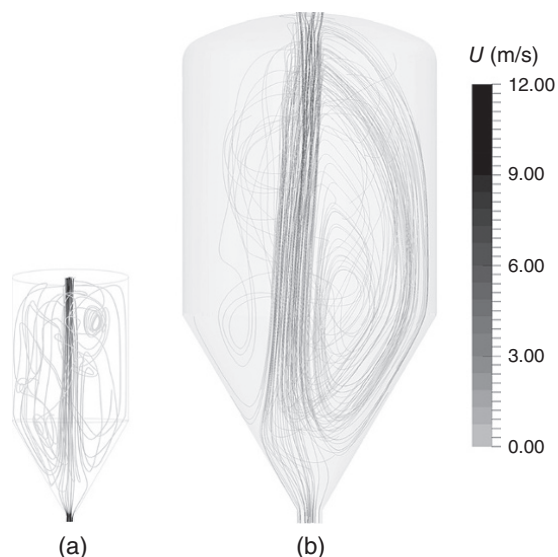


FIGURE 12.13 Comparison of the velocity profiles in the two spray dryer scales. Streamlines colored with velocity magnitude U starting from the chamber inlet of the spray dryer and showing the central drying gas jet surrounded by low velocity flow recirculation. (a) Small scale. (b) Large scale.

velocity presents a twofold increase and the breakup length a fourfold decrease between the small- and larger-scale dryer.

The results of the simulations are discussed below. In all figures, except where indicated, the gas flow and spray are present even if purposefully not shown for clarity.

12.7.2.1 Gas Profile In Figure 12.13, a typical gas flow pattern in the drying chambers is shown. In this characteristic spray drying geometry, the drying chamber diameter is much larger than the diameter of the gas disperser outlet, and consequently most of the flow lies on a central hot gas jet. The momentum diffusion from the central jet into the surrounding gas in the drying chamber induces its motion in a pattern usually denoted as recirculation. The velocity of the drying gas jet dictates, to a first approximation, the time the droplets have to dry prior to exiting the drying chamber (also known as residence time). The velocity in the surrounding recirculation zone is about an order of magnitude smaller and in the opposite direction and thus can transport droplets and particles upward. For a given spray dryer model from a given manufacturer, the design across the different scales is usually close to being geometrically similar, which leads to comparable gas velocity profiles and residence times proportional to the chamber length.

12.7.2.2 Temperature Profile The corresponding temperature profile is shown to illustrate that the recirculation zone is typically at lower temperature than the central hot jet and closer to the outlet SD temperature (Figure 12.14).

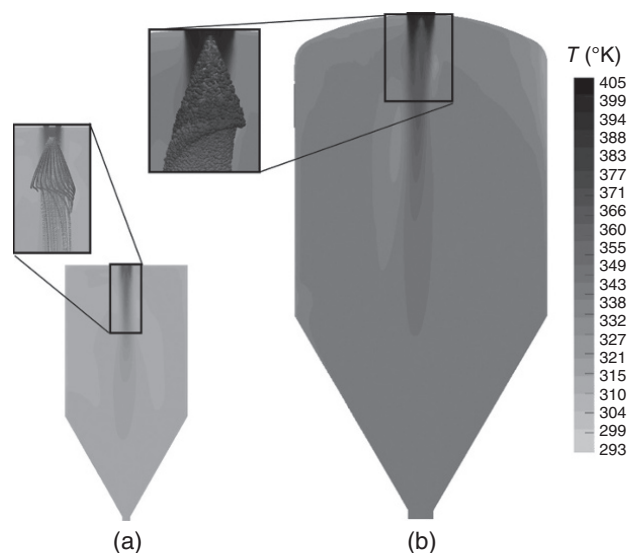


FIGURE 12.14 Comparison of the temperature profiles (K) in the two spray dryer scales ($y = 0$ plane). The temperature contours with the spray droplets overlaid are shown in the insets. (a) Small scale. (b) Large scale.

This highlights why spray droplets escaping the central hot jet have slower drying kinetics and can lead to deposits at the top of the drying chamber near the nozzle itself, which is a common challenge encountered when developing a spray drying process. In the insets of Figure 12.14, the temperature profile near the injection zone, caused by the interaction of the evaporation kinetics of the spray and dynamics of the drying gas jet, is highlighted. This can be scale dependent as it depends on the diameter of the drying gas jet as well as the injection velocity and spray density, as discussed in the beginning of the section.

To a first approximation, the temperature profile is similar across scales, just like the velocity profile. However, a finer analysis suggests that in the larger scale the combination of a cooler temperature profile in the spray plume plus the faster droplet velocity results in a cooler/slower drying process. Conversely, in the recirculation region, the smaller scales exhibit lower temperatures due to the larger heat losses per unit volume, leading to cooler/slower drying process for the fraction of droplets that escape the central hot jet.

12.7.2.3 Solvent Evaporation Profile The impact of the factors discussed in Section 12.7.2.2 on the evaporation kinetics is also observable on the spatial distribution of the evaporated solvent in the gas as shown in Figure 12.15. Although a mass balance suffices to determine the fraction of solvents in the drying gas as it exits the chamber, the case study at hand allows an illustration of more subtle effects caused by scale and the interaction between the spray and the drying gas. Noticeably, the solvent fraction of the recirculating gas for the smaller scale is larger than that of the

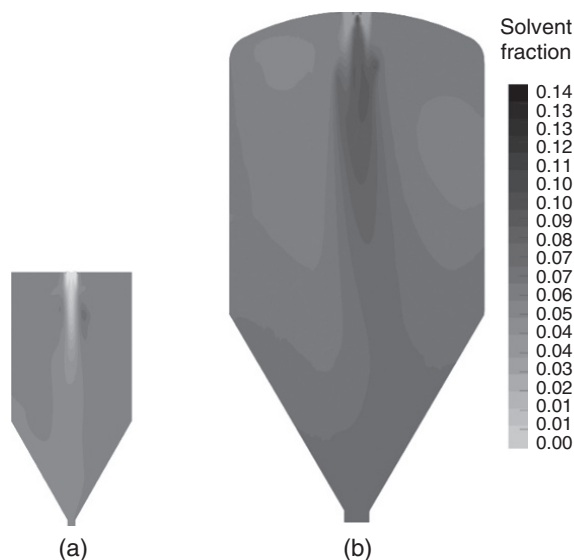


FIGURE 12.15 Comparison of the solvent fraction profiles (w/w) in the drying gas flow for the two spray dryer scales ($y = 0$) plane. (a) Small scale. (b) Large scale.

larger scale, which is likely a consequence of a larger fraction of droplets drying outside the thinner hot jet in the smaller dryer. This fraction of droplets experiences very different drying rates than those drying in the central hot jet, which directly impacts their final size and density, as discussed in Section 12.7.1. Consequently, this fraction of dried particles can contribute to a wider span in the overall size distribution or even lead to bimodal distributions. Note that a similar variation in the solvent's fraction of the recirculating gas could have been caused by a change in the spray angle, droplet size, or spray velocity even for the same scale, which emphasizes the need for a careful scale-up to maintain product attributes across scales as similar as possible.

12.7.2.4 Spray Plume The length and width of the spray plume for the two scales, as well as the size evolution of the droplets during drying, is shown in Figure 12.16. The two plumes are similar owing to the fact that similar droplet size distributions were used and the gas distributors yield comparable drying gas jet velocities entering the drying chamber. However, variations in the choice of processing conditions (e.g. increased gas flow rates compared to the nominal setting in order to increase throughput) and geometrical differences between gas distributors can have an impact on the spray plume and on the drying kinetics.

These simulations allow a qualitative illustration of the challenges and compromises that need to be overcome during process scale-up, but these are, by no means, comprehensive. Product buildup on the equipment surfaces [83, 84], hotter wall regions in the spray dryer, and risk to product quality and collection efficiency are some of the examples where

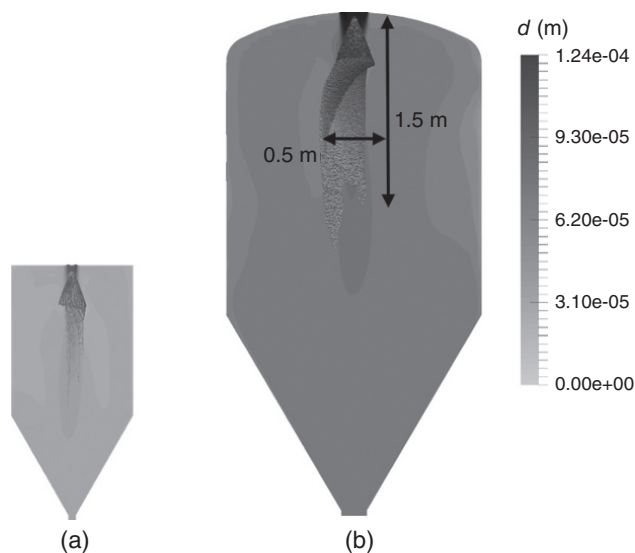


FIGURE 12.16 Comparison of the spray plume for the two spray dryer scales. The droplets are colored by their diameter (m) overlaid to the temperature profiles ($^{\circ}K$) presented in Figure 12.14. (a) Small scale. (b) Large scale.

computational modeling has also a great potential to streamline process development and facilitate quality to be inbuilt on the process design.

12.8 PROCESS DESIGN AND QUALITY BY DESIGN

When a spray drying process is transferred between different equipment scales, the intention is to maintain the drying profile for any individual droplet such that it will form an equivalent particle. This includes maintaining the composition and size of the droplet, the rate at which solvent is evaporated from its surface, the temperature it reaches during drying, and the duration of the drying process. The size of the droplets (as well as the droplet size distribution) is determined by the selection of the nozzle, the feed rate, and (for two-fluid nozzle) the atomization gas pressure. There are several approaches to measuring or estimating the droplet size, which range from empirical correlations to direct laboratory measurement using light diffraction (such as phase Doppler particle analysis). However, at this time, no in-line droplet measurement technique is readily available for spray drying.

Since droplets rapidly entrain in the drying gas, the residence time in the equipment is primarily determined by the equipment design (chamber size, as well as flow pattern created by the gas disperser) and the gas flow rate. Assuming sufficient residence time, the temperature of the particle will equilibrate with the process gas at approximately the outlet temperature (or slightly cooler, due to small levels of evaporative cooling after the particle shell forms). Therefore, the

outlet temperature is maintained between scales in order to maintain the heat flux into the particle.

As discussed in Sections 12.6 and 12.7, the rate of solvent evaporation from the droplet/particle surface is impacted by a number of process parameters. While numerous combinations of these parameters can match a single response parameter (such as relative saturation at the outlet or temperature at the spray dryer inlet), it is difficult to maintain all possible responses that might impact the particle drying profile. Therefore, the standard approach to preserving the drying rate is to use a constant condenser temperature, outlet temperature, and liquid to gas flow rate ratio. These terms are independent of equipment scale, and in an ideal system (adiabatic and at equilibrium), conserve the inlet temperature and solvent concentration at the outlet.

However, some equipment will be constrained by its design or qualified ranges, making it impossible to maintain all parameters between scales (such as minimum condenser temperature achievable with the plant utilities or maximum solution flow rate achievable with a pump's capacity). In practice, even if all parameters are achievable, simply maintaining them still may not always be sufficient to maintain particle properties between scales [85]. The droplet sizes achievable vary with scale – it is difficult to supply enough energy to form very small droplets at large flow rates. Conversely, due to equipment dimensions, the shorter residence time at small scale prevents drying of very large droplets. Achieving sufficient residence time for sufficient particle drying to occur can also be a consideration at small scale when processing with slow drying rates. At larger scales, the residence time is often much longer than required for formation of a particle; the additional residence time typically has little impact on the particles beyond slight drying of residual solvent out from the particle interior.

No equipment is perfectly adiabatic, so the thermodynamics in the process will be impacted by the heat losses; these losses tend to be less at larger scales, which have less surface area relative to the gas flow rate. Even subtle differences in instrumentation, such as placement of thermocouples, can result in small changes in the drying rates occurring in the process.

One of the most difficult scale effects to account for is the difference in the mixing between the droplets and the drying gas between different spray dryers (Figure 12.17). This effect is not accounted for by thermodynamic models (which assume complete mixing and equilibrium between the gas and liquid phases), but can be seen through CFD models. There are several sources that affect the air and liquid mixing near the nozzle. Most pharmaceutical spray dryers use a single nozzle within the spray chamber, despite over an order of magnitude difference in the spray rate between laboratory-, pilot-, and commercial-scale dryers. This results in a higher density of droplets in the spray plume at larger scales. Additionally, the type of nozzle (pressure, two-fluid, rotary, etc.)

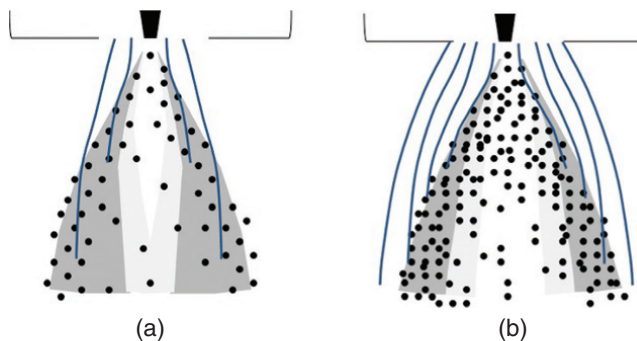


FIGURE 12.17 Schematic of differences in droplet and drying gas mixing at different scales. (a) Smaller-scale spray/gas mixing. (b) Large-scale spray/gas mixing.

and the design of the gas disperser will impact how readily the drying gas can mix with the spray plume. Even after the inlet gas and the spray plume mix, the chamber shape and gas flow rate will determine the overall flow pattern within the chamber. These spatial differences within the chamber mean that particles will experience different drying profiles, temperature profiles, and residence times depending on where they are located within the flow field.

The impact of these scale effects can vary between products and equipment scales, making it difficult to predict without prior knowledge of the equipment and similar formulations. Mitigation of scale effects may require minor adjustment of process parameters, depending on the extent to which the particle properties are impacted. However, there are some general trends on how spray drying processes are effected as the process scales up, some of which can be favorable and others may pose challenges. Examples of these are presented in Table 12.2.

12.8.1 Process Scale-Up Case Study

The previous sections highlighted various aspects of spray drying process scale-up with a strong emphasis on the manufacture and characterization of ASDs. We complement the discussion with a scale-up case study between a pilot-scale spray dryer with a nominal drying gas flow rate of 400 kg/h and a commercial unit with four times that capacity. At both scales pressure-swirl atomization systems are used.

Manufacturing in a larger scale is desirable for a commercial process since the increased process throughput is often required to meet commercial demand while typically leading to a reduction in the cost of goods.

For the present case study, the scale-up objective was a three- to fourfold increase in process throughput while matching the product attributes, namely, a median particle size between 36 and 48 μm (preferably 42 μm) and a bulk density between 0.20 and 0.24 g/ml (preferably 0.22 g/ml).

TABLE 12.2 Shift in Spray Drying Process Performance with Scale-Up

Potential Improvement with Scale-Up	Potential Challenge with Scale-Up
<p><i>Inlet temperature</i></p> <ul style="list-style-type: none"> • May decrease due to less heat loss, decreasing risk of thermal degradation near the gas inlet <p><i>Residence time</i></p> <ul style="list-style-type: none"> • Typically longer, allowing for drying of larger droplets or at higher relative saturation conditions <p><i>Product yield</i></p> <ul style="list-style-type: none"> • Typically increases with longer batches and less relative equipment surface area versus throughput, but a few products may decrease due to lower cyclone efficiency <p><i>Particle size</i></p> <ul style="list-style-type: none"> • Larger droplets become possible with longer residence time, which often improve product flow <p><i>Campaign length</i></p> <ul style="list-style-type: none"> • Typically increases from less fouling observed at larger scales, from larger relative surface area and larger diameter pipes 	<p><i>Atomization</i></p> <ul style="list-style-type: none"> • Difficult to deliver sufficient energy (solution pressure or atomization pressure) to create small droplets <p><i>Development cost</i></p> <ul style="list-style-type: none"> • Process development requires larger quantities of materials <p>Uncertain Impact with Scale-Up</p> <p><i>Density</i></p> <ul style="list-style-type: none"> • Less efficient mixing between the drying gas and spray plume can result in denser particles, which may be desirable or detrimental to the formulation <p><i>Cycle time</i></p> <ul style="list-style-type: none"> • Can increase or decrease, as dependent primarily on relative batch size, as wide range of spray times possible on all equipment sizes

Data collected in the pilot scale unit from carefully planned development trials reveal the sensitivity of the critical product attributes to the process conditions. As shown in Figure 12.18, the particle size was influenced by both the outlet temperature and the feed solution flow rate, while the tapped density is primarily influenced by outlet temperature only. Larger drying chamber outlet temperatures lead to lighter particles, and larger feed flow rates lead to smaller particles. This is the expected behavior from the arguments presented in Sections 12.6 and 12.7, where slower drying tends to lead to denser and shiveled particles and higher flow rates require larger atomization pressure and tend to lead to smaller particles. The data also show that larger temperatures yield smaller particles, which may seem contradictory to the particles being less dense. This is likely a consequence of secondary effects such as the spray drying gas interactions and droplet–droplet interactions discussed in Section 12.6. In Figure 12.19, the optimization analysis summarizes the range of the process parameters that yield product within the target product particle size and density (mean particle size of 36–48 μm and a bulk density of 0.20–0.24 g/ml).

The scale-up to the commercial unit was supported by a second set of pilot-scale experiments where the process conditions were based on the scale-independent parameters discussed in Section 12.6, namely, the condenser and chamber outlet temperature as well as the liquid feed to drying gas ratio. The nozzle scale-up was supported by mechanistic modeling of atomization to ensure similar ranges of median droplet size at both scales. As noted in Section 12.7, maintaining the same droplet size and spray angle leads to different breakup lengths and injection velocities, which may impact the final product attributes and will ultimately

contribute to the differences in the response of the product properties to the process parameters.

In Figures 12.18 and 12.19, the data from the development trials at the commercial scale are compared against the pilot-scale results. From the outset it can be appreciated that the chosen process conditions for the commercial-scale trials yielded similar product attributes, indicating an adequate scale-up strategy. However, it can also be noted that the response of the product attributes to the process parameters differs between scales. This is particularly noticeable in the response of the particle size where, e.g. a steeper change of the particle size with the outlet temperature and shallower change with the feed flow rate occurs in the commercial scale, whereas in the pilot scale the converse is true. These are examples of secondary scale effects pertaining to the additional level of complexity discussed in Section 12.6.

Nevertheless, for the current case study, these effects are not very pronounced, rendering the development trials at the commercial scale mostly confirmatory. However, not all products and spray drying processes exhibit such mild scale effects. State-of-the-art computational tools and mechanistic understanding of the process can aid in assessing early on the scale-up risk in meeting the target product attributes and provide tailored scale-up criteria to meet the scale-up objectives.

12.9 SUMMARY

Spray drying is increasingly used in the pharmaceutical industry to generate ASDs, which have been shown to enhance the bioavailability of drugs with low aqueous solubility. The composition of the solid dispersion is designed to

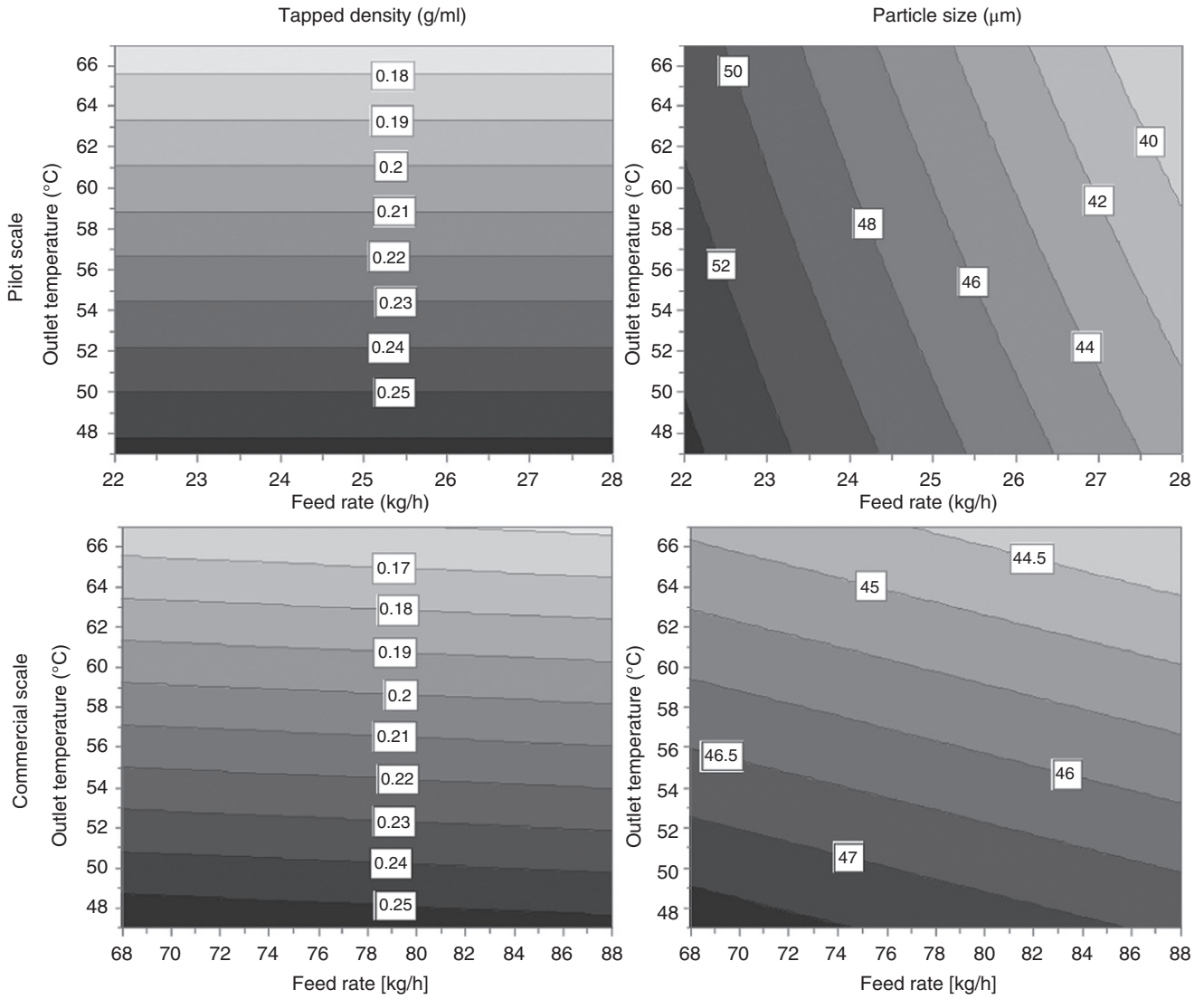


FIGURE 12.18 Response surface of the product attributes to key process parameters based on carefully planned development trials and fitted with a multilinear regression model.

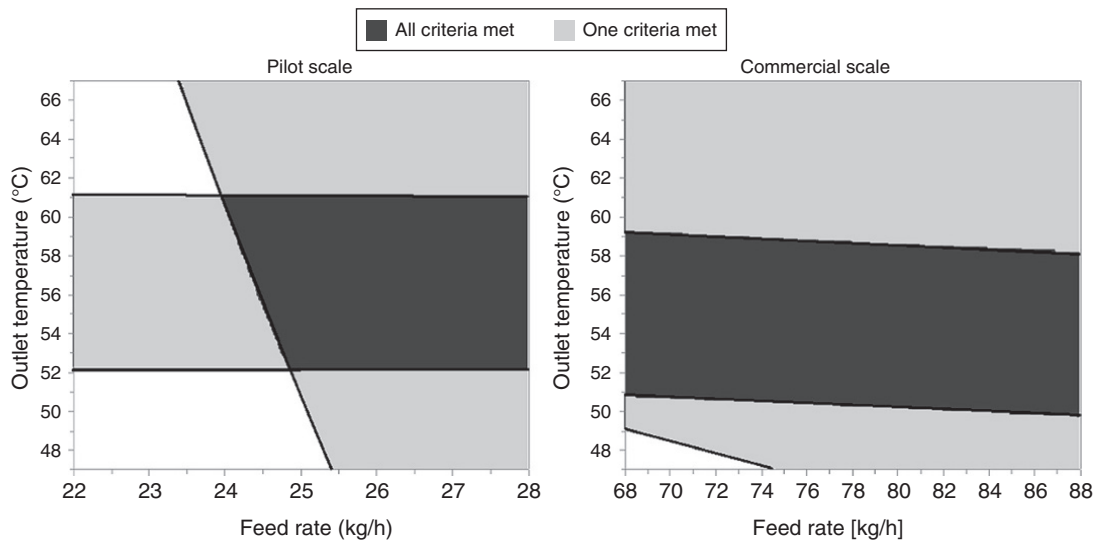


FIGURE 12.19 Optimization analysis based on the response surface model derived from the development trials.

generate a material that will both increase the solubility of the drug in solution and maintain the amorphous form of the drug throughout shelf life. The stability of the ASD is strongly linked to concepts from polymer sciences, such as reduced molecular mobility of drug molecules in a molecular dispersion with a polymer with a higher glass transition temperature. The spray drying unit operations used to generate these ASDs can be understood through fundamental principles across a range of scales. On the scale of a single particle, the rate of solvent evaporation is a balance between mass and energy transfer, and it defines the resulting structure of the SDD particle. On the scale of the spray drying chamber, CFD models combined with single-particle drying models are used to inform how the rate of solvent evaporation differs across the population of particles. On the scale of the entire spray drying equipment, global mass and energy balances provide information on the average temperatures and relative saturation that particles experience during drying, highlighting combinations of process parameters that are more likely to result in common failure modes. This understanding of the fundamental principles behind the spray drying process allows for more informed experimental designs with SDD attributes, which are often determined by the interactions of several process parameters. The final goal is to establish a process design space that is able to meet the desired SDD attributes of downstream compactibility, physical stability, and dissolution.

REFERENCES

- Vasconcelos, T., Marques, S., das Neves, J., and Sarmento, B. (2016). Amorphous solid dispersions: rational selection of a manufacturing process. *Advanced Drug Delivery Reviews* 100: 85–101.
- Vo, C.L.-N., Park, C., and Lee, B.-J. (2013). Current trends and future perspectives of solid dispersions containing poorly water-soluble drugs. *European Journal of Pharmaceutics and Biopharmaceutics* 85 (3): 799–813.
- Masters, K. (1991). *Spray Drying Handbook*, 5e. New York, NJ: Wiley.
- Newman, A. (ed.) (2015). *Pharmaceutical Amorphous Solid Dispersions*. Hoboken, NJ: Wiley.
- Paudel, A. (2013). Manufacturing of solid dispersions of poorly water soluble drugs by spray drying: formulation and process considerations. *International Journal of Pharmaceutics* 453: 253–284.
- Templeton, A., Byrn, S.R., Haskell, R.J., and Prisinzano, T.E. (eds.) (2014). *Discovering and Developing Molecules with Optimal Drug-Like Properties*, AAPS Advances in the Pharmaceutical Sciences Series, vol. 15. New York: Springer-Verlag.
- Takagi, T., Ramachandran, C., Bermejo, M. et al. (2006). A provisional biopharmaceutical classification of the top 200 oral drug products in the United States, Great Britain, Spain, and Japan. *Molecular Pharmaceutics* 3: 631–643.
- Dahan, A., Wolk, O., Kim, Y.H. et al. (2013). Purely in silico BCS classification: science based quality standards for the world's drugs. *Molecular Pharmaceutics* 10: 4378–4390.
- Fong, S.Y.K., Bauer-Brandl, A., and Brandl, M. (2017). Oral bioavailability enhancement through supersaturation: an update and meta-analysis. *Expert Opinion on Drug Delivery* 14: 403–426.
- He, Y. and Ho, C. (2015). Amorphous solid dispersions: utilization and challenges in drug discovery and development. *Journal of Pharmaceutical Sciences* 104: 3237–3258.
- Mitra, A. and Kesiosoglou, F. (2013). Impaired drug absorption due to high stomach pH: a review of strategies for mitigation of such effect to enable pharmaceutical product development. *Molecular Pharmaceutics* 10: 3970–3979.
- Sotthivirat, S., McKelvey, C., Moser, J. et al. (2013). Development of amorphous solid dispersion formulations of a poorly water-soluble drug, MK-0364. *International Journal of Pharmaceutics* 452: 73–81.
- Friesen, D.T., Shanker, R., Crew, M. et al. (2008). Hydroxypropyl methylcellulose acetate succinate-based spray-dried dispersions: an overview. *Molecular Pharmaceutics* 5: 1003–1019.
- Harmon, P., Galipeau, K., Xu, W. et al. (2016). Mechanism of dissolution-induced nanoparticle formation from a copovidone-based amorphous solid dispersion. *Molecular Pharmaceutics* 13: 1467–1481.
- Newman, A., Knipp, G., and Zografi, G. (2012). Assessing the performance of amorphous solid dispersions. *Journal of Pharmaceutical Sciences* 101: 1355–1377.
- Alonzo, D.E., Zhang, G.G., Zhou, D. et al. (2010). Understanding the behavior of amorphous pharmaceutical systems during dissolution. *Pharmaceutical Research* 27: 608–618.
- Kanaujia, P., Lau, G., Ng, W.K. et al. (2011). Nanoparticle formation and growth during in vitro dissolution of ketoconazole solid dispersion. *Journal of Pharmaceutical Sciences* 100: 2876–2885.
- Hancock, B. and Zografi, G. (1997). Characteristics and significance of the amorphous state in pharmaceutical systems. *Journal of Pharmaceutical Sciences* 86: 1–12.
- Duda, J.L. (1985). Molecular-diffusion in polymeric systems. *Pure and Applied Chemistry* 57: 1681–1690.
- Janssens, S. and Van den Mooter, G. (2009). Review: physical chemistry of solid dispersions. *Journal of Pharmacy and Pharmacology* 61: 1571–1586.
- Yu, L. (2001). Amorphous pharmaceutical solids: preparation, characterization and stabilization. *Advanced Drug Delivery Reviews* 48: 27–42.
- Kawakami, K. and Pikal, M. (2005). Calorimetric investigation of the structural relaxation of amorphous materials: evaluating validity of the methodologies. *Journal of Pharmaceutical Sciences* 94 (5): 948–965.
- Adrjanowicz, K., Kaminski, K., Wojnarowska, Z. et al. (2010). Dielectric relaxation and crystallization kinetics of ibuprofen at ambient and elevated pressure. *Journal of Physical Chemistry B* 114: 6579–6593.
- Marsac, P.J., Shamblin, S.L., and Taylor, L.S. (2006). Theoretical and practical approaches for prediction of drug-polymer

- miscibility and solubility. *Pharmaceutical Research* 23: 2417–2426.
25. Gordon, M. and Taylor, J. (1952). Ideal copolymers and the second order transitions of synthetic rubbers. I. Non-crystalline copolymers. *Journal of Applied Chemistry* 2: 493–500.
 26. Hancock, B.C. and Zografi, G. (1994). The relationship between the glass-transition temperature and the water-content of amorphous pharmaceutical solids. *Pharmaceutical Research* 11: 471–477.
 27. Konno, H. and Taylor, L. (2006). Influence of different polymers on the crystallization tendency of molecularly dispersed amorphous felodipine. *Journal of Pharmaceutical Sciences* 95: 2692–2705.
 28. Rumondor, A., Ivanisevic, I., Bates, S. et al. (2009). Evaluation of drug-polymer miscibility in amorphous solid dispersion systems. *Pharmaceutical Research* 26: 2523–2534.
 29. Baghel, S., Cathcart, H., and O'Reilly, N.J. (2016). Polymeric amorphous solid dispersions: a review of amorphization, crystallization, stabilization, solid-state characterization, and aqueous solubilization of biopharmaceutical classification system class ii drugs. *Journal of Pharmaceutical Sciences* 105: 2527–2544.
 30. Baird, J.A. and Taylor, L.S. (2012). Evaluation of amorphous solid dispersion properties using thermal analysis techniques. *Advanced Drug Delivery Reviews* 64: 396–421.
 31. Couchman, P.R. (1984). The composition-dependent glass-transition: relations between temperature, pressure, and composition. *Polymer Engineering and Science* 24 (2): 135–143.
 32. Fox, T. (1956). Influence of diluent and of copolymer composition on the glass transition temperature of a polymer system. *The Bulletin of the American Physical Society* 1: 123–132.
 33. Hancock, B.C. and Zografi, G. (1993). The use of solution theories for predicting water-vapor absorption by amorphous pharmaceutical solids: a test of the Flory-Huggins and Vrentas models. *Pharmaceutical Research* 10: 1262–1267.
 34. Miyazaki, T., Yoshioka, S., Aso, Y., and Kojima, S. (2004). Ability of polyvinylpyrrolidone and polyacrylic acid to inhibit the crystallization of amorphous acetaminophen. *Journal of Pharmaceutical Sciences* 93: 2710–2717.
 35. Bhattacharya, S. and Suryanarayanan, R. (2009). Local mobility in amorphous pharmaceuticals-characterization and implications on stability. *Journal of Pharmaceutical Sciences* 98: 2935–2953.
 36. Sun, Z.H., Dong, L.S., Zhuang, Y.G. et al. (1992). Beta-relaxation in polyimides. *Polymer* 33: 4728–4731.
 37. Hikima, T., Hanaya, M., and Oguni, M. (1999). Microscopic observation of a peculiar crystallization in the glass transition region and beta-process as potentially controlling the growth rate in triphenylethylene. *Journal of Molecular Structure* 479: 245–250.
 38. Vyazovkin, S. and Dranca, I. (2007). Effect of physical aging on nucleation of amorphous indomethacin. *The Journal of Physical Chemistry B* 111: 7283–7287.
 39. Leane, M., Sinclair, W., Qian, F. et al. (2013). Formulation and process design for a solid dosage form containing a spray-dried amorphous dispersion of ibipinabant. *Pharmaceutical Development and Technology* 18: 359–366.
 40. Singh, A., Van Humbeeck, J., & Van den Mooter, G. (2009). A new twist in the old story-can compression induce mixing of phase separated solid dispersions? A case study of spray-dried miconazole-PVPVA64 solid dispersion systems. *Pharmaceutical Research*, 26, 2523–2534.
 41. Démuth, B., Nagy, Z., Balogh, A. et al. (2015). Downstream processing of polymer-based amorphous solid dispersions to generate tablet formulations. *International Journal of Pharmaceutics* 486 (1–2): 268–286.
 42. Koganti, V., Carroll, F., Ferraina, R. et al. (2010). Application of modeling to scale-up dissolution in pharmaceutical. *AAPS PharmSciTech* 11 (4): 1541–1548.
 43. Lipp, C.H. (2012). *Practical Spray Technology: Fundamentals and Practice*. Lake Jackson, TX: Lake Innovation LLC.
 44. Graham, L.J., Taillon, R., Mullin, J., and Wigle, T. (2010). Pharmaceutical process/equipment design methodology case study: cyclone design to optimize spray-dried-particle collection efficiency. *Computers and Chemical Engineering* 34 (7): 1041–1048.
 45. Pinho, R.L. (2003). Pilot- and industrial-scale experimental investigation of numerically optimized cyclones. *Industrial and Engineering Chemistry Research* 42 (1): 145–154.
 46. Roggo, Y., Chalus, P., Maurer, L. et al. (2007). A review of near infrared spectroscopy and chemometrics in pharmaceutical technologies. *Journal of Pharmaceutical and Biomedical Analysis* 44: 683–700.
 47. Maltesen, M., Weert, M.v., and Grohgan, H. (2012). Design of experiments-based monitoring of critical quality attributes for the spray-drying process of insulin by NIR spectroscopy. *AAPS PharmSciTech* 13: 747–755.
 48. Rahman, Z., Siddiqui, A., Bykadi, S., and Khan, M.A. (2014). Near-infrared and Fourier transform infrared chemometric methods for the quantification of crystalline tacrolimus from sustained-release amorphous solid dispersion. *Journal of Pharmaceutical Sciences* 103: 2376–2385.
 49. Chan, L.W., Tan, L.H., and Heng, P.W. (2008). Process analytical technology: application to particle sizing in spray drying. *AAPS PharmSciTech* 9: 259–266.
 50. Tomba, E., Facco, P., Bezzo, F. et al. (2012). Combining fundamental knowledge and latent variable techniques to transfer process monitoring models between plants. *Chemometrics and Intelligent Laboratory Systems* 116: 67–77.
 51. Garcia-Munoz, S. and Settell, D. (2009). Application of multivariate latent variable modeling to pilot-scale spray drying monitoring and fault detection: monitoring with fundamental knowledge. *Computers and Chemical Engineering* 33: 2106–2110.
 52. Billinge, S. and Kanatzidis, M. (2004). Beyond crystallography: the study of disorder, nanocrystallinity and crystallographically challenged materials with pair distribution functions. *Chemical Communications* 4: 749–760.
 53. de Araujo, G., Benmore, C., and Byrn, S. (2017). Local structure of ion pair interaction in lapatinib amorphous dispersions characterized by synchrotron X-ray diffraction and pair distribution function analysis. *Scientific Reports* 7: 46367.
 54. Hedoux, A. (2016). Recent developments in the Raman and infrared investigations of amorphous pharmaceuticals and

- protein formulations: a review. *Advanced Drug Delivery Reviews* 100: 133–146.
55. Karmwar, P., Boetker, J.P., Graeser, K.A. et al. (2011). Investigations on the effect of different cooling rates on the stability of amorphous indomethacin. *European Journal of Pharmaceutical Sciences* 44: 341–350.
 56. Taylor, L.S. and Zografi, G. (1997). Spectroscopic characterization of interactions between PVP and indomethacin in amorphous molecular dispersions. *Pharmaceutical Research* 14: 1691–1698.
 57. Karmwar, P., Graeser, K., Gordon, K.C. et al. (2011). Investigation of properties and recrystallisation behaviour of amorphous indomethacin samples prepared by different methods. *International Journal of Pharmaceutics* 417: 94–100.
 58. Saerens, L., Ghanam, D., Raemdonck, C. et al. (2014). In-line solid state prediction during pharmaceutical hot-melt extrusion in a 12 mm twin screw extruder using Raman spectroscopy. *European Journal of Pharmaceutics and Biopharmaceutics* 87: 606–615.
 59. Karavas, E., Georgarakis, M., Docoslis, A., and Bikiaris, D. (2007). Combining SEM, TEM, and micro-Raman techniques to differentiate between the amorphous molecular level dispersions and nanodispersions of a poorly water-soluble drug within a polymer matrix. *International Journal of Pharmaceutics* 340: 76–83.
 60. Fraga, I., Montserrat, S., and Hutchinson, J. (2007). TOPEM, A new temperature modulated DSC technique: application to the glass transition of polymers. *Journal of Thermal Analysis and Calorimetry* 87: 119–124.
 61. Luebbert, C., Huxoll, F., and Sadowski, G. (2017). Amorphous-amorphous phase separation in API/polymer formulations. *Molecules* 22 (2): 296.
 62. Vehring, R. (2008). Pharmaceutical particle engineering via spray drying. *Pharmaceutical Research* 25: 999–1022.
 63. Hogeckamp, S. and Pohl, M. (2003). Porosity measurement of fragile agglomerates. *Powder Technology* 130: 385–392.
 64. Poozesh, S., Setiawan, N., Arce, F. et al. (2017). Understanding the process-product-performance interplay of spray dried drug-polymer systems through complete structural and chemical characterization of single spray dried particles. *Powder Technology* 320: 685–695.
 65. Dobry, D.E., Settell, D.M., Baumann, J.M. et al. (2009). A model-based methodology for spray-drying process. *Journal of Pharmaceutical Innovation* 4: 113–142.
 66. Green, D.W. and Perry, P. (2008). *Perry's Chemical Engineers' Handbook*. New York: McGraw-Hill.
 67. Ivey, J.W., Bhambri, P., Church, T.K. et al. (2017). Humidity affects the morphology of particles emitted from beclomethasone dipropionate pressurized metered dose inhalers. *International Journal of Pharmaceutics* 520: 207–215.
 68. Bayvel, L. and Orzechowski, Z. (1993). *Liquid Atomization*. Washington, DC: Taylor & Francis.
 69. Senecal, P., Schmidt, D., Nouar, I. et al. (1999). Modeling high-speed viscous liquid sheet atomization. *International Journal of Multiphase Flow* 25: 1073–1097.
 70. Schmidt, D.P., Nouar, I., Senecal, P.K. et al. (1999). Pressure-swirl atomization in the near field. SAE Paper, 01-0496.
 71. Holterman, H.J. (2003). *Kinetics and Evaporation of Water Drops in Air*. Wageningen: IMAG.
 72. Brenn, G. (2005). Concentration fields in evaporating droplets. *International Journal of Heat and Mass Transfer* 48: 395–402.
 73. Williamson, R.E. and Dale Threadgill, E. (1974). A simulation for the dynamics of evaporating spray droplets in agricultural spraying. *Transactions of the ASAE* 17: 254.
 74. George, O.A., Xiao, J., Rodrigo, C.S. et al. (2017). Detailed numerical analysis of evaporation of a micrometer water droplet suspended on a glass filament. *Chemical Engineering Science* 165: 33–47.
 75. Law, C.K. and Law, H.K. (1982). A d2-Law for multicomponent droplet vaporization and combustion. *AIAA Journal* 20: 522–527.
 76. Handscomb, C.S., Kraft, M., and Bayly, A.E. (2009). A new model for the drying of droplets containing suspended solids after shell formation. *Chemical Engineering Science* 64: 228–246.
 77. Mezhericher, M., Levy, A., and Borde, I. (2007). Theoretical drying model of single droplets containing insoluble or dissolved solids. *Drying Technology* 25: 1025–1032.
 78. Hecht, J.P. and King, C.J. (2000). Spray drying: influence of developing drop morphology on drying rates and retention of volatile substances. 2. Modeling. *Industrial and Engineering Chemistry Research* 39: 1766–1774.
 79. Tran, T.T.H., Jaskulski, M., Avila-Acevedo, J.G., and Tsotsas, E. (2017). Model parameters for single-droplet drying of skim milk and its constituents at moderate and elevated temperatures. *Drying Technology* 35: 444–464.
 80. Chen, X.D. (2008). The basics of a reaction engineering approach to modeling air-drying of small droplets or thin-layer materials. *Drying Technology* 26: 627–639.
 81. Wilcox, D.C. (1998). *Turbulence Modeling for CFD*. La Canada, CA: DCW Industries.
 82. Albrecht, H.D. (2013). *Laser Doppler and Phase Doppler Measurement Techniques*. Berlin, London: Springer Science & Business Media.
 83. Francia, V., Martin, L., Bayly, A.E., and Simmons, M.J. (2016). The role of wall deposition and re-entrainment in swirl dryers. *AIChE Journal* 61 (6): 1804–1821.
 84. Francia, V., Martin, L., Bayly, A.E., and Simmons, M.J. (2017). Agglomeration during spray drying: airborne clusters or breakage at the walls? *Chemical Engineering Science* 162: 284–299.
 85. Thybo, P., Hovgaard, L., Lindeløv, J.S. et al. (2008). Scaling up the spray drying process from pilot to production scale. *Pharmaceutical Research* 25 (7): 1610–1620.

13

THE FREEZE DRYING PROCESS: THE USE OF MATHEMATICAL MODELING IN PROCESS DESIGN, UNDERSTANDING, AND SCALE-UP

VENKAT KOGANTI

Celgene Corporation, Summit, NJ, USA

SUMIT LUTHRA

Pfizer Worldwide Research and Development, Andover, MA, USA

MICHAEL J. PIKAL

School of Pharmacy, University of Connecticut, Storrs, CT, USA[†]

13.1 INTRODUCTION

Freeze drying, also termed “lyophilization,” is a drying process employed to convert solutions of labile materials into solids of sufficient stability for distribution and storage. A typical production-scale freeze dryer consists of a drying “chamber” containing temperature-controlled shelves that is connected to a “condenser” chamber via a large valve. The condenser chamber houses a series of plates or coils capable of being maintained at very low temperature (i.e. $<-50^{\circ}\text{C}$). One or more vacuum pumps in series are connected to the condenser chamber to achieve pressures in the range of 0.03–0.3 Torr in the entire system during operation. A commercial freeze dryer may have 10–20 shelves with a total load on the order of 50 000 10 cc vials. The objective in a freeze drying process is to convert most of the water into ice in the “freezing stage,” remove the ice by direct sublimation in the “primary drying stage,” and finally remove most of the unfrozen water in the “secondary drying” stage by desorption. The water removed from the product is reconverted into ice by the condenser.

In a typical freeze drying process, an aqueous solution containing the drug and various formulation aids, or “excipients,” is filled into glass vials, and the vials are loaded

into the temperature-controlled shelves. The shelf temperature is reduced, typically in several stages, to a temperature in the vicinity of -40°C , thereby converting nearly all of the water into ice. Some excipients, such as buffer salts and mannitol, may partially crystallize during freezing, but most “drugs,” particularly proteins, remain amorphous. The drug and excipients are typically converted into an amorphous glass also containing large amounts of unfrozen water (15–30%) dissolved in the solid (i.e. glassy) amorphous phase. Thus, most of the desiccation actually occurs during the freezing stage of the freeze drying process. After all water and solutes have been converted into solids, the entire system is evacuated by the vacuum pumps to the desired control pressure, the shelf temperature is increased to supply energy for sublimation, and primary drying begins. Due to the large heat flow required during primary drying, the product temperature runs much colder than the shelf temperature. The removal of ice crystals by sublimation creates an open network of “pores” that allows pathways for escape of water vapor out of the product. The ice–vapor boundary (i.e. the boundary between frozen and “dried” regions) generally moves from the top of the product toward the bottom of the vial in roughly planar fashion as primary drying proceeds. Primary drying is normally the longest part of the freeze drying process. Primary drying times on the order of days are not uncommon, and in rare cases, weeks may be required for a

[†] Deceased.

combination of poor formulation and suboptimal process design. While some secondary drying does occur during primary drying (i.e. desorption of water from the amorphous phase occurs to a limited extent once the ice is removed from that region), the start of secondary drying is normally defined, in an operational sense, as the end of primary drying (i.e. when all ice is removed). Of course, since not all vials behave identically, some vials enter secondary drying, while other vials are in the last stages of primary drying. When the judgment is made that all vials are devoid of ice, the shelf temperature is typically increased to provide the higher product temperature required for efficient removal of the unfrozen water. The final stages of secondary drying are normally carried out at shelf temperatures in the range of 25–50 °C for several hours. Here, since the demand for heat is low, the shelf temperature and the product temperature are nearly identical.

Since freeze drying plants are very expensive and process times are often long, a freeze-dried dosage form is relatively expensive to produce. Indeed, both because of cost and ease of use, a “ready-to-use” solution is the preferred option for a parenteral dosage form, particularly if the solution can withstand terminal heat sterilization. However, most parenteral drugs undergo excessive degradation during terminal sterilization. Even if sterility requirements may be satisfactorily met without terminal sterilization (i.e. sterile processing), many drugs do not have sufficient stability in the solution state to allow the long-term storage required for pharmaceutical products. Certainly, terminal sterilization is not an option for a protein product. Indeed, many proteins are insufficiently stable in aqueous solution, even when refrigerated, to allow storage for more than a few months without suffering significant degradation. Of course, some proteins are quite stable in aqueous solution, insulin being the classic example of a “solution stable” protein product [1]. When an aqueous solution does not have sufficient stability, the product must be produced in solid form. At least for small molecules, stability normally increases in the order solution \ll glassy solid $<$ crystalline solid [2–4], likely a result of restricted motion in solids with the high degree of order in the crystalline solid limiting reactivity even further. It should also be noted that the stability upon crystallization of a solid noted for small molecules may not extend to proteins. Although a direct experimental comparison is limited to one example, that of insulin, crystalline insulin is actually significantly less stable than amorphous freeze-dried insulin [5]. Since pharmaceutical proteins cannot generally be produced on a commercial scale by crystallization, a glassy solid is usually the only solid-state option.

Freeze drying [6–9] and spray drying [10–12] are drying methodologies in common use in the pharmaceutical industry that are suitable for the production of glassy solids. Freeze drying is basically a low-temperature process. In general, a protein formulation can be dried to on the order of 1% water or less without any of the product exceeding 30 °C. Thus, conventional wisdom states that freeze drying is less likely

to cause thermal degradation than a “high-temperature” process, such as spray drying. However, it must be noted that due to self-cooling as the water evaporates, the product temperature in a spray drying process is far less than the input air temperature and residence times in the dryer are very short. Indeed, it has been shown that suitably formulated, stable protein glasses may be produced by direct evaporative drying (i.e. drying without freezing) [13, 14]. Such direct evaporative drying process may involve spray drying or alternate new technologies [15]. While some of the factual material in Ref. 14 has been challenged [16], it must be admitted that freeze drying is not the only process by which protein solutions may be successfully converted into “stable” glasses.

Historically, freeze drying is the method of choice for products intended for parenteral administration. Sterility and relative freedom from particulates are critical quality attributes for parenterals. Largely because the solution is sterile filtered immediately before filling into the final container and further processing is relatively free of exposure to humans, a freeze drying process maintains sterility and “particle-free” characteristics of the product much easier than processes that must deal with dry powder handling issues, such as dry powder filling of a spray-dried or bulk crystallized powder. Indeed, with modern robotic automatic loading systems [17], humans can be removed from the sterile processing area entirely, at least in principle. Furthermore, since the vials are sealed in the freeze dryer, moisture and headspace gas can easily be controlled, an important advantage for products whose storage stability is adversely affected by residual moisture and/or oxygen. Since the critical heat and mass transfer characteristics for freeze drying are nearly the same at the laboratory scale as in full production, resolution of scale-up problems tends to be easier for a freeze drying process than for spray drying, at least in our experience. Also, development of a freeze-dried product requires less material for formulation and process development, a particularly important factor early in a project.

While freeze drying has a long history in the pharmaceutical industry as a technique for stabilization of labile drugs, including proteins, many proteins suffer irreversible change, or degradation, during the freeze drying process [18–22]. Even when the labile drug survives the freeze drying process without degradation, the resulting product is rarely found perfectly stable during long-term storage, particularly when analytical techniques with a sensitivity to detect low levels of degradation (i.e. $\approx 0.1\%$) are employed. Both small molecules [2–4, 23] and proteins [24–27] show degradation during storage of the freeze-dried glass. In many cases, instability is serious enough to require refrigerated storage [24, 25, 28].

Stability problems are most often addressed by a combination of formulation optimization and attention to process control. Lyoprotectants are added to stabilize the protein during the freeze drying process as well as to provide storage stability, and the level and type of buffer is optimized. Optimization of the freezing process may be critical, control of

product temperature during drying is critical for products that tend to suffer cake “collapse” during primary drying, and control of residual moisture is nearly always critical for storage stability. Formulation and process are interrelated in that the process design depends on formulation and process variations, particularly freezing variations, can change the physical state of the formulation. A bad formulation can be nearly impossible to freeze dry, and even with a well-designed formulation, a poorly designed process may require more than a week to produce material of suboptimal quality. While blind empiricism may, in time, yield an acceptable formulation and process, an appreciation for the materials science of amorphous systems and some understanding of heat and mass transfer relevant to freeze drying are needed for efficient development of freeze-dried pharmaceuticals. Obviously, one also requires at least a phenomenological understanding of the major degradation pathways specific to the protein under consideration.

Once a suitable formulation and process are developed in the laboratory, hopefully at least close to optimized, the laboratory process needs to be transferred to manufacturing. While freeze drying is, in many ways, a relatively easy process to scale-up since the volume and nature of the primary container system is independent of scale (i.e. the same fill volume and vial in both lab and manufacturing), there do exist a number of differences between laboratory and manufacturing. First, the time scales are often far different for some stages of the process, the most notable being the time required for loading a manufacturing dryer being much longer than the corresponding time in the laboratory. Thus, particularly if the relative humidity in the vicinity of the freeze dryer is high, one may experience condensation on the shelves during loading a production dryer but not during the corresponding operation in the laboratory. Such condensation may cause freezing variations that have consequences for the drying process as well. In addition, it is often possible to change shelf temperature very rapidly in the laboratory, but due to greater thermal mass in the manufacturing environment, the production dryer may be unable to match the laboratory dryer’s performance. Secondly, there may exist heat and mass transfer differences between the laboratory dryer and the production dryer that require a slightly different shelf temperature profile with time for manufacturing than was used for the laboratory process [29]. Since the objective is to maintain the same product temperature history during the process in manufacturing that was validated in the laboratory, the shelf temperature vs. time program in manufacturing may well be different than found optimal for the laboratory [29]. In addition, the maximum heat and mass transfer that the dryer can handle without losing control varies with the dryer design, and one may find that a process that runs very well in the laboratory dryer may overload the manufacturing dryer and cause loss of chamber pressure control [29]. However, the most important and most troublesome scale-up issue is often the difference in freezing behavior that

one experiences in the class 100 “clean” environment of the production operation relative to the relatively high particulate content in the laboratory air. Thus, typically [30] supercooling is greater in manufacturing than in the laboratory, and since the size of the ice crystals (and resulting pores in the dry cake) decreases with increasing degree of supercooling, the laboratory produces larger ice crystals, larger pores, and less resistance to vapor transport. The net result is that the primary drying time is shorter in the laboratory than in manufacturing and the product temperature runs colder in the laboratory than in manufacturing. The effect can be significant, with primary drying running 10–30% longer in manufacturing [30–32]. Thus, the material being freeze-dried in the laboratory is typically not representative of the material being dried in production, simply because the freezing behavior is different. Frequently, the process is arbitrarily adjusted in an attempt to compensate for this anticipated difference. That is, the duration of primary drying is extended so that one may be reasonably confident that in the production batch all vials are devoid of ice when the shelf temperature is increased to facilitate secondary drying. However, one is never sure of how much extension is necessary, so there are better solutions to the problem, as discussed in more detail in Section 13.3.

Freeze drying is one of the few unit operations where the underlying physics is relatively well understood, and the theoretical models are not mostly empirical. Thus, modeling can be used very effectively for process design and scale-up using simple steady-state models [33] or more elaborate non-steady-state models that address both sublimation and desorption drying [34]. Use of the theoretical models, with appropriate input data to define heat and mass transfer parameters, allows greater insight into the impact of changes in operating conditions on product quality and brings much greater efficiency into robustness testing than possible with the empirical testing characteristic of pure statistical approaches.

This chapter begins with a detailed discussion of the freezing process, with suggestions for circumventing the problems arising from freezing differences between laboratory and manufacturing, and includes a brief discussion of the few attempts to model freezing behavior. We then continue with a discussion of drying behavior, which includes a detailed discussion of the utility of the various theoretical models that can be effectively utilized in process design and scale-up. We conclude with a brief discussion of modeling vapor flow within the freeze dryer, a topic of importance but one that has received little attention in the literature.

13.2 FREEZING PROCESSES

Freezing is the first step in the lyophilization process where most of the water is separated from the solute as ice crystals. It is also a very important step as the structure of the frozen

matrix governs the rate of sublimation and desorption. Figure 13.1 shows the shelf and product temperature profile of product frozen in vials on a laboratory-scale shelf freeze dryer. The data shown are for 5% w/w sucrose solution. During the cooling step the solution remains liquid well below the equilibrium freezing temperature ($-13\text{ }^{\circ}\text{C}$ for the example shown in Figure 13.1). Ice nucleation occurs in the supercooled liquid in a stochastic manner and proceeds rapidly. Nucleation is quickly followed by crystal growth. Ice nucleation can occur by two mechanisms: homogeneous or heterogeneous. Homogeneous nucleation is defined as aggregation of pure material (in this case water, which undergoes homogeneous nucleation at $-40\text{ }^{\circ}\text{C}$) [35]. Heterogeneous nucleation, on the other hand, occurs when aggregates form on foreign solids such as dirt, container wall, etc. Homogeneous ice nucleation does not occur in pharmaceutical freeze drying operations due to inevitable presence of foreign solids. The foreign surface acts as a catalyst, thus reducing the surface free energy for formation of the nuclei. This lowering of free energy means that degree of supercooling is less as compared with homogeneous nucleation. The size of the ice crystals is dependent on the nucleation and growth rate, which are both governed by the degree of supercooling. A higher nucleation rate would result in smaller ice crystals and vice versa. The stochastic nature of nucleation would mean different degrees of supercooling and thus differences in ice crystal size and distribution across the batch and significant differences between batches, particularly when laboratory and production batches are compared [30]. In general, it

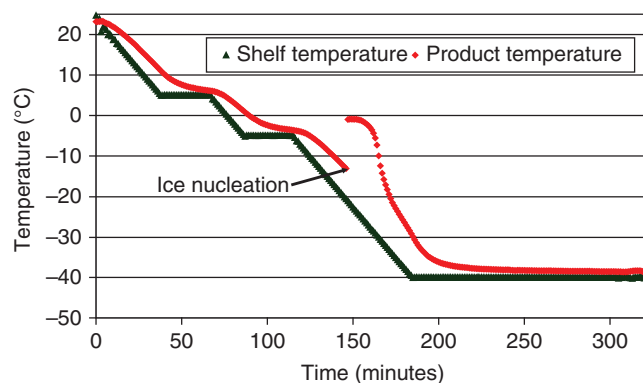


FIGURE 13.1 Typical shelf and product temperature profile during freezing for a formulation containing amorphous solute. The data are for 5% w/w sucrose solution lyophilized in a laboratory-scale freeze dryer. The product temperature profile is recorded with a thermocouple located at the bottom center of the vial. The product temperature profile shows significant supercooling before ice nucleation. Ice nucleation causes an increase in product temperature up to approximately the equilibrium melting point. The product temperature stays high as ice crystallization proceeds, and it eventually starts to drop and follow the shelf temperature.

is desirable to have larger ice crystals and narrow size distribution. Therefore, control over degree of supercooling is desirable. However, to date the methods to control ice nucleation are only in the experimental stage and have not yet been applied in commercial setting. These issues are discussed in more detail later.

13.2.1 Cooling Rate

In a typical freezing process, the only variable that is under direct control is the cooling rate of the shelf fluid. Further, the range of cooling rates achievable is not particularly large. At most, the samples can be cooled at $\approx 2\text{ }^{\circ}\text{C}/\text{min}$. It has been noted that a cooling rate of $1\text{ }^{\circ}\text{C}/\text{min}$ is generally optimal as it provides moderate supercooling and reasonably fast freezing rate [36].

13.2.2 Solute Concentration and Phase Changes

When liquid water is removed as ice crystals as freezing proceeds, the solutes (active pharmaceutical ingredient (API) and excipients) are concentrated in the unfrozen region between the ice crystals. The concentration continues until the solute crystallizes or converts into an amorphous glassy system. The physical nature of the solutes and their properties has a profound impact on the rest of the process and merits further discussion. Upon completion of freezing the solute matrix may be completely amorphous, or crystalline, or a mixture of amorphous and crystalline phases. The importance of this matrix is in the mechanical structure in that it provides for efficient drying and formation of an elegant product. However, to maintain this solid matrix, it is important that the product temperature during primary drying should not increase above a critical temperature known as the collapse temperature. In completely amorphous matrices, the collapse temperature (T_c) is related to the glass transition temperature of the frozen concentrate (T'_g) and is generally about $2\text{ }^{\circ}\text{C}$ higher than T'_g [37]. In a completely crystalline matrix, T_c is equal to the temperature of the eutectic melt (T_{eu}).¹ The relationship in a mixed amorphous and crystalline system is governed by the relative ratio of amorphous and crystalline phase [38]. In most cases, mixed systems are dominated by the crystalline phase by design, so the effective collapse temperature is close to the eutectic melt even though the collapse temperature (and T'_g) of the amorphous phase may be much lower. These crystalline phases designed into

¹A rigorous definition of “eutectic melting” in ternary systems requires all the components to exist in the crystalline state. This is not the case in the majority of pharmaceutical systems for freeze drying. However, we have retained the term “eutectic” to describe solute + ice melting, as is the practice in the pharmaceutical community.

the product are called bulking agents and are added to provide mechanical strength and/or raise the effective collapse temperature. Mannitol is the most commonly used bulking agent. Figure 13.2 shows the thermogram for a frozen solution containing amorphous and crystalline solutes. The heating curve shows the presence of T'_g and T_{eu} followed by the ice melting endotherm.

13.2.3 Annealing

Annealing is carried out by holding samples isothermal above the T'_g for a few hours. It is relatively common to employ an annealing step to facilitate solute crystallization of API or bulking agents [39]. Figure 13.3 shows the shelf and product temperature profile for a solution containing solute that crystallized during freezing. The data shown are for 5% w/w mannitol solution freeze-dried on a laboratory-scale freeze dryer. The shelf temperature profile includes an annealing step at -20°C in this case. A distinct feature of the product temperature profile in this case is the appearance of bumps during the annealing step. These bumps are likely a result of heat released due to crystallization of solute and water during annealing. However, ice nucleation in adjacent

vials could be another explanation for the observed bumps. Annealing is also currently the only commercially viable method to modify ice morphology. Annealing is believed

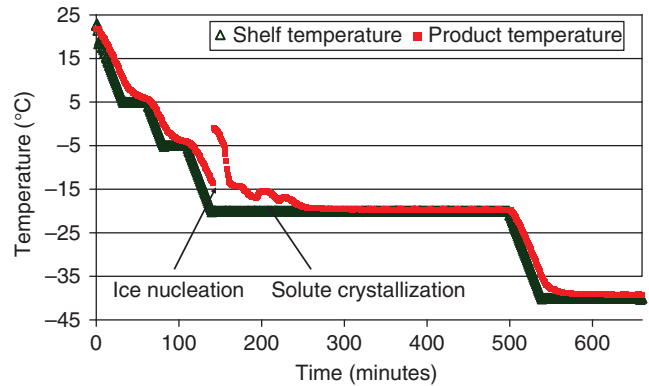


FIGURE 13.3 Typical shelf and product temperature profile for a formulation containing a crystallizable solute. The data are for 5% mannitol w/w solution freeze-dried in a laboratory-scale freeze dryer. The shelf temperature profile incorporates an annealing step at -20°C . In addition to the ice nucleation event observed in Figure 13.1, thermal event related potentially to solute crystallization is apparent in the product temperature profile.

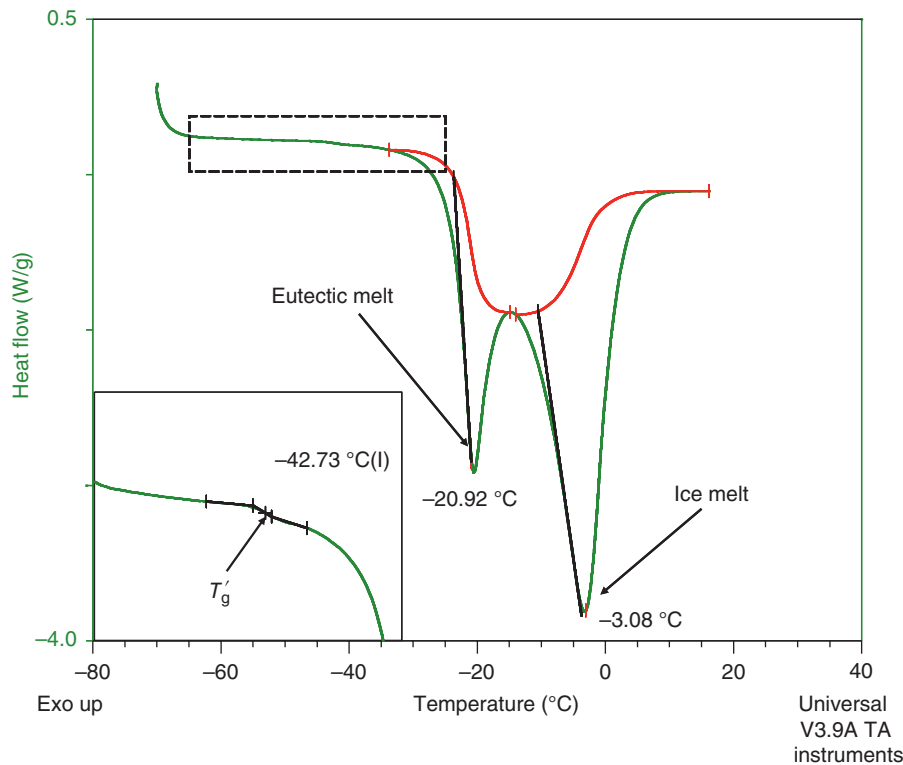


FIGURE 13.2 Thermogram of a frozen solution containing amorphous and crystalline solutes obtained by differential scanning calorimetry. The heating curve shown displays the glass transition temperature of the amorphous frozen concentrate (T'_g), eutectic melt of crystalline component (T_{eu}), and the ice melt event. The data show that there is approximately 20°C window between T'_g and T_{eu} where annealing could be carried out.

to result in increased mean crystal size and narrower distribution due to Ostwald ripening [40]. When the frozen matrix is heated above the T'_g , the ice crystals below a critical size decrease in size and effectively “melt,” and those larger than the critical size grow in a diffusion-dependent manner. The increase in mean size and narrower distribution are both advantageous as larger ice crystals mean larger pores for water vapor to escape during primary drying and narrowing of size distribution results in batch homogeneity. The changes in primary drying time that have been reported are significant, such as a 3.5-fold increase in primary drying rate reported in one study [39]. This increase in primary drying rate is correlated with a decrease in mass flow resistance of the dry layer, as reported in another study [41]. However, there have been reported cases (and unpublished experience of the authors) where annealing has not led to a decrease in mass flow resistance or increase in primary drying rate [42]. There are other reasons why annealing may not be suitable in all cases. The increase in the size of ice crystals will reduce the surface area, which decreases desorption rate during secondary drying and/or prolong reconstitution time. The residual water content of the dried product may therefore be higher [43]. Annealing may also promote phase separation resulting in unintended crystallization of a solute such as a buffer, which would produce a large pH shift, or phase separation within the amorphous phase [44]. Such phase separations may adversely affect the in-process or storage stability of lyophilized protein products where lyoprotectants are used to improve physical and chemical stability and are required to be in the same phase as the protein to impart the stabilization effect. Thus, the benefit of annealing in reducing primary drying time and improving batch homogeneity may be offset by stability problems that might arise from phase separation, and in some cases, the additional time required for annealing and secondary drying may negate the reduction in primary drying time. Certainly, optimization of the annealing step is required to achieve the greatest benefit.

13.2.4 Methods to Control Ice Nucleation

There are other methods that are currently under development to control the ice nucleation temperature. Ice fog technology is based on introducing cold nitrogen gas in the drying chamber to create an ice fog at the desired temperature of nucleation [45]. The concept was further [32] developed to evaluate the impact of nucleation temperature on mass flow resistance of the dry layer and primary drying time. An empirical direct correlation between specific surface area and mass flow resistance was also established. It was shown that a lower degree of supercooling, i.e. nucleation occurring at a higher temperature, led to smaller specific surface area and faster primary drying. It was noted that the technique required more development to be a viable method for nucleation control.

The use of ultrasound to induce ice nucleation has also been published [46]. Ice nucleation is thought to occur through bubble cavitation. A direct correlation between ice nucleation temperature and sublimation rate during primary drying was observed. Induction of ice nucleation through electric high voltage pulses has also been published and has been shown to impact primary drying rate [47, 48]. The method known as electrofreezing was only successful for solutions containing nonionic species. None of these methods have actually been applied to a commercial process. Commercial application may require retrofitting of existing units, which would be expensive in the least but may also be impractical in other cases. Nonetheless, these methods do demonstrate that control of ice nucleation temperature has great potential benefit. An option that would not require any changes to the existing equipment would be preferable, such as induction of freezing by vacuum. A vacuum freezing method to induce top-down freezing arising from evaporative cooling has been published [49, 50]. The method would require further development to address concerns with secondary drying kinetics and residual water content similar to annealing. In addition, top-down freezing may increase the chances of vial breakage, and the technique needs to be carefully controlled to avoid boiling and the resulting splattering of the solution.

13.2.5 Modeling of the Freezing Process

Freezing processes relevant to lyophilization, i.e. in vials, have been discussed in the literature [51]. For the purpose of modeling, the process has been divided into two parts: cooling and freezing. The cooling step has been modeled using Fourier's second law and is straightforward as the input parameters density, heat capacity, cooling rate, and thermal conductivity of the solution can be easily determined from published literature and experimentation. Modeling of the freezing step becomes more complicated as one would expect because the theories of nucleation and crystallization are less well defined than simple heat transfer. Also, the terms involving phase changes introduce additional parameters that at best are difficult to estimate or measure. Nonetheless this approach was used to estimate mean ice crystal size and mass flow resistance of the dry layer. It is clear that modeling the freezing step is not as advanced as the drying steps and is not at a level where a scientist would determine the input parameters in the lab and plug the values into a simple model that produces useful quantitative results. However, the published data do confirm some general concepts that nucleation temperature and cooling rate both impact the mean ice crystal size and distribution and that increasing the nucleation temperature and lowering the cooling rate result in larger ice crystals, thus lowering mass flow resistance of the dry layer. Additional progress in modeling the freezing process is needed so that the impact of process variations, e.g. annealing, on the mean ice crystal size and distribution could be

accessed and the mass flow resistance parameter for modeling the drying process could readily be obtained.

13.2.6 Scale-Up Issues

The manufacturing of clinical or commercial pharmaceutical products is conducted in sterile, particulate-free Class 100 areas, whereas lyophilization cycle development is generally conducted in laboratories that are not at all “particle-free.” The particulate load, i.e. heterogeneous nucleation sites, thus varies when scaling up from laboratory to clinical or commercial site. This means that the degree of supercooling in a production environment would be greater, leading to smaller ice crystals and higher resistance to mass flow through the dried layer. Therefore, primary drying parameters developed in the laboratory would generally lead to higher product temperature and longer drying time in the production environment [30].

13.2.7 Rational Freezing Process Development

A development scientist should take (essentially) the following approach to develop a rational freezing process:

1. *Selection of cooling rate:* The impact of cooling rate parameter has been discussed in Section 13.2.1. In summary, a cooling rate between 0.5 and 1 °C/min would generally lead to moderate supercooling, which is uniform intra-vial and inter-vial, while providing a sufficiently fast freezing rate to avoid phase separation.
2. *Selection of final freezing temperature and time:* The lowest temperature during the freezing stage depends on the solidification temperature of the system. The temperature should be at least be 2 °C lower than the solidification temperature [36]. However, it is common to see -40 to -50 °C as empirical choice for lowest temperature, and while often not necessary it also does not normally cause a significant problem. The hold time at the lowest temperature is a function of the fill volume. During development the product temperature from thermocouple should be monitored as a function of time to determine the time required to reach the lowest desired temperature. In general one to two hours hold time is sufficient for fill volumes less than 1 cm, and two to four hours hold time is sufficient for fill volumes more than 1 cm. Fill volumes greater than 2 cm are generally not recommended.
3. *Selection of annealing conditions:* Choosing annealing conditions remains somewhat empirical, and existing literature provides limited guidance on this aspect. As discussed above, annealing may be carried out with different objectives. If the formulation contains a component that must be crystallized, it has been suggested

that annealing should be carried out at a temperature between T'_g and the onset of the ice melt endotherm [36]. However, the temperature differential between the two may be quite large. The example shown in Figure 13.1 demonstrates this point where the difference between T'_g and T_{cu} is about 20 °C. The exact choice of annealing temperature and time need to be optimized experimentally. Annealing must be carried out at a temperature that (i) is above the T'_g and (ii) falls on the crystal growth curve. Therein lies the difficulty as the crystal growth curves are not readily available. For commonly used bulking agents such as mannitol and glycine, annealing could be carried out from -25 to -20 °C for several hours to maximize crystallization, provided the fraction of crystallizing solute is high. If annealing is desired to bring about change in ice morphology and improving batch homogeneity for a totally amorphous system, one needs annealing temperatures relatively close to the onset of melting. It was shown for model amorphous systems such as sucrose and hydroxyethyl starch (HES) that annealing at temperatures between -10 and -2.4 °C for 5–10 hours was needed for maximum change in primary drying rate and ice morphology [40].

4. *Addressing scale-up issues:* The difference in nucleation temperature between lab and production scale due to change in environmental particulate load could be eliminated if development work were carried out in the same environment. However, this is normally not practical. Therefore, at the present time, annealing is the only method that has been used to minimize or eliminate supercooling effects in a manufacturing environment. Further development of techniques to control ice nucleation may change this scenario in the future.

13.3 DRYING PROCESSES

Mathematical representation of the drying problem can be described as follows: once the solution is frozen, the vials are heated by raising the shelf temperature, resulting in sublimation of frozen ice initially and desorption of unfrozen water later. As drying proceeds from top to the bottom of the vial, the dried layers of the cake offer resistance to the water vapor flow due to sublimation of the ice from the layers underneath. The freeze drying problem is hence a heat transfer (to the vial from shelf and surroundings) and mass transfer (transport of water vapor through porous dried layers and then from the main chamber to the condenser) problem that can be modeled utilizing the fundamentals of heat and mass transfer processes. Further details of the current state of knowledge in modeling this process are described below.



FIGURE 13.4 Image of typical commercial-scale freeze dryer. This freeze dryer has a total shelf surface area of 39 m². It has 24 shelves and 6 trays per shelf. This image is obtained from Pfizer Kalamazoo manufacturing facility.

Mathematical modeling of the drying process provides methodology that streamlines experimental screening approaches for developing optimal freeze drying cycles that produce a quality product in a robust process. A particularly important application of drying process modeling is in the area of freeze drying process scale-up. A typical scale-up from laboratory- to commercial-scale freeze drying will increase the shelf surface area available for freeze dryer from 4.5 to 220 ft² [52]. Figure 13.4 shows an image of a typical commercial-scale freeze dryer. Heat transfer may differ, and differences in heterogeneous nucleation normally produce significant differences in ice structure and therefore in pore structure of the cake, which impact mass transfer within the dry layer. Also, differences in dryer design may lead to differences in transport properties between laboratory and commercial dryers. A clear understanding of the dependence of drying kinetics on the heat and mass transfer characteristics of the vial and dried cake, respectively, and the impact of the differences between lab and commercial scale will facilitate rational scale-up of the freeze drying process, avoiding expensive failures, and hence will result in efficient development of a robust process, thereby decreasing the cost and time of development [53].

13.3.1 Steady-State Heat and Mass Transfer Modeling

Although most reported modeling work uses non-steady-state modeling techniques, there are some examples

describing the use of simple steady-state theory to model the primary drying process. Using the pseudo-steady-state approximation, solution of the heat and mass transfer equations has been obtained at several stages during primary drying phase, thus evaluating temperature and pressure profiles as a function of time [33]. Using this simple model, the authors studied the effect of the product temperature on drying time, effects of shelf temperature and chamber pressure were evaluated, and the optimum vial size to minimize primary drying time was identified. This simple model was also utilized to evaluate the effect of process nonuniformities (e.g. variability of vial heat transfer coefficient within the same freeze drying run, nonuniform shelf temperature, product resistance variation) on the drying times and product temperature during primary drying.

The following equations describe a typical freeze drying process where the solution to be freeze-dried is in a vial, which is placed on top of a temperature-controlled shelf. The steady-state approximation is used, meaning that all of the heat supplied from the shelf is utilized in subliming the ice from the interface [33]. See the nomenclature section for nomenclature details.

$$\frac{dm}{dt} = \frac{A_p \times (P_0 - P_c)}{R_p} \quad (13.1)$$

$$\Delta H_s \times \frac{dm}{dt} = \frac{dQ}{dt} \quad (13.2)$$

$$\frac{dQ}{dt} = A_v \times K_v \times (T_s - T_b) \quad (13.3)$$

$$\Delta H_s \times \frac{P_0 - P_c}{R_p} \times \frac{A_p}{A_v \times K_v} = T_s - T \quad (13.4)$$

EXAMPLE PROBLEM 13.1

Problem Statement

Sucrose is common excipient used in parenteral formulations. Determine the product temperature during primary drying and length of primary drying time during lyophilizing a sucrose solution at the conditions described below. After building the model, utilize the model to perform an *in silico* robustness test at shelf temperatures that are ± 3 °C relative to shelf temperature set point and chamber pressure ± 50 mTorr relative to the chamber pressure set point.

Given:

1. Shelf temperature set point, T_{shelf} : -25 °C.
2. Chamber pressure set point, P_{chamber} : 100 mTorr.
3. Heat of sublimation of ice – 660 cal/g.
4. Average dry layer resistance, R_p – $3 \text{ cm}^2 \cdot \text{h} \cdot \text{Torr} / \text{g}$.

5. Overall heat transfer coefficient between vial and surroundings, $K_v = 0.00042 \text{ cal}/(\text{s}\cdot\text{cm}^2\cdot\text{K})$.
6. Vial dimensions – inner cross-sectional area = 5.85 cm^2
Outer cross-sectional area = 7.08 cm^2 .
7. Formulation details: solids concentration = $0.05 \text{ g solid/g liquid}$, fill volume = 5 ml .

Solution

Part A

Equation (13.4) can be solved for the unknown interface temperature, T , using the parameters given in the problem statement. The solution can be obtained manually or by using the Solver feature in Excel. Note that the interface temperature, T , and interface vapor pressure, P_0 , are related by the following relationship:

$$P_0 = 2.698 \times 10^{10} \exp\left(\frac{-6144.96}{T}\right),$$

where

T is in K.

P_0 is the vapor pressure of ice in Torr.

Following the above procedure, one finds that the average interface temperature is -34.5°C and the primary drying time is 31.5 hours. Once P_0 is obtained, dm/dt can be calculated using Eq. (13.1). Total amount of ice to be sublimed (Δm) can be calculated from the fill volume and solids concentration. Using these two quantities, time required to sublime all the ice (primary drying time) can be calculated as $\frac{\Delta m}{(dm/dt)}$.

In a laboratory-scale freeze drying run (same operating conditions described in this example) with 5 wt % sucrose as the solution (in same vial whose dimensions are described in this example), the product temperature was measured by placing a thermocouple at the bottom of the vial during drying. It was found that the product temperature was -35°C and the primary drying time was 28 hours. The vial was

located in the center of the freeze dryer, and the end point of primary drying was considered to be the point when the product temperature reading starts to increase from the steady-state value to reach the shelf temperature set point. One simplification in the above equations is to neglect the difference between the sublimation interface temperature and bottom temperature difference. To obtain more accurate results, this difference can also be accounted for [54].

Part B

We now have a mathematical model that describes the primary drying process of sucrose. During a manufacturing process operation set point, deviations may occur due to a variety of reasons. The model can be used to predict the effect of set point deviations and therefore test process robustness. Table 13.1 lists the product temperatures and drying times for four extreme deviations from the set points. Also listed in the table is the change in product temperature and drying time when compared with original set point ($T_{\text{shelf}} = -25^\circ\text{C}$ and $P_{\text{chamber}} = 100 \text{ mTorr}$). As noted above the model predictions can differ from the experimental values. Therefore understanding the relative changes in product temperature and drying time as a result of process deviations is one of the useful output from model predictions. Model predictions for the original set point are also listed in Table 13.1 for reference. This example demonstrates one of the several ways in which a successful model can be used in lyophilization cycle development to aid in choosing the operating conditions at the lab scale and also investigate the robustness of the process.

Recently an interactive modeling tool has been proposed, assuming quasi-steady-state heat transfer in frozen layer and in dry product region as well as quasi-steady-state mass transfer in the dried layer [55]. This software is based on a one-dimensional heat and mass transfer model that describes both the primary and the secondary drying stages and also describes the transition region between primary and secondary drying. Using this interactive tool, the user can optimize the shelf temperature and/or chamber pressure profile to achieve desired product temperatures.

TABLE 13.1 Evaluation of Effect of Change in Shelf Temperature and Chamber Pressure Changes on Product Temperature During Primary Drying and Primary Drying Time

$T_{\text{shelf}} (^\circ\text{C})$	$P_{\text{chamber}} (\text{mTorr})$	Product Temperature ($^\circ\text{C}$)	ΔT (Product Temperature at New Condition – Product Temperature at Original Conditions) ($^\circ\text{C}$)	Drying Time (h)	ΔTime (Drying Time at New Condition – Drying Time at Original Conditions) (h)
-25	100	-34.5	0	31.5	0
-22	50	-35.5	-1	22	-9.5
-22	150	-32	2.5	30	-1.5
-28	50	-37.5	-3	31.5	0
-28	150	-33.5	1	53	21.5

13.3.2 Non-steady-state Heat and Mass Transfer Modeling

While the simple steady-state models quantitatively describe primary drying, desorption drying (i.e. secondary drying) cannot be accurately described by such models, and several researchers have developed non-steady-state models of sublimation and desorption. Some advantages of non-steady-state models include residual moisture prediction as a function of time and describing the non-steady-state parts of primary drying (immediately after a change in shelf temperature) [34]. Liapis et al. have presented a sorption–sublimation model to describe the effect of operating conditions on drying times [56, 57]. The initial model that takes into account heat transfer only from the top surface of the frozen cake [56] has been extended to a more pharmaceutically representative case of heat transfer from both top and bottom surfaces of the frozen cake [58]. One-dimensional energy and material balance equations for frozen and dried layers were solved. Utilizing this mathematical model, the authors could calculate the sorbed water concentration profiles at the end of primary drying at different positions in the cake as a function of different operating conditions. A further comparison of the performance of different mathematical models for predicting sorbed water concentrations as a function of time was also provided. The authors utilize the one-dimensional mathematical model to predict optimal operating conditions in a freeze drying cycle. This mathematical model of sublimation and desorption drying was further extended to model the primary and secondary drying stages in vial lyophilization that is a more realistic representation of a typical pharmaceutical freeze drying process [59, 60]. A further application of modeling was in understanding the mechanism of bound water removal utilizing one-dimensional non-steady-state modeling of both primary and secondary drying stages [61]. Using the mathematical model coupled with experimental confirmation, they conclude that the removal of bound water during primary drying portion of freeze drying is negligible.

A further improvement of the one-dimensional model was made by extending it to a two-dimensional system [62]. A finite element method was used to solve two-dimensional heat and mass transfer equations for the frozen and dried layers. Accounting for the removal of ice and sorbed water, this model predicts the position and geometric shape of the moving interface, thus modeling the entire primary and secondary drying stages. A physical rationale for the choice of boundary conditions used was provided elsewhere [34]. The authors further evaluate the model predictions using experimental results suggesting the usefulness of the utilizing mathematical modeling in freeze drying process development [34]. An example for non-steady-state model equations is shown below. The equations provided below are intended to give the reader a sense for the non-steady-state formulation of the freeze drying problem. Equation (13.5) describes the

water vapor flow in the dry layer. This equation can be summarized as the sum of change in water concentration in the gas phase and the change in water concentration in the solid phase equals the flux of water out of the system. Equations (13.6) and (13.7) describe the change in water content of the solid phase and molar water flux in the dried region. Equation (13.6) assumes that the rate of change in water content of the solid phase is proportional to the difference between the water content and the equilibrium water content at the surrounding water activity, a_w , at temperature T , denoted $C \times (a_w, T)$ where k_g is a “rate constant” assumed to exhibit Arrhenius temperature dependence. The terms on the right-hand side of Eq. (13.7) represent the contributions of both diffusion and bulk fluid flow to the water flux. Bulk fluid flow is mostly Knudsen flow in usual pharmaceutical freeze drying applications. Further simplification of Eq. (13.7) is possible by considering the fact that vapor composition in the cake and in the drying chamber during practical primary drying situation is nearly 100% water vapor; we can ignore diffusion in the dry layer during primary drying, leaving only bulk flow or Knudsen flow as primary contributor to water vapor flux. These simplifications are described in corresponding Ref. [34].

Equations (13.9)–(13.11) describe the heat transfer in dried layer and frozen layer. Equation (13.9) describes the heat transfer in dry layer that is conservation of energy in dry layer. Energy conservation for frozen layer leads to Eq. (13.10). Equation (13.11) is the conservation of energy at the sublimation interface, where rates of heat flow into the interface from dry and frozen layers are compensated by the rates of heat removed by gas flow and sublimation. As mentioned above complete details and further simplifications of these equations can be found in corresponding Ref. [34].

Non-steady-state model equations representing freeze drying (see Section 13.4) are as follows:

$$\varepsilon \frac{\partial C_{w,g}}{\partial t} + \rho_I \frac{\partial C_{w,s}}{\partial t} = -\nabla \cdot N_w \quad (13.5)$$

$$\frac{\partial C_{w,s}}{\partial t} = -k_g (C_{w,s} - C^*(a_w, T)) \quad (13.6)$$

$$N_w = -k_1 \nabla C_{w,g} - k_2 C_{w,g} \nabla P \quad (13.7)$$

$$N_w = -\frac{M_w}{RT} \cdot K_w \left(1 + \frac{C_{01} P_w}{K_w \mu_{mx}} \right) \nabla P_w \approx -\frac{M_w}{RT} \cdot K_w \nabla P_w \quad (13.8)$$

$$\rho_I C_{pI} \frac{\partial T}{\partial t} = k_I \nabla^2 T + \rho_I \Delta H_v \frac{\partial C_{w,s}}{\partial t} - C_{p,g} \nabla (N_I T) \quad (13.9)$$

$$\rho_{II} C_{pII} \frac{\partial T}{\partial t} = k_{II} \nabla^2 T \quad (13.10)$$

$$-k_I \left(\frac{\partial T}{\partial n} \right)_I + k_{II} \left(\frac{\partial T}{\partial n} \right)_{II} = -N_{in} C_{p,g} T - N_{wn} \Delta H_s \quad (13.11)$$

Utilization of mathematical modeling techniques described above will result in greater understanding of the effect of process variables on the quality of the product. These modeling tools provide an excellent opportunity to apply the quality by design principles to ensure that quality of the freeze-dried product is built into the process. One such practical industrial application of mathematical modeling of freeze drying is demonstrated in freeze drying of azithromycin solution [63]. The authors have confirmed the model predictions of the PASSAGE² software [34, 62] and further utilized the mathematical model to predict the operating conditions at pilot scale to achieve a product temperature profile that is equivalent to the profile achieved for an optimized laboratory-scale lyophilization cycle. This work demonstrated the utility of the numerical models in the area of scale-up and optimization of lyophilization cycles at commercial scale [63].

Several other modeling approaches and applications have been reported in the literature that describe the freeze drying process utilizing fundamental heat and mass transfer models [64–70]. These reports demonstrate that the fundamental understanding of the drying stage has evolved extensively and scientists can utilize these techniques to rapidly develop optimized lyophilization cycles.

13.3.3 Determination of Modeling Parameters: Dry Layer Resistance and Heat Transfer Coefficient

The importance of heat transfer coefficient and mass transfer resistance during drying, especially for modeling purposes, should now be evident to the reader. The manometric temperature measurement (MTM) procedure is one of the techniques that are useful in estimation of the mass transfer resistance offered by the dried layer and stopper.

In the MTM method the flow of water vapor from the product chamber to condenser is momentarily interrupted during primary drying by quickly closing the valve separating the chamber and condenser, resulting in an increase in chamber pressure due to sublimation [71]. This transient increase in chamber pressure is modeled by considering the several factors contributing to the pressure rise. A curvilinear regression estimates the parameters such as vapor pressure of ice and resistance offered by dried layer. The capabilities of this modeling technique have been further examined in estimation of product temperature [54], measurement of dry layer resistance [72], and heat and mass transfer measurements, including the vial heat transfer coefficient [73].

An “expert system” (SMART) that will allow development of an optimized freeze drying process during

laboratory-scale development in one experiment is another application of the MTM technique [74]. SMART is an excellent example of the utilization of heat and mass transfer theory and modeling to facilitate the development of freeze drying cycle conditions. Measurement of product resistance and vapor pressure of ice by MTM allows calculation of the sublimation rate [71], which is then used to optimize the shelf temperature settings. Figure 13.5 summarizes the SMART concept. The expert systems algorithms control the different parts of the freeze drying. Freezing conditions are chosen based on the input parameters regarding formulation details. Primary drying conditions are chosen based on the MTM feedback. Also, the algorithm chooses conditions for secondary drying based on input parameters. A detailed description can be found in the corresponding Ref. [75]. A variation of the original MTM approach, denoted “pressure rise analysis (PRA),” was proposed as an improvement to the original MTM algorithm. Using this PRA model, the authors estimate the same parameters as obtained by MTM, sublimation front temperature, resistance offered by dry layer, and the overall vial heat transfer coefficient. Utility of SMART and MTM during the lyophilization cycle development has been demonstrated in measuring product resistance, predicting product temperature and primary drying time [76].

Another method, based on an analysis of the normal product temperature history during primary drying, has been suggested as a method to obtain product resistance data without doing special experiments, such as the MTM experiment [77]. This technique once again utilizes the understanding of heat and mass transfer mechanisms to determine desired parameters (mass transfer resistance) of the formulation during freeze drying. Other techniques utilize similar applications of heat and mass transfer fundamentals to estimate mass transfer resistances during freeze drying [78–80]. Recently, another variation of the original MTM approach was suggested as a more rigorous model for describing the pressure rise during the valve closing procedure [81]. The stated final goal of using this advanced approach is to develop an online tool for controlling the heating strategy during freeze drying, and the same group recently reported an online monitoring system (and hence control) for primary drying phase of lyophilization as a further application of the mathematical modeling [82]. This system provides in-line control signals (adjusting shelf heating fluid temperature throughout primary drying), utilizing simple heat and mass transfer algorithm in conjunction with measurements made during the freeze drying process to achieve desired product and process performance. While this general procedure is identical to that employed by the MTM procedure within the SMART freeze dryer technology, the details of the heat and mass transfer analysis do differ. Further details can be found in the corresponding Ref. [82].

Another important input parameter for models is the heat transfer coefficient of the vial. This parameter can be

²PASSAGE is commercially available freeze drying software capable to solving unsteady-state mass and heat transfer equations. This is commercially available from Technalysis Inc.

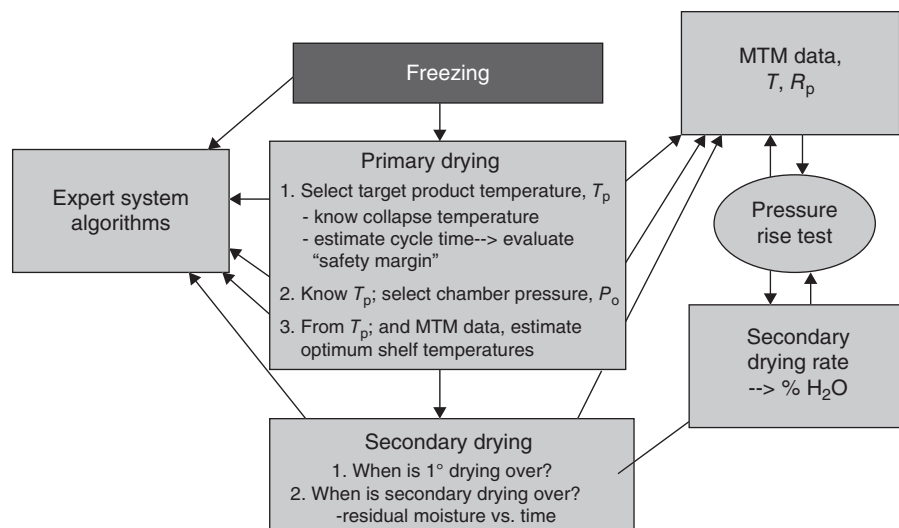


FIGURE 13.5 Summary of the “SMART” freeze dryer concept. *Source:* From Tang et al. [74]. Reproduced with permission from Springer Nature.

determined by performing simple water sublimation experiments in the laboratory as a function of chamber pressure [83], as well as by MTM measurements (or variations on the MTM technique) as noted above.

Successful model development for the drying process depends on the accuracy of the modeling parameters. Dry layer resistance and heat transfer coefficient are two critical parameters. While one might estimate these quantities from an existing database, ideally these parameters should be measured for the application of interest, particularly with the dry layer resistance that varies considerably between products and is relatively difficult to quantitatively predict.

13.3.4 Issues in Scale-Up of Freeze Drying Process

The heat transfer coefficient of a given vial depends not only on the bottom contour of the vial but also on the location of the vial within the vial array in a given freeze dryer. Heat is transferred between shelf and vial due to three mechanisms: heat transfer from the contact area between the vial and the shelf, conduction of gas molecules between vial bottom and shelf surface, and radiation. While the first two modes of heat transfer can be considered independent of the location of the vial, the radiative heat transfer depends on the location of the vial. Edge vials have a greater area of exposure to radiation heat transfer (i.e. part of the side as well as top and bottom), and often the chamber surfaces that are responsible for side radiation are hotter than the shelf surface. These effects result in differences in heat transfer coefficients between center and edge vials [84, 85]. Understanding the impact of such heat transfer coefficient variation on the drying performance will help the scientist design the process so as to achieve optimal product irrespective of its location in

the freeze dryer. A multidimensional unsteady-state modeling was utilized to determine the effect of location of the vials in freeze dryer on the overall drying time temperature distribution [86]. This study shows the importance of wall temperature in influencing the drying characteristics in vials located at different positions in a freeze dryer.

Another important challenge in the freeze drying process is the scale-up of lyophilization cycles from laboratory to pilot to commercial-scale dryers. The edge vial effect can be scale dependent. The heat and mass transfer issues during freeze drying process development have been summarized in an excellent review [87]. In addition to providing a review of heat and mass transfer mechanisms to be considered during freeze drying cycle development, the authors also discuss the related scale-up issues. As mentioned above, radiative heat transfer varies due to the location of the vial in the freeze dryer. This radiation effect is also different from a laboratory dryer to pilot- or commercial-scale dryer. Differences in percentage of edge vials and differences in wall temperatures and differences in emissivities between lab and manufacturing scale introduce scale-up differences that can be significant. For example, a front vial in a laboratory freeze dryer can receive ≈ 1.8 times greater heat transfer than a corresponding vial in a manufacturing freeze dryer [29]. This additional heat input will directly affect the product temperature and drying time. Hence, understanding these differences between freeze dryers is essential to proper scale-up. Another important scale-based difference is the temperature distribution across the freeze dryer shelf. Differences between the shelf temperature set point and measured shelf surface temperature as a function of the sublimation rate are reported [29]. These differences can be dryer specific in that dryer design and heat transfer characteristics of the fluid can cause

shelf surface temperature differences between different freeze dryers even at identical thermal loads. A properly designed shelf mapping study can determine the magnitude of the expected effects [87]. These differences between laboratory- and production-scale dryers have been highlighted, and a step-by-step systematic approach to correlate dryers at two scales leading toward a successful scale-up is discussed in literature [52].

It should be obvious that the application of engineering principles of heat and mass transfer modeling and scale-up adjustments are essential to the successful design of a freeze drying process for manufacturing. Utilizing these engineering principles will facilitate a rational lyophilization cycle development and scale-up effort. The “general rules” for successful process design and scale-up can be summarized as follows:

1. Select the optimal freezing conditions that results in an ice morphology that is uniform within a given vial, uniform between vials in the same batch, uniform between batches, and uniform from laboratory to manufacturing. In addition, a structure composed of larger ice crystals is advantageous in that such a structure produces lower product resistance and faster primary drying.
2. Utilize the understanding of the heat and mass transfer mechanisms during drying to select the primary and secondary conditions to achieve maximal drying rate (minimal drying time) while maintaining the product temperature below the critical product temperature:
 - a. Utilize the techniques described in literature to measure heat transfer coefficient of the vial of interest.
 - b. Determine the resistance offered by dry layer during the drying stage.
 - c. Utilizing mathematical models obtain initial estimates for the first laboratory cycle and then optimize the cycle conditions using the SMART freeze dryer methodology (or equivalent) and/or a few experiments to confirm results.
3. Understand the freeze dryer differences with respect to the heat and mass transfer mechanisms during scale-up and utilize mathematical models to estimate the cycle conditions at larger scale.

Some examples of application of these concepts for rational scale-up have been reported in literature [63, 88, 89]. Heat and mass transfer models in conjunction with limited lyophilization runs were utilized to successfully determine heat transfer coefficients and to evaluate the robustness of the lyophilization cycle at different operating conditions [88]. A systematic approach by utilizing lab-scale experiments to determine heat and mass transfer coefficients and mathematical modeling to predict operating conditions at

pilot scale, thus minimizing the number of pilot-scale runs, has also been reported [63]. Hopefully, increased use of theoretical modeling will be a norm in the future as increased emphasis is placed on “quality by design.”

13.3.5 Mass Flow from Chamber to Condenser

Many advancements in modeling the sublimation and desorption of water vapor within the vial have been made as discussed in Sections 13.3.1 and 13.3.2. The literature on modeling the flow of water vapor once it leaves the vial, i.e. from chamber to condenser, is, however, sparse. The lyophilization process can be limited by mass transfer within the freeze dryer at high sublimation load [90]. In addition, freeze dryer design differences at different scales may lead to different product temperature profiles that may not be captured in vial modeling. Therefore, it is important that these factors be captured in the models for freeze drying. Some recent publications have described the use of various tools to model vapor flow in the freeze dryer. Most of the work cited has used computational fluid dynamics (CFD) for this purpose. The effects of some geometrical parameters of the drying chamber such as clearances between the shelves and position of the duct between the chamber and the condenser on the global fluid dynamics of the sublimated vapor in both small- and industrial-scale drying chamber were investigated as a function of the sublimation rate [91]. It was concluded that local pressure differentials existed in the freeze dryer and contributed to heterogeneity in sublimation rate in addition to the commonly known effects such as radiation effects. These effects were more pronounced in larger-scale freeze dryers. However, the magnitude of these effects was not explicitly discussed. It has been documented that the duct between the drying chamber and the condenser may be a bottleneck at high sublimation rates [90]. This effect has been called choked flow, and CFD has been employed to identify critical variables impacting choked flow [92]. It was determined that ratio of chamber to condenser pressure determined the onset of choked flow. Recently, a more comprehensive model that includes the drying chamber, duct, and condenser using CFD has been published [93]. The Knudsen number (Kn), a dimensionless parameter, has been used to define the flow regime, and various equations that are valid in different flow regimes have been described. Kn is the ratio of the molecular mean free path (λ) and characteristic length scale of flow (L). For example, Navier–Stokes equation has been used where Kn is below 0.01 and continuum hypothesis is valid, i.e. one is dealing with fluid flow or “viscous flow.” The Boltzmann equation has been used in rarefied or “free molecular flow” regime ($Kn > 0.1$). A new approach to solving the Boltzmann equation using a statistical direct simulation Monte Carlo (DSMC) method has been described. Differences between a lab- and industrial-scale freeze dryer, especially the impact of clean-in-place/steam-in-place

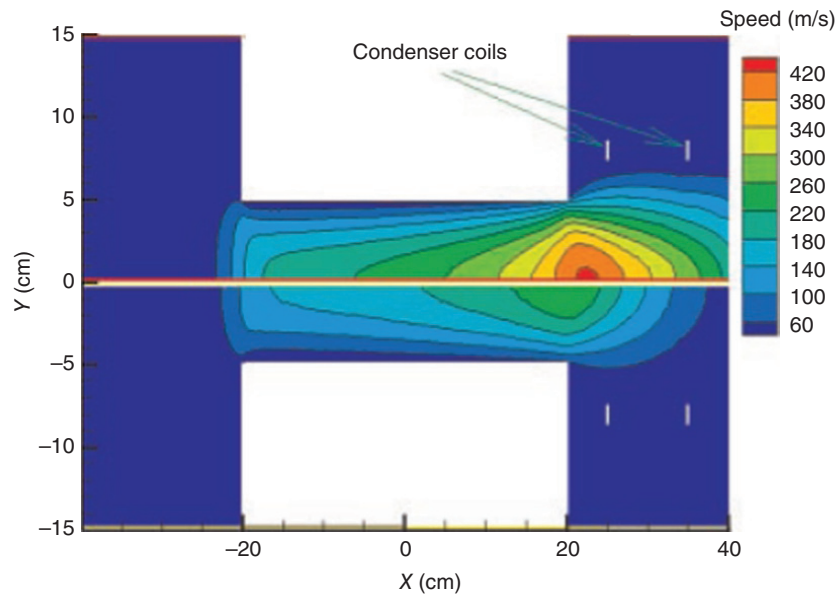


FIGURE 13.6 Velocity profile of water vapor in laboratory-scale freeze dryer. The top half shows that the results for 50 mTorr chamber pressure and choked flow conditions at the condenser entrance are apparent. The lower half shows simulation for 30 mTorr. The vapor velocity in this case remains below the Mach I limit. The data show potential choked flow conditions for some operating parameters. *Source:* From Alexeenko et al. [93]. Reprinted with Permission of Wiley-Liss, Inc. a subsidiary of John Wiley & Sons, Inc.

(CIP/SIP) line in industrial freeze dryer, on the fluid flow were presented. Simulation of water vapor flow in the freeze dryer also showed that choked flow conditions may exist at the point of maximum flow velocity, which is along the axis of the duct at the entrance to the condenser. Figure 13.6 shows the simulation for two different chamber pressures. The top half shows that the results for 50 mTorr chamber pressure and choked flow conditions at the condenser entrance are apparent. The lower half shows simulation for 30 mTorr. The vapor velocity in this case remains below the Mach I limit. It has therefore been argued that computational fluid flow studies may also be useful in the design of freeze dryers.

Following the guidelines presented in this chapter and also the references should result in systematic approach to freeze drying process development and scale-up. Increased use of modeling wherever applicable will be consistent with the “quality by design” philosophy.

13.4 NOMENCLATURE

Steady-State Model Equations

$\frac{dm}{dt}$	Sublimation rate
A_p	Inner cross-sectional area of the vial
P_0	Vapor pressure of ice at interface temperature
P_c	Chamber pressure

R_p	Dry layer resistance
$\frac{dQ}{dt}$	Rate of heat transfer
A_v	Outer cross-sectional area of the vial
K_v	Heat transfer coefficient between vial and the surroundings (includes heat transfer from shelf contact and also radiation)
T_s	Shelf temperature
T_b	Product temperature at the bottom of the vial
T	Product temperature at the sublimation interface
ΔH_s	Heat of sublimation of ice

Non-steady-state Model Equations

ϵ	Void fraction in the dried region
ρ_1	Density of dry region
$C_{w,s}$	Concentration of sorbed water
N_w	Molar flux of water
k_g	Mass transfer coefficient for desorption
C^*	Equilibrium concentration of sorbed water
a_w	Water activity
T	Temperature
k_1	Bulk diffusivity constant
k_2	Self-diffusivity constant
P	Total pressure
C_{pl}	Heat capacity of dry layer
k_1	Thermal conductivity of dry layer
ΔH_v	Heat of vaporization of sorbed water

ρ_{II}	Density of the frozen region
C_{PII}	Heat capacity of frozen region
k_{II}	Thermal conductivity of frozen layer
N_{w_n}	Molar flux of water, normal component
ΔH_s	Heat of sublimation of ice

REFERENCES

- Brange, J. (1987). *Galenics of Insulin*, 1e. Berlin: Springer.
- Pikal, M.J., Lukes, A.L., and Lang, J.E. (1977). Thermal decomposition of amorphous β -lactam antibacterials. *J. Pharm. Sci.* 66 (9): 1312–1316.
- Pikal, M.J., Lukes, A.L., Lang, J.E., and Gaines, K. (1978). Quantitative crystallinity determinations for β -lactam antibiotics by solution calorimetry: correlations with stability. *J. Pharm. Sci.* 67 (6): 767–773.
- Pikal, M.J. and Dellerman, K.M. (1989). Stability testing of pharmaceuticals by high-sensitivity isothermal calorimetry at 25°C: cephalosporins in the solid and aqueous solution states. *Int. J. Pharm.* 50 (3): 233–252.
- Pikal, M.J. and Rigsbee, D.R. (1997). The stability of insulin in crystalline and amorphous solids: observation of greater stability for the amorphous form. *Pharm. Res.* 14 (10): 1379–1387.
- Pikal, M.J. (1990). Freeze-drying of proteins. Part II: Formulation selection. *BioPharm (Eugene, Oreg.)* 3 (9): 26–30.
- Pikal, M.J. (1990). Freeze-drying of proteins. Part I: Process design. *BioPharm* 3 (8): 18–20, 22–24, 26–28.
- Pikal, M.J. (1994). Freeze-drying of proteins: process, formulation, and stability. *ACS Symp. Ser.* 567, (Formulation and Delivery of Proteins and Peptides): 120–133.
- MacKenzie, A.P. (1965). Factors affecting the transformation of ice into water vapor in the freeze-drying process. *Ann. N.Y. Acad. Sci.* 125: 522–547.
- Broadhead, J., Rouan, S.K.E., and Rhodes, C.T. (1992). The spray drying of pharmaceuticals. *Drug Dev. Ind. Pharm.* 18 (11–12): 1169–1206.
- Mumenthaler, M., Hsu, C.C., and Pearlman, R. (1994). Feasibility study on spray-drying protein pharmaceuticals: recombinant human growth hormone and tissue-type plasminogen activator. *Pharm. Res.* 11 (1): 12–20.
- Masters, K. (ed.) (1991). Applications of spray-drying: the food industry; the pharmaceutical-biochemical industry. In: *Spray-Drying Handbook*, 491–676. Essex: Longman Scientific and Technical.
- Franks, F. (1993). Solid aqueous solutions. *Pure Appl. Chem.* 65 (12): 2527–2537.
- Roser, B. (1991). Trehalose drying: a novel replacement for freeze-drying. *BioPharm (Eugene, Oreg.)* 4 (8): 47–53.
- Abdul-Fattah, A.M., Kalonia, D.S., and Pikal, M.J. (2007). The challenge of drying method selection for protein pharmaceuticals: product quality implications. *J. Pharm. Sci.* 96 (8): 1886–1916.
- Levine, H. and Slade, L. (1992). Another view of trehalose for drying and stabilizing biological materials. *BioPharm (Eugene, Oreg.)* 5 (4): 36–40.
- Pregolato, F. and Curto, P. (1992). Parental lyophilization facilities: an innovative approach to loading and unloading operations. *Proceedings of the International Congress, Advanced Technologies for Manufacturing of Aseptic and Terminally Sterilized Pharmaceuticals and Biopharmaceuticals*, Basel, Switzerland (17–19 February 1992), pp. 4–30.
- Carpenter, J.F., Crowe, J.H., and Arakawa, T. (1990). Comparison of solute-induced protein stabilization in aqueous solution and in the frozen and dried states. *J. Dairy Sci.* 73 (12): 3627–3636.
- Tanaka, K., Takeda, T., and Miyajima, K. (1991). Cryoprotective effect of saccharides on denaturation of catalase by freeze-drying. *Chem. Pharm. Bull.* 39 (5): 1091–1094.
- Hellman, K., Miller, D.S., and Cammack, K.A. (1983). The effect of freeze-drying on the quaternary structure of L-asparaginase from *Erwinia carotovora*. *Biochim. Biophys. Acta Protein Struct. Mol. Enzymol.* 749 (2): 133–142.
- Ressing, M.E., Jiskoot, W., Talsma, H. et al. (1992). The influence of sucrose, dextran, and hydroxypropyl β -cyclodextrin as lyoprotectants for a freeze-dried mouse IgG2a monoclonal antibody (MN12). *Pharm. Res.* 9 (2): 266–270.
- Izutsu, K.-I. and Yoshioka, S. (1995). Stabilization of protein pharmaceuticals in freeze-dried formulations. *Drug Stab.* 1 (1): 11–21.
- Bell, L.N. and Hageman, M.J. (1994). Differentiating between the effects of water activity and glass transition dependent mobility on a solid state chemical reaction: aspartame degradation. *J. Agric. Food Chem.* 42 (11): 2398–2401.
- Pikal, M.J., Dellerman, K.M., Roy, M.I., and Riggan, R.M. (1991). The effects of formulation variables on the stability of freeze-dried human growth hormone. *Pharm. Res.* 8 (4): 427–436.
- Pikal, M.J., Dellerman, K., and Roy, M.L. (1992). Formulation and stability of freeze-dried proteins: effects of moisture and oxygen on the stability of freeze-dried formulations of human growth hormone. *Dev. Biol. Stand.* 74: 21–38.
- Townsend, M.W. and DeLuca, P.P. (1988). Use of lyoprotectants in the freeze-drying of a model protein, ribonuclease A. *J. Parenter. Sci. Technol.* 42 (6): 190–199.
- Prestrelski, S.J., Pikal, K.A., and Arakawa, T. (1995). Optimization of lyophilization conditions for recombinant human interleukin-2 by dried-state conformational analysis using Fourier-transform infrared spectroscopy. *Pharm. Res.* 12 (1259): 1250–1259.
- Roy, M.L., Pikal, M.J., Rickard, E.C., and Maloney, A.M. (1992). The effects of formulation and moisture on the stability of a freeze-dried monoclonal antibody-vinca conjugate: a test of the WLF glass transition theory. *Dev. Biol. Stand.* 74: 323–340.
- Rambhatla, S., Tchessalov, S., and Pikal Michael, J. (2006). Heat and mass transfer scale-up issues during freeze-drying. III: Control and characterization of dryer differences via operational qualification tests. *AAPS PharmSciTech* 7 (2): E39.
- Roy, M.L. and Pikal, M.J. (1989). Process control in freeze drying: determination of the end point of sublimation drying by an electronic moisture sensor. *J. Parenter. Sci. Technol.* 43 (2): 60–66.
- Searles, J.A., Carpenter, J.F., and Randolph, T.W. (2001). The ice nucleation temperature determines the primary drying rate

- of lyophilization for samples frozen on a temperature-controlled shelf. *J. Pharm. Sci.* 90 (7): 860–871.
32. Rambhatla, S., Ramot, R., Bhugra, C., and Pikal Michael, J. (2004). Heat and mass transfer scale-up issues during freeze drying: II. Control and characterization of the degree of supercooling. *AAPS PharmSciTech* 5 (4): e58.
 33. Pikal, M.J. (1985). Use of laboratory data in freeze drying process design: heat and mass transfer coefficients and the computer simulation of freeze drying. *J. Parenter. Sci. Technol.* 39 (3): 115–139.
 34. Pikal, M.J., Cardon, S., Bhugra, C. et al. (2005). The nonsteady state modeling of freeze drying: in-process product temperature and moisture content mapping and pharmaceutical product quality applications. *Pharm. Dev. Technol.* 10 (1): 17–32.
 35. Gilra, N.K. (1968). Homogeneous nucleation temperature of supercooled water. *Phys. Lett. A* 28 (1): 51–52.
 36. Tang, X. and Pikal, M.J. (2004). Design of freeze-drying processes for pharmaceuticals: practical advice. *Pharm. Res.* 21 (2): 191–200.
 37. Pikal, M.J. and Shah, S. (1990). The collapse temperature in freeze drying: dependence on measurement methodology and rate of water removal from the glassy phase. *Int. J. Pharm.* 62 (2–3): 165–186.
 38. Chatterjee, K., Shalaev, E.Y., and Suryanarayanan, R. (2005). Partially crystalline systems in lyophilization: II. Withstanding collapse at high primary drying temperatures and impact on protein activity recovery. *J. Pharm. Sci.* 94 (4): 809–820.
 39. Milton, N. and Nail, S.L. (1996). The physical state of nafcillin sodium in frozen aqueous solutions and freeze-dried powders. *Pharm. Dev. Technol.* 1 (3): 269–277.
 40. Searles, J.A., Carpenter, J.F., and Randolph, T.W. (2001). Annealing to optimize the primary drying rate, reduce freezing-induced drying rate heterogeneity, and determine T_g' in pharmaceutical lyophilization. *J. Pharm. Sci.* 90 (7): 872–887.
 41. Chouvenec, P., Vessot, S., and Andrieu, J. (2006). Experimental study of the impact of annealing on ice structure and mass transfer parameters during freeze-drying of a pharmaceutical formulation. *PDA J. Pharm. Sci. Technol.* 60 (2): 95–103.
 42. Lu, X. and Pikal Michael, J. (2004). Freeze-drying of mannitol-trehalose-sodium chloride-based formulations: the impact of annealing on dry layer resistance to mass transfer and cake structure. *Pharm. Dev. Technol.* 9 (1): 85–95.
 43. Pikal, M.J., Shah, S., Roy, M.L., and Putman, R. (1990). The secondary drying stage of freeze drying: drying kinetics as a function of temperature and chamber pressure. *Int. J. Pharm.* 60 (3): 203–217.
 44. Heller, M.C., Carpenter, J.F., and Randolph, T.W. (1997). Manipulation of lyophilization-induced phase separation: implications for pharmaceutical proteins. *Biotechnol. Prog.* 13 (5): 590–596.
 45. Rowe, T.D. (1990) A technique for nucleation of ice. *International Symposium on Biological Product Freeze-Drying and Formulation*, Geneva, Switzerland (26 October 1990).
 46. Nakagawa, K., Hottot, A., Vessot, S., and Andrieu, J. (2006). Influence of controlled nucleation by ultrasounds on ice morphology of frozen formulations for pharmaceutical proteins freeze-drying. *Chem. Eng. Process.* 45 (9): 783–791.
 47. Petersen, A., Rau, G., and Glasmacher, B. (2006). Reduction of primary freeze-drying time by electric field induced ice nucleus formation. *Heat Mass Transf.* 42 (10): 929–938.
 48. Petersen, A., Schneider, H., Rau, G., and Glasmacher, B. (2006). A new approach for freezing of aqueous solutions under active control of the nucleation temperature. *Cryobiology* 53 (2): 248–257.
 49. Liu, J., Viverette, T., Virgin, M. et al. (2005). A study of the impact of freezing on the lyophilization of a concentrated formulation with a high fill depth. *Pharm. Dev. Technol.* 10 (2): 261–272.
 50. Kramer, M., Sennhenn, B., and Lee, G. (2002). Freeze-drying using vacuum-induced surface freezing. *J. Pharm. Sci.* 91 (2): 433–443.
 51. Nakagawa, K., Hottot, A., Vessot, S., and Andrieu, J. (2007). Modeling of freezing step during freeze-drying of drugs in vials. *AIChE J.* 53 (5): 1362–1372.
 52. Kuu, W.Y., Hardwick, L.M., and Akers, M.J. (2005). Correlation of laboratory and production freeze drying cycles. *Int. J. Pharm.* 302 (1–2): 56–67.
 53. Sadikoglu, H., Ozdemir, M., and Seker, M. (2006). Freeze-drying of pharmaceutical products: research and development needs. *Dry. Technol.* 24 (7): 849–861.
 54. Tang, X., Nail Steven, L., and Pikal Michael, J. (2006). Evaluation of manometric temperature measurement, a process analytical technology tool for freeze-drying: part I, product temperature measurement. *AAPS PharmSciTech* 7 (1): E14.
 55. Trelea, I.C., Passot, S., Fonseca, F., and Marin, M. (2007). An interactive tool for the optimization of freeze-drying cycles based on quality criteria. *Dry. Technol.* 25 (5): 741–751.
 56. Litchfield, R.J. and Liapis, A.I. (1979). An adsorption-sublimation model for a freeze dryer. *Chem. Eng. Sci.* 34 (9): 1085–1090.
 57. Millman, M.J., Liapis, A.I., and Marchello, J.M. (1985). An analysis of the lyophilization process using a sorption-sublimation model and various operational policies. *AIChE J.* 31 (10): 1594–1604.
 58. Liapis, A.I. and Marchello, J.M. (1984). Advances in the modeling and control of freeze-drying. *Adv. Dry.* 3: 217–244.
 59. Liapis, A.I. and Bruttini, R. (1995). Freeze-drying of pharmaceutical crystalline and amorphous solutes in vials: dynamic multi-dimensional models of the primary and secondary drying stages and qualitative features of the moving interface. *Dry. Technol.* 13 (1): 43–72.
 60. Sheehan, P. and Liapis, A.I. (1998). Modeling of the primary and secondary drying stages of the freeze drying of pharmaceutical products in vials: numerical results obtained from the solution of a dynamic and spatially multi-dimensional lyophilization model for different operational policies. *Biotechnol. Bioeng.* 60 (6): 712–728.
 61. Sadikoglu, H. and Liapis, A.I. (1997). Mathematical modeling of the primary and secondary drying stages of bulk solution freeze-drying in trays: parameter estimation and model discrimination by comparison of theoretical results with experimental data. *Dry. Technol.* 15 (3 & 4): 791–810.

62. Mascarenhas, W.J., Akay, H.U., and Pikal, M.J. (1997). A computational model for finite element analysis of the freeze-drying process. *Comput. Methods Appl. Mech. Eng.* 148: 105–124.
63. Kramer, T., Kremer, D.M., Pikal, M.J. et al. (2009). A procedure to optimize scale-up for the primary drying phase of lyophilization. *J. Pharm. Sci.* 98 (1): 307–318.
64. Nastaj, J. and Witkiewicz, K. (2004). Numerical model of freeze drying of random solids at two-region conductive-radiative heating. *Inz. Chem. Procesowa* 25 (1): 109–121.
65. Boss, E.A., Maciel Filho, R. II, and Vasco de Toledo, E.C. III (2002). Freeze drying process: real time model. *Tech. Pap. ISA* 426: 59–70.
66. Tu, W., Chen, M., Yang, Z., and Chen, H. (2000). A mathematical model for freeze-drying. *Chin. J. Chem. Eng.* 8 (2): 118–122.
67. Kisakurek, B. and Celiker, H. (1984). A modified moving boundary model for freeze-dryers. *Proc. Int. Drying Symp., 4th FIELD Full Journal Title 2*: 420–424.
68. Hottot, A.I., Peczkalski, R., Vessot, S.v., and Andrieu, J. (2006). Freeze-drying of pharmaceutical proteins in vials: modeling of freezing and sublimation steps. *Dry. Technol.* 24 (5): 561–570.
69. Boss, E.A., Filho, R.M., and de Toledo, E.C.V. (2004). Freeze drying process: real time model and optimization. *Chem. Eng. Process.* 43 (12): 1475–1485.
70. Velardi, S.A. and Barresi, A.A. (2008). Development of simplified models for the freeze-drying process and investigation of the optimal operating conditions. *Chem. Eng. Res. Design* 86 (1): 9–22.
71. Milton, N., Pikal, M.J., Roy, M.L., and Nail, S.L. (1997). Evaluation of manometric temperature measurement as a method of monitoring product temperature during lyophilization. *PDA J. Pharm. Sci. Technol.* 51 (1): 7–16.
72. Tang, X.C., Nail Steven, L., and Pikal Michael, J. (2006). Evaluation of manometric temperature measurement, a process analytical technology tool for freeze-drying: part II, measurement of dry-layer resistance. *AAPS PharmSciTech* 7 (4): 93.
73. Tang, X.C., Nail Steven, L., and Pikal Michael, J. (2006). Evaluation of manometric temperature measurement (MTM), a process analytical technology tool in freeze drying: part III, heat and mass transfer measurement. *AAPS PharmSciTech* 7 (4): 97.
74. Tang, X., Nail, S.L., and Pikal, M.J. (2005). Freeze-drying process design by manometric temperature measurement: design of a smart freeze-dryer. *Pharm. Res.* 22 (4): 685–700.
75. Chouvenc, P., Vessot, S., Andrieu, J., and Vacus, P. (2004). Optimization of the freeze-drying cycle: a new model for pressure rise analysis. *Dry. Technol.* 22 (7): 1577–1601.
76. Gieseler, H., Kramer, T., and Pikal, M.J. (2007). Use of manometric temperature measurement (MTM) and SMART freeze dryer technology for development of an optimized freeze-drying cycle. *J. Pharm. Sci.* 96 (12): 3402–3418.
77. Kuu, W.Y., Hardwick, L.M., and Akers, M.J. (2006). Rapid determination of dry layer mass transfer resistance for various pharmaceutical formulations during primary drying using product temperature profiles. *Int. J. Pharm.* 313 (1–2): 99–113.
78. Pikal, M.J., Shah, S., Senior, D., and Lang, J.E. (1983). Physical chemistry of freeze-drying: measurement of sublimation rates for frozen aqueous solutions by a microbalance technique. *J. Pharm. Sci.* 72 (6): 635–650.
79. Kuu, W.-Y., McShane, J., and Wong, J. (1995). Determination of mass transfer coefficients during freeze drying using modeling and parameter estimation techniques. *Int. J. Pharm.* 124 (2): 241–252.
80. Zhai, S., Su, H., Taylor, R., and Slater, N.K.H. (2005). Pure ice sublimation within vials in a laboratory lyophiliser; comparison of theory with experiment. *Chem. Eng. Sci.* 60 (4): 1167–1176.
81. Velardi, S.A., Rasetto, V., and Barresi, A.A. (2008). Dynamic parameters estimation method: advanced manometric temperature measurement approach for freeze-drying monitoring of pharmaceutical solutions. *Ind. Eng. Chem. Res.* 47 (21): 8445–8457.
82. Barresi, A.A., Pisano, R., Fissore, D. et al. (2009). Monitoring of the primary drying of a lyophilization process in vials. *Chem. Eng. Process. Process Intensif.* 48 (1): 408–423.
83. Pikal, M.J., Roy, M.L., and Shah, S. (1984). Mass and heat transfer in vial freeze-drying of pharmaceuticals: role of the vial. *J. Pharm. Sci.* 73 (9): 1224–1237.
84. Rambhatla, S. and Pikal, M.J. (2003). Heat and mass transfer scale-up issues during freeze-drying. I: atypical radiation and the edge vial effect. *AAPS PharmSciTech* 4 (2): 111–120.
85. Brülls, M. and Rasmuson, A. (2002). Heat transfer in vial lyophilization. *Int. J. Pharm.* 246 (1–2): 1–16.
86. Gan, K.H., Bruttini, R., Crosser, O.K., and Liapis, A.I. (2005). Freeze-drying of pharmaceuticals in vials on trays: effects of drying chamber wall temperature and tray side on lyophilization performance. *Int. J. Heat Mass Transfer* 48 (9): 1675–1687.
87. Rambhatla, S. and Pikal, M.J. (2004). Heat and mass transfer issues in freeze-drying process development. *Biotechnol. Pharm. Asp.* 2, (Lyophilization of Biopharmaceuticals): 75–109.
88. Tsinontides, S.C., Rajniak, P., Pham, D. et al. (2004). Freeze drying: principles and practice for successful scale-up to manufacturing. *Int. J. Pharm.* 280 (1–2): 1–16.
89. Tchessalov, S., Dixon, D., and Nick, W. (2007). Principles of lyophilization scale up. *Am. Pharm. Rev.* 10 (3): 88–93.
90. Searles, J. (2004). Observation and implications of sonic water vapor flow during freeze-drying. *Am. Pharm. Rev.* 7 (2): 58, 60, 62, 64, 66–68, 75.
91. Rasetto, V.M., Daniele, L., and Baressi, A. (2008). Computational fluid dynamics in freeze-drying technology. In: *Freeze Drying of Pharmaceuticals and Biologicals*. Breckenridge, CO August 6–9.
92. Patel, S.M., Swetaprovo, C., and Pikal, M.J. (2008). Choked flow and importance of Mach I in freeze-drying process design. In: *Freeze Drying of Pharmaceutical and Biologicals*. Breckenridge, CO August 6–9.
93. Alexeenko, A.A., Ganguly, A., and Nail, S.L. (2009). Computational analysis of fluid dynamics in pharmaceutical freeze-drying. *J. Pharm. Sci.* 98 (9): 3483–3494.

14

STERILIZATION PROCESSES IN THE PHARMACEUTICAL INDUSTRY

PIERO M. ARMENANTE

Otto H. York Department of Chemical and Materials Engineering, New Jersey Institute of Technology, University Heights, Newark, NJ, USA

OTUTE AKITI

BlackThorn Therapeutics, San Francisco, CA, USA

14.1 INTRODUCTION

Industrial sterilization processes can be defined as those operations having as objective the destruction, permanent inactivation, or physical removal of all microorganisms, such as bacteria and other biological agents, initially present in equipment or devices (e.g. a bioreactor), process materials (e.g. air sparged in a fermenter), or products (e.g. parenteral drugs, syringes). A number of approaches can be utilized for this purpose (e.g. high temperature, radiation), but all required not only to inactivate microorganisms but also to preserve as much as possible the integrity of the material undergoing sterilization.

Sterility can be defined as “the quality or condition of being sterile,” where *sterile* is defined in this contest as being “free from living organisms and especially pathogenic microorganisms.” From here it follows that sterilization refers to any process that eliminates, removes, kills, or deactivates all forms of life present in a specified region, such as a medication or culture media (WHO, Glossary). So, although sterility seems to be an absolute concept, the reality is that is not possible to ensure *absolute* sterility in any real system through *actual* sterilization processes, because of the initial presence of an enormous number of microorganisms that must be typically inactivated and the excessively harsh conditions that should be achieved uniformly in the material to be sterilized (almost inevitably damaging the equipment or product) in order to fully achieve sterility. In fact, and more sensibly, the Code of Federal Regulations (21 CFR Part 600.3 [1]) states that “sterility is interpreted to mean freedom

from viable contaminating microorganisms, as determined by § 610.12...” i.e. *through appropriate tests*. On the other hand, what is achievable is to put in place sterilization processes that ensure that the probability of survival of undesired microorganisms, typically bacteria, is reduced to an arbitrary but predefined extremely low level.

By necessity any sterilization process is designed to create a harsh environment for microorganisms, e.g. through the establishment of high temperatures or the exposure to toxic compounds, which often result in collateral damage to the material or product that is intended to be made sterile. Therefore, it is essential to recognize that any well-designed sterilization process is always the result of a trade-off between achieving the desired sterilization objective on the one hand and, on the other hand, ensuring that the material/product to be sterilized is still intact, or at least damaged to an acceptable level, at the end of the sterilization process. This inevitably results in some sterilization processes being preferred to others, depending on the specific circumstances. For example, sterilizing a large molecule therapeutic protein product using a typically cost-effective high-temperature sterilization process would be highly inappropriate since it would effectively ensure that the protein is completely deactivated at the end. Similarly, gaseous toxic chemicals are routinely used to sterilize surgical instruments, but not used to sterilize drug products since they would leave behind toxic residues that could eventually end up in the patient’s body. Finally, the cost of the sterilization process, in terms of both capital expenditure and operation expenses, is obviously an important factor that must be accounted for

in the selection of the most appropriate and cost-effective sterilization method.

14.2 PRELIMINARY MICROBIOLOGY CONCEPTS RELEVANT TO STERILIZATION

Microorganisms can be defined as all living organisms that are visible only under the microscope. Therefore, before examining sterilization processes, it is important to briefly examine the main intended targets of sterilization, i.e. the microorganisms themselves and their characteristics. Modern taxonomy broadly classifies all living organisms into three domains: eukaryota, bacteria, and archaea (the last two cumulatively being referred to as “prokaryotic organisms” or prokaryotes). The first domain includes all eukaryotic organisms (eukaryotes), i.e. unicellular or multicellular organisms whose cells are large (typically 10–20 μm) and contain a well-defined nucleus and other structures (organelles) enclosed within membranes. Eukaryotic organisms include all animals, plants, and fungi, but only some of them are human pathogens, such as protozoans and yeasts. However, they are not typically critical to sterilization processes, because any sterilization process inactivating or removing more resistant microorganisms such as prokaryotes will also do the same to eukaryotes.

Prokaryotes are unicellular organisms characterized by the absence of a membrane-bound nucleus, mitochondria, or any other membrane-bound organelle. They are relatively simple, more primitive, and smaller (0.5–3 μm) if compared with eukaryotes. They are extremely widespread in the environment, are capable of duplicating rapidly (typically within 20 minutes), and are very adaptable and versatile in their acceptance of a large variety of nutrients, temperature conditions, and environments. Like all cells, they are provided with a cell membrane (about 70 \AA thick), but they are often also surrounded by a rigid cell wall (about 200 \AA thick). Bacteria are the most significant of the prokaryotic microorganisms, and they are by far the most important and abundant class of microorganisms present in the environment. Depending on their morphology, visible under the microscope, they are classified as cocci (spherical), bacilli (straight rods), vibrio (curved rods), spirillum (spiral), and spirochete (corkscrewed shape).

Most human pathogenic bacteria, the most relevant as far as industrial sterilization is concerned, have evolved to live optimally at temperatures between 30 and 37 $^{\circ}\text{C}$ and at a $\text{pH} \approx 7.0$. All actively growing bacteria, including pathogens, are typically present in the so-called *vegetative* state, i.e. the normal state in which they actively metabolize nutrients and can rapidly duplicate and form bacterial colonies. However, while in this state, they have also relatively limited resistance to high temperatures (they are typically and rapidly inactivated even at mild temperatures as low as 70 $^{\circ}\text{C}$) or the action

of even mild chemical disinfectants (e.g. alcohol). Unfortunately for sterilization, some bacterial species (typically Gram-positive bacteria such as *Bacillus* and *Clostridium*), including those responsible for very serious diseases such as anthrax and tetanus, can form *endospores* (often refer to simply as “spores” in this context) under adverse conditions. A spore-forming bacterium can morph itself into a dormant, highly resistant, and nonreproductive spore structure according to the scheme shown in Figure 14.1. Endospores consist simply of the bacterium’s DNA, ribosomes, and large amounts of dipicolinic acid. Bacteria colonies start forming endospores in order to protect themselves against unfavorable environments, including very high or low temperatures, absence of water (desiccation), high radiation intensities, and the presence of toxic chemicals. Figure 14.2 shows the micrograph of a bacteria population undergoing sporulation (i.e., spore formation) [2]. As an endospore, a bacterium can remain dormant for extended periods of times, even years, and even in the total absence of nutrients or water. However, upon the reestablishment of more favorable conditions around it, an endospore can morph itself back into the vegetative bacterial state and resume normal functions, including nutrient uptake and duplication. Clearly the presence of bacterial endospores poses significant challenges to the achievement of sterility because of their resistance.

Finally, and just for completeness, one should mention viruses, small infectious agents that can be classified as intracellular parasites, capable of replicating only inside the living cells. Their size is in the range 0.02–0.2 μm , i.e. one to two orders of magnitude smaller than bacteria. Unlike prokaryotic and eukaryotic cells, viruses have no cell structure (i.e. no nucleus, cytoplasm, membrane, or cell wall). They only consist of a nucleic acid core containing the virus genetic material (either RNA or DNA) surrounded by a protein coating (capsid) and, typically, a lipid-containing coating. Unlike living parasitic organisms that utilize the host cell to get nutrients, viruses only alter the metabolism of the host cell so that new viruses (instead of cellular material) are produced. Outside the cell viruses are completely inert. Viruses do not reproduce by dividing or by forming spores. However, once inside the cell, their genetic material utilizes the cellular machinery to manufacture more viruses, which are then released after the cell breaks (i.e. it lyses). A number of viruses can be pathogenic to humans if ingested or if they enter the blood stream. In general, virus inactivation is virus specific, but some viruses can be inactivated by some of the same sterilization methods used for microorganisms [3]. However, sterilization processes target only microorganisms, not viruses, and virus inactivation is not usually a requirement of sterilization processes. However, some typical sterilization processes also result in the deactivation or removal of some classes of viruses. In addition, a virus clearance step can now be found in the biomanufacturing of some biological products to specifically reduce the viral load.

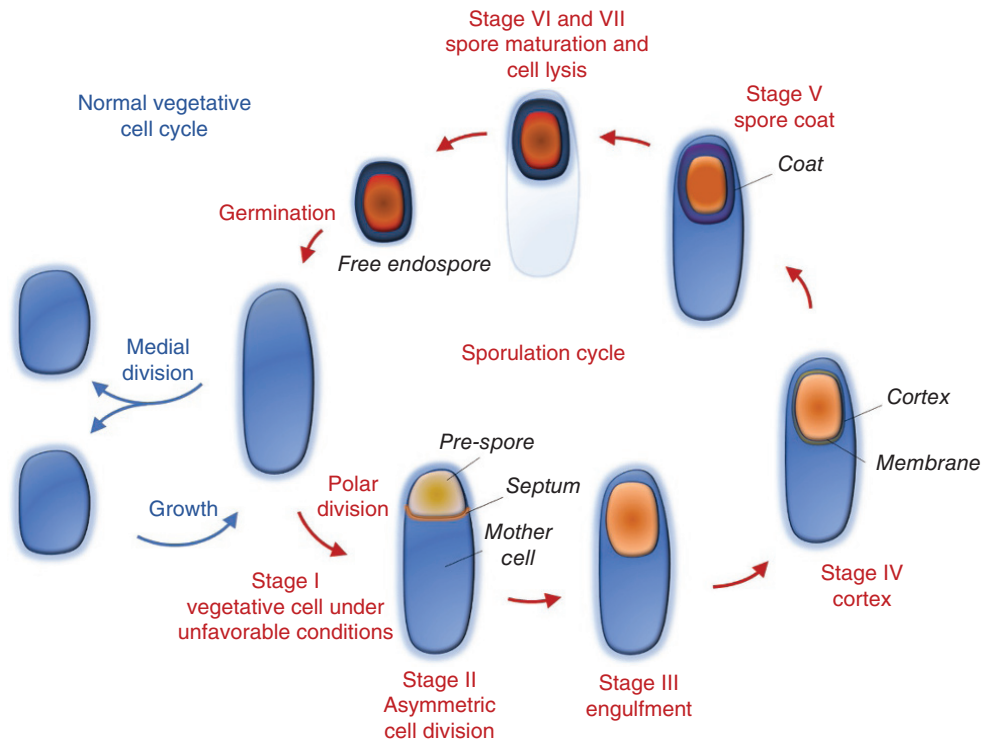


FIGURE 14.1 Typical sporulation cycle for an endospore-forming microorganism. *Source:* Reproduced with permission of Springer Nature.

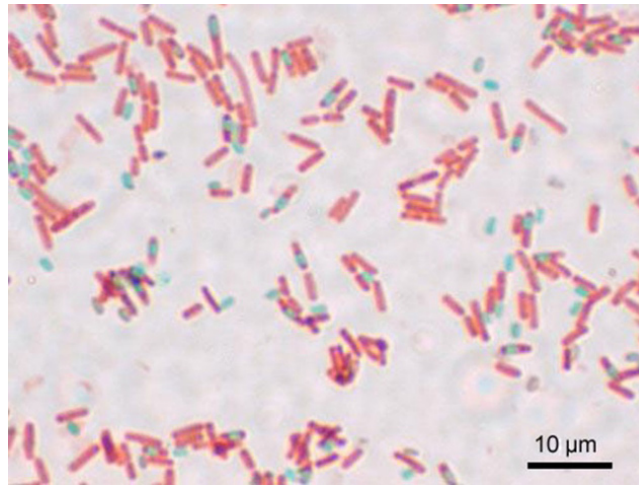


FIGURE 14.2 Microscopic image of the bacterial spore formation of *Bacillus subtilis* (ATCC 6633). Spore staining, magnification : 1000. (Black) spores, (gray) vegetatives. *Source:* Wikipedia.

14.3 OVERVIEW OF STERILIZATION METHODS AND INDUSTRIAL STERILIZATION PROCESSES

Having reviewed some very basic microbiology, it is now possible to examine the sterilization methods commonly used in the industrial practice. One way to classify sterilization methods is by examining the method through which

microbial inactivation is achieved. Accordingly, sterilization methods can be classified as follows [4–9]:

1. **Microbial destructive methods.** These methods achieve complete obliteration of the microbial biomass by transforming biological cell components (proteins, amino acids, nucleic acids, etc.) into water, carbon dioxide, and inorganic chemical species. This can be

achieved only by imposing extreme temperatures on the biomass such as those found in incinerators, by directly flaming inoculation loops in the laboratory, or by using highly concentrated oxidizing acids. This type of approaches has only very limited use since many products can undergo such an extreme treatment without damage. Therefore, these methods are not commonly used industrially other than for the incineration of wastes.

2. **Permanent microbial inactivation methods.** These methods rely on the use of physical or chemical treatments to alter key components of the biological machinery of the cell to make the cells, including endospores, nonviable. As a result, at the end of the treatment, the basic conformation of the cells can possibly still be observed under a microscope (Figure 14.3), but the cells' biochemistry has been so severely disrupted that the cells are no longer capable of conducting any biological process (i.e. the cells are no longer viable). These processes are very commonly used industrially and include the following:

a. **Thermal sterilization methods**, such as steam (moist) sterilization and dry heat sterilization, consisting of keeping the microorganisms at a sufficiently high temperature for a sufficiently long period of time. Thermal sterilization is based on the disruptive effects of temperature on the chemistry of life and is therefore the oldest, most

dependable, and most common sterilization method. In addition, it is the most important method used industrially to sterilize high volume liquids, including fermentation media, and it is often one of the most cost effective.

- b. **Irradiation and high energy sterilization methods**, relying on the use of high energy radiation (e.g. gamma radiation) or high energy physical state (plasma). High-intensity penetrating ionizing radiations, such as the gamma rays emitted from the decay of radioactive materials, have a significant sterilizing effect. They are now used in many industrial applications including all those cases in which thermal sterilization is not applicable because of the thermal lability of the product (e.g. drug) or the devices (e.g. syringes) undergoing sterilization. In addition, lower intensity radiation emitters are often utilized in maintaining asepsis (i.e. the continuous state of exclusion of harmful organisms) in particular areas such as laboratories, sterile cabinets, water-for-injection loops, and equipment. Ultraviolet (UV) radiation is especially useful for this purpose. Because of its limited penetration power, UV light is used especially for maintaining sterility in air, on walls, and on other flat surfaces without creating severely hazardous working conditions.
- c. **Chemical sterilization methods**, consisting of exposing the microorganisms to some highly toxic

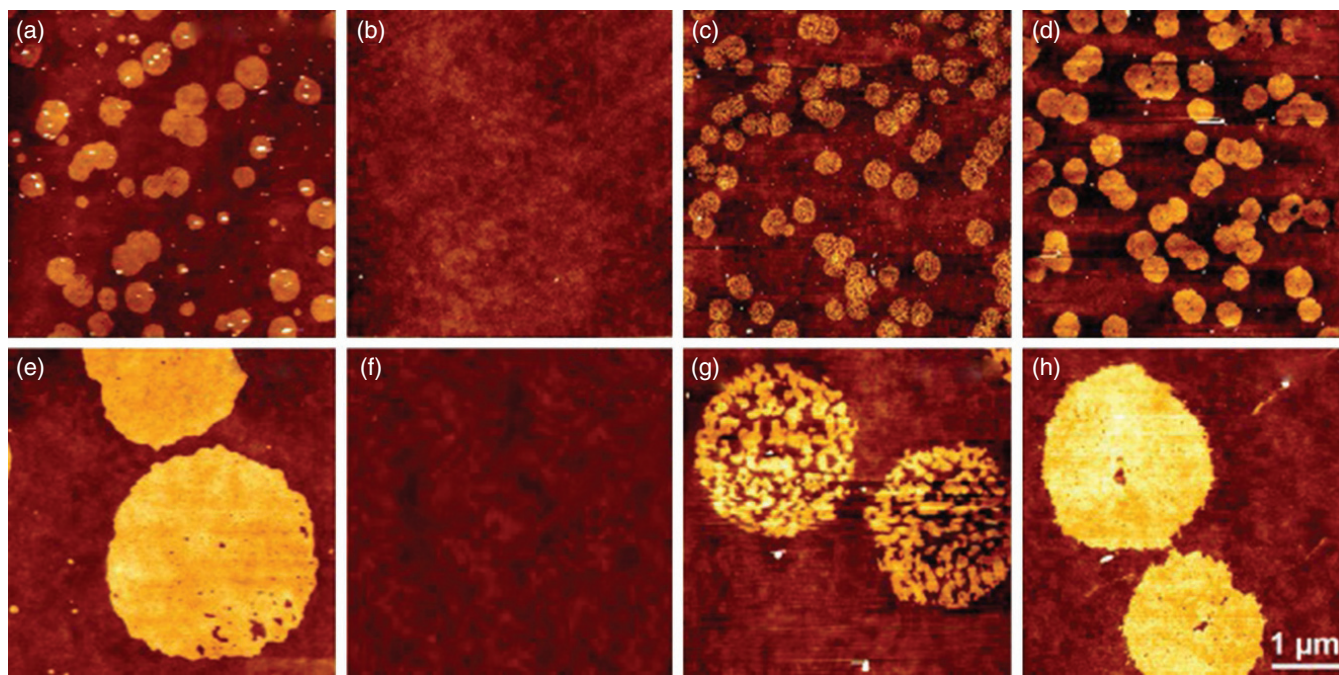


FIGURE 14.3 Microscope images showing the effects of different sterilization treatments on cells: (a) unreacted, (b) steam, (c) plasma, (d) gamma, (e) unreacted, (f) steam, (g) plasma, and (h) gamma [10]. *Source:* Reproduced with permission of John Wiley & Sons.

liquids or, much more often, sterilant gases. The most commonly used sterilizing gaseous chemicals are chlorine, ethylene oxide, propylene oxide, hydrogen peroxide, ozone, and chlorine dioxide. They are primarily used to sterilize equipment or medical supplies. Their main disadvantages are that they have a relatively low penetration effectiveness, may not be particularly effective against spores and their residuals may interfere with subsequent processes or with the product. Furthermore, their chemical reactivity makes them a potential health hazard for the operators, and special precautions must be taken to avoid or minimize human exposure. This, in turn, introduces handling problems. Chemical sterilants are also extensively used as disinfectants in laboratories, clean rooms, and other fixed installations in which the primary aim is a significant reduction of the microbial count (disinfection) rather than the complete inactivation of all microorganisms (sterilization). Toxic liquids such as solutions of dissolved chlorine, formaldehyde, quaternary ammonium salts, ozone, hydrogen peroxide, and peracetic acid are used primarily for disinfection.

3. **Methods based on the physical removal/exclusion of microorganisms (filtration methods).** These methods rely on the use of some type of filters, such as depth filters and, especially and most importantly, membrane sterilizing filters, allowing the liquid or gas to be sterilized to pass but preventing the microorganisms from going through. Sterile filtration is achieved by using membrane filters with uniformly small pore sizes (0.22 or 0.11 μm) capable of blocking even the smallest known bacterium, *Brevundimonas diminuta*. Clearly, both the filter and the system downstream of it, such as a receiving container or apparatus, must have been previously sterilized using one of the other sterilization methods. Filtration is extensively used for the sterilization of gases (especially air) because of the large volumes of fluid involved, low initial bacterial count, and the lower energy consumption and costs associated with it. Liquid filtration, and especially sterile filtration, finds important applications especially in biopharmaceutical manufacturing, but it imposes stringent demands on the filtering element (e.g. the ability to withstand high-pressure differences across the filter element, high filtration area, resistance to conventional sterilization method, such as thermal sterilization, in order to sterilize the filter). In addition, the in-process absolute proof of filtration effectiveness is hard to establish despite the many improvements made in the fabrication of the filters (e.g. well-defined pore size distribution, predictable pressure differentials, improved chemical compatibility).

The choice of one of the above sterilization methods and, from this, the selection of the specific sterilization process to be used in each case will depend on the type of material to be sterilized, its intended use, the effect of the sterilization method on the product, and the equipment and operating costs. In the pharmaceutical and biotechnology industries, different sterilization methods are applied to materials, equipment, and products of different chemical composition and physical characteristics. A list of the most common applications is provided in Table 14.1 [11]. The corresponding, most appropriate sterilization methods will depend on such properties and, especially, the effect of the sterilization method on the product.

In summary, nearly all sterilization processes currently used in the industrial practice fall into one of the following categories:

- Thermal sterilization processes
- Radiation sterilization processes
- Chemical sterilization processes
- Sterile filtration processes

A more detailed description of these methods is provided in the following sections.

14.4 THERMAL STERILIZATION PROCESSES

Thermal sterilization is the most commonly used sterilization method in the pharmaceutical and biotechnology industries, especially in large-scale processes such as biotechnology and fermentation processes or in the thermal sterilization of glass vials and ampoules [4, 12]. The method relies on the exposure of the material to be sterilized to a *sufficiently high temperature* and for a *sufficiently long period of time* to produce the desired level of sterility. All life forms, including microorganism spores (the most heat resistant of all), are temperature sensitive and can be permanently inactivated by thermal sterilization methods, resulting in the loss of viability of the microorganisms that are initially present in the material, without necessarily producing their physical destruction. Since the achievement of sterile conditions also implies the inactivation of spores, longer sterilization times and higher temperatures are required than those needed for vegetative organisms, thus imposing significantly stricter conditions on sterilization processes than on milder processes.

Thermal sterilization is often the most economical and effective way of achieving sterilization, and well-defined levels of sterility assurance can be achieved using thermal processes, which further contributes to its widespread use. Nevertheless, thermal sterilization is invariably associated with the partial deterioration of the materials or equipment, such as nutrients, that were originally intended to be protected by the sterilization process. However, many products or materials are degraded at rates that are much slower than the inactivation rate of even the most heat-resistant

TABLE 14.1 Typical Industrial Sterilization Methods

Material, Equipment, or Product to Be Sterilized	Type of Material, Equipment, or Product	Physical State of Material to Be Sterilized	Typical Sterilization Method
Fermentation vessels	Permanent equipment	Solid	Thermal
Continuous sterilizers	Permanent equipment	Solid	Thermal
Liquid dispensers and other pharmaceutical equipment	Permanent equipment	Solid	Thermal
Sterile rooms	Permanent equipment	Solid	Chemical
Microbiological equipment and apparatus	Permanent or disposable equipment	Solid	Thermal, chemical, radiation
Syringes	Disposable supplies	Solid	Thermal, radiation
Sterile room clothing	Disposable supplies	Solid	Thermal, radiation
Disposable filters	Disposable equipment	Solid	Thermal, radiation
Vials	Disposable supplies	Solid	Thermal
Injectable solid drugs	Product	Solid	Thermal, filtration
Fermentation media	Nutrient	Liquid	Thermal, filtration
Fermentation nutrient oils	Nutrient	Liquid	Thermal, filtration
Silicone oils	Supply	Liquid	Thermal, filtration
Antifoam	Nutrient	Liquid	Thermal, filtration
Fats	Nutrient	Liquid	Thermal, filtration
Water for injection	Product/supply	Liquid	Thermal
Sterile process water	Supply	Liquid	Thermal
Fermentation air	Supply	Gas	Filtration
Sterile room air	Supply	Gas	Filtration
Laminar flow air	Supply	Gas	Filtration
Nitrogen blanket in sterile area	Supply	Gas	Filtration

Source: Data from Richards [11].

organisms, implying that, in most cases, only limited product degradation occurs during sterilization. In those cases when this is not true, as with many biological products, thermal sterilization cannot be used, and other sterilization methods must be employed.

14.4.1 Moist Heat (Steam) Sterilization Processes

When microorganisms are exposed to moist heat at a sufficiently high temperature (i.e. by adding steam directly and/or by adequately heating an aqueous system containing the microorganisms), denaturation of critical cellular components, mainly proteins, begins. Since such proteins are essential for microbial growth and reproduction, microbial death follows if enough time is allowed for sufficient level of protein damage to occur. Therefore, direct exposure of the material undergoing sterilization to high-temperature saturated steam is an extremely effective and rapid way of achieving sterility. There are a number of additional reasons for this. Condensing steam releases a significant amount of energy because of the very large latent heat of water. The heat transfer coefficient of condensing steam is among the highest achievable, which maximizes the likelihood that heat will penetrate and be released into all parts of the material undergoing sterilization. In addition, microorganisms live in

aqueous environments, and they largely consist themselves of aqueous solutions. As a result, they are inactivated more quickly in the presence of an aqueous physical sterilizing agent such as steam. Steam sterilization processes are associated with high sterility assurance, typically requiring monitoring only for time, temperature, and pressure.

The main disadvantage of steam sterilization is that water boils at only 100 °C. This temperature is too low to be of any practical use for sterilization since it would result in unacceptably long sterilization times to be effective against endospores. Therefore, all steam sterilization processes are carried out under pressure, in autoclaves or similar pressure vessels, in order to achieve higher temperatures and shorter sterilization times.

14.4.1.1 Batch Steam Sterilization via Autoclaving

Autoclaves are pressure vessels commonly used to sterilize equipment, materials, and/or supplies using steam (Figure 14.4). They are typically used in a variety of sterilization operations, but they are primarily employed to sterilize homogeneous liquids or small solids in small-, medium-, or even large-scale operations. Sterilization of lab equipment, medical instruments, or sealed ampoules filled with liquids is a typical process utilizing autoclaves.



FIGURE 14.4 Examples of commercial steam autoclaves. Top panels: FOF steam sterilizer (Courtesy of Fedegari Group). Lower panel: AMSCO Evolution Floor Loader steam sterilizer (Courtesy of Steris).

A steam autoclave consists of a pressurizable chamber provided with a removable lid or door (Figure 14.5). The material to be sterilized, typically stacked on trays in larger systems, is charged to the chamber. Then the door is closed, and the process begins. Autoclaves are designed so as to allow the air initially present in the chamber to be removed prior to sterilization. Failure to do so will result in the presence of air pockets in which moist heat will penetrate more slowly, thus possibly resulting in inadequate sterilization. Two methods are typically used to remove air. The first relies on gravity displacement of air by steam as the steam enters the chamber from the top. Hot steam does not readily mix with air in the chamber. Hence, air can be removed through

an opening at the bottom of the autoclave. Alternatively, air can be removed using a vacuum pump prior to sterilization. Most autoclaves use repeated vacuum cycles to eliminate the air initially contained within the autoclave. After the air purging cycle is completed, the actual sterilization step begins, which consists of closing all venting valves and admitting steam to bring the autoclave to the sterilization temperature. Steam condensate is removed at the bottom of the autoclave through a steam trap. Enough time must be allowed to enable all parts of the autoclave to reach the sterilization temperature. Then, this temperature is maintained for the required time period (typically 20 minutes at a sterilization temperature of 121 °C).

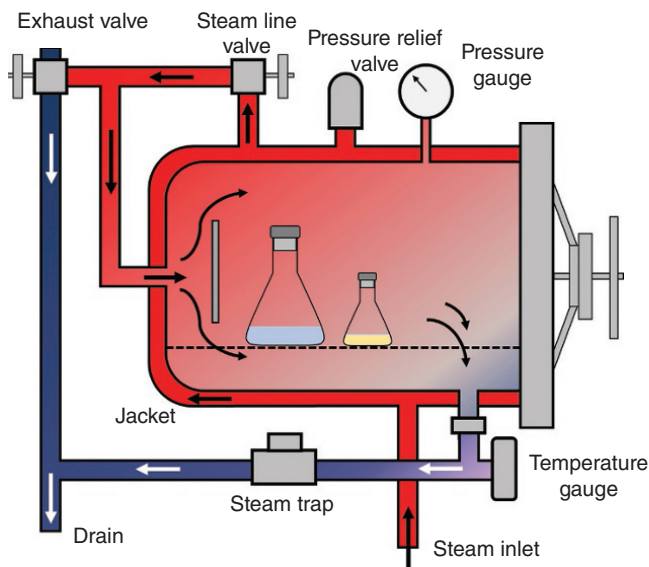


FIGURE 14.5 Schematic of a typical autoclave.

Microprocessors are commonly used to monitor and control the entire autoclaving cycle. The control strategy relies on the use of sensors placed in different parts of the autoclave, most typically the drain [13]. If solids or stagnant liquids are present, sensors should be placed inside them in order to ensure that the sterilization temperature is reached even in the solid core or in the bulk of the liquid before the actual sterilization cycle begins.

Validation of sterilization effectiveness involves the achievement of the required sterilization temperature and its maintenance for the required sterilization time for a given load of material to be sterilized. If a different material is to be sterilized in the same autoclave, validation of the sterilization procedure must be repeated unless it is possible to prove that the new material is clearly more easily sterilized than the material for which the original validation was conducted (e.g. if sterilization in an autoclave has been validated for a certain number of vials of given size, the validation process must not be repeated if a new batch of smaller vials is to be sterilized).

Biological or chemical indicators can be used to periodically test the effectiveness of the sterilization process. Biological indicators (BIs) are test systems typically consisting of sealed vials containing a pure strain of microorganisms with a well-known, defined resistance to a specific sterilization process. Endospores of selected microorganisms particularly resistant to sterilization are primarily used as biosensors since they are the most difficult to inactivate. To make the sterilization test appropriately stringent, the microbial species chosen for a given sterilization process is the most resistant to that specific process. For the case of steam sterilization, *Geobacillus stearothermophilus* spores are known to have high resistance to steam and are therefore

used as BIs to validate these sterilization processes. Vials with strips containing these organisms are placed in the autoclave, retrieved after the sterilization cycle is complete, and then cultured to determine microbial survival and thus the effectiveness of the sterilization process. Chemical indicators consisting of indicator strips or “challenge packs” rely instead on change in color or physical changes to indicate their exposure to a predetermined temperature for a predetermined time. For example, exposure to steam may melt the specific chemical contained at the bottom of a paper strip. The chemical then penetrates the strip at a known rate, thus providing information of its system temperature and exposure time. Different types of chemical indicators exist. Both biological and chemical indicators provide critical assurance that the necessary conditions for effective microbial inactivation were met during the sterilization process.

14.4.1.2 Batch Steam Sterilization of Liquid Culture Media in Bioreactors

Batch sterilization of liquid media prior to inoculation is a common operation in any biological process, from cell culture to full-scale antibiotic production in large-scale bioreactors. The culture medium is first prepared in batching tanks and then charged to the bioreactor that is heated up to the sterilization temperature by passing steam in the jackets and coils and by sparging steam directly through the air sparger. Any live steam directly added to the bioreactor produces a dilution effect that must be taken into account when preparing the medium. Once the sterilization temperature has been achieved, steam supply is shut off, and the system is allowed to remain at that temperature for the time period required for sterilization to take place. Then, cooling water and/or brine passed through the jacket and coil are used to cool down the bioreactor prior to inoculation. Batch sterilization is an energy-intensive process especially in large-scale operations since significant energy must be used to heat up the liquid batch to the sterilization temperature and then the same amount of heat must be removed during the cooling process.

14.4.1.3 Continuous Steam Sterilization of Liquid Culture Media

The rate of microbial endospore inactivation increases rapidly with temperature above a critical threshold value. As a result, sterilization processes that would normally require 20–30 minutes to complete at 121 °C require only a few seconds at temperature as high as 145 °C. The rate of thermal degradation of drug substances and nutrients also increases with temperature, but often not at as rapidly as spore inactivation. Therefore, sterilizing at higher temperature for shorter times is typically advantageous since it results in a process that ensures the same level of sterility with reduced product degradation. However, very rapidly heating and cooling phases are difficult to achieve in batch processes because of the thermal inertia of the systems and the significant time required to deliver and remove the significant

amounts of heat needed to change temperature. Such short sterilization times can only be achieved in continuous sterilizers in which the liquid undergoing sterilization flows through a series of heat exchangers where the temperature of the medium is rapidly raised in a heat exchanger, maintained such a high sterilization temperature as the medium flows, unheated in a holding tube, and then rapidly lowered in a cooling heat exchanger.

Two types of continuous sterilizers are used in the industrial practice. In the first one, steam is directly and continuously injected into a feed stream of concentrated unsterile medium to rapidly bring the medium temperature to the desired (high) sterilization temperature (dilution of the feed by the condensing steam must be taken into account to achieve the final desired nutrient concentration). The medium then flows through an unheated holding pipe whose length is such that the residence time of the medium in the pipe is the same as that needed to achieve the desired level of sterilization. The now sterilized medium then flows continuously through a cooling heat exchanger, where its temperature is rapidly lowered, and then is fed to the previously sterilized bioreactor [14–16]. While this process has some of the advantages of continuous sterilization process (mainly reduced product degradation), it still has the same high energy requirements as the batch process since all the energy that has been supplied by the steam must be removed in the cooling section and cannot be recovered. Therefore, a second type of continuous sterilizers is often used especially in large-scale industrial processes. In the process shown in Figure 14.6, unsterile medium from the batching area is pumped from the holding tank through a heat recovery exchanger where it is preheated at the expenses of the hot medium that has just undergone sterilization. The preheated medium is then brought to its sterilization temperature in a steam heater

and then passed through the holding tube. As before, the holding tube is just an insulated pipe whose length is such to ensure that the medium flowing through it will spend a pre-determined period of time to produce the required level of sterilization. After the holding tube, a significant fraction of the energy content of the sterilized medium is recovered to preheat the incoming medium, as described above. The sterile medium is then cooled down to the final temperature in a cooler exchanger and charged to the presterilized fermenters.

Continuous sterilization is a common and valid alternative to batch sterilization of culture media or products such as milk (continuous processes relying on high sterilization temperatures are typically labeled VHT processes, for “very high temperature”). Continuous sterilization has a number of advantages over batch sterilization. Since short sterilization times are required, continuous sterilization can be achieved when the medium is charged to presterilized bioreactors, drastically reducing the bioreactor downtime, especially in large-scale operations. Energy savings can also be significant since the thermal energy content of the just sterilized high-temperature medium can be partially utilized to preheat the incoming unsterile medium. In addition, the surge capacity for steam is reduced. Furthermore, the process can be easily monitored, controlled, and provided with a number of safeguards to avoid charging unsterile medium to the fermenters.

However, if solids are suspended in the liquid medium; this may result in an additional issue since heat transfer through solid particles proceeds at a slower rate than that in the liquid in which the particles are immersed. As a result, the temperature in the core of the particles may be lower than that of the liquid during a short time period commonly achieved in continuous sterilizers. This implies that those microorganisms that are lodged inside the particles may undergo a less

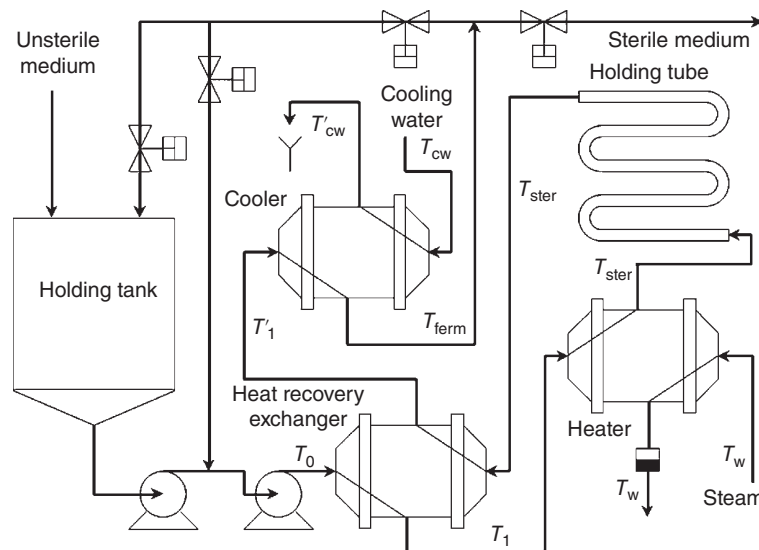


FIGURE 14.6 Typical continuous sterilization system [17]. Source: Reproduced with permission of John Wiley & Sons.

effective sterilization process than those in the liquid bulk [4, 17–20]. This effect is typically negligible in most batch processes since the long sterilization time (typically 20 minutes or longer) and the long heating and cooling times usually provide ample time for small particles to thermally equilibrate with the surrounding liquid, thus ensuring a similar level of microbial reduction in both the particles and the liquid. However, the same is not true for continuous processes because of the much shorter sterilization time. In such cases, large temperature differences between the solid core and the surrounding liquid can be predicted for short sterilization times, with significant consequences for sterilization effectiveness. Therefore, solid particles should be excluded, whenever possible, from the medium prior to continuous sterilization. This is the case for most pharmaceutical formulations that have very stringent requirements on the maximum size of the solids (typically of the order of μm). However, most commercial fermentation media have a high concentration of suspended solids [21]. In such cases the impact of the presence of solids on sterilization effectiveness should be accounted for in the design of the sterilization process.

14.4.2 Dry Heat Sterilization Processes

Dry heat sterilization relies on the achievement of high temperatures in a closed container (often at atmospheric pressure) to sterilize the material in it. Dry heat is the method of choice for sterilizing items that tolerate high temperatures but are incompatible with steam. As a result, dry heat processes are used in the pharmaceutical industry to sterilize glassware (such as vials and ampoules used as sealed containers for drugs), nonvolatile thermostable powders (such as zinc oxide, talc, sulfamerazine), or oily solvents and drugs (e.g. corn oil, cottonseed oil, sesame oil, desoxicorticosterone).

A typical dry heat batch sterilizer is a conceptually simple apparatus resembling a common convection oven and consisting of a chamber through which hot air is circulating so as to heat the material undergoing sterilization, maintain it at the sterilization temperature as long as required, and then cool it down (Figure 14.7). The operation of the sterilizer is also similarly simple, consisting of loading the material, conducting the heating cycle, and then removing the sterilized material after cooling.

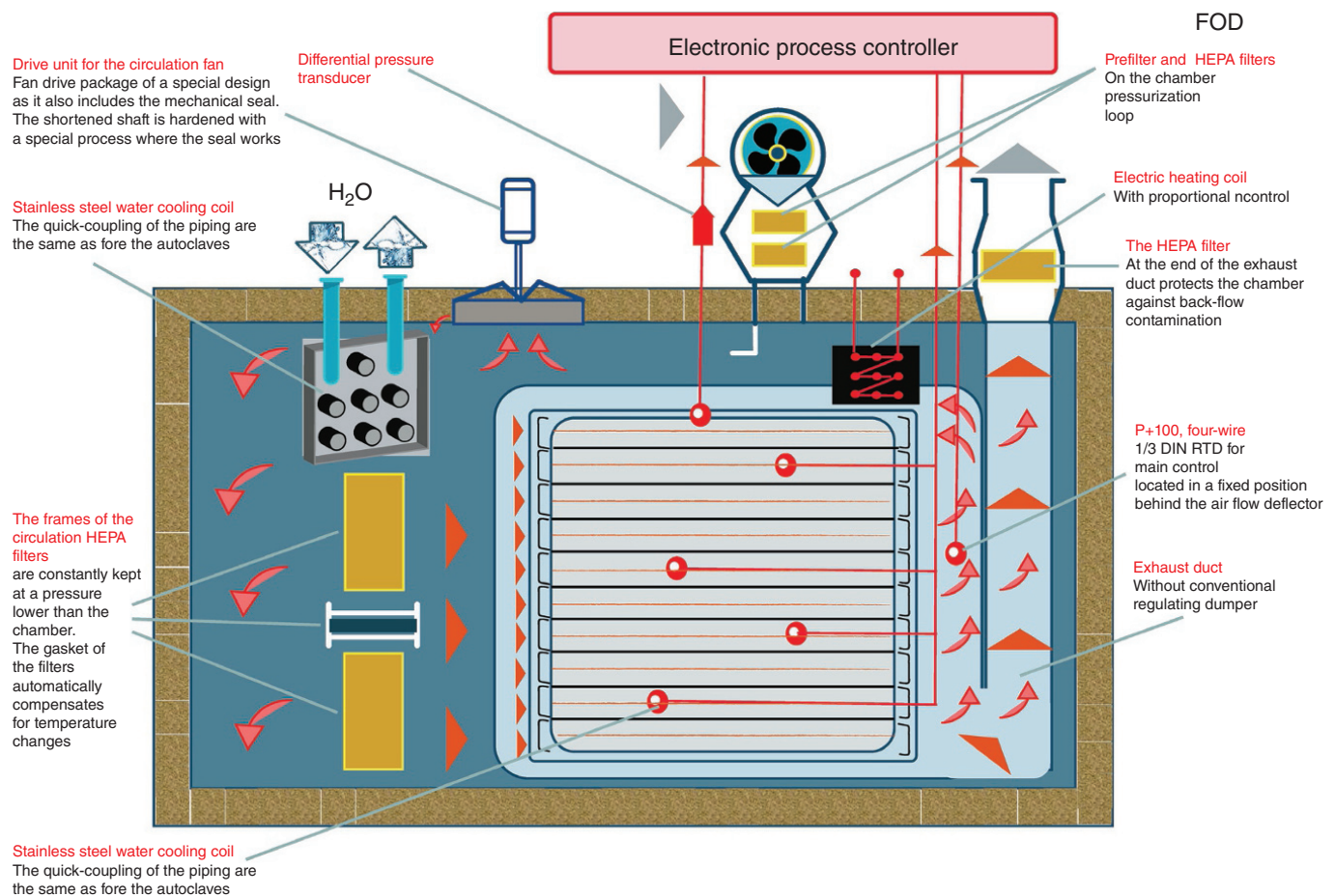


FIGURE 14.7 Schematic of a dry heat sterilizer (FOD_dry-heat_sterilizer; Courtesy of Fedegari Group).

Dry heat sterilizer can be operated at atmospheric pressure, and this is clearly an advantage over steam sterilization. However, since air is the sterilizing medium, there are a number of disadvantages to dry sterilization, mainly the fact that air has low density, low sensible heat, and low heat transfer coefficient. Therefore, heat transfer is low, although heat radiation and conduction through the metal surfaces of the sterilizer also contribute to heat delivery. In addition, microorganisms live in aqueous environments, and dry heat can additionally affect the heat transfer through the cell wall. More importantly, the rate of microbial death under dry conditions is much slower than when the medium is aqueous or when sterilization is conducted in the presence of steam, even at the same temperature, since the mechanism of temperature inactivation is different when the microorganisms are not in an aqueous medium because of the lack of water penetrating the cell. Whereas microbial inactivation via steam is caused by denaturation and coagulation of essential cell proteins, dry heat produces microbial death from dehydration followed by protein oxidation [22]. All this implies that dry heat sterilization processes are typically conducted at much higher temperatures and for longer times (e.g. 120–180 minutes at 160 °C or 90–120 minutes at 170 °C) than steam sterilizations.

In the biotechnology industry dry heat sterilization is primarily carried out when the liquid undergoing sterilization is nonaqueous and water immiscible and therefore cannot be exposed to steam without forming another phase. This is the case for materials such as antifoam oil, nutrient oils, or fats, which have to be sterilized independently from the main fermentation medium. The materials are often sterilized in batch tanks heated and cooled indirectly through the jacket or coil. Small continuous sterilizers can also be used in alternative.

14.4.3 Dry Heat Depyrogenation Processes

Pyrogens are all those substances that can cause transient fever and, in the worst case, death due to shock, when introduced into the blood stream. Endotoxins are fragments of the cell walls of Gram-negative bacteria that are toxic to humans and can induce severe pyrogenic (fever-inducing) activity. Therefore, parenteral products must be free of endotoxins or have levels below accepted limits. Being only cell fragments, endotoxins are much more resistant to heat treatments than living cells and endospores. In fact, the temperatures typically used in moist heat sterilization are much too low to affect endotoxins. Therefore, they can only be destroyed by dry heat treatments (depyrogenation processes) conducted at even higher temperature and for longer times than those used for dry heat sterilization. As a result, batch depyrogenation is typically conducted for 60–90 minutes at 230 °C or 30–60 minutes at 250 °C. Clearly any material that undergoes depyrogenation is also sterilized. The reverse is not true.

Glass vials and ampoules used for the sterile manufacturing of injectable drugs are typically washed with deionized water and a final rinse of pyrogen-free water. Then they are dried and depyrogenated (and hence sterilized) in hot air ovens using dry heat. Alternatively, depyrogenation is conducted in continuously operated dry heat tunnels (Figure 14.8) connecting the non-sterile area of the plant, where the clean but non-sterile vials are charged, to the sterile area [21].

14.4.4 Batch Sterilization of Solids

In all heat treatment processes, the desired sterilization level in the material to be sterilized is achieved *only* if the temperature is maintained for the desired time *everywhere* within

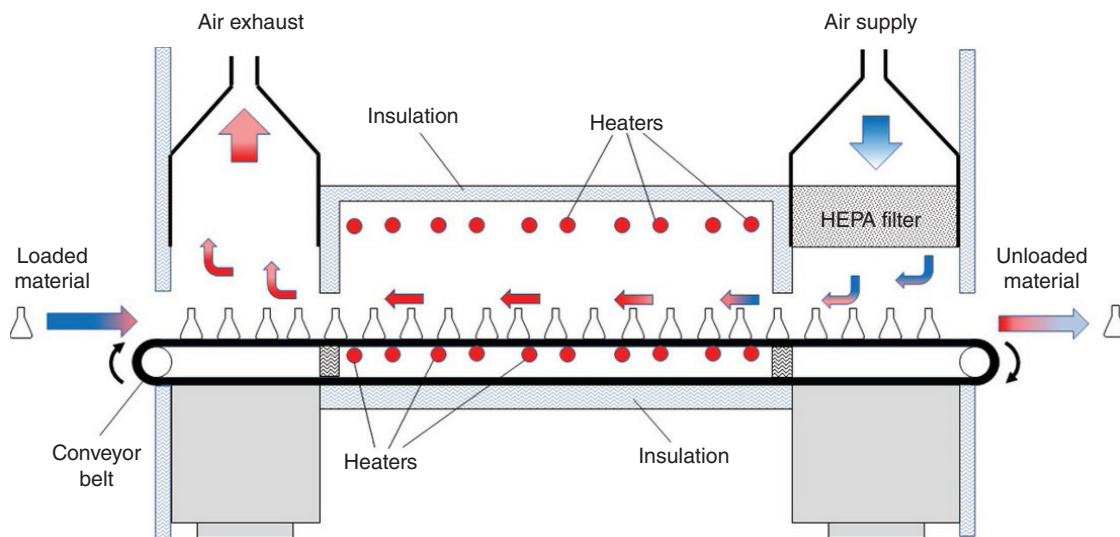


FIGURE 14.8 Schematic of a dry heat tunnel sterilizer (not drawn to scale).

that material. Convection and conduction are the primary mechanisms for heat transfer in sterilizers. Depending on the process, these mechanisms may be more or less effective in achieving temperature uniformity. For example, sparging steam in a bioreactor is a very effective method to ensure uniform temperature in the medium undergoing sterilization. Similarly, the temperature in stirred liquids is typically uniform in the liquid bulk. When moist heat is used, for example in autoclaves, steam penetrates even small, loose solids and releases heat rapidly and effectively throughout. However, there are situations when heat transfer is less than optimal, as in the case when the heating medium is a low-density fluid with a low heat transfer coefficient such as air or, in all cases, when solids form mass lumps or are present in sealed containers. In such cases, the route by which heat is transported within the solid is not convection but conduction, which is typically a slower mechanism, especially if the solid has a low heat conductivity. This implies that even if the temperature around the solid is quickly raised to the sterilization temperature, it will still take time for the heat to travel to the core of the solid and sterilize its content. If the solids are large and/or the sterilization time is too short, it may well be that the solid center never reaches the sterilization temperature or the core is maintained at the sterilization temperature for too short a time period to ensure adequate sterility at the end of the process. This can be easily visualized by considering how long it takes for an egg to cook when placed in boiling water. The cooking process (irreversible denaturation of egg proteins) occurs much more quickly near the outer egg layer where the temperature raises more rapidly and much more slowly in the inner yolk, whose temperature raises much more slowly because of the low heat conduction of the egg white. Therefore, protein denaturation in deeper layers does not effectively start until the temperature profile throughout the egg is more uniform. Cracking the shell and dropping the egg content in boiling water result in the rapid denaturation of the outer egg white and the more slowing cooking of its yolk content (poached egg). Finally, scrambling the egg content in a bowl and dropping it in boiling water (egg drop soup) results in dispersing it and breaking it up in small agglomerates that cook much more quickly since heat penetrates more rapidly throughout because of the small distance it has to travel to reach their cores.

All this implies that although the underlying inactivation mechanism to sterilize the microbial content in solids is the same as that as for the sterilization of the same microorganisms in well-stirred liquids, the heat transfer process may be much less effective. Therefore, whenever (larger) solids or solid aggregates are sterilized, one must ensure that the sterilization temperature at the core of the solids is achieved and maintained for the required time, as in all other cases. In many practical instances, this additionally requires monitoring the temperature not just inside the autoclave but also in the solid core (as done with meat thermometers when

cooking a roast, just to add to the previous cooking example). For instance, when sterilizing a large number of sealed vials placed on stacked trays, penetration thermocouples are inserted in sample vials placed at critical locations on the trays, and the temperatures of their contents are recorded to ensure that sterilization is achieved in every other vial in that batch. This may be needed also when a larger volume of stagnant liquid in larger ampoules is sterilized, since liquid mixing inside them is limited to the that achievable only by natural convection inside the ampoules.

14.5 ANALYSIS OF THERMAL STERILIZATION PROCESSES

The quantitative analysis of sterilization processes is based on the use of the traditional chemical engineering tools, such as mass balances, energy balances, and reactor engineering, to model different processes and predict their performance, primarily in terms of final probability of microbial contamination at the end of the process and the impact of the process on the product. In this section we will examine how *thermal* sterilization processes are analyzed. The starting point is the quantification of the initial and final levels of microbial contamination to be attained through the sterilization process. Then kinetic theory can be used to quantify the rates of microbial death and product degradation, which, when combined with the appropriate mass balances, can yield quantitative information on the sterilization process and be the basis for a rational approach to process design.

14.5.1 Specific Objectives of Sterilization Processes

In order to design any sterilization process, one must first determine, or estimate, the initial microbial content of the product to be sterilized and quantify the final objective as far as microbial inactivation is concerned [23]. The initial point requires knowledge of initial microbial “bioburden,” i.e. the number of microorganisms initially present in, or on, the product and their type, which determines how resistant they are to a specific sterilization method. Several approaches can be used to determine the bioburden. The simplest and most conservative (“overkill”) approach is to assume that the bioburden is made of an appropriately large initial number of bacteria in the product and that they are made up of the “worst” (i.e. most difficult to sterilize) kind of microorganisms [24]. For example, *G. stearothermophilus* (formerly *Bacillus stearothermophilus*) is an endospore-forming bacterium that is particularly resistant to steam sterilization. Therefore, if a steam autoclave is used, inactivating this type of microorganism ensures that all other types of bacteria are also inactivated. The overkill approach is the safest as far as achievement of sterilization is concerned, but it is also the one that results in the harshest sterilization process

(e.g. higher temperatures and/or longer sterilization times during thermal sterilization), which may have a more negative impact on the product. At the opposite extreme, one can determine the actual number and composition of the bioburden and design the sterilization process under the assumption that only those microorganisms in those numbers will be inactivated. This approach results in a less demanding sterilization process (and less damaging for the product), but it also carries a greater risk that the sterilization objectives may not be achieved.

Quantitatively, if $n_N(t)$ indicates the number of microorganisms (or spores) in the product at any time t , then n_{N_0} is the initial bioburden at the beginning of the sterilization process, i.e. at $t = 0$. In many cases the microorganisms are present in a liquid (e.g. a vial or a bioreactor), and the number $n_N(t)$ and concentration $N(t)$ at any time t are simply related to each other as

$$N(t) = \frac{n_N(t)}{V} \quad \text{and} \quad n_N(t) = N(t) \cdot V \quad (14.1)$$

where

V is the liquid volume.

Similarly, the initial bioburden concentration, N_0 , is

$$N_0 = \frac{n_{N_0}}{V} \quad \text{and} \quad n_{N_0} = N_0 \cdot V \quad (14.2)$$

In practice, N_0 can vary from nearly zero for biotoxic compounds (e.g. antibiotic solutions) to 10^6 – 10^8 microorganisms/ml for easily metabolized compounds (e.g. sugar solutions).

Designing a sterilization process additionally requires the quantification of the risk of *not* achieving sterility, i.e. the definition of the endpoint of the process, which can be established as follows. Leaving just *one* (or more) living microorganisms (or spores) behind after a sterilization process implies that the process has failed since that microorganism will rapidly reproduce and render the batch (or vial, container, or device) non-sterile. The *sterility assurance level* (SAL) is defined as the probability of survival of microorganisms after a sterilization process, even if it is just one microorganism. SAL is a predictor of the efficacy of the sterilization process: the smaller the SAL value, the more likely the product will be sterile.

If n_{N_f} is defined as the *final* number of microorganism in the batch at the end of the sterilization process, then

$$\text{if } n_{N_f} \geq 1 \Rightarrow \text{Non-sterile product} \Rightarrow \text{SAL} = 1 \quad (14.3)$$

implying that there is 100% probability that such a system is not sterile. Therefore, to make SAL reasonably small implies reducing n_{N_f} . Since microorganisms only come as discrete units, the only logic alternative to $n_{N_f} \geq 1$ would be $n_{N_f} = 0$, which would imply absolute sterility and SAL = 0.

However, absolute sterility can never be guaranteed 100% since sterilization is intrinsically a probabilistic process. Therefore, in order to relate n_{N_f} to SAL in a meaningful but simple fashion and without using a more correct probabilistic approach, one can reason as follows. If one sterilizes *one* single batch (e.g. the liquid in a container, such as the growth medium in a bioreactor) and the process leaves *one* microorganism in the container, then

$$\text{if } n_{N_f} = \frac{1 \text{ microorganism}}{1 \text{ container}} \Rightarrow \text{SAL} = 1 \quad (14.4)$$

i.e. the container is 100% non-sterile. If one repeats the same process using 10 containers and, on average, one microorganism every 10 containers remains alive after the sterilization process, one container out of 10 will be non-sterile, i.e.

$$\text{if } n_{N_f} = \frac{1 \text{ microorganism}}{10 \text{ containers}} \Rightarrow \text{SAL} = 0.1 \quad (14.5)$$

i.e. 90% of the time the process will be successful, but it will fail in 10% of the cases. The same logic would apply if one processes the same container 10 times instead of 10 containers simultaneously. Similarly, if one repeats the same sterilization process 100 times (or conducts the process on 100 containers at the same time) and one in 100 containers is non-sterile, then

$$\text{if } n_{N_f} = \frac{1 \text{ microorganism}}{100 \text{ containers}} \Rightarrow \text{SAL} = 0.01 \quad (14.6)$$

Following the same logic for an arbitrarily large number of containers, x , (or for the same large number of sterilization processes, x , in the same container), one can conclude that n_{N_f} is equal to

$$n_{N_f} = \frac{1 \text{ microorganism}}{x \text{ containers}} \Rightarrow \text{SAL} = \frac{1}{x} \quad (14.7)$$

and hence that

$$\begin{aligned} n_{N_f} &= [1 \text{ microorganism}] \times \left[\begin{array}{l} \text{probability of survival of 1} \\ \text{microorganism in 1 container} \end{array} \right] \\ &= [1 \text{ microorganism}] \times [\text{SAL}] \end{aligned} \quad (14.8)$$

In practice, n_{N_f} has the *same* numerical value as the desired SAL for the sterilization process. Again, this statement does not imply that there is a fractional number of live microorganisms in each container (an impossibility), but rather the probability of finding a single viable microorganism in a container at the end of the sterilization process is equal to the SAL.

Different SAL values can be required. The typical SAL value for terminal sterilization is 10^{-6} , which, when applied, for example, to vials of product, would imply 1 non-sterile vial in 1 000 000. For other types of sterilization processes, the SAL can be much larger, i.e. the probability of non-sterility is much higher (e.g. 10^{-4} or 1 in 10 000 for heat-resistant medical devices; 10^{-3} or 1 in 1000 for filtration and aseptic processing).

Therefore, the sterilization objective can now be quantified in terms of ratio of the number of viable organisms at the end of the sterilization process to the corresponding initial number, i.e. as the ratio n_{N_f}/n_{N_0} . In addition, since this ratio is typically a very small number, it is convenient to introduce the *log10-reduction ratio*, n (often simply refer to as *log reduction*), as follows:

$$\frac{n_{N_f}}{n_{N_0}} = \frac{\text{SAL}}{n_{N_0}} \Rightarrow n \equiv -\log_{10}\left(\frac{n_{N_f}}{n_{N_0}}\right) = -\log_{10}\left(\frac{\text{SAL}}{n_{N_0}}\right) \quad (14.9)$$

From its definition it follows that “ n ” is the number of orders of magnitude by which the microbial population has been reduced, and it should **not** to be confused with the n_N , the actual number of viable microorganisms. Clearly, the value of log reduction n depends on both the initial number of microorganisms in the batch undergoing sterilization (n_{N_0}) and the desired low level probability of non-sterility (SAL) at the end of the process.

Since liquids are often sterilized, it is also important to relate the reduction number of microorganisms to the corresponding *reduction in microbial concentration*, i.e. to additionally determine the ratio N_f/N_0 , where ($N_f = n_{N_f}/V$)

$$\frac{n_{N_f}}{n_{N_0}} = \frac{N_f \cdot V}{N_0 \cdot V} = \frac{\text{SAL}}{n_{N_0}} \Rightarrow n \equiv -\log_{10}\left(\frac{N_f \cdot V}{N_0 \cdot V}\right) = -\log_{10}\left(\frac{\text{SAL}}{N_0 \cdot V}\right) \quad (14.10)$$

One should be careful *not* to confuse the SAL with the log reduction, as the equations above and the following example show.

EXAMPLE PROBLEM 14.1

The growth media in a 10-l fermenter and a 10 000 l fermenter undergo sterilization. In both cases, the desired SAL is 10^{-4} (1 failure in 10 000 fermentations). Both fermenters have the same initial microbial concentration ($N_0 = 10^6$ microorganisms/l). Determine the required log reduction n in both cases.

Using Eq. (14.10) it is

$$n = -\log_{10}\left(\frac{\text{SAL}}{N_0 \cdot V}\right) = -\log_{10}\left(\frac{10^{-4}}{10^6 \frac{1}{1} \cdot 101}\right) = -\log_{10}(10^{-11}) \\ = 11(\text{small fermenter})$$

$$n = -\log_{10}\left(\frac{\text{SAL}}{N_0 \cdot V}\right) = -\log_{10}\left(\frac{10^{-4}}{10^6 \frac{1}{1} \cdot 10\,0001}\right) \\ = -\log_{10}(10^{-14}) = 14(\text{large fermenter})$$

The sterilization requirements are more stringent for the larger fermenter (larger log reduction) since, although the initial microbial concentration (N_0) and the endpoint are the same ($n_{N_f} = \text{SAL}$) in both cases, the larger volume in the large fermenter makes the initial number of microorganisms $n_{N_0} = N_0 V$ much larger. Therefore, the larger initial bioburden coupled with the requirement to obtain the same final number of microorganisms (and hence the same SAL) makes the sterilization process in the larger system much more demanding.

So far, we have examined the sterilization objectives as far as the achievement of a sufficiently low probability of contamination is concerned. However, this requirement needs to be balanced against possible concurrent damage to, or degradation of, the product, which also occurs during the same sterilization process. To determine this, one must determine whether the product content (expressed as amount of product) or, alternatively, its concentration decreases too significantly following the sterilization process. If one assumes that m_0 and m_f are, respectively, the initial and final amounts of product left in the batch undergoing sterilization and C_{m_0} and C_{m_f} are the corresponding initial and final concentrations, then the value of the following ratios determines if the selected process is acceptable:

$$\text{Fraction of remaining product} = \frac{m_f}{m_0} = \frac{C_{m_f} \cdot V}{C_{m_0} \cdot V} = \frac{C_{m_f}}{C_{m_0}} \quad (14.11)$$

For example, if the product to be preserved is a vitamin added to a fermentation medium, thermal sterilization may partially degrade it, possibly too significantly. Therefore, if a given sterilization process results in the achievement of the desired microbial sterilization objective at the expense of an unacceptable level of vitamin degradation, then the overall process will not be adequate, and it will have to be modified either in terms of the operating conditions of a similar thermal process (e.g. by varying the sterilization temperature and time) or by changing the sterilization process altogether (e.g. by sterilizing the vitamin separately using another sterilization method, such as sterile filtration).

In conclusion, a viable sterilization process must be able to achieve its intended microbial reduction objective while preserving the product that is undergoing the same process.

14.5.2 Kinetics of Microbial Thermal Death and Thermal Degradation of Product

As already mentioned, thermal sterilization relies on the fact that key cellular components such as the proteins necessary to sustain life in microorganisms are inactivated at high temperature. Therefore, the analysis of thermal sterilization processes is based on the well-founded assumption that the degradation process can be assumed to proceed, like any other chemical reaction, at a rate that can be obtained from standard kinetic theory based on power-law models, orders of reaction, and Arrhenius theory, which will be reviewed here first. However, and additionally, sterilization processes are often and more commonly analyzed using a simplified version of the microbial death kinetics (using parameters such as the D -value and Z -value), which is less rigorous but easier to use. They will be examined second.

14.5.2.1 Microbial Thermal Death Kinetics The key assumption for the use of classic reaction kinetic theory for thermal microbial death is that the permanent microbial inactivation process (microbial death) is the result of the inactivation of a single enzyme or a single class of enzymes within the microorganisms. In addition, the process is assumed to follow *first-order kinetics*, i.e. the microbial death rate is assumed to proceed at a rate that is proportional to the number concentration of microorganisms present at any given time in the system. In quantitative terms this implies that the rate of microbial death rate can be expressed as [25, 26]

$$r_N = K_d \cdot N \quad (14.12)$$

where

r_N is the specific microbial death rate, $(\text{ml s})^{-1}$.

N is the number concentration of viable microorganisms, ml^{-1} .

K_d is the proportionality parameter, s^{-1} .

The parameter K_d is the specific kinetic rate of microbial thermal death and can be obtained from the Arrhenius equation

$$K_d = K_{d0} e^{-E_d/R_G T} \Rightarrow \ln(K_d) = \ln(K_{d0}) - \frac{E_d}{R_G} \cdot \frac{1}{T} \quad (14.13)$$

where

K_{d0} is the Arrhenius constant of the thermal death kinetic process, s^{-1} .

E_d is the activation energy of the thermal death kinetic process, J/mol .

R_G is the ideal gas constant, $\text{J}/(\text{mol K})$.

T is the absolute temperature, K .

From this equation it follows that any increase in sterilization temperature produces a reduction in the absolute value of the argument in the exponential term, thus resulting in an increase in the value of K_d and a faster sterilization process. In the same equation the activation energy term, E_d , determines how rapidly the specific sterilization rate, K_d , varies as a function of temperature. For example, if the activation energy is high, then K_d will increase very rapidly when the temperature is increased of even just a few degrees, thus resulting in a faster sterilization process.

The numerical values for K_{d0} and E_d depend on the type of microorganisms and their intrinsic capability to withstand high temperatures. However, spore-forming microorganisms can produce spores with a significantly greater ability to resist high temperatures than the same microorganisms in the vegetative state. For most spores the value of the activation energy is usually quite high, typically in the range 200 000–400 000 J/mol [27]. Some typical values for these constants are given in Table 14.2. This table shows that *G. stearothermophilus* is one of the most heat-resistant organisms commonly encountered. Therefore, thermal sterilization calculations are often made using the death kinetic parameters for this microorganism as a worst-case scenario.

The first-order kinetic model for microbial death given above can be used to determine the degree of sterility achieved during a simple sterilization process in which the microorganisms are exposed to a *constant* high temperature for a given time, as when a vial containing microorganisms is rapidly immersed in a high-temperature liquid, left there for a fixed time t , and then removed and rapidly quenched to room temperature. Quantitatively, for a given homogenous microbial suspension occupying a constant volume V , the mass balance for the microorganisms during

TABLE 14.2 Numerical Values of the Arrhenius Constants and Activation Energies for the Thermal Death of Different Types of Microorganisms

Microorganism	K_{d0} (min^{-1})	E_d (J/mol)
<i>Geobacillus stearothermophilus</i>	7.9×10^{38a}	2.876×10^{5a}
<i>G. stearothermophilus</i>	9.5×10^{37b}	2.834×10^{5b}
<i>G. stearothermophilus</i>	9.8×10^{37c}	2.876×10^{5a}
<i>Bacillus subtilis</i>	$3.6 \times 10^{42} - 5.2 \times 10^{42d}$	3.182×10^{5e}
<i>Clostridium botulinum</i>	–	3.433×10^{5e}
<i>Escherichia coli</i>	1.5×10^{84a}	5.317×10^{5a}

^aAiba et al. [27].

^bDeindoerfer and Humphrey [28].

^cExtracted from Moss and Smith [29] using Deindoerfer and Humphrey's [28] value for the activation energy.

^dExtracted from Moss and Smith [29] using Wang et al.'s [26] value for the activation energy.

^eWang et al. [26].

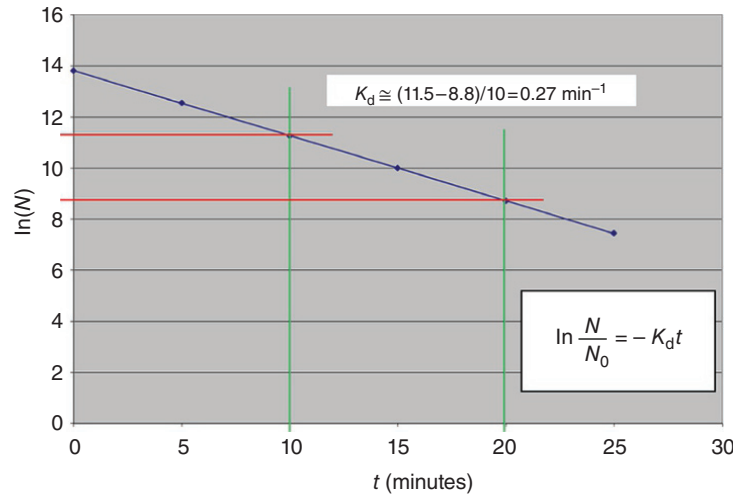


FIGURE 14.9 Semilogarithmic plot of N vs. t at constant temperature.

a batch sterilization process can be described by the following microbial balance:

$$V \frac{dN}{dt} = -Vr_N = -VK_d \cdot N \quad (14.14)$$

which states that the change in the viable number of organisms in the system (left-hand side) is equal to the rate at which the organisms are inactivated as a result of thermal death (right-hand side). Since the sterilization temperature is kept constant in this case, K_d is also constant, and the simple integration of this differential equation using separation of variables yields

$$\ln \frac{N}{N_0} = -K_d t \quad (14.15)$$

where

N_0 is the initial number concentration of viable organisms, ml^{-1} .

The ratio N/N_0 is the fraction of viable organisms that have survived sterilization at time t . This equation shows that a semilogarithmic plot of N vs. t results in a straight line whose slope is a function of temperature, which was indeed found to be the case for most microorganisms and their spores. An example of the semilogarithmic relationship between the fraction of surviving organisms N/N_0 and time is shown in Figure 14.9.

EXAMPLE PROBLEM 14.2

We want to determine the sterilization time required to reduce the concentration of *G. stearothermophilus* so that the final probability of contamination in a 100-l bioreactor is 1 in

one million knowing that the initial microbial concentration is 4×10^7 microorganisms/ml, and to do so at two sterilization temperatures, i.e. 121 and 115 °C, assuming in both cases that the temperature is instantaneously raised to the sterilization temperature (i.e. the sterilization process is isothermal).

From Eq. (14.10) we can calculate that

$$\frac{\text{SAL}}{N_0 \cdot V} = 2.5 \times 10^{-19} \Rightarrow n = -\log_{10} \left(\frac{\text{SAL}}{N_0 \cdot V} \right) = 18.602$$

i.e. this process will require reducing the microbial concentration by 18.6 orders of magnitude. Using Eq. (14.10) we can additionally determine that $N_f = N = 1 \times 10^{-11}$ microorganisms/ml. Equation (14.15) can now be used to calculate the sterilization time t provided that we know K_d , which depends on the temperature. Entering the temperatures (in degree K) in Eq. 14.13 and using $K_{d0} = 9.5 \cdot 10^{37} \text{ min}^{-1}$ and $E_d = 2.876 \times 10^5 \text{ J/mol}$ from Table 14.2 as input, we can find that

$$K_d = 2.586 \text{ min}^{-1} \text{ at } T = 121^\circ\text{C} = 394.15 \text{ K}$$

$$K_d = 0.679 \text{ min}^{-1} \text{ at } T = 115^\circ\text{C} = 388.15 \text{ K}$$

The corresponding sterilization times from Eq. (14.15) are

$$t = 16.6 \text{ minutes at } T = 121^\circ\text{C} = 394.15 \text{ K}$$

$$t = 63.1 \text{ minutes at } T = 115^\circ\text{C} = 388.15 \text{ K}$$

One can clearly see that the temperature has a dramatic effect on the sterilization time and that only a 6 °C difference nearly quadruples the sterilization time.

14.5.2.2 Simplified Expressions for Microbial Death Kinetics: D-Value, F-Value, and Z-Value Approach The above expressions for the kinetics of microbial thermal death have a sound thermodynamic and kinetic theory basis. However, a simplified approach is widely encountered in the industrial practice to describe the kinetics of microbial death. This approach quickly conveys the idea of how effective a specific sterilization process is. However, when pushed beyond its boundary of applicability, this approach can result in oversimplifications and less correct predictions of the sterilization effectiveness of a process.

The *decimal reduction time*, D , is defined as the sterilization time required to produce a reduction of one order of magnitude (i.e. a 10-fold change) in the microbial population number or concentration for a given sterilization temperature [30, 31]. For example, if the temperature of a vial containing 10^6 microorganisms of the same species is suddenly brought to 121°C and if the D -value for that species at this temperature is two minutes, then only 10^5 microorganisms can be expected to be found in the vial after a two-minute exposure. Waiting for two more minutes will reduce the microbial population to 10^4 microorganisms and so on. Clearly, this approach makes sterilization calculation very simple, at least in constant temperature processes. Because of its definition, the D -value is directly and univocally related to K_d and is therefore a measure of the kinetic rate of the sterilization process. However, unlike K_d , the D -value and its unit (typically minutes) immediately convey the idea of how fast the sterilization process proceeds at any given temperature: the greater the D -value, the more resistant the organism will be to sterilization, and the slower the sterilization process will be. For this reason, the D -value and its multiple described below, the F -value, are extremely popular in the industrial practice.

From the definition of the D -value, one can easily establish the relationship between the D -value and K_d using Eq. (14.15). The resulting equation is

$$\ln \frac{10N}{N} = -K_d D \quad (14.16)$$

in which it is assumed that the sterilization process is conducted batchwise at constant temperature. From this equation it follows that

$$D = \frac{\ln(10)}{K_d} = \frac{2.303}{K_d} \quad (14.17)$$

If the sterilization temperature is kept constant, by combining Eq. (14.15) with Eq. (14.17), it follows that

$$n = \log_{10} \frac{N}{N_0} = -\frac{t}{D} \quad (14.18)$$

i.e. a plot of $\log_{10}(N)$ vs. time gives a straight line with slope $-1/D$, as shown in Figure 14.10. Again, “ n ” here is the *log10-reduction ratio*, as previously described, i.e. the number of orders of magnitude by which the microbial population has been reduced.

Table 14.3 gives the values for K_d and D for different organisms at 115 and 121°C . This table shows that most heat-resistant organisms have D -values of the order of minutes at temperatures near the common sterilization temperature of 121°C . The same table also shows that organisms such as *Escherichia coli* are much more easily inactivated than heat-resistant organisms such as *G. stearothermophilus* since *E. coli* does not form spores, while *G. stearothermophilus* does.

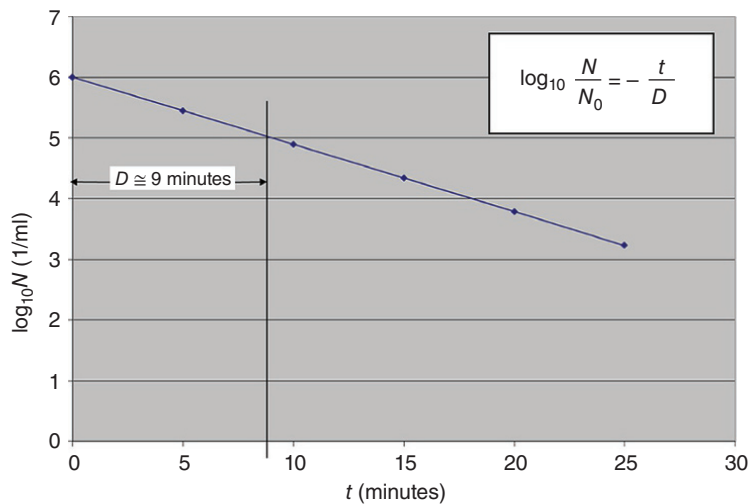


FIGURE 14.10 Plot of $\log_{10}(N)$ vs. t at constant temperature.

TABLE 14.3 K_d , D -Values, and Z -values for Different Organisms at Two Different Temperatures

Microorganism	K_d (min^{-1}) at $T = 121^\circ\text{C}$	D (min) at $T = 121^\circ\text{C}$	D (min) at $T = 115^\circ\text{C}$	Z -value ($^\circ\text{C}$)
<i>Bacillus stearothermophilus</i>	0.4–1.5	1.5–6	10–24	6–7
<i>Clostridium botulinum</i>	1.6–2.9	0.8–1.4	2.8–3.6	—
<i>Bacillus subtilis</i>	3.3–5.8	0.4–0.7	2.2	8–13
<i>Escherichia coli</i>	5.3×10^{13}	4.3×10^{-14}	5×10^{-13}	—

Source: Partially readapted from Wallhäusser [32].

EXAMPLE PROBLEM 14.3

We want to determine the D -value for *Bacillus subtilis* spores 5230 (ATCC 35021) at $T = 117^\circ\text{C}$. To do so, experiments are conducted by inserting vials, each containing 10^6 spores, in a BIER apparatus (a test device designed to raise the temperature of the vials nearly instantaneously). The vials are removed at different times and immediately cooled. The vials are then opened, and their content is incubated in petri dishes to determine the number of surviving microorganisms. The data are shown in the table.

Time, t (min)	Number of Surviving Microorganisms
0	1.0×10^6
5	2.8×10^5
10	7.8×10^4
15	2.2×10^4
20	6.1×10^3
25	1.7×10^3

We can use Eq. (14.18) rewritten as

$$\log_{10} \frac{N}{N_0} = \log_{10} \frac{N \cdot V}{N_0 \cdot V} = -\frac{t}{D} \Rightarrow \log_{10}(N \cdot V) - \log_{10}(N_0 \cdot V) = -\frac{1}{D}t$$

Therefore, plotting $\log_{10}(\text{surviving microorganisms})$ vs. t will result in a straight line with slope $-1/D$ and intercept $\log_{10}(\text{initial microorganisms})$, similar to Figure 14.10. After regressing the data, the resulting value for D at $T = 117^\circ\text{C}$ is found to be 9.03 minutes, which means that it takes about nine minutes to reduce the microbial concentration by one order of magnitude (for example, from 10^6 to 10^5 , as graphically shown in Figure 14.10).

Another parameter often encountered in sterilization calculations is the thermal death time, or F -value, defined as the time required to produce a specific decrease in the number of surviving organisms (typically by a 10^{-6} factor, i.e. 6 orders of magnitude). Because of its definition the F -value is just a multiple of the D -value. For example, if a reduction of 10^{-n} is desired, then from Eq. (14.15) it is

$$\ln \frac{N}{10^{-n}N} = \ln \frac{10^n N}{N} = -K_d F \quad (14.19)$$

from which

$$F = \frac{\ln(10^n)}{K_d} = \frac{2.303 \cdot n}{K_d} = nD \quad (14.20)$$

Because of their definitions the D -value and the F -value are implicitly dependent on the temperature. Therefore, another parameter, the Z -value, is used to describe the susceptibility of microbial death to temperature. The Z -value is defined as [30, 33]

$$Z = \frac{T_2 - T_1}{\log_{10} D_1 - \log_{10} D_2} \quad (14.21)$$

where

D_1 and D_2 are the values of D in correspondence of temperatures T_1 and T_2 for which $D_1 = 10 D_2$.

Because of its definition the Z -value represents the temperature increase corresponding to a 10-fold reduction in the D -value. Typical values for the Z -value are given in Table 14.3. A more mathematically precise definition of the Z -value is

$$Z = -\frac{dT}{d(\log_{10} D)} \quad (14.22)$$

Therefore, assuming that Z is a constant (which is only approximately correct as explained below), a plot of $\log_{10} D$ vs. T should give a straight line with slope $-1/Z$, as shown in Figure 14.11. Integrating this equation between a reference temperature T_{ref} and the desired temperature T , assuming again that Z is constant, results in

$$D = D_{\text{ref}} \cdot 10^{\frac{T_{\text{ref}} - T}{Z}} \quad (14.23)$$

from which D at any temperature can be calculated knowing the Z -value and the value of D (D_{ref}) at any one temperature (T_{ref}).

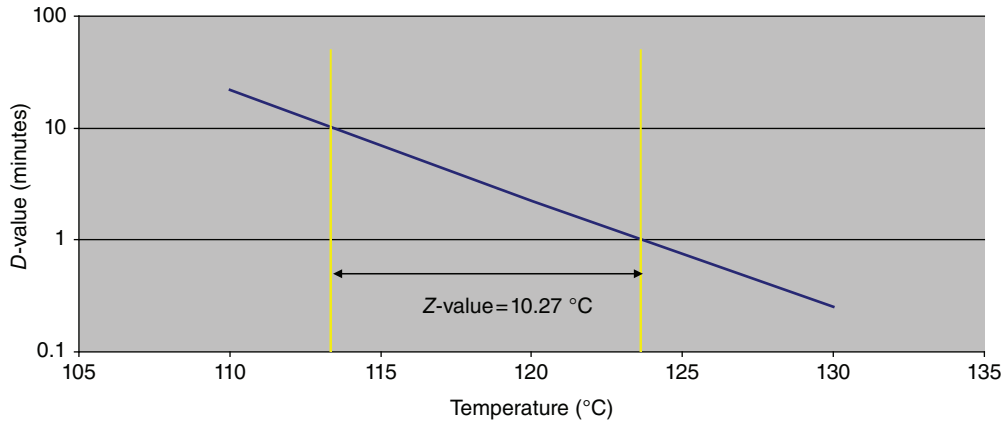


FIGURE 14.11 Plot of $\log_{10}(D)$ vs. temperature.

EXAMPLE PROBLEM 14.4

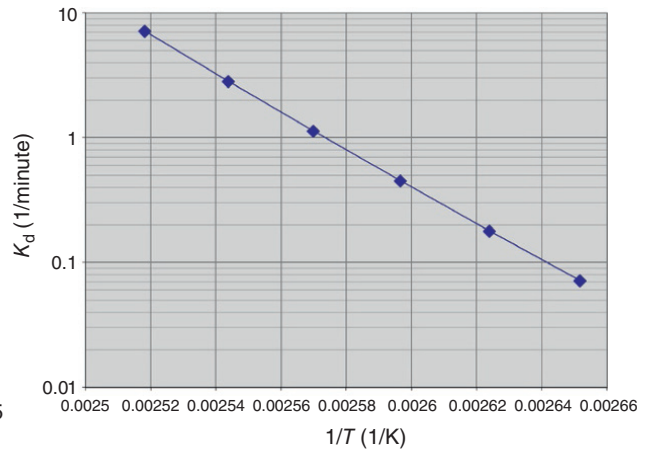
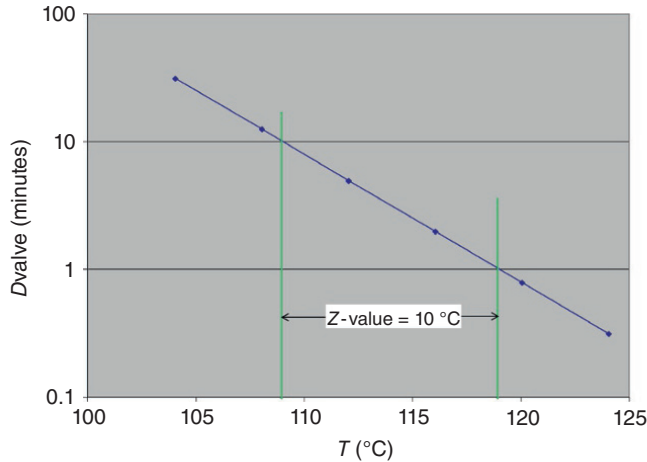
We want to determine the Z -value for a microorganism from a series of thermal inactivation tests conducted on a microbial population at different temperatures in a BIER apparatus. The results of the experiments are reported in the table as “input data.” Each one of D -value in this table was obtained as in the previous example. To find the Z -value Eq. (14.23) can be rewritten as follows:

$$\log_{10}D - \log_{10}D_{\text{ref}} = \frac{T_{\text{ref}}}{Z} - \frac{1}{Z}T$$

Therefore, a regression of the $\log_{10}(D)$ data vs. T will result in a straight line with slope equal to $-1/Z$, as shown in the example figure on the left, resulting in $Z = 9.99\text{ }^\circ\text{C}$, i.e. about $10\text{ }^\circ\text{C}$. It should be remarked that in this case T does not need to be expressed using an absolute temperature scale since Eq. (14.23) shows that only the temperature differences are relevant.

Input Data		Derived Data		
T ($^\circ\text{C}$)	D -Value (min)	T (K)	$1/T$ (1/K)	K_d (min^{-1})
104	31.85	377.16	0.002 651	0.072 295
108	12.68	381.16	0.002 624	0.181 592
112	5.03	385.16	0.002 596	0.459 598
116	2.01	389.16	0.002 570	1.145 565
120	0.80	393.16	0.002 543	2.878 231
124	0.32	397.16	0.002 518	7.240 834

In addition, using Eq. (14.17) it is possible to use the same input data to calculate the K_d value for each D -value as a function of $1/T$ (using an absolute temperature scale, i.e. K in this case). These derived data are also shown in the table. Then by regressing $\ln(K_d)$ vs. $1/T$ as in Eq. (14.13), it is possible to obtain $\ln(K_{d0})$ from the intercept (for $1/T \rightarrow 0$) and $-E_d/R_G$ from the slope to obtain $K_{d0} = 3.62 \times 10^{38}\text{ min}^{-1}$ and $E_d = 2.867 \times 10^5\text{ J/mol}$ (see example figure on the right).



As already mentioned, it is often assumed that $\log_{10}D$ is a linear function of the temperature T , as shown in Figure 14.11. If this were truly valid, then the Z -value would be a constant. However, this assumption is not really correct, as one can verify by substituting the expression for D from Eq. (14.17) into Eq. (14.21). After some simplifications one can easily obtain Z as a function of the generic absolute temperature T from the equation

$$Z = \frac{2.303R_G}{E_d} T^2 \quad (14.24)$$

(Remark: Whereas, according to its definition, the Z -value can be expressed in any kind of temperature scale, absolute temperatures must be used in Eq. (14.24).) Equation (14.24) is clearly a function of temperature and not a constant, as often stated. Therefore, caution should be exerted when Z -values are used to predict the behavior of the system at temperatures much higher than those used to experimentally determine the Z -value in the first place. On the other hand, if the Z -value is used to predict the change in sterilization effectiveness resulting from small temperature changes within the typical temperature range used in commercial sterilization applications, then the term T^2 in Eq. (14.24) is approximately constant, and so is the Z -value. In quantitative terms, if the Z -value at $T = 121^\circ\text{C}$ is known and is assumed to be equal to Z_{121} , then the Z -value at 140, 150, 160, and 170°C will be, respectively, 9.9, 15.3, 20.8, and 26.4% greater than Z_{121} . These limited variations of the Z -value (at least in the common sterilization temperature range) partially justify the popular use of the Z -value in the sterilization calculations.

Once the D -value and the Z -value (assuming it is a constant) have been determined at a given temperature, it is possible to estimate the effect of both time and the actual sterilization temperature on the reduction in microbial population. To do this the initial F -value, F_{ref} (calculated at the reference temperature, T_{ref} , for which the data are available), is first determined using Eq. (14.20), after an initial microbial reduction level has been selected (e.g. 10^{-6}). This will give the time required for the sterilization process. Then the effect of temperature can be obtained from Eq. (14.21), which can be rewritten in terms of F instead of D and used to determine the F -value at a generic temperature, T . The resulting equation for the final reduction in microbial population is [32]

$$F = F_{\text{ref}} \cdot 10^{\frac{T_{\text{ref}} - T}{Z}} \quad (14.25)$$

Since the Z -value is not really a constant but a function of temperature (as shown above), Eq. (14.25) represents only an approximation. If the variability of the Z -value with temperature is accounted for, then a slightly different but more correct expression can be obtained. Such an expression is [34]

$$F_T = F_{T_{\text{ref}}} \cdot 10^{\frac{(T_{\text{ref}} - T) T_{\text{ref}}}{Z_{\text{ref}} T}} \quad (14.26)$$

which is identical to Eq. (14.25) except for the term T_{ref}/T appearing in the exponent. Tabulations of the F -value calculated with this equation are available elsewhere [34].

14.5.2.3 Kinetics of Thermal Degradation of the Product

During the thermal sterilization of a product, microorganisms are not the only ones to be affected by high temperature. The product, for example the organic nutrient components of culture media, also undergoes chemical degradation that can result in the partial loss of some of the valuable characteristics of the product. This is especially true for organic materials and especially protein-containing products since they are likely to be denatured by the temperature just as the proteins that make up the microorganisms are.

The expression for the specific product degradation depends on the underlying degradation mechanism (decomposition, oxidation, polymerization, etc.). However, a common way to quantify it is by using a power-law expression in which the degradation rate is a function only of the concentration of the product itself and where the order of the reaction depends on the degradation mechanism, i.e.

$$r_m = K'_m C_m^\beta \quad (14.27)$$

where

r_m is the specific rate of product inactivation g/(l·s).

C_m is the concentration of product, g/l.

β is the order of the reaction.

K'_m is the proportionality parameter, (l/g) $^{\beta-1}$ s $^{-1}$.

For most products, the rate of inactivation can be assumed to follow a first-order expression similar to that used above to describe the microbial death process:

$$r_m = K_m C_m \quad (14.28)$$

where

K_m is the proportionality parameter, s $^{-1}$.

As usual, the parameter K_m is the specific kinetic rate of product degradation and can be expressed as

$$K_m = K_{m_0} e^{-E_m/R_G T} \quad (14.29)$$

where

K_{m_0} is the Arrhenius constant (frequency factor) of the nutrient degradation process, s $^{-1}$.

E_m is the activation energy of the nutrient degradation process, J/mol.

R_G is the ideal gas constant, J/(mol K).

14.5.3 Analysis of Batch Steam Sterilization Processes for Liquids Using Microbial Death Kinetics

Here we assume that a well-mixed culture of volume V containing, among the others, a specific microbial species especially resistant to steam sterilization (e.g. *G. stearothermophilus*) with an initial N_0 microbial concentration undergoes steam sterilization conducted in batch mode. For the case in which the entire sterilization process was conducted at constant temperature (difficult to achieve in practice), the reduction that we already found for the microbial concentration N at any time is given by

$$\ln \frac{N}{N_0} = -K_d t \quad (\text{for } T = \text{constant}) \quad (14.30)$$

which can be very useful, in first approximation, to determine the sterilization time t needed to achieve the desired microbial reduction (N/N_0).

However, actual batch sterilization processes are conducted in three steps, as shown in Figure 14.12, consisting of heating the batch to the sterilization temperature, holding it at this temperature for as long as needed, and then cooling it down. For example, batch sterilization of culture media is commonly encountered in large-scale bioprocesses. Concentrated non-sterile medium is charged to the bioreactor, which is typically provided with an external jacket and/or internal coils for heating and cooling. The bioreactor is heated up to the sterilization temperature by passing steam in the jackets and coils and/or sparging steam directly through the air sparger. The added live steam produces a dilution effect that must be taken into account when preparing the medium. Once the sterilization temperature has been achieved, the steam supply is shut off, and the bioreactor is allowed to remain at that temperature for the time period required for sterilization to take place (holding period). Then, cooling water and/or brine are used to cool down the bioreactor prior to inoculation.

Heating and cooling can last a long time, especially in large systems, and microbial inactivation occurs even during

these phases, thus adding an additional degree of sterility to that achieved just during the holding phase. While this may increase the effectiveness of the sterilization process and thus be a conservative approach to ensure final sterility, the heating and cooling phases also contribute to product degradation. Therefore, in many cases it is critical to make sure that the overall sterilization process (heating, holding, and cooling) does not result in an “oversterilization” of the batch at the expense of unacceptable product degradation. For this reason, it is often important to determine the *actual* microbial degradation level achieved through the *entire* sterilization process.

As before, the starting point for the analysis is the differential mass balance for the microorganisms in a differential time interval dt , resulting in

$$V \frac{dN}{dt} = -VK_d(T, t)N = -VK_{d0}e^{-E_a/R_G T(t)}N \quad (14.31)$$

where the microbial death rate K_d is a function of not only of temperature but also of time since the temperature in the system changes with time. If $N = N_0$ at $t = 0$, then integration yields

$$\begin{aligned} \ln \left(\frac{N_0}{N} \right)_{\text{non-isothermal cycle}} &= \int_0^t K_d(T, t') dt' = \int_0^t K_{d0} e^{-E_a/R_G T(t')} dt' \end{aligned} \quad (14.32)$$

The overall microbial reduction can be calculated from this equation only if the temperature dependence with time for the entire sterilization process is known since $T(t)$ appears in the exponential term under the integral sign.

The integral in this equation can be more easily calculated numerically if one knows the temperature profile for the entire cycle, as in Figure 14.12. However, it is often convenient to apply the same equation separately to each of the heating, holding, and cooling steps in Figure 14.12 and then calculate the microbial reduction produced by each one of them from

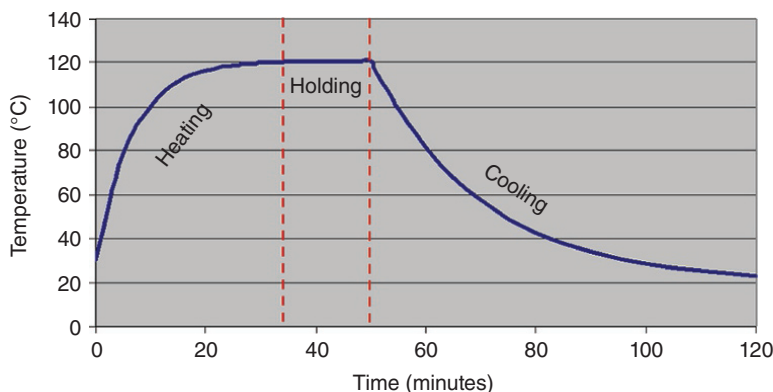


FIGURE 14.12 Temperature profile during a typical steam sterilization process.

$$\ln \frac{N_0}{N_1} = K_{d0} \int_0^{t_1} \exp\left(\frac{-E_d}{R_G T_h(t)}\right) dt \quad (\text{heating phase}) \quad (14.33)$$

$$\begin{aligned} \ln \frac{N_1}{N_2} &= K_{d0} \int_{t_1}^{t_2} \exp\left(\frac{-E_d}{R_G T_q(t)}\right) dt \\ &= K(T_q) \cdot (t_2 - t_1) \quad (\text{holding phase}) \end{aligned} \quad (14.34)$$

$$\ln \frac{N_2}{N} = K_{d0} \int_{t_2}^t \exp\left(\frac{-E_d}{R_G T_c(t)}\right) dt \quad (\text{cooling phase}) \quad (14.35)$$

Then, the final ratio of microbial concentration reduction is given by

$$\ln\left(\frac{N_0}{N}\right)_{\text{non-isothermal cycle}} = \ln \frac{N_0}{N_1} + \ln \frac{N_1}{N_2} + \ln \frac{N_2}{N} \quad (14.36)$$

Clearly, in order to achieve an adequate sterility level, this value for N/N_0 must be lower than the value set through Eq. (14.10).

In these equations N_1 and N_2 are the microbial concentrations at the end of the heating and holding cycles, respectively, and $T_h(t)$, $T_q(t)$, and $T_c(t)$ are the temperature profiles for the heating, holding, and cooling cycle, respectively. Notice that $T_q(t)$ is maintained constant in the vast majority of cases. Therefore, $T_q(t)$ in such cases is just equal to the sterilization temperature, and integration of Eq. (14.34) becomes very simple. From these results one can see that all phases of the sterilization cycle and not only the “sterilizing period” (i.e. the holding cycle) contribute to the overall sterilization. Clearly, if heating and cooling times are short, they will contribute little to overall sterilization, and the time spent at the sterilization temperature will control the extent of microbial reduction. However, if they are long, the batch will spend a considerable amount of time at temperatures not as high as the sterilization temperature but close enough, and this will add not only to the effectiveness of microbial death but also to product degradation.

The integration of these equations requires the knowledge of the initial condition for each step (i.e. the residual microbial concentration at the beginning of each step, obtained from the previous step) as well as the temperature change as a function of time during each step. This, in turn, depends on the way in which the heating or cooling cycles are conducted, i.e. on how quickly heat is supplied or removed. Table 14.4 shows the equations for the most common temperature profiles for the heating and cooling cycles. These functions can be substituted in the previous equations to obtain N_0/N_1 , N_1/N_2 , and N_2/N . The integrated expression for each one of these functions is given elsewhere [35, 36]. More simply, or in all other cases, integration can be conducted numerically once the temperature profiles are known using the equation in Table 14.4 or are obtained experimentally from trial runs.

14.5.4 Analysis of Batch Steam Sterilization Processes for Liquids Using Simplified Expressions for Microbial Death Kinetics

We can now repeat the same analysis conducted in the previous section for the case in which the D -value, F -value, or Z -value approach is used. The starting point is the determination of the temperature profile as a function of time that the process is undergoing. Again, typically batch sterilization processes are conducted in the three steps, shown in Figure 14.12 (heating, holding, and cooling). The mass balance for the microorganisms during a batch sterilization process, shown in Eq. (14.14), can be rewritten using the equation relating the D -value to K_d (Eq. 14.17) as follows:

$$\frac{dN}{dt} = -K_d(T) \cdot N = -\frac{\ln(10)}{D(T)} N = -\frac{2.303}{D(T)} N \quad (14.37)$$

where the notation $D(T)$ and $K_d(T)$ is simply to remind that both are functions of the temperature, which is not constant. Integrating this equation (with initial condition $N = N_0$ at $t = 0$) and recalling how D can be expressed as a function of temperature (Eq. 14.23) yields

TABLE 14.4 Temperature Profiles in Batch Sterilization of Liquids

Type of Heat Transfer	Equation for Temperature Profile of Medium	Type of Function
Electric heating	$T(t) = T_0 + \frac{q}{V\rho_1 C_{pl}} t$	Linear
Steam heating	$T(t) = T_w - (T_w - T_0) \exp\left(-\frac{UA}{V\rho_1 C_{pl}} t\right)$	Exponential
Water cooling	$T(t) = T_{co} + (T_0 - T_{co}) \exp\left[\frac{Q_c C_{pc}}{V\rho_1 C_{pl}} \left(1 - e^{-\frac{tUA}{Q_c C_{pc}}}\right)\right]$	Exponential

where

- T_0 is the initial temperature of medium, °C
- q is the rate of heat transfer, kcal/s
- V is the volume of medium to be sterilized, m³
- ρ_1 is the density of medium, kg/m³
- C_{pl} is the heat capacity of medium, kcal/(kg·°C)
- T_w is the steam temperature, °C
- U is the heat transfer coefficient, kcal/(m²·s·°C)
- A is the heat transfer area, m²
- T_{co} is the initial temperature of cooling water, °C
- Q_c is the flow rate of cooling water, m³/s
- C_{pc} is the heat capacity of cooling water, kcal/(kg·°C)

Source: Adapted from Bailey and Ollis [35].

$$n = \log_{10} \left(\frac{N_0}{N} \right)_{\text{non-isothermal cycle}} \tag{14.38}$$

$$= \int_0^t \frac{1}{D(T, t')} dt' = \frac{1}{D_{\text{ref}}} \int_0^t \left[10^{\frac{T(t') - T_{\text{ref}}}{Z}} \right] dt'$$

which can be solved only if $T(t)$ is known for the entire heating, holding, and cooling stages, as in the rigorous approach design case. The lethal rate, L , also found in the literature, is nothing more than the argument of the integral, and it is defined as

$$L(t) \equiv 10^{\frac{T(t) - T_{\text{ref}}}{Z}} \tag{14.39}$$

It is often convenient to break up the above integral into three components (heating, holding, and cooling), i.e.

$$n = \log_{10} \left(\frac{N_0}{N} \right)_{\text{non-isothermal cycle}} \tag{14.40}$$

$$= \frac{1}{D_{\text{ref}}} \int_0^{t_{\text{h}}} 10^{\frac{T_{\text{h}}(t) - T_{\text{ref}}}{Z}} dt' + \frac{1}{D_{\text{ref}}} \int_{t_{\text{h}}}^{t_{\text{S}}} 10^{\frac{T_{\text{S}}(t) - T_{\text{ref}}}{Z}} dt' + \frac{1}{D_{\text{ref}}} \int_{t_{\text{S}}}^{t_{\text{c}}} 10^{\frac{T_{\text{c}}(t) - T_{\text{ref}}}{Z}} dt'$$

which implies that the final ratio of microbial concentration reduction is given by

$$n = \log_{10} \left(\frac{N_0}{N} \right)_{\text{non-isothermal cycle}} \tag{14.41}$$

$$= \log_{10} \left(\frac{N_0}{N_1} \right) + \log_{10} \left(\frac{N_1}{N_2} \right) + \log_{10} \left(\frac{N_2}{N} \right)$$

EXAMPLE PROBLEM 14.5

A non-isothermal sterilization process is conducted over a 30 minute period in a bioreactor. The temperature is measured over time and is reported in the table as “input data.” The target microorganism to be inactivated has a Z -value of 9°C and a D -value at 119°C equal to 1.38 minutes. We want to determine the microbial \log_{10} -reduction ratio, n , for this process. In this case we can integrate Eq. (14.38) numerically by using any appropriate software package to find that $n = 12.04$.

To do the same integral calculation using a spreadsheet, we can approximate the integral in Eq. (14.38) as a summation of trapezoids, as follows:

$$n = \log_{10} \left(\frac{N_0}{N} \right)_{\text{non-isothermal cycle}} \tag{14.42}$$

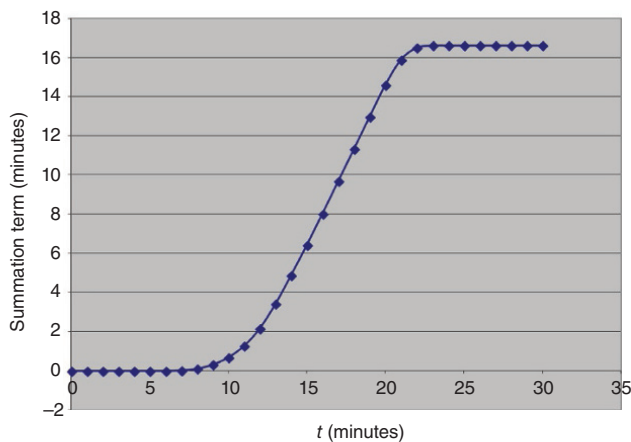
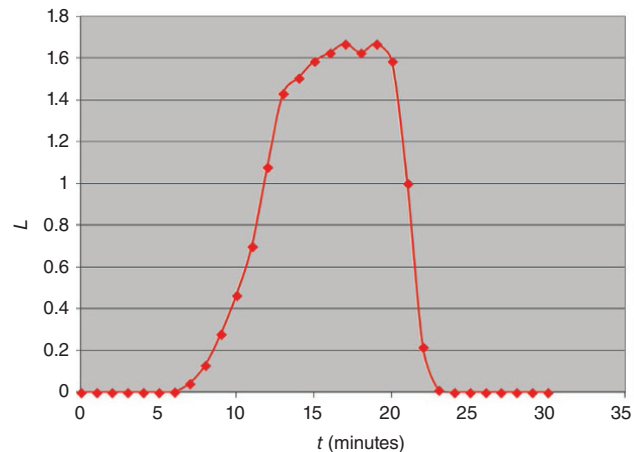
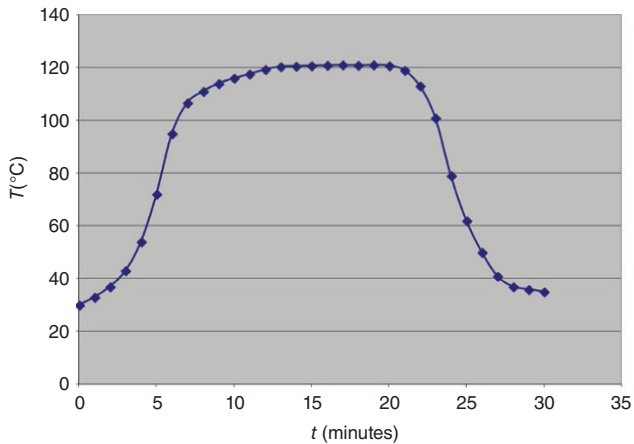
$$\cong \frac{1}{D_{\text{ref}}} \sum_{i=1}^m \left[\frac{L(t_i) + L(t_{i-1})}{2} \cdot (t_i - t_{i-1}) \right]$$

where

$$L(t) = \left[10^{\frac{T(t) - T_{\text{ref}}}{Z}} \right]$$

In this calculation $T_{\text{ref}} = 119^\circ\text{C}$ and $D_{\text{ref}} = 1.38$ minutes. The results are in the “derived data” columns in the table. When the final summation (16.62 minutes) is divided by D_{ref} , the result is $n = 12.04$, once again. Remark: The final summation term (16.62 minutes) is actually the F -value at 119°C . Plots of $T(t)$, $L(t)$, and the summation term as t increases are presented in the example figures below. Notice that the lethality L is practically zero if the temperature is too low. Also, the third plot (summation) is just the area under the curve of $L(t)$.

i	Input Data		Derived Data			
	t (min)	T (°C)	L	Δt (min)	L × Δt (min)	Summation (min)
0	0	30	1.29E-10	0.00	0.00E+00	0
1	1	33	2.78E-10	1.00	2.04E-10	2.04E-10
2	2	37	7.74E-10	1.00	5.26E-10	7.30E-10
3	3	43	3.59E-09	1.00	2.18E-09	2.91E-09
4	4	54	5.99E-08	1.00	3.18E-08	3.47E-08
5	5	72	5.99E-06	1.00	0.00	3.06E-06
6	6	95	2.15E-03	1.00	0.00	0.00
7	7	106.7	4.30E-02	1.00	0.02	0.02
8	8	111	1.29E-01	1.00	0.09	0.11
9	9	114	2.78E-01	1.00	0.20	0.31
10	10	116	4.64E-01	1.00	0.37	0.68
11	11	117.6	6.99E-01	1.00	0.58	1.27
12	12	119.3	1.08E+00	1.00	0.89	2.16
13	13	120.4	1.43E+00	1.00	1.26	3.41
14	14	120.6	1.51E+00	1.00	1.47	4.88
15	15	120.8	1.58E+00	1.00	1.55	6.42
16	16	120.9	1.63E+00	1.00	1.61	8.03
17	17	121.0	1.67E+00	1.00	1.65	9.68
18	18	120.9	1.63E+00	1.00	1.65	11.32
19	19	121.0	1.67E+00	1.00	1.65	12.97
20	20	120.8	1.58E+00	1.00	1.63	14.60
21	21	119	1.00E+00	1.00	1.29	15.89
22	22	113	2.15E-01	1.00	0.61	16.50
23	23	101	1.00E-02	1.00	1.13E-01	16.61
24	24	79	3.59E-05	1.00	5.02E-03	16.62
25	25	62	4.64E-07	1.00	1.82E-05	16.62
26	26	50	2.15E-08	1.00	2.43E-07	16.62
27	27	41	2.15E-09	1.00	1.18E-08	16.62
28	28	37	7.74E-10	1.00	1.46E-09	16.62
29	29	36	5.99E-10	1.00	6.87E-10	16.62
30	30	35	4.64E-10	1.00	3.93E-09	16.62
			16.62			



14.5.5 Equivalent Sterilization Cycles

In many circumstances one may want to determine under what operating conditions two sterilization cycles generate the same final level of sterility. Two cycles are said to be equivalent when they achieve the same final level of sterility, i.e.

$$\left(\frac{N}{N_0}\right)_{\text{cycle1}} = \left(\frac{N}{N_0}\right)_{\text{cycle2}} \Rightarrow \ln\left(\frac{N}{N_0}\right)_{\text{cycle1}} = \ln\left(\frac{N}{N_0}\right)_{\text{cycle2}} \quad (14.42)$$

which also implies that

$$n_{\text{cycle1}} \equiv \log_{10}\left(\frac{N_0}{N}\right)_{\text{cycle1}} = \log_{10}\left(\frac{N_0}{N}\right)_{\text{cycle2}} \equiv n_{\text{cycle2}} \quad (14.43)$$

$$n_{\text{cycle1}} = n_{\text{cycle2}} = n \quad (14.44)$$

i.e. the number of orders of magnitude, n , by which the microbial population has been reduced in equivalent cycles is the same. From this definition, it follows that two isothermal cycles, one conducted at constant temperature T_1 and the other at constant temperature T_2 , are equivalent if the corresponding sterilization times, t_1 and t_2 , are related to each other as follows:

$$\frac{t_2}{t_1} = \frac{F_2}{F_1} = 10^{\frac{T_1 - T_2}{Z}} \quad (14.45)$$

More interesting is the case in which one wants to determine the equivalence between a non-isothermal cycle and an isothermal cycle conducted at a reference temperature T_{ref} . By rearranging the previous equations and recalling that $F = nD$, it is

$$n = \log_{10}\left(\frac{N_0}{N}\right)_{\text{non-isothermal cycle}} = \frac{1}{D_{\text{ref}}} \int_0^t 10^{\frac{T(t') - T_{\text{ref}}}{Z}} dt' = \log_{10}\left(\frac{N_0}{N}\right)_{T_{\text{ref}}} = \frac{F_{\text{ref}}}{D_{\text{ref}}} \quad (14.46)$$

i.e.

$$F_{\text{ref}} = nD_{\text{ref}} = \int_0^t 10^{\frac{T(t') - T_{\text{ref}}}{Z}} dt' \quad (14.47)$$

where

F_{ref} is the time needed to sterilize at a constant temperature T_{ref} and resulting in a sterilization cycle equivalent to that obtained with the non-isothermal cycle.

Similar equations can be obtained using the more rigorous kinetic approach.

EXAMPLE PROBLEM 14.6

We want to determine which sterilization cycles conducted isothermally are equivalent to the non-isothermal examined in the previous example. For example, what is an equivalent cycle conducted isothermally at 119 °C? This will require calculating the F -value at 119 °C, i.e. the time that our batch should spend isothermally at this temperature to achieve the same sterilization level achieved in the previous non-isothermal example, i.e. the same n . However, we

already went through this calculation in the previous example. That cycle resulted in $n = 12.04$. Then, since the D -value 119°C is 1.38 minutes, $F(119^\circ\text{C}) = n \times D(119^\circ\text{C}) = 12.04 \times 1.38 \text{ minutes} = 16.62 \text{ minutes}$, as before.

Similarly, one could ask what the F -value would be if the isothermal cycle is conducted at 121°C . Clearly n will be the same, based on the definition of equivalent cycle, but we now need D at 121°C . Knowing the Z -value (9°C) and D at 119°C , we can calculate D at 121°C using Eq. (14.23) and find out that this value is 0.827 minutes. Then, $F(121^\circ\text{C}) = n \times D(121^\circ\text{C}) = 12.04 \times 0.827 \text{ minutes} = 9.96 \text{ minutes}$.

Finally, what is the F -value if the cycle is conducted at 128°C ? Repeating the same process, we can calculate $D(128^\circ\text{C}) = 0.138 \text{ minutes}$ and $F(128^\circ\text{C}) = 1.66 \text{ minutes}$. Clearly, increasing the sterilization temperature has a dramatic effect on the sterilization time needed to achieve the same level of microbial reduction.

In summary, the previous non-isothermal cycle and all these isothermal cycles result in the same sterilization effects (i.e. microbial reduction) and are thus equivalent.

14.5.6 Analysis of Product Degradation During Batch Steam Sterilization Processes of Liquids

The determination of the extent of product degradation during batch sterilization can be achieved by performing a mass balance for the product similar to that conducted for microorganisms to determine microbial inactivation. If one assumes that the product degradation follows a first-order kinetics with respect to the product concentration, then the resulting mass balance is

$$V \frac{dC_m}{dt} = -VK_m \cdot C_m \quad (14.48)$$

where the concentration at $t = 0$ is C_{m_0} . Integration of this equation gives

$$\ln \frac{C_m}{C_{m_0}} = \int_0^t K_m(T, t') dt' = \int_0^t K_{m_0} e^{-E_a/R_G T(t')} dt' \quad (14.49)$$

So, once again, the determination of the residual fraction of product at the end of the sterilization process, C_m/C_{m_0} , will depend on the temperature profile required by the process to achieve the desired sterilization effects, as shown in Figure 14.12. If the entire process is conducted at constant temperature, this equation simplifies to

$$\ln \frac{C_m}{C_{m_0}} = -K_m t \quad (14.50)$$

In general, if the rate of product degradation is faster than the rate of microbial death, then thermal sterilization will result in significant or possibly complete degradation of the product (as when the product contains proteins) at the end of the sterilization cycle. Fortunately, many products as well as components of culture media have thermal degradation rates that are lower than the microbial death rate, which means that thermal sterilization is often a viable option. If this is not the case, then another sterilization method must be selected.

In addition, the temperature at which the sterilization process is conducted is also critical. The activation energies for the degradation of many products are typically much lower than those associated with microbial thermal inactivation. This implies that if the sterilization temperature is increased (and the sterilization time is appropriately reduced to achieve the same level of sterility), the kinetics of the microbial death process increases much more rapidly than the kinetics of product degradation. This results in a sterilization process in which the product degradation is more limited than if the same process had been conducted at a lower temperature but for a longer time.

As an example, one may consider the case of thiamine degradation during heat sterilization, assuming that the medium must be sterilized against *G. stearothermophilus* [26, 37]. The specific kinetic rates for thermal death and thiamine inactivation are reported in Figure 14.13 as a function of the inverse of the sterilization temperature. The slope of each of these curves represents the activation energy of the process. If the sterilization took place at 100°C , then the time

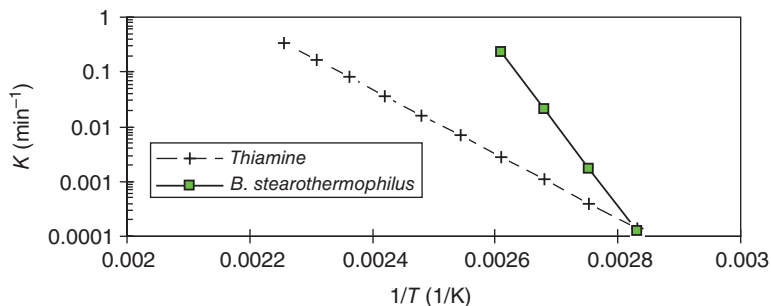


FIGURE 14.13 Specific kinetic rates for thermal death of *Geobacillus stearothermophilus* and thiamine degradation as a function of $1/T$ (temperature in K).

required to reduce the concentration of the organisms by 12 orders of magnitude would be about 1300 minutes. Even discounting the impracticality of such a long sterilization process, more than 75% of the thiamine would be degraded during this period.

However, if the sterilization is conducted at 140 °C, then the same degree of sterility would be achieved in less than 12 seconds, and the loss of thiamine would be only about 0.7%. This is one of the reasons (in addition to the considerably shorter operating time and reduced overhead) why continuous sterilization processes are conducted at high temperatures and for a short period of time, as already discussed above.

EXAMPLE PROBLEM 14.7

A 2000 bioreactor undergoes sterilization at 121 °C. The medium is estimated to contain an initial microbial concentration equal 1500 spores/ml that should be reduced so that the failure rate should be no more than 1 in 10 000 sterilizations. The medium also contains 40 mg/l of a heat-sensitive critical nutrient. Let us determine the amount of the nutrient left at the end of the sterilization process. Given:

$$K_{d_0} = 9.5 \times 10^{37} \text{ min}^{-1} \quad \text{and} \quad E_d = 2.876 \times 10^5 \text{ J/mol (spores)}$$

$$K_{m_0} = 1.5 \times 10^4 \text{ min}^{-1} \quad \text{and} \quad E_m = 4.4 \times 10^4 \text{ J/mol (nutrient)}$$

We need first to determine n and the final value of N at the end of the sterilization process. Following the procedure used before, it is

$$\frac{\text{SAL}}{N_0 \cdot V} = 3.33 \times 10^{-14} \Rightarrow n = -\log_{10} \left(\frac{\text{SAL}}{N_0 \cdot V} \right) = 13.48$$

Using Eq. (14.10) we can additionally determine that $N_f = N = 5 \times 10^{-11}$ spores/ml.

Entering the temperature (in °K) in Eq. (14.13), we can find that

$$K_d = 2.586 \text{ min}^{-1} \text{ at } T = 121 \text{ °C} = 394.15 \text{ K}$$

Therefore, the corresponding sterilization time from Eq. (14.15) is

$$t = 12 \text{ minutes at } T = 121 \text{ °C} = 394.15 \text{ K}$$

As for the nutrient, K_m at the sterilization temperature is found to be

$$K_m = 0.022 \text{ min}^{-1} \text{ at } T = 121 \text{ °C} = 394.15 \text{ K}$$

Eq. (14.50) can now be used to find out the effect of the sterilization on the nutrient knowing the required sterilization time:

$$\ln \frac{C_m}{C_{m_0}} = -K_m t = -0.022 \frac{1}{\text{min}} 12 \text{ minutes} = -0.265$$

$$\Rightarrow \frac{C_m}{C_{m_0}} = 76.7\%$$

Therefore, of the 40 mg/l of nutrient in the initial batch, only 30.68 mg/l are left after sterilization at 121 °C.

If the sterilization process is conducted instead at 128 °C, then similar calculations show that achieving the same sterilization level will require only 2.65 minutes (practically achievable only through continuous sterilization) and that the concentration of the nutrient left in the batch will be 37.14 mg/l i.e. 92.9% of the initial value. Clearly sterilizing at a higher temperature is advantageous in this case since the kinetics of microbial inactivation increases faster with temperature than the kinetics of nutrient degradation.

14.5.7 Analysis of Continuous Sterilization Processes for Liquid Culture Media

14.5.7.1 Continuous Sterilization of Liquids Assuming Piston Flow For the continuous sterilization system shown in Figure 14.6, a simple but conservative design of a continuous sterilizer can be obtained assuming that the desired level of sterility must be achieved in the holding tube alone, thus neglecting any contribution of the other sections [38]. In addition, one can further assume that the medium moves in piston flow in the holding tube (this assumption will be later removed to account for axial dispersion). As a result, each slug of medium flowing through the holding section will not be affected by the presence of other slugs of material preceding or following it. Since the medium enters the holding section at the desired sterilization temperature and leaves it at the same temperature after a time, t_{res} , equal to the residence time in the holding tube, then the level of sterility can be calculated using an equation similar to Eq. (14.15), i.e.

$$\ln \frac{N}{N_0} = -K_0 e^{-E_d/RT_{\text{ster}}} t_{\text{res}} \quad (14.51)$$

where

$$t_{\text{res}} = \frac{L}{v} = \frac{LS'}{Q} \quad (14.52)$$

where

t_{res} is the residence time of medium in holding tube, second.

L is the length of holding tube, m.

v is the velocity of medium in holding tube, m/s.

S' is the cross-sectional area of holding tube, m².

Q is the volumetric flow rate of medium, m³/s.

In the more general case in which the contributions of all the sections of the sterilizing system are considered, the equations to calculate the rate of microbial reduction produced by each section can be obtained using a method similar to that used in the batch sterilization process case. The only

TABLE 14.5 Temperature–Time Profiles in Continuous Sterilization Systems

Type of Heat Transfer	Equation for Temperature Profile	Type of Function
Electric heating	$T(t) = T_0 + \frac{q}{S\rho_l C_{pl}L}t$	Linear
Steam heating	$T(t) = T_w - (T_w - T_0)\exp\left(-\frac{UA}{S\rho_l C_{pl}L}t\right)$	Exponential
Countercurrent heat exchange	$T(t) = T_{c1} + \frac{(T_0 - T_{c1})}{a-1} \left\{ 1 - \exp\left[\frac{(a-1)UA}{S\rho_l C_{pl}L}t\right] \right\}$	Exponential
Countercurrent heat exchange with $a = 1$	$T(t) = T_0 + \frac{UA}{S\rho_l C_{pl}L}(T_{c1} - T_0)t$	Linear

$$\text{with } a = \frac{Q\rho_l C_{pl}}{Q_c \rho_c C_{pl}}$$

where

q is the rate of heat transfer, kcal/s

ρ_l is the density of medium, kg/m³

C_{pl} is the heat capacity of medium, kcal/(kg·°C)

L is the length of exchanger, m

T_w is the steam temperature, °C

U is the heat transfer coefficient, kcal/(m²·s·°C)

A is the heat transfer area, m²

T_{c1} is the exit temperature of service fluid, °C

Q and Q_c are flow rate of medium and service fluid, respectively, m³/s

C_p and C_{pc} are heat capacity of medium and service fluid, respectively, kcal/(kg·°C)

difference is that the temperature–time profiles will now be obtained by making the appropriate energy balance in each heat exchanger so as to obtain the temperature of each infinitely thin slug of medium as it travels through it. The resulting temperature–time profiles are reported in Table 14.5 [17, 18, 28, 39]. These temperature profiles can then be inserted in the following equation:

$$\ln\left(\frac{N_0}{N}\right)_{\text{exchanger}} = K_{d0} \int_0^t \exp\left(\frac{-E_d}{R_G T(t)}\right) dt \quad (14.53)$$

in which $T(t)$ is one of the expressions given in Table 14.5 to obtain the rate of microbial reduction produced in each exchanger. The final microbial reduction can be obtained by an equation similar to Eq. (14.36) in which all contributions are accounted for.

The same approach can be used to determine the extent of product degradation. Accordingly, the mass balance for the product moving through each section of the continuous sterilizer is

$$\ln\left(\frac{C_{m0}}{C}\right)_{\text{exchanger}} = \int_0^t K_m(T, t') dt' = \int_0^t K_{m0} e^{-E_d/R_G T(t')} dt' \quad (14.54)$$

from which the final value for C_m/C_{m0} can be obtained.

It was already mentioned above that increasing the sterilization temperature and reducing the sterilization time accordingly to achieve the same level of sterility is often beneficial to products since the activation energies of many product degradation reactions are smaller than those for microbial inactivation reaction. This is one of the main reasons to use very high-temperature short-time processes such as in continuous sterilization. This improvement cannot only be predicted but also be verified in practice. Data on the performance of fermentation processes using batch and continuous sterilization processes have been reported [26]. For example, the production of riboflavin has been shown to improve by more than a factor of 7.4 in one case and 70 in another when the process was switched from batch sterilization (at 121 °C for 25 or 45 minutes) to continuous sterilization (at 135 °C for five minutes). Similar improvements (from 50 to 1200%) were observed in the fermentation of vitamin B12 and were attributed to the preservation of the nutrient when continuous sterilization was adopted.

14.5.7.2 Effect of Axial Dispersion on Continuous Sterilization of Liquid Media

In the previous section the assumption was made that the medium flows through all parts of the continuous sterilizer in piston flow. In practice, this is quite difficult to accomplish. Firstly, the medium is always subjected to a number of flow expansions, contractions,

and direction changes (as, for example, when it flows through a plate-and-frame heat exchanger), which introduce some degree of back mixing. Secondly, even assuming that the flow occurs in a straight circular pipe, the mean velocity of the fluid is different from the velocities of the single fluid elements [40, 41].

The type of flow in a circular pipe depends on the Reynolds number, Re , defined as $vD\rho/\mu$ where v is the average velocity, D is the pipe diameter, ρ is the fluid density, and μ is the fluid viscosity [42]. For $Re < 2100$ the flow in the pipe will be laminar, i.e. the fluid elements will all move in the same direction of the flow. For $Re > 4000$ the flow will be turbulent and will be characterized by the presence of turbulent eddies superimposed on the average longitudinal flow. If the flow is laminar, the distribution of velocities will be parabolic with respect to the radial position. This implies that the fluid elements near the pipe wall will travel much more slowly than, say, the fluid element in the center of the pipe (having a velocity twice the average velocity). If the flow is turbulent, radial mixing will also occur, and the average velocity will typically be approximately 25% greater than the average velocity. Therefore, piston flow constitutes only the limit case of a fluid flowing under very turbulent flow conditions.

The existence of velocity distributions implies that if the average velocity is used in the design of the sterilizer, then those fluid elements traveling at higher velocities will be associated with a lower degree of sterility. A possible remedy is to determine the velocity of the fastest moving element and use that in the design calculation instead of the average velocity. This is of course a conservative approach since all fluid elements except the fastest one will be oversterilized.

A better approach is to model the phenomenon by assuming that an axial diffusion effect (produced by the flow turbulence and back mixing) is present. This effect can be superimposed on the axial translation effect produced by the average fluid velocity. Then, a steady-state microbial balance can be written for an infinitely thin element of medium maintained at a constant sterilization temperature [43]. In the derivation of the following equations, it is assumed that the temperature is constant. Therefore, this approach can only be used for the design of the holding tube. Then, the microorganism balance takes the form [27, 44]

$$v \frac{dN}{dx} - D_x \frac{d^2N}{dx^2} = -K_d N \quad (14.55)$$

where

D_x is the axial dispersion coefficient of the microorganisms, m^2/s .

x is the axial coordinate, m

In this equation the first term represents the net number of organisms entering the element because of the average fluid

flow, the second term represents the contribution to the mass balance of axial dispersion, and the last term represents the microbial reduction due to the thermal death process.

The boundary conditions for this equation can be established at the beginning and at the end of the holding section. They are

$$vN_0 = \left(vN - D_x \frac{dN}{dx} \right) \Big|_{x=0} \quad (14.56)$$

and

$$\frac{dN}{dx} \Big|_{x=L'} = 0 \quad (14.57)$$

where

L' is the length of the holding tube, m .

Integration of this equation yields

$$\frac{N(x=L')}{N_0} = \frac{4\zeta e^{Pe/2}}{(1+\zeta)^2 \exp\left[\frac{Pe\zeta}{2}\right] - (1-\zeta)^2 \exp\left[-\frac{Pe\zeta}{2}\right]} \quad (14.58)$$

where

$$\zeta = \sqrt{1 + \frac{4Da}{Pe}} \quad (14.59)$$

This solution contains two important nondimensional number, i.e. the Peclet number, Pe , and the Damköhler number, Da , defined, respectively, as

$$Pe = \frac{vL'}{D_x} \quad (14.60)$$

and

$$Da = \frac{K_d L'}{v} \quad (14.61)$$

The solutions of Eq. (14.58) are graphically displayed elsewhere [27, 35]. For large values of Pe (i.e. for small deviation from piston flow), Eq. (14.58) can be rewritten as [35]

$$\frac{N(x=L')}{N_0} \cong \exp\left[-Da + \frac{Da^2}{Pe}\right] \quad (14.62)$$

It should be remarked, once again, that the equations derived in this section are valid only for isothermal systems. Therefore, unlikely the equation obtained in the previous

section, they can only be used for the design of the holding section and not for the design of the heat exchangers.

14.5.7.3 Continuous Sterilization of Liquids Containing Solid Particles The presence of solid particles, such as agglomerates or lumps, in industrial liquid media for large-scale fermentation is quite common [21]. As explained in the next sections in greater details, heat transfer in solids is a slow process if the solids have poor heat conductivity. Consequently, the temperature in the core of suspended solids can be appreciably lower than that of the surrounding liquid during the short sterilization times required by continuous sterilization. This may result in the microorganisms lodged inside the particles to be less effectively inactivated during sterilization process than those in the liquid bulk [19, 20]. By contrast, this is not typically an issue in most batch processes since the longer sterilization time (typically 20 minutes) and longer heating and cooling times usually provide ample time for small particles to thermally equilibrate with the surrounding liquid, thus ensuring a similar level of microbial reduction in both the particles and the liquid. At any rate, solid particles should be excluded, whenever possible, from the liquid prior to continuous sterilization. If not, the impact of the presence of solids on sterilization effectiveness should be accounted for in the design of the sterilization process.

If solids are present in the medium moving in a continuous sterilization, a design approach accounting for sterilization inside the solids was described in the literature and can be used [17, 18].

14.5.8 Analysis of Dry Heat Sterilization and Depyrogenation Processes

Dry heat sterilization processes can be analyzed in a similar fashion as moist heat processes but with a major difference: the rate of microbial death under dry conditions is much slower than when the medium is aqueous or if sterilization is conducted in the presence of steam. If temperature were the only factor affecting thermal sterilization, this should not be the case. However, the mechanism of microbial inactivation is different for dry heat (dehydration and protein oxidation as opposed to denaturation of cell proteins, as in the moist heat case). Therefore, one should expect that the kinetic parameters for inactivation reaction would be different. In addition, heat transfer in air is much less efficient than steam condensation, and this may result in local temperature inhomogeneities that slow down the overall inactivation process.

In any case, the design equations for dry heat processes are the same as for moist heat. For example, for constant temperature dry heat sterilization, the microbial reduction ratio is given by

TABLE 14.6 D-Values and Z-Values for Microorganisms During Dry Heat Sterilization

Microorganism	D-Value (min)	Z-Value (°C)
<i>Bacillus subtilis</i>	30 (at 120 °C)	18.3–23.3
	2.7 (at 140 °C)	
	2 (at 150 °C)	
	1.3 (at 160 °C)	
	0.5 (at 180 °C)	
<i>Bacillus stearothermophilus</i>	0.3 (at 160 °C)	14.4–24.4
	0.15 (at 180 °C)	
<i>Clostridium sporogenes</i>	6 (at 150 °C)	18–33
	2 (at 160 °C)	
	0.25 (at 180 °C)	

Source: Data from Wallhäuser [32].

$$\log_{10} \frac{N}{N_0} = -\frac{t}{D_{\text{dry heat}}} \quad (14.63)$$

However, sterilization parameters such as the *D*-values and *Z*-values are significantly higher for dry heat sterilization than for steam sterilization, as shown in Table 14.6. All other factors being equal, this results in higher sterilization temperatures and/or longer sterilization times, since the approach to determine the degree of sterilization these processes is the same as that outlined for moist heat sterilization.

A mitigating factor in the design of dry heat sterilization processes for oily substances is the fact that the initial bioburden is typically much lower than that of aqueous media, and it is often 100–200 microorganisms/ml (as opposed to 10⁴–10⁹ microorganisms/ml in aqueous media) [32]. In addition, the volumes to be dry sterilized are typically small.

If dry heat is used for depyrogenation, the inactivation kinetic parameter (*D*-values and *Z*-values) are even higher, thus requiring higher temperature and/or longer times. For example, ampoules and vials are routinely depyrogenated, at a minimum of 250 °C for 30 minutes or 200 °C for 60 minutes.

14.5.9 Analysis of Sterilization Processes for Solids

When a solid or a stagnant liquid in a container is heat sterilized, the main route by which heat is transported within the solid (or liquid) is conduction rather than convection. Convection is a notoriously slower heat transfer mechanism than conduction. This implies that even if the temperature around the solid is quickly raised to the sterilization temperature, it will still take some time for the heat to travel to the core of the solid and sterilize its content. If the solid is large and/or the sterilization time too short, it may well be that the solid center never reaches the sterilization temperature or the core is maintained at the sterilization temperature for too short a time period to ensure adequate sterility at the end of the process.

These problems can be analyzed using unsteady-state heat transfer equations. For the case in which the solid is a slab or a square prism, the unsteady-state heat transfer equation in three dimensions can be written as [45]

$$\rho_s C_{ps} \frac{\partial T_s}{\partial t} = k_s \left(\frac{\partial^2 T_s}{\partial x^2} + \frac{\partial^2 T_s}{\partial y^2} + \frac{\partial^2 T_s}{\partial z^2} \right) \quad (14.64)$$

where

- T_s is the temperature inside the solid, °C.
- ρ_s is the density of the solid, kg/m³.
- C_{ps} is the heat capacity of the solid, J/(kg·°C).
- k_s is the heat conductivity of the solid, J/(m h·°C).

For the simpler case in which heat transfer occurs only (or primarily) in one direction, as in a long (semi-infinite) slab of fixed thickness heated from both sides and having a uniform initial temperature, the previous equation becomes

$$\rho_s C_{ps} \frac{\partial T_s}{\partial t} = k_s \frac{\partial^2 T_s}{\partial x^2} \quad (14.65)$$

with initial condition $t = 0$, $T = T_{s_0}$

and boundary condition $x = \pm b$, $T = T_{s_1}$

where

- b is the 1/2 of the thickness of the slab, m.
- T_{s_0} is the initial temperature of the slab, °C.
- T_{s_1} is the temperature at the surface of the slab, °C.
- x is the axial distance, m.

The solution of this equation is [45, 46]

$$\frac{T_{s_1} - T_s}{T_{s_1} - T_{s_0}} = 2 \sum_{n=0}^{\infty} \frac{(-1)^n}{\left(1 + \frac{1}{2}\right)\pi} e^{-(n+\frac{1}{2})^2 \pi^2 \frac{at}{b^2}} \cos \left[\left(n + \frac{1}{2} \right) \frac{\pi x}{b} \right] \quad (14.66)$$

where

$$\alpha = k_s / \rho_s C_{ps}$$

Although this equation may look complex because of the presence of the infinite summation, the solution typically converges very rapidly using only the first few terms (sometimes even only one) of the summation. The temperature at the center of the slab can be obtained by imposing that $x = 0$. The solution for this equation has been tabulated in graphs for convenient use [42]. In addition, if a finite square prism is considered, the solution can still be obtained as the product of three solutions such as that represented by Eq. (14.66) for the three dimensions [46, 47].

If the solid is a sphere with homogeneous initial temperature and internal composition and is heated by keeping its external surface at constant temperature, the following equation applies:

$$\rho_s C_{ps} \frac{\partial T_r}{\partial t} = k_s \left(\frac{\partial^2 T_r}{\partial r^2} + \frac{2}{r} \frac{\partial T_r}{\partial r} \right) \quad (14.67)$$

where

- r is the radial distance from the center of the sphere, m.
- T_r is the temperature in the solid at a position r inside the particle, °C.

with boundary and initial conditions

$$\left. \frac{\partial T_r}{\partial r} \right|_{r=0} = 0 \quad (14.68)$$

$$T_r(r=R, t) = T_{r_1} \quad (14.69)$$

$$T_r(r, t=0) = T_{r_0} \quad (14.70)$$

where

- R is the radius of the sphere, m.
- T_{r_0} is the initial temperature of the particle, °C.
- T_{r_1} is the temperature at the surface of the particle, °C.

The solution of this equation takes the general form [46]

$$T_r(r, t) = T_{r_1} + \frac{2R(T_{r_1} - T_{r_0})}{\pi r} \sum_{n=1}^{\infty} \frac{(-1)^n}{n} e^{-\left(\frac{n\pi}{R}\right)^2 at} \sin \frac{n\pi r}{R} \quad (14.71)$$

At the center of the sphere, this equation reduces to

$$T_r(r=0, t) = T_{r_1} + 2(T_{r_1} - T_{r_0}) \sum_{n=1}^{\infty} (-1)^n e^{-\left(\frac{n\pi}{R}\right)^2 at} \quad (14.72)$$

EXAMPLE PROBLEM 14.8

A spherical solid mass with a 12 mm diameter (~1/2 in.) initially at room temperature (20 °C) is dropped in boiling water (100 °C). Let us determine the temperature at the center of this object after two minutes. Data (similar to those of a cellulosic material):

- Heat conductivity of solid material, $k_s = 0.208$ W/(m·°C).
- Density of solid material, $\rho_s = 825$ kg/m³.
- Heat capacity of solid material, $C_{ps} = 2385$ J/(kg·°C).

- External heat transfer coefficient between solid object and water, $h = 600 \text{ W}/(\text{m}^2 \cdot ^\circ\text{C})$ (this value is very high, thus keeping the surface of the object always nearly identical to that of the surrounding water).

We can use Eq. (14.72) to solve this problem. In this case $\alpha = k_s/(\rho_s C_{ps}) = 0.105 \times 10^{-6} \text{ m}^2/\text{s}$ and $R = 6 \text{ mm}$. The use of this equation seems formidable because of the infinite summation. However, one should realize that the values of the exponential terms drop very rapidly as n increases. For example, for $t = 2$ minutes the first term in the summation (with $n = 1$) is equal to -0.031 , the second term (with $n = 2$) is equal to 9.089×10^{-7} (i.e. sign aside, 5 orders of magnitude smaller than the first), and the third term (with $n = 3$) is equal to -2.551×10^{-14} (i.e. 12 orders of magnitude smaller than the first). Higher terms will be even smaller. So, in practice, only the first few terms (or even just the first one in this case) in the summation are typically sufficient to achieve convergence in most cases. This calculation can be easily made with available computer software or even with a spreadsheet. In this problem, Eq. (14.72) yields $T = 95.06 \text{ }^\circ\text{C}$. This means that even after two minutes the core of this small object has not even approximated the external boiling water temperature. Given the extreme sensitivity of the microbial inactivation reaction to temperature, this can have a shielding effect on the microorganisms in this object with obvious impact on the achievement of the desired sterility level in actual sterilization processes.

Once the temperature profile is known as a function of time and position within the solid, it is possible to calculate the residual fraction of microorganisms by integrating the sterilization equation over the entire volume of the solid and for the time period of sterilization. For the case of a sphere in which the initial concentration of microorganisms is constant and equal to N_0 , this integral takes the form [17, 18]

$$\frac{N(t)}{N_0} = \frac{3}{4\pi R^3} \int_0^R \exp \left[\int_0^t k_{d_0} \exp \left(\frac{E_d}{R_g T_r(r,t)} \right) dt \right] 4\pi r^2 dr \quad (14.73)$$

Because of the presence of the temperature function $T(r,t)$ (as given by Eq. (14.71)), this integral can only be integrated numerically.

14.6 RADIATION STERILIZATION PROCESSES

The use of radiation as a way to sterilize products or equipment is not new, but in recent years this approach has become more widely used as an alternative to more conventional sterilization processes, such as heat treatments, especially if the product is heat sensitive (e.g. biological products).

14.6.1 Basic Radiation Sterilization Concepts

The type of radiations used in the biopharmaceutical industry for this purpose can be classified into ionizing or nonionizing. Ionizing radiations are radiations that contain enough energy to result in the ejections of electrons from the receiving atoms or molecules, which are then ionized. Ionizing radiations can be further classified into corpuscular, i.e. energetic subatomic particles moving at high speeds (typically at 1% of the speed of light or larger), and electromagnetic, i.e. high energy, high-frequency electromagnetic waves with high penetration value, such as X-rays or gamma (γ)-rays. Electromagnetic radiations, such as radio waves, infrared, visible and UV light, X-rays, and gamma rays, refer to the electromagnetic waves (or their quanta, i.e. photons) propagating through space and carrying radiant energy. The shorter the radiation wavelength, the higher the energy of each photon, and the greater its penetration power will be. For this reason, gamma rays (with a 0.000 01-nm wavelength) are high energy ionizing and penetrating waves, whereas UV light (with wavelength of about 10–400 nm) is not.

When high energy radiation interacts with matter, electron ejection results in the formation of highly reactive free radicals producing double-bond formation, cross-linking, chemical bond breakup, and many other types of chemical reactions. Interaction of high energy radiation with microorganism leads to major DNA breakage, cell membrane damage, and other types of cellular damage, resulting in microbial inactivation (Figure 14.14). Therefore, high energy radiations, especially electron beams and, more commonly, gamma rays, are used industrially to achieve sterilization. By contrast, lower energy nonionizing electromagnetic radiations, such as UV light, do not carry enough energy to ionize molecules, but they can only excite electrons, i.e. move them to a higher energy state. This may result in the possible acceleration of reactions, which, in microorganisms, may lead to the formation of abnormal bonds in the DNA but not typically its breakage. In such a case, DNA damage occurs mainly through the formation of pyrimidine dimers, which can be formed at wavelengths in most of the UV range, including much of the band that is in the nonionizing range. UV light, even if nonionizing, can produce some free radicals that induce cellular damage. Therefore, nonionizing radiations cannot be used for sterilization purposes but are used for disinfection and the maintenance of asepsis.

14.6.2 Sterilization Processes Using Ionizing Radiations

Industrial radiation sterilization processes utilize electron beams (e-beams), gamma rays, or X-rays. The first and second radiation types are the most common [48–50]. E-beam systems generate a high energy electron sheet (with electron energies on the order of 8–10 megaelectron volts (MeV))

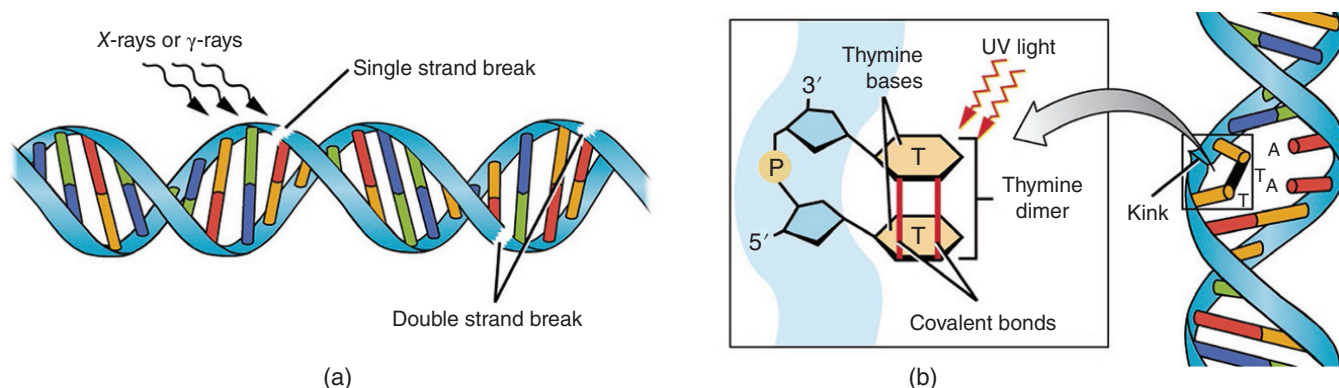
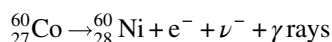


FIGURE 14.14 Effect of ionizing and nonionizing radiation on DNA. (a) Ionizing radiation (b) non-ionizing radiation. *Source:* Courtesy of McGraw-Hill.

using a cathodic electron gun and a magnetic electron accelerator. This sheet then irradiates the product moving under it on a conveyor belt (Figure 14.15) and sterilizes its content, provided that it is made of low-density material ($<0.30 \text{ g/cm}^3$), such as plastic syringe packages, since e-beams have relatively low penetration levels. In case the container to be sterilized is too large, a two-pass (front and back) e-beam irradiation approach is used (Figure 14.15).

Gamma ray sterilization utilizes instead radioactive isotopes, typically cesium-137 (^{137}Cs) or, more commonly, cobalt-60 (^{60}Co), as gamma ray sources. Cobalt-60 undergoes spontaneous beta decay (β -decay) with a half-life of 5.272 years in which a neutron is transformed into a proton inside the atomic nucleus according to the following nuclear reaction:



resulting in the formation of nickel, an electron e^{-} , an anti-neutrino ν^{-} , and high energy ionizing gamma rays with 1.17 and 1.33 MeV energies.

In radiation sterilization facilities, Cobalt-60 slugs are assembled in wall-like modular racks (Figure 14.16) forming the irradiation cell that is placed at the center of a thick-walled concrete structure to avoid dispersing radiation outside the facility (Figure 14.17). The material undergoing sterilization is slowly moved by a conveyor belt around the irradiation cell, thus receiving the prescribed radiation dose necessary to achieve sterilization. The process is routinely used to sterilize drug substances and products such as syringes [51, 52], is highly reliable, and has a high SAL since it can be monitored by placing dosimeters (devices used to measure an absorbed dose of ionizing radiation) within the material undergoing sterilization.

E-beams and ionizing radiations not only sterilize but can also promote product and material degradation, as well as discoloration of containers, as shown in Figure 14.18 for the case of glass vials. These effects are typically small and

acceptable, but tests must be conducted in advance to ensure that this is the case. It should be stressed however that radiation sterilization does *not* make the material being sterilized radioactive since radiation energy levels much higher than those achievable by e-beams or Cobalt-60 decay and needed for this purpose. Induced radioactivity in irradiated material appears only if the energy level of the fast electron source is much higher than 10 MeV or if a source of gamma rays with energy levels higher than 5 MeV is available. The radioactive decay of Cobalt-60 cannot generate such high energy level radiations.

14.6.3 Disinfection Processes Using Nonionizing Radiations

These types of processes utilize primarily UV radiation to achieve their objectives since UV radiation has low penetration power and is nonionizing and thus capable of achieving disinfection but not full sterilization in most cases [54]. Although UV radiation includes radiations approximately in the range 10–400 nm, the wavelength range used for disinfection is typically 240–310 nm. UV disinfection results mainly from the reaction between two molecules of thymine (one of the key DNA nucleobases) within the DNA structure to form thymine dimers. When enough damage of this type is produced, the microorganism is inactivated. DNA absorbs UV radiation most effectively at a wavelength of 265 nm commonly used in UV units. Significant level of disinfection can be obtained with UV systems in pharmaceutical water treatment plants (>99.9%).

UV technology was first used industrially for the disinfection of water in municipal water treatment plant, but UV lamp units are now commonly found in the pharmaceutical industry to disinfect process water especially in ultrapure water systems. In such cases UV lamps are typically found after prefilters, prior to the reverse osmosis unit, and in recirculation water loops found in purified water storage and distribution systems. UV radiation lamps are also commonly used to



FIGURE 14.15 Sterilization system using E-beam: front irradiation (top panel) and back irradiation (bottom panel) of large packages. *Source:* Courtesy of Rhodotron, IBA.

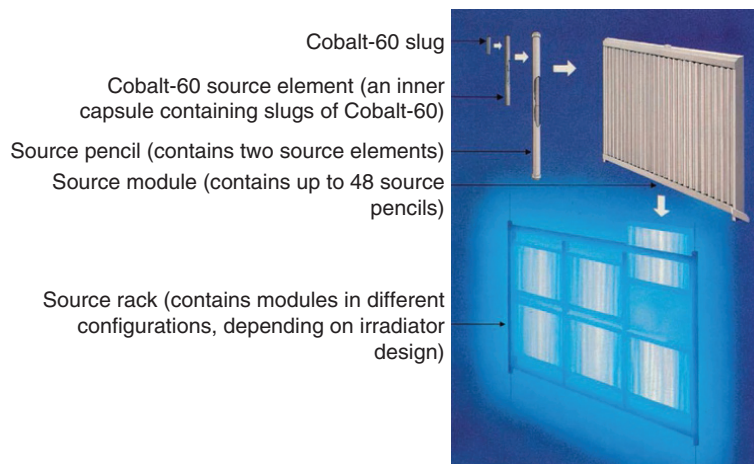


FIGURE 14.16 Cobalt 60 slugs in modular rack. *Source:* Courtesy of Nordion.

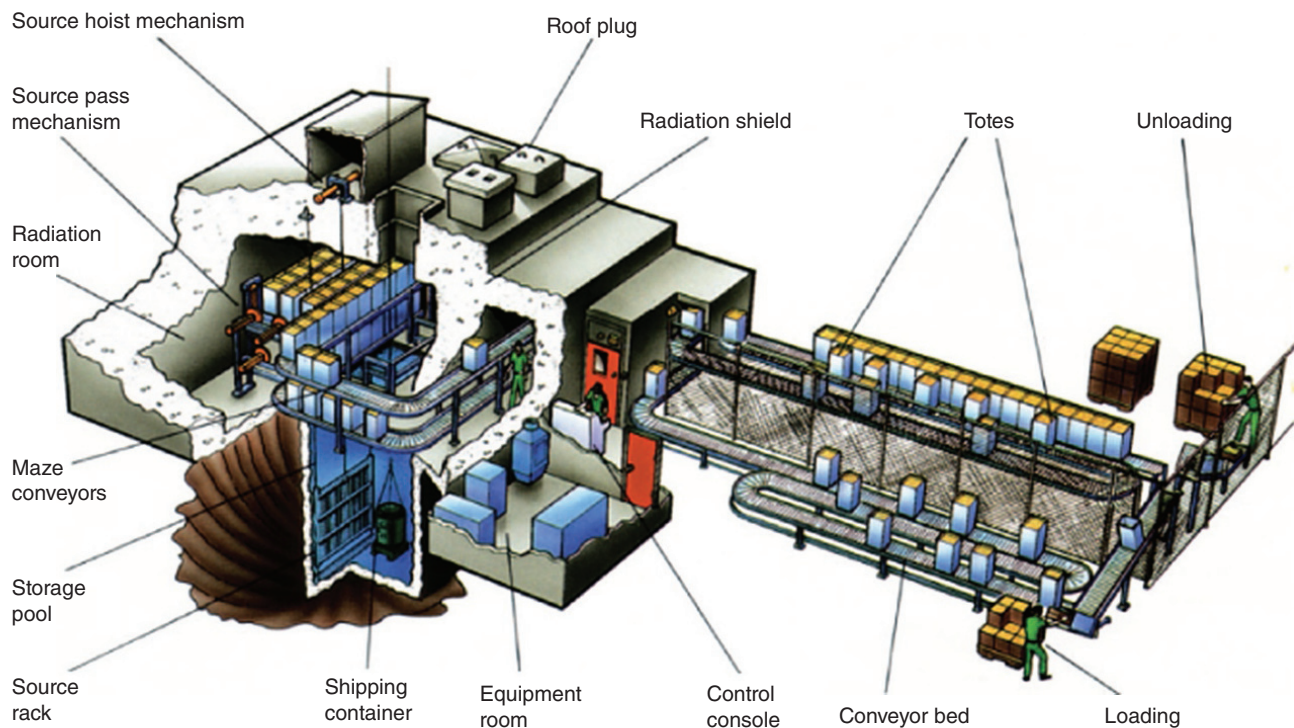
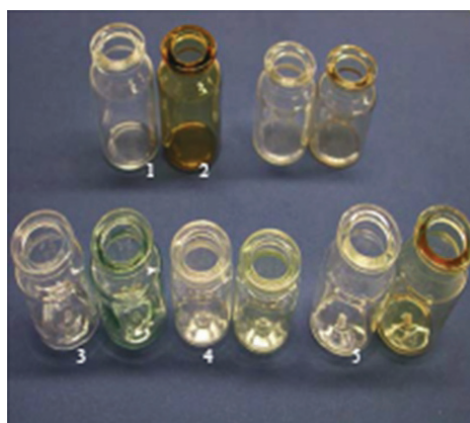


FIGURE 14.17 Schematic of a gamma irradiation facility. *Source:* Courtesy of Nordion.



Left: unirradiated
Right: irradiated

- 1 Type I glass
- 2 CeO glass
- 3 COC plastic
- 4 CZ plastic
- 5 CZ plastic

FIGURE 14.18 Effect of irradiation on packing material [53].

inactivate infectious airborne microorganisms, especially in HVAC systems, in hospitals, and in laboratories. UV lamps are typically used as upward-facing position for purifying air or in a downward position for sanitizing laboratory work surfaces or the surfaces of liquids. Finally, in the biotechnology and fermentation industry, UV light is routinely used to promote induced mutagenesis aimed at improving the productivity of selected microbial strains capable of synthesizing desired biomolecules.

14.7 ANALYSIS OF RADIATION STERILIZATION PROCESSES

As before, this analysis is based on the assumption that the microorganisms to be sterilized are in a liquid volume V and their concentration N changes during the irradiation process. However, it could be easily modified to use instead the number of microorganisms present on a surface or a container.

An unsteady-state balance for the number of surviving microorganism subject to irradiation, similar to Eq. (14.14), is

$$V \frac{dN}{dt} = -Vr_{N\text{radiation}} \quad (14.74)$$

where the right-hand term is the specific rate of microbial death caused by radiation, which can be assumed to be proportional to the microbial concentration, the rate of the radiation absorption by the microorganisms, and a proportionality constant that depends on a number of factors including type of ionizing radiation, target microorganisms, temperature, and shielding effects. In addition, given the typical short time of the radiation sterilization processes, the effect of ionizing radiation can be expected to be cumulative. In other terms, any new dose of radiation delivered to the microorganisms contributes to reducing the microbial population. When a sufficiently high number of "radiation hits" reach the cell DNA or enzymes, microbial death follows. With these assumptions, the rate of microbial death caused by radiation is

$$r_{N\text{radiation}} = K_{dY} \left(\frac{dY}{dt} \right) \cdot N \quad (14.75)$$

where

r_N is the specific microbial death rate, $(\text{ml} \cdot \text{s})^{-1}$.

N is the number concentration of viable microorganisms, ml^{-1} .

Y is the dose of radiation energy absorbed by microorganisms ($\text{J/kg} = 1 \text{ Gray}$).

(dY/dt) is the rate of radiation absorbed by the microorganisms dY/dt (Gray/s).

K_{dY} is the proportionality parameter, $1/\text{Gray}$.

The unit of radiation absorbance is the gray = 1 J of ionizing radiation absorbed by the biomass per unit mass of biomass (or its multiple the kilogray (kGy)). Substituting this rate expression in the microbial mass balance gives

$$V \frac{dN}{dt} = -Vr_N = -VK_{dY} \left(\frac{dY}{dt} \right) N \quad (14.76)$$

which can be simplified and rewritten as

$$\frac{dN}{dY} = -K_{dY}N \quad (14.77)$$

The parameter K_{dY} depends mainly on the type of ionizing radiation, the microbial species, and the temperature. The dependence on temperature can be expressed by the usual Arrhenius equation, but since the temperature is low and constant during radiation sterilization, this is usually not a factor.

Therefore, for any given radiation type, irradiating a specific microbial species K_{dY} is a constant, and Eq. (14.77) can be integrated with the usual initial condition to give the design equation for radiation sterilization processes:

$$\ln \frac{N}{N_0} = -K_{dY}Y \quad (14.78)$$

This equation clearly shows that any additional dose of radiation absorbed by the microorganisms contributes to the log reduction of microbial population. Similar to what is done for thermal sterilization, a D -value can be introduced (D_Y), defined here as the radiation dose (typically in kGy) required to reduce the microbial population to 10% of the initial value (i.e. by a one order of magnitude reduction). From this definition it follows that

$$D_Y = \frac{\ln(10)}{K_{dY}} = \frac{2.303}{K_{dY}} \quad (14.79)$$

and

$$\log_{10} \frac{N_0}{N} = \frac{Y}{D_Y} \quad (14.80)$$

Therefore, plotting the decimal reduction ratio on the left-hand side as a function of radiation dose absorbed gives a straight line with slope $-1/D_Y$, as shown in Figure 14.19. Typical values of D_Y for a number of microorganisms are shown in Table 14.7. Clearly, the greater the D_Y value, the stronger the radiation dose required to inactivate the microorganisms will be.

The damage inflicted to any product during radiation sterilization can be similarly quantified assuming a similar type of degradation kinetics as for the microorganisms. The resulting equation is

$$\ln \frac{C_m}{C_{m_0}} = -K_{mY}Y \quad (14.81)$$

where

K_{mY} is the product degradation rate constant for radiation exposure.

14.8 CHEMICAL STERILIZATION PROCESSES

Chemical sterilization, i.e. the sterilization of materials and equipment using toxic chemicals, is typically used for system that cannot be sterilized by other methods or for which other sterilization approaches would be impractical. For example, thermoplastic materials that are sensitive to the high temperatures used in thermal sterilization process are often chemically sterilized. The same applies to rubbers and some

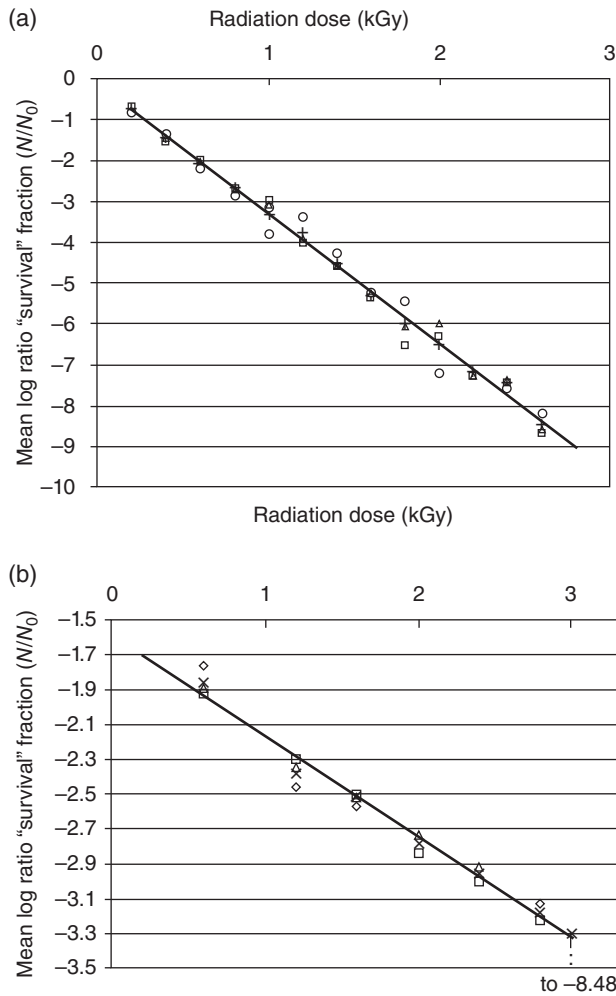


FIGURE 14.19 Survival ratios of *Staphylococcus aureus* cells as a function of radiation dose (a) γ -ray and (b) electron beam. *Source:* From Shehata et al. [49]. <http://www.acmicrob.com/microbiology/effects-of-gamma-and-electron-beam-irradiation-on-viability-and-dna-elimination-of-staphylococcus-aureus.php?aid=181>. Licensed under CCBY 4.0.

plastic material that can become more brittle or discolored after irradiation. Large systems, such as aseptic lines or clean rooms, can be practically sterilized only using chemical sterilants.

Microbial death typically occurs when microorganisms are exposed for sufficient time to a toxic chemical agent. Typically, such agents work by denaturing cell membranes, enzymes, and proteins or by altering membrane permeability. However, the specific mechanism of microbial inactivation depends on the specific chemical being used and is not always easy to ascertain, especially since each sterilant may interact with a specific cell constituent or more than one.

A number of chemical sterilants are used in a variety of applications in the pharmaceutical industry, as both sterilizing and disinfecting agents. They include the following:

- *Gaseous or vaporized chemical agents*, including ethylene oxide, hydrogen peroxide, ozone, and chlorine dioxide.
- *Liquid chemical agents*, including dissolved chlorine solutions, iodine solutions, dissolved stabilized formaldehyde (formalin), glutaraldehyde, quaternary ammonium salts, and solutions of ozone, hydrogen peroxide, or peracetic acid [56]. Additionally, liquids such as alcohols and solution of phenolic compounds are often used as disinfectants.

Gaseous or vaporized agents are by far the most common chemicals used in chemical sterilization because of their relative ease of use, effectiveness against endospores, high penetration in the material undergoing sterilization, and their usability in large and complex machinery or spaces. Therefore, they are commonly used to sterilize not only small items such plastic syringes, petri plates, medical equipment and instruments (all surgical sharps, fiber-optic endoscopes, surgical telescopes, ophthalmic instruments, plastic instruments, specula, respirators, and inhalation therapy supplies), sutures,

TABLE 14.7 D_{γ} -Values for Different Microorganisms

Microorganism	Classification	D_{γ} (kGy)	Condition
<i>Clostridium botulinum</i> spores	Bacterium	2.9	Phosphate buffer, 0 °C
<i>Clostridium botulinum</i> spores	Bacterium	4.6	Meat, 0 °C
<i>Clostridium botulinum</i> spores	Bacterium	3.9	Phosphate buffer, -196 °C
<i>Clostridium botulinum</i> spores	Bacterium	6.8	Meat, -196 °C
<i>Aspergillus flavus</i>	Fungus	0.6	Aerated water, 20 °C
<i>Aspergillus niger</i>	Fungus	0.42	Aerated water, 20 °C
<i>Cladosporium cladosporioides</i>	Fungus	0.03–0.25	Aerated water, 20 °C
<i>Curvularia geniculata</i>	Fungus	2.42–2.90	Aerated water, 20 °C
<i>Coxsackievirus B-2</i>	Virus	5.3	Water, -90 °C
<i>Coxsackievirus B-2</i>	Virus	7.0	Meat, 16 °C
<i>Coxsackievirus B-2</i>	Virus	8.1	Meat, -90 °C
<i>HIV</i>	Virus	8.8	Bone, -78 °C

Source: Data extracted from Aquino [55]. <https://www.intechopen.com/books/gamma-radiation/sterilization-by-gamma-irradiation>. Licensed under CC BY 3.0.

artificial heart valves, rubber and plastic tubing (e.g. catheters), and electrical equipment (e.g. drills, pumps, motors) but also large items and equipment such as clean rooms, barrier isolators (formerly referred to as glove boxes), locally controlled environments, biological safety hoods, aseptic processing lines, centrifuges, freeze dryers, filling lines, and other process equipment. By contrast, liquid agents are mainly used as sterilants for small items such as medical instruments (e.g. heat-sensitive reusable endoscopes) or as disinfectants.

Chemical sterilization has also drawbacks. Material compatibility is a critical issue: one must ensure that the material undergoing sterilization is compatible with the specific chemical sterilant, since different chemical sterilizing agents can have different interactions with the material being sterilized. Sterilization occurs only if the sterilant reacts with microbial cell components; therefore, it is critical that the sterilizing agent penetrates and diffuses in every nook and cranny of the equipment undergoing sterilization or through the wrapping surrounding the medical instruments being sterilized and that adequate time is allowed for this to happen.

Chemicals that are toxic to microorganisms are also toxic to humans; hence adequate operating procedures must be in place, and protection for the operators must be ensured to minimize human exposure. Some agents (e.g. ethylene oxides) are explosive in contact with air, and they must be diluted with inert gases, such as carbon dioxide, during the process. In addition, the unreacted sterilant must be completely removed at the end of the process using appropriate

procedures. The potential for contamination with residuals of the sterilizing agent in the material to be sterilized prevents chemical sterilization from being used to sterilize product such as APIs and limits its use to devices, supplies, and equipment, as mentioned above.

The processes associated with chemical sterilization using the most common sterilizing agents are now reviewed.

14.8.1 Sterilization with Ethylene Oxide

Ethylene oxide (EtO) is a colorless and flammable gas with a molecule consisting of a three-membered ring containing one oxygen atom and two carbon atoms (C_2H_4O). It is a cyclic ether and the simplest epoxide, and because it is a strained ring, it can easily participate in alkylation reactions resulting from the opening of its ring. When microorganisms are exposed to it, EtO reacts with proteins in cell membrane, resulting in membrane rupture and cell lysis, thus making EtO a very effective sterilant.

EtO sterilization processes are conducted in chambers such as those shown in Figure 14.20. Since EtO is potentially explosive in the presence of more than 3% of air, commercial mixtures of EtO in CO_2 are commonly used (typically 10% EtO–90% CO_2 or 85% EtO–15% CO_2).

The sterilization process consists of loading the chamber with the material undergoing sterilization (e.g. surgical instruments), locking the chamber door, heating the chamber to the desired temperature (typically 40–60 °C simply to increase the reaction rate and allow spores to morph back



FIGURE 14.20 EtO sterilization chambers. Source: Courtesy of Sterile Technologies, Inc.

to the vegetative state when they are more vulnerable), repeatedly applying vacuum to remove all air, conducting a vacuum hold leak test, humidifying by steam injection (to promote EtO reaction), loading the EtO mixture from gas cylinders to achieve EtO concentrations in the range 400–1200 ppm, and then conducting the sterilization step. EtO penetrates most plastic and paper wrapping, but since it must directly contact microorganisms to be effective, enough time must be allowed for EtO to penetrate the material to be sterilized. Diffusion–sterilization times vary from 150 to 180 minutes for 10% EtO mixtures to 90–120 minutes for 15% mixtures, although longer exposure periods up to 24 hours are possible. Once the sterilization step is completed, the unreacted EtO mixture is evacuated from the chamber and sent to a catalytic burner, air/vacuum pulses are conducted in the chamber, a vacuum hold is maintained to allow residual EtO to diffuse out of the load, and finally the vacuum is broken, and the load is removed.

EtO is very toxic not only to microorganisms but also to humans since it is carcinogenic and teratogenic. Therefore, OSHA mandates that operator exposure to EtO is limited to 1 ppm, measured as an eight hour time-weighted average. Exposure may not exceed the short-term excursion limit of 5 ppm EtO averaged over any 15 minute sampling period [57].

14.8.2 Sterilization with Hydrogen Peroxide and Hydrogen Peroxide–Plasma Systems

Hydrogen peroxide (H_2O_2) is a pale-blue clear liquid that can be easily vaporized. Hydrogen peroxide readily decomposes to form water and highly reactive nascent oxygen, and it is therefore a powerful oxidizing agent that can be used as disinfectant and sterilant against microorganisms.

By itself, vaporized hydrogen peroxide (VHP) in low concentration (typically 10 ppm) is used in sterilization processes for systems that can be difficult to sterilize otherwise, such as isolator systems or fill lines [58, 59]. These systems, such as that shown in Figure 14.21, are designed for aseptic processing of material, and their enclosure can be filled with VHP to achieve sterility prior to their use. Other process equipment, such as freeze dryers and fill lines, can also be sterilized by VHP.

Small items such as medical instruments can also be sterilized using VHP using low-temperature sterilizers (Figure 14.21). These devices consist of vaporization unit in which liquid H_2O_2 is converted into vapor, which is then injected in the chamber containing the load. Prior to H_2O_2 admission, the air in the chamber is removed by vacuuming it in order for hydrogen peroxide to be able to contact all surfaces and penetrate lumens. Several H_2O_2 addition and removal cycles are typically performed, each lasting only a few minutes. After sterilization, the vaporized H_2O_2 is vacuumed from the chamber and converted into water and oxygen.

More recently, hydrogen peroxide has also been used in combination with gas plasma to achieve sterilization. Gas plasma is the fourth state of matter, consisting of clouds of ions, neutrons, and electrons created by the application of a strong electric or magnetic field. VHP gas plasma sterilizers operate on the principle that when VHP is exposed to a high magnetic field, the ensuing cold-temperature plasma results in the breakage of the H_2O_2 molecule creating free radicals and UV radiation and inducing rapid microbial sterilization. Devices such as the unit shown in Figure 14.22 operate on this principle. The material to be sterilized is charged to the chamber, which is evacuated by applying vacuum. Then liquid hydrogen peroxide is admitted to the chamber where it rapidly vaporizes. Sufficient time is allowed for the VHP to diffuse through the chamber and permeate the load. Once the diffusion–permeation phase is complete, the pressure in the chamber is reduced, and a strong magnetic field is imposed, breaking apart the H_2O_2 molecule, generating the plasma, and achieving sterilization through a combination of chemical oxidation and generation of UV radiation. Once the magnetic field is removed, the process stops, and any residual unreacted species recombines to form molecular oxygen, water, and sterilization products. This entire process may be repeated one or more times to ensure the achievement of the desired sterility level. Then sterile air is admitted, and the load is removed. The entire cycle can last less than an hour and is conducted at room temperature, but only small-volume materials and devices can be fitted in a typical sterilization unit of this kind.

14.8.3 Sterilization with Ozone

Ozone (O_3) is a triatomic oxygen molecule consisting of O_2 with a loosely bonded third oxygen atom that is readily available to attach to, and oxidize, other molecules. The presence of the additional oxygen atom and the highly instability of the resulting molecule (half-life: ~22 minutes at room temperature) makes ozone a powerful oxidant. Ozone is a pale-blue gas with a characteristic pungent odor that is industrially produced when O_2 is energized and split into two monatomic (O_1) molecules that may collide with O_2 molecules to form ozone O_3 . Ozone is commonly used as an oxidant, disinfectant, and sterilizing agent. It has been used for medicinal applications since the late 1800s, and it has been utilized for years as a drinking water disinfectant.

Sterilization with ozone is typically accomplished by generating ozone *in situ* using different methods, the most common being corona discharge [60]. Accordingly, pure oxygen gas flowing through a tube is passed through a high voltage gap where the energy released by the corona discharge breaks apart the oxygen molecules generating a mixture of oxygen and ozone that is admitted to the sterilization chamber. In some commercial applications, the ozone generator is placed directly in the sterilization chamber (Figure 14.23). In some



FIGURE 14.21 Barrier isolator sterilized by vaporized hydrogen peroxide. Top panels: FCIS pharmaceutical isolator (left, closed; right, open). *Source:* Courtesy of Fedegari Group; bottom panel: V-PRO[®] maX 2 low-temperature sterilization system using vaporized hydrogen peroxide chemistry. *Source:* Courtesy of Steris.

other applications, dual sterilants, i.e. VHP and ozone, are used in combination to achieve low-temperature (41 °C) terminal sterilization of heat- and moisture-sensitive medical devices [61].

14.8.4 Sterilization with Nitrogen Dioxide

Nitrogen dioxide (NO₂) is one of the several nitrogen oxides. It has a boiling point of 21.2 °C at atmospheric pressure, implying that it has a relatively high saturated vapor pressure at ambient temperature, which also means that liquid NO₂

from commercially readily available cylinders may be used as a convenient source for the vapor. Nitrogen dioxide is a reddish-brown vapor, but above its boiling point, it is in equilibrium with its dimer, dinitrogen tetroxide (N₂O₄), a colorless vapor, although NO₂ is heavily favored at low temperatures. Nitrogen dioxide is a very effective sterilant against a wide range of microorganisms, the primary mechanism of lethality being DNA degradation [62]. In addition, there is some evidence that it may also act as a depyrogenating agent.

Nitrogen dioxide can be used in sterilizing chambers or in barrier isolators. In the first application, the material to be



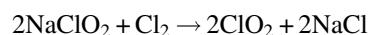
FIGURE 14.22 VHP gas plasma sterilizer. *Source:* Courtesy of Tuttnauer.

sterilized is loaded to the chamber that is then evacuated, as already explained above, before admitting a metered dose of nitrogen dioxide to achieve a pressure of less than 1% of saturation pressure so as to obtain a concentration of 10–20 mg of NO_2 per liter of chamber. Humidification is also recommended to promote microbial inactivation (55–75% relative humidity). After the diffusion–sterilization time has elapsed, the vacuum is broken, sterile air is admitted, and the excess nitrogen oxide is removed by passing the air stream over a solid scrubber material typically containing NaHS and

NaOH. When used in isolators, no vacuum is applied, and the required amount of nitrogen dioxide is added directly to the atmospheric air in the isolator, thus requiring longer cycle times of about an hour, including diffusion–sterilization and aeration times.

14.8.5 Sterilization with Chlorine Dioxide

Chlorine dioxide (CD), i.e. ClO_2 , is a yellowish-green gas, which, because of his strong oxidizing properties, is used in a number of industrial applications such as pulp and paper bleaching. In addition, it also has marked bactericide, virucide, and sporicide properties, and it is therefore used in water treatment and in sterilization applications. CD has a number of advantages. Unlike chlorine gas, it does not form organic chlorinated products (chloromethanes, chloramines) when reacting with organic materials (e.g. microorganisms). It has a relative lower human toxicity since it is noncarcinogenic and non-mutagenic, low residual CD concentration in the material treated, and significant material compatibility. However, it cannot be easily compressed and transported in cylinders, and it is therefore typically generated, *in situ* and on demand, using different synthetic routes depending on its final use. For sterilization purposes the most significant is the reaction of dry sodium chlorite and chlorine:



achieved by passing chlorine gas from cylinders over a dry bed of sodium chlorite pellets. The sterilization process is similar to those used with other gaseous sterilants, and CD can be effectively used in sterilization chambers such as those shown in Figure 14.24 or in barrier isolators, filling isolators, or similar systems. The reaction temperature for sterilization is 30–32 °C, and the CD concentration range is typically 10–30 ppm. Humidity is also required for CD to be effective (70–85% relative humidity) and a contact time of about 60 minutes is often used. At the end of the process, the unreacted CD is removed by passing it over a sodium thiosulfate scrubber.

14.8.6 Sterilization Using Liquid Agents

Liquid agents are infrequently used as sterilants since they wet the material being sterilized, are difficult to remove at the end of the process, have limited penetration capability with respect to vapors and gases, and cannot be used in large systems such as barrier isolators. Solutions of peracetic acid (0.26%) are very effective agents against both microorganisms and their spores. They have the advantage of not generating harmful decomposition products, thus leaving no residues, and are effective even in the presence of organic matter although at high acid concentration. Therefore, if they are used as sterilizing agents, they are used in specific,



FIGURE 14.23 Ozone sterilization system. *Source:* Courtesy of Genlantis.



FIGURE 14.24 Left: chlorine dioxide sterilization unit. *Source:* Courtesy of Consolidated Sterilizer Systems. Right: combination of chlorine dioxide gas/steam sterilizer. *Source:* Steridox-Steamer, ClorDiSys Solutions Inc.



FIGURE 14.25 Liquid sterilizer for endoscope based on with peracetic acid chemistry-based agent. *Source:* Courtesy of Steris.

small-scale applications such as for chemically sterilizing medical (e.g. endoscopes, arthroscopes) and surgical instruments [63]. One such example is the sterilization unit shown in Figure 14.25 using a peracetic acid chemistry-based liquid

agent as sterilant for reusable endoscopes and duodenoscopes used in endoscopic procedures, which is the only rapid (23 minute cycle time) and practical solution for this type of medical devices.

14.8.7 Disinfection Using Liquid Agents

Liquid agents are commonly used as disinfectants. A number of them are used in the industrial practice, and some of them are briefly reviewed here.

14.8.7.1 Solutions of Chlorine Compounds Solutions of hypochlorites, such sodium hypochlorite (bleach) or calcium hypochlorite, are very commonly used disinfectants in concentration of 5.25–6.15% of sodium hypochlorite. In water, these compounds undergo hydrolysis and the formation of undissociated hypochlorous acid (HOCl), which is the main disinfecting agent. The resulting concentration of HOCl depends on the pH: increasing the pH decreases the concentration of undissociated HOCl and hence the disinfection effectiveness. Hypochlorite solutions have a broad spectrum of antimicrobial activity, are fast acting, do not leave toxic residues, remove biofilms from surfaces, and have only limited toxicity, but they corrode metals and are inactivated by organic matter. Other compounds such as sodium dichloroisocyanurate and chloramine-T are also used as disinfectants.

14.8.7.2 Hydrogen Peroxide Solutions Aqueous solutions containing hydrogen peroxide in concentration greater than 3% inactivate most microorganisms [64, 65] including spores [66] and viruses [67]. Therefore, they are widely used as surface disinfectant in laboratories and for small processing equipment.

14.8.7.3 Alcohols Alcohols have typically limited biocidal activity and do not affect spores. Protein denaturation is their main mechanism of action. Ethanol is the only industrially significant alcohol commonly used for surface cleaning and disinfection purposes [68]. In the biotechnology industry, ethanol is used to sterilize small glassware and metallic devices but only because these objects are first immersed in alcohol and then flamed them. Therefore, sterilization is actually achieved through incineration in this case.

14.8.7.4 Aldehyde Solutions Aldehydes in general are highly reactive chemicals. They can inactivate microorganisms by reacting with the amide group of amino acids and proteins. The simplest aldehyde, formaldehyde, is used mainly as a water-based solution (formalin), containing 37% formaldehyde by weight. Formaldehyde has been used as preservative and disinfectant for a long time, but because of its toxic and carcinogenic nature, its use as a disinfectant is diminishing.

Solutions (2%) of glutaraldehyde (a saturated dialdehyde) are more commonly used as disinfectants and chemical sterilants but mainly for medical equipment such as endoscopes, since they are noncorrosive to metal, do not attack rubber and most plastics, but are very effective against bacteria, spores, mycobacteria, fungi, and viruses. Glutaraldehyde solutions

are commonly used in the biopharmaceutical industry as surface disinfectants and in a number of applications [8], such as when sterile needles are used to pierce through rubber stoppers to add nutrients/inoculum to sterile bioreactors.

Ortho-phthalaldehyde solutions containing 0.55% 1,2-benzenedicarboxaldehyde (OPA) are extremely effective disinfectants.

14.8.7.5 Phenolic Solutions Phenol has been used for decades as a disinfectant in hospitals. However, other phenol derivatives (alkyl, phenyl, benzyl, halogen) are now used because of their improved action as antimicrobial agents. However, they are not considered high level disinfectants and their use is limited to noncritical applications such as cleaning laboratory surfaces and noncritical medical devices.

14.8.7.6 Solutions of Quaternary Ammonium Compounds Quaternary ammonium salts or quaternary ammonium compounds are salts of quaternary ammonium cations, such as alkyl dimethyl benzyl ammonium chloride and alkyl dodecyl dimethyl ammonium chloride, in which the hydrogen in ammonium ion has been replaced with alkyl or aryl groups and halide, sulfate, or similar radical occupies the fifth valence of the nitrogen atom. Their solutions have been shown to have significant antimicrobial activity and are therefore widely used to disinfect and deodorize floors, walls, and surgical instruments. In the biopharmaceutical industry they are used to clean, wipe, and disinfect surfaces that do not come in direct contact with raw materials.

14.9 ANALYSIS OF CHEMICAL STERILIZATION PROCESSES

This analysis can start from the typical degradation kinetic rate expression used for the disinfection of water in the water and wastewater industry (typically using chlorine in the form of hypochlorous acid (HOCl) as the disinfectant), although the same approach can be applied to the kinetics of sterilization reactions. Accordingly, the rate of microbial inactivation (death) can be expressed as

$$r_N = K'_d C^\beta \cdot N \quad (14.82)$$

where

C is the concentration of disinfecting or sterilizing chemical agent (mol/l or g/l).

β is the order of the reaction for the sterilant (nondimensional exponent).

K'_d is the die-off constant ($1/(\text{mol/l})^\beta \cdot \text{s}$) or $1/((\text{g/l})^\beta \cdot \text{s})$.

This expression simply indicates that the rate of inactivation depends not only on the microbial concentration, N , but also on the concentration of the chemical agent, C .

Depending on the interaction of this agent with the microorganisms, the exponent β may vary. Therefore, it needs to be determined from experiments. As usual, the parameter K'_d is the specific kinetic rate of product degradation, which, following the Arrhenius equation, can be expressed as

$$K'_d = K'_{d_0} e^{-E_d/R_G T} \quad (14.83)$$

Substituting the above rate expression in the differential mass balance for the microorganisms and simplifying results in

$$\frac{dN}{dt} = - \left(K'_{d_0} e^{-E_d/R_G T} \right) C^\beta N \quad (14.84)$$

This equation can be integrated only if the temperature profile and the evolution of the concentration of the sterilant are known as a function of time. In practice, one can make very reasonable assumptions to proceed. In a microbial chemical sterilization process, the sterilizing agent is consumed as it reacts with, and inactivates, the microorganisms. In most cases of relevance to the pharmaceutical industry, the microbial concentration is smaller by orders of magnitude than the concentration of the chemical agent. Therefore, the chemical agent is only minimally depleted during the process, and its concentration can be safely assumed to be constant throughout the process and equal to the initial concentration. Secondly, in most cases the temperature is constant throughout the process, implying that K'_d is also a constant. Therefore, with these assumptions and with the usual initial condition, the above equation can be integrated to give the design equation

$$\ln \frac{N}{N_0} = - [K'_d \cdot C^\beta] t \quad (T = \text{constant}; C \cong \text{constant}) \quad (14.85)$$

This equation, also known as the Chick–Watson equation, shows that microbial reduction is a function of the exposure

time, t , the type of chemical agent and the operating temperature (through K'_d), and the agent concentration C , whose sterilizing effectiveness depends on the exponent β . This equation was obtained for microorganisms in a liquid suspension (hence the use of the microbial concentration N). However, very similar equations would result if N is replaced by the total microbial count or the surface concentration of microorganisms, even when wrapped in paper. This is in fact the type of challenge test implemented by using BIs for EtO sterilization, such as strips of filter paper carrying bacterial spores and sealed within a peel-open envelope. These strips also undergo the sterilization process, thus simulating the exposure of microorganisms on the surface of similarly wrapped materials. They are then tested for spore survival at the end of the sterilization process.

Equation (14.85) can be used to interpret experimental data. For example, the data reported in Figure 14.26 show how microbial death proceeds as a function of time for different concentration of ethylene oxide at constant temperature. Clearly, increasing the sterilant concentration reduces the time required to achieve the same level of microbial reduction. The method to obtain the parameters K'_d and β from these data consists of a two-step approach, as follows:

Step 1. For each sterilant concentration C , a plot of $\ln(N/N_0)$ vs. t (or $\log_{10}(N/N_0)$ vs. t) should give a straight line with slope $-y$. By repeating the same experiments at different sterilant concentrations, it would be possible to obtain a number of such slopes. For each set it is that $y = K'_d C^\beta$.

Step 2. Since from $y = K'_d C^\beta$ it follows that $\ln(y) = \ln(K'_d) - \beta \ln(C)$, then plotting $\ln(y)$ vs. $\ln(C)$ will produce a straight line with slope $-\beta$ and intercept $\ln(K'_d)$ from which K'_d at that temperature and β are obtained.

If the same procedure is additionally repeated at different temperatures, it is possible to obtain the effect of T on K'_d using the Arrhenius equation. Other factors may play a role

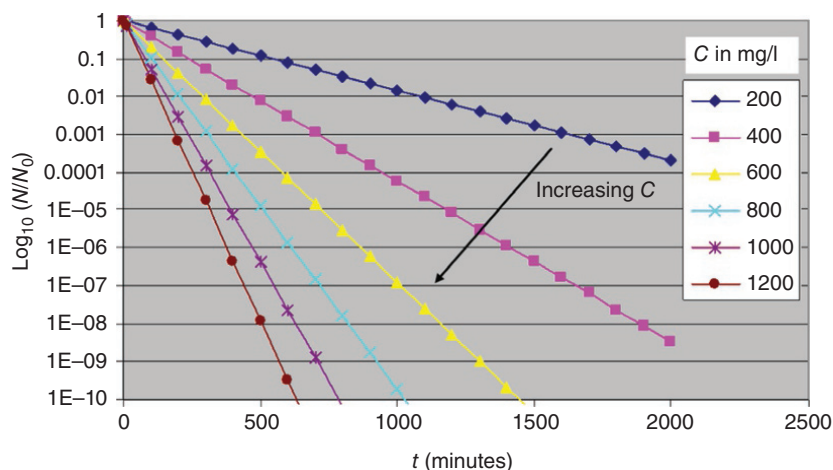


FIGURE 14.26 Microbial survival during a chemical sterilization batch process with ethylene oxide. Source: Data from Rodriguez et al. [69].

in the kinetics of the inactivation process, such as the humidity at which the process is conducted, which should be treated as a separate parameter.

Although the Chick–Watson equation is commonly used, especially in water treatment, deviations are possible, including initial lag or shoulder effect, delaying the start of the inactivation process if the chemical agent first reacts with other non-microbial constituents, or tailing effects, in which the microorganisms are partially shielded from the agent if they are embedded in solid particles. The chemical inactivation process occurs only after the chemical agent has diffused throughout the material being sterilized. This is the reason for the initial repeated vacuum cycles during sterilization with gaseous agents. This initial diffusion and load preconditioning steps are in addition to the reaction time predicted with the Chick–Watson equation.

14.10 STERILE FILTRATION PROCESSES

Filtration is the process by which solid particles dispersed in a fluid are separated from the fluid by passing the solid–fluid suspension through porous septum or some type of medium capable of entrapping the particles. Sterilization by filtration relies on the physical removal of microorganisms from the fluid by passing it through a membrane through which the microorganisms cannot pass. Clearly, this requires that

the sterile filtration unit transfers the fluid to a downstream zone that was previously sterilized by some other means, such as heat treatment or chemical sterilization.

In the biopharmaceutical industry sterile filtration is used for two primary purposes:

1. Filtration sterilization of gases and especially air, typically in cell culture, fermentation, and other biological processes. This often involves treating large volumes of gas but low initial bacterial count.
2. Filtration sterilization of liquids and especially filtration of medicinal products, such as biologics, which cannot be sterilized by more conventional but damaging processes such as heat sterilization.

Generally speaking, filtration (*not* necessarily sterile filtration) can be accomplished using three types of filters, each relying on different filtration mechanisms:

- *Depth filters*, in which the suspension is passed through a bed made of a filtering medium such as a bed of particulates (e.g. sand in sand filters used in wastewater treatment) or a matrix made of randomly oriented intertangled strands or mats (e.g. synthetic materials, plastic, glass filaments), as in a typical HEPA filter (Figure 14.27). The openings in the bed are not uniform in size, and they are often larger than the particles being filtered. Solids become trapped in the interstices of the

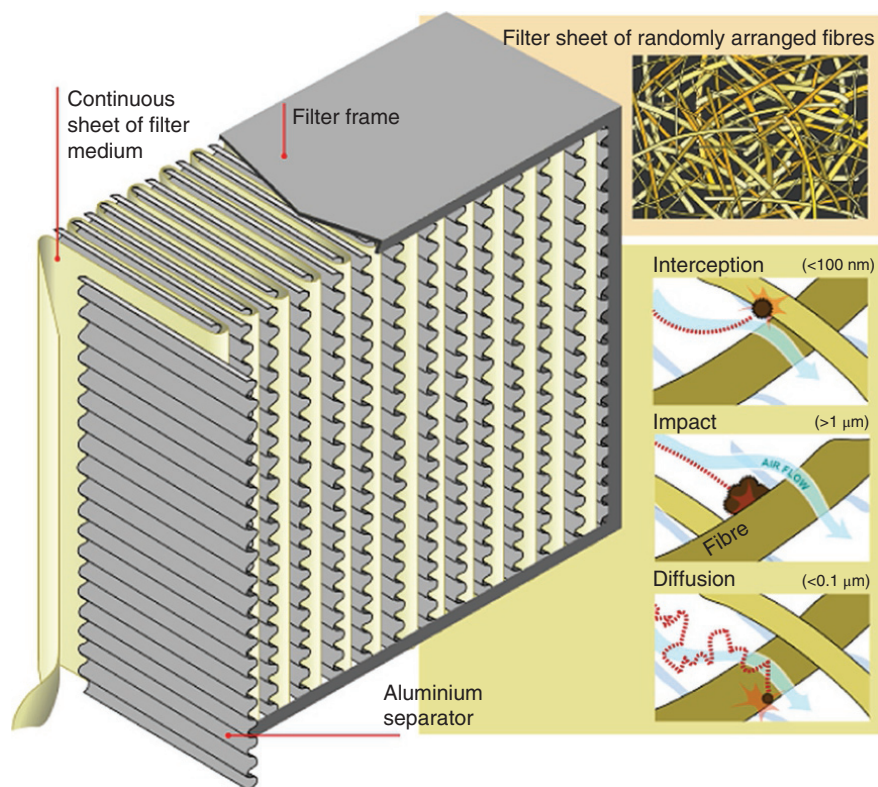


FIGURE 14.27 HEPA filter. Source: Wikipedia.



FIGURE 14.28 Lenticular style depth filter cartridges (Millistak+® Depth Filters: left) and their assemblage (right). *Source:* Both images Courtesy of MilliporeSigma.

medium not only as a result of a sieving action but also because of other separation mechanisms such as particle sedimentation and inertial impaction on the bed material, interception, adhesion, diffusion, and flocculation (Figure 14.27). Depth filters such as those shown in Figure 14.28 are often used as filtration and prefiltration units in pharmaceutical manufacturing. These filters can separate a significant fraction of the initial particle load but cannot ensure that all microbial contaminants, especially small bacteria, can be removed. Therefore, they cannot be classified as sterilizing filters, although they are often used as prefilters in sterile filtration processes [70].

- *Cake filters*, in which the suspension is passed through a porous septum (e.g. a woven plastic or metal wire) initially retaining, possibly only partially, the solids that then pile up, forming an increasingly thicker “cake.” From this point on, the cake of previously filtered solids becomes the primary filtering medium, and the newly filtered solids accumulate on the surface of the cake and increase its depth. Thus, the cake itself, i.e. the filtered solids retained on the septum, provides most of filtering action. In cake filtration the cake, not the septum, is the real filtering element. Cake filters (e.g. Nutsche filters) (Figure 14.29) are commonly found in the pharmaceutical industry especially in those operations, such as crystallization of intermediates or drug substances, in which the particles constitute the product and not the contaminants and should be recovered as such at the end – something that would not be easily achievable with depth filters. However, cake filters

are not able to retain all microorganisms in the fluid and cannot be used as sterilizing filters.

- *Membrane filters*, consisting of a very thin (~100–250 μm) microporous structure typically made of organic polymers, with a uniform and specific small pore size (Figure 14.30). Because of their uniform structure, microporous membranes can retain all particles or microorganisms larger than their pore size, primarily by surface capture. Membrane filters operate essentially as sieves, although accumulation of particles around or inside the pores (bridging and caking) also exerts an additional filtration action and even allows particles smaller than the pore size to be removed from the fluid by a cake filtration mechanism. Membrane filters with pore size of 0.22 μm or smaller (e.g. 0.11 μm) have openings smaller than the smallest microorganisms such as *B. diminuta*, which have a size of about 0.45 μm (viruses, which are only 0.04–0.1 μm in size, are not living organisms, and their removal is not a typical requirement for any sterilization processes, not just filtration sterilization processes). Consequently, membrane filters with pore sizes equal to or less than 0.22 μm can filter out all living microorganisms, including spores, and are therefore classified as sterilizing filters. Different types of polymers are used in the manufacture of sterilizing membrane, and their material properties are critical. Hydrophilic membranes are spontaneously wet by water, which easily penetrates them, and are therefore used for the filtration (sterile or not) of aqueous liquids. They are typically made of different types of cellulosic materials (cellulose nitrate,

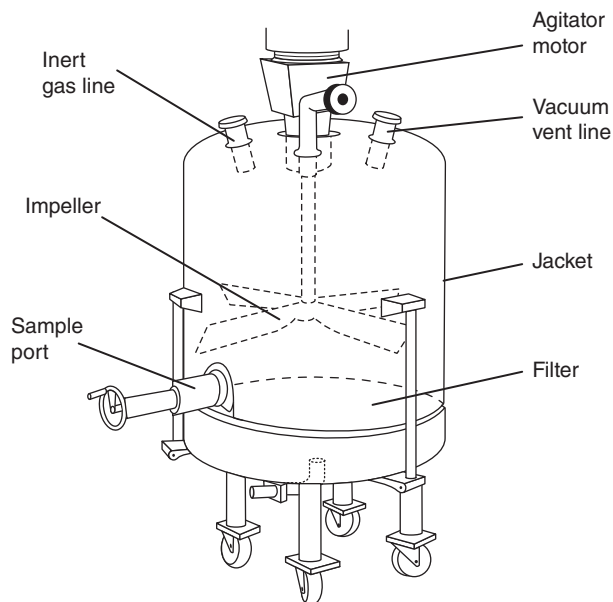
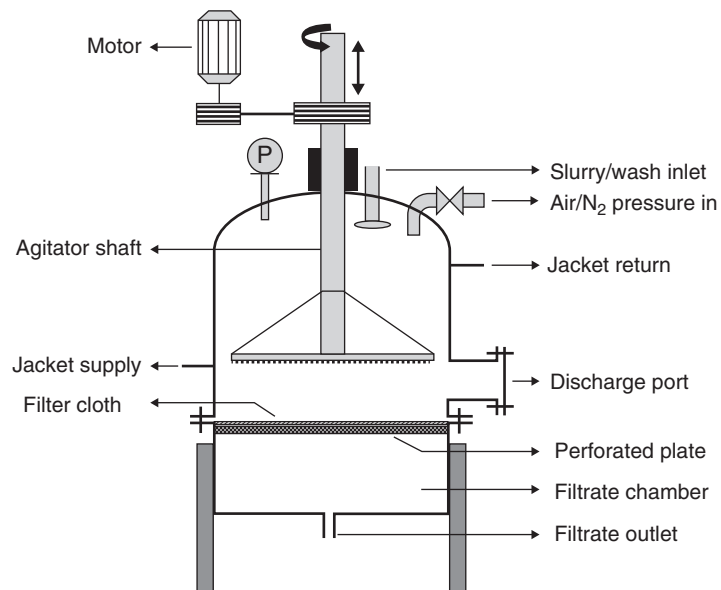


FIGURE 14.29 Nutsche-type cake filters. *Source:* From Murugesan, Sharma, and Tabora. Design of Filtration and Drying Operations, reprinted with permission from John Wiley & Sons 2010.

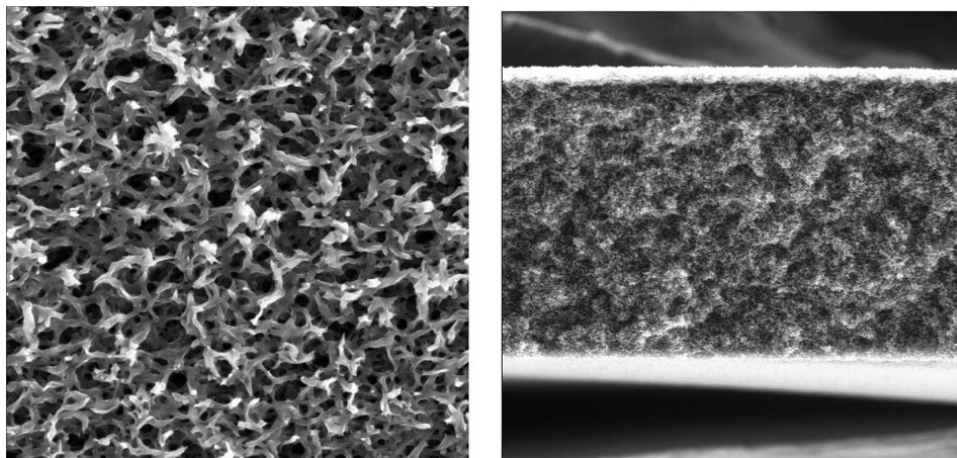


FIGURE 14.30 Membrane filter: submicron membrane structure (left). Magnified membrane cross section (right). *Source:* Courtesy of MilliporeSigma.

cellulose ester, cellulose acetate, regenerated cellulose), polycarbonates, modified polyvinylidene fluoride (PVDF), polyethersulfone (PES), nylon, and others. By contrast, membranes made of hydrophobic (water-repelling) materials are not spontaneously wet by water, and they resist water penetration and tend to decrease their surface area in contact with the water. Therefore, they are used for the sterile filtration of organic liquids and, more importantly, of gases (as for the inlet and exhaust filtration of bioreactors, tank vents, etc.), since their water-repellent characteristic makes them inhospitable for microorganisms, which require an aqueous environment to grow. These filters are typically made of polytetrafluoroethylene (PTFE), polypropylene, polysulfone, polycarbonate, polyvinylidene fluoride (PVDF), and other materials. All sterilizing membranes are additionally highly porous (up to 99%) to achieve a high flow rate with reduced pressure drop, capable of mechanically withstanding the resulting pressure differential across them, resistant to heat sterilization (since they and the equipment downstream of them must be sterilized prior to its use), and resistant to chemical attack (most membranes accept liquids with a pH ranging from 0 to 14).

In summary, sterile filtration processes utilize both depth filters, as prefilters, in order to remove the larger amount of particle and microbial contaminants in the fluid and membrane filters, as the final sterilizing filters. The main characteristics of these two types of filters are summarized in Table 14.8.

14.10.1 Sterile Filtration Systems

14.10.1.1 Sterile Filtration Systems for Gases As already pointed out, sterile filtration of gases and especially air for biological systems is extremely common. A typical aerobic fermentation requires a significant amount of sterile air to be sparged in the bioreactor. For example, in antibiotic

fermentation processes air supply rates range from 0.1 to 1 unit volume of air per unit volume of culture medium per minute (vvm). Therefore, for an industrial 10 000-l fermenter flow rates of sterile air as high as 600 m³/h may be required. Air typically contains 1 000–10 000 microorganisms/m³ [27], which means that as many as 6 million microorganisms per hour will need to be removed from the air to avoid bacterial contamination of the batch. This bacterial load must be removed by the filtration system in which the sterile membrane filter element is the last component of a train of filters, the other components being one or more depth filters, from coarser (and less expensive) to finer, which remove the bulk of the microorganisms prior to the last membrane sterilizing step. A typical air sterilization unit is shown in Figure 14.31, and it consists of a compressor and its ancillary components, followed by a number of depth filters (coalescing filters, particulate filter, carbon filters) and, finally, by a sterilizing membrane filter.

The key component of such a system is the sterilizing filter cartridge mounted inside a filter housing (Figure 14.32). The membrane in the typical cartridge is assembled similarly as in Figure 14.33 (although the membrane in this image is of the non-sterilizing type). This figure shows not just the multiple filtering layers but also the rigid inner and outer plastic supports to mechanically supporting the cartridge assembly and the membrane.

Filters used for air filtration are always made of hydrophobic materials (e.g. PTFE) to minimize bacterial growth on the sterile filtering element. In addition, sterile air filters are mounted on the vent of bioreactors to prevent releasing microorganisms growing inside it in the environment. A schematic of the typical biotech plant with the location of critical filter units is shown in Figure 14.34.

14.10.1.2 Non-sterile Filtration Systems for Gases As already mentioned, in addition to sterilizing membrane filters, depth filters are heavily used in pharmaceutical facility, primarily to reduce the microbial load. In particular,

TABLE 14.8 Main Characteristics of Depth Filters and Membrane Filters

Depth Filters	Membrane Filters
Cannot be used as sterilizing filters	Can be used as sterilizing filters if pore opening is 0.22 μm or less
Filtration occurs throughout the filter	Filtration occurs at the surface
Can retain larger amounts of solid impurities (mainly throughout the filter)	Can retain a small amount of solid impurities (only at the surface)
Can retain some or most of the particles smaller than the filter openings as a result of a number of filtration mechanisms	Retain particles smaller than the rating as a result of cake filtration effects at the surface
Pressure buildup is gradual as impurities accumulate	Pressure buildup can be rapid as pore become clogged with impurities
Microorganisms that are retained on the “dirty” side may migrate to the “clean” side over time and grow there	If the membrane has the correct pore size and is integral, microorganisms cannot migrate across membrane
Filtering element is typically deep, and fluid holdup is relatively large	Filtering element is very thin (10–250 μm), and fluid holdup is relatively small
Less expensive	More expensive

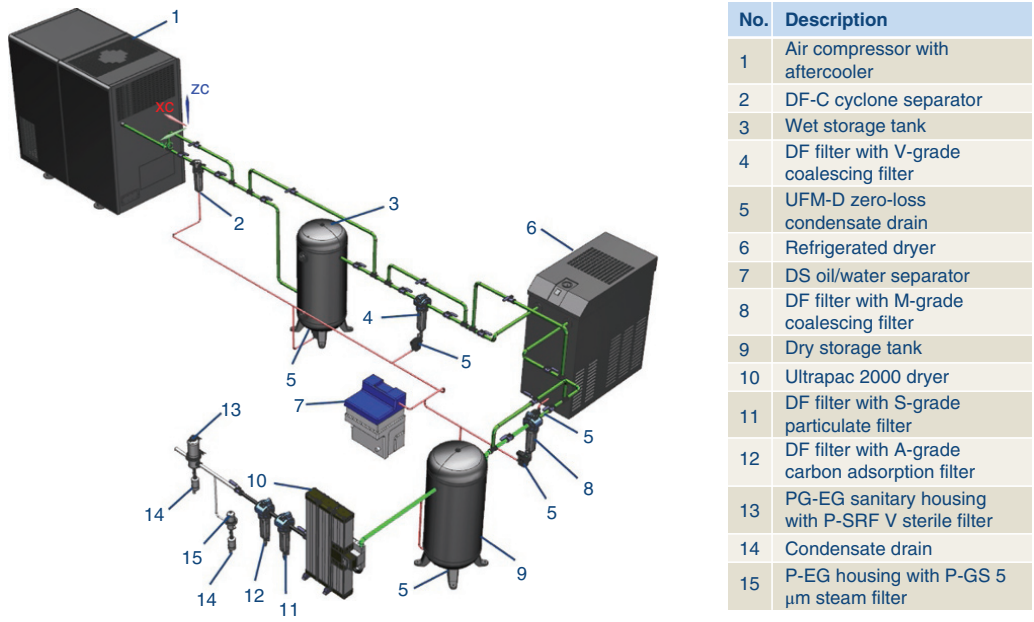


FIGURE 14.31 Air sterilization system. Source: Courtesy of Donaldson Company, Inc.

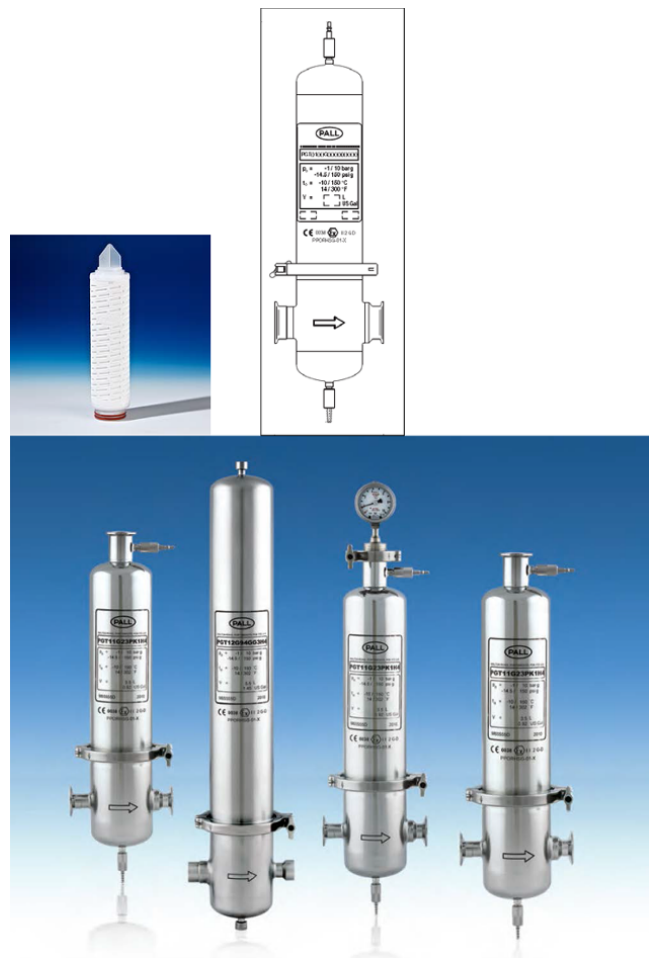


FIGURE 14.32 Sterilizing-grade filter cartridge (Supor EKV, Pall; top left) and filter housings (PGT Filter Housings, Pall Corp.; top right and bottom). Source: Courtesy of Pall Corp.



FIGURE 14.33 Pleated all-polypropylene non-sterilizing filter cartridge (Poly-Fine II Series Filter, Pall). *Source:* Courtesy of Pall Corp.

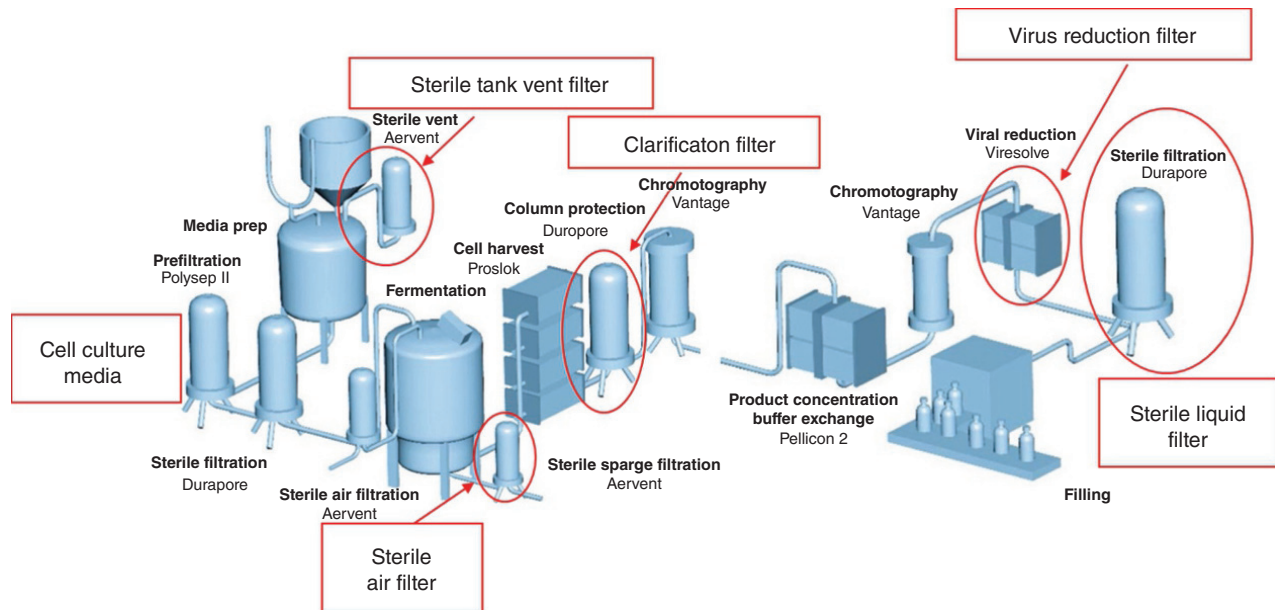


FIGURE 14.34 Typical biotech facility showing the location of critical filtration units. *Source:* Courtesy of MilliporeSigma.

high-efficiency particulate-arresting (HEPA) filters are very commonly used throughout in pharmaceutical manufacturing plants, packaging areas, clean rooms laboratories, etc. to remove microorganisms in air to very small values (although they cannot sterilize the air). HEPA filters are depth filters consisting of banks fibrous material strands separated by corrugated paper or metallic sheets. A metal grill holds the filter

and helps sustain the laminar air flow, which is typically moving air with an air velocity of 230 ± 90 cm/min. HEPA filters must remove at least 99.97% of dust and particulate matter with a size of at least $0.3 \mu\text{m}$ and above from air streams. Similarly, ultralow particulate air (ULPA) filters must remove at least 99.999% of dust and particulate matter with a size of at least $0.1 \mu\text{m}$ and above from air streams.

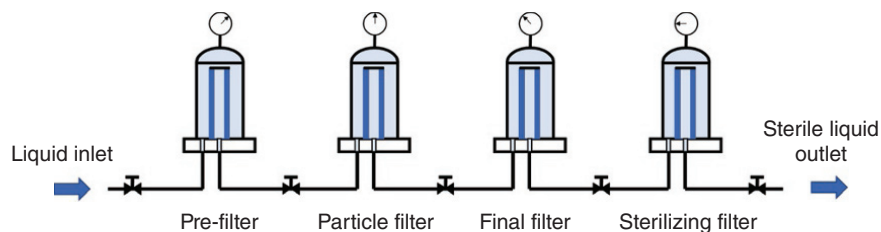


FIGURE 14.35 Sterile liquid filtration system.

14.10.1.3 Sterile Filtration Systems for Liquids In all those cases in which a liquid contains a product that is heat-sensitive, non-high-temperature-based sterilization methods must be used. Filtration sterilization is often the most common and economical of them. In general, filtration of liquids is more demanding than that of gases because of the higher viscosity. However, the same overall approach used for gas filtration can be applied, i.e. the last membrane sterilizing filter must be preceded by one or more prefilters that remove larger particles and in larger amounts, as shown in Figure 14.35. Since most of the liquids that must be filtered are aqueous solutions, hydrophilic depth filters and especially sterilizing cartridges made of hydrophilic materials such as polyethersulfone (PES) are used. In the manufacturing of small molecule APIs, non-sterilizing filters such as those shown in Figure 14.28 are also often used.

In the biotech industry, liquid filtration finds wide applications since most biological products are heat labile. In a typical process, the cell culture from the bioreactor must be first filtered to remove the cells. This is typically accomplished by utilizing depth filters, often disposable, where the cells are harvested and, in subsequent clarification steps, also using some type of depth filter, as shown in Figure 14.34, to remove residual cells and impurities. Other steps may follow, including a virus reduction step, also using membrane filters, prior to the final sterile filtration step, typically accomplished with some type of 0.22 μm sterilizing membrane made of hydrophilic materials such as polyvinylidene fluoride (PVDF).

A potential problem with sterile filtration of liquids is the possible leaching of extractable materials from the membrane cartridge assembly. This issue is more challenging when organic solvents are sterile filtered (in which case the membrane material must be hydrophobic and compatible with the solvent in the first place), as during the manufacturing of active pharmaceutical ingredients containing methylene chloride, acetone, isopropyl alcohol, ethyl acetate, and other solvents, since some components of the cartridge may be more soluble in organic solvents than in water. However, filter manufacturers have recognized this issue and have significantly addressed this problem by manufacturing cartridges with low levels of extractables.

14.10.2 Integrity Testing of Sterile Filtration Systems

In heat sterilization, sterility assurance can be established by ensuring that the process has been conducted at the proper temperature for the proper time, as determined by an established protocol. No such approach is available in filtration sterilization, and failure to achieve sterilization cannot be easily detected. Two issues are associated with the possible failure of filtration sterilization systems: one pertains to the proper installation and usage of the filtration system, and the other to the sterilizing performance of the sterile filter membrane. Improper installation and operation can compromise the integrity of a filter. Therefore, strict adherence to the assembling and operation protocols must be assured. As for the second issue, a number of tests can be used to ensure that sterilizing filter membranes maintain their sterile filtration properties and that no microorganisms can cross them. To this end, (i) typically new lots of membrane filters are statistically tested to ensure that they are free of defects, and (ii) filters in operation are tested to prove that they did not lose their integrity during the filtration process. These tests can be classified as destructive and nondestructive, and they are reviewed below.

14.10.2.1 Microbial Challenge Test This destructive test consists of passing a culture of the smallest known microorganism (*B. diminuta*) through a sterilizing filter, collecting the effluent, and determining if it contains the microorganism. Clearly, if this is the case, the filter fails. The test is summarized in Figure 14.36 and its works as follows. A culture medium containing *B. diminuta* (ATCC 19146) is prepared and then filtered through the sterile filter undergoing the test (e.g. a 0.22 μm -pore filter disk). A minimum of 10^7 microorganisms/cm² of filter must be filtered. This filtrate is then passed through a second 0.45 μm filter disk, which is incubated on a sterile TSA agar (nutrient) plate and monitored for formation of bacterial colonies. If colonies are formed, the filter has failed.

This test is used primarily by filter manufacturers to check the integrity of one or more filters in a lot of newly manufactured filters. Clearly, once a filter undergoes this test, it would not be prudent to reuse it, and it is therefore discarded.

14.10.2.2 Viral Challenge Test This destructive test is much less common than the microbial challenge test since

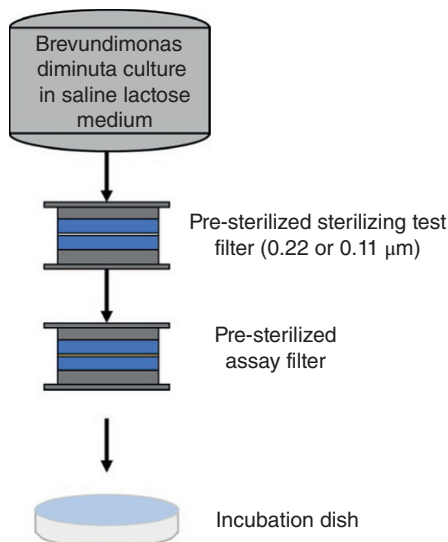


FIGURE 14.36 Bacterial challenge test.

viruses are much smaller than bacteria (with sizes between 0.04 and 0.1 μm , depending on the virus) and thus much more difficult to remove or inactivate. In addition, not being living organisms, their removal is not currently a requirement for sterility processes. However, the number of many viruses was found to be significantly reduced during sterile filtration processes designed to remove bacteria. The viral challenge test is similar to the microbial challenge test but bacteriophages (or phages, i.e. viruses that infects and replicates within bacteria) or other types of viruses are used instead of bacteria. Many filter manufacturers test their filters for phage challenge with a procedure similar to that for the bacterial challenge test.

14.10.2.3 Bubble Point Test The bubble point test is a nondestructive and very common method used to test the integrity of membrane filters of uniform pore sizes. When a material is placed in contact with a wetting liquid (e.g. water on the flat surface of a cellulosic material), the fluid preferentially wets the material and tends to spread over it (the opposite is true if the liquid does not wet the material, as when raindrops "bead up" in contact with the waxy surface of a polished car). If the wetted material is made of a fine matrix, the wetting liquid will tend to penetrate into the matrix as a result of the capillary effect, i.e. the same effect that results in the capillary rise of a wetting liquid in a small tube: the smaller the tube, the higher the rise. Similarly, the smaller the pore in a matrix, the more difficult it will be to dislodge the liquid from that pore once it is in it (as in a sponge), i.e. a higher external pressure will have to be applied across that pore of the matrix to remove the liquid from that pore. While this is true in general, it is particularly useful for the case of a thin membrane made of a matrix of pores of

uniform sizes, since, once fully impregnated with a specific wetting liquid with a given surface tension, all the pores will require the same pressure differential across the membrane to dislodge the liquid inside them. Therefore, the bubble point test consists in wetting a membrane with the appropriate liquid, placing it between two regions containing gas (air), and gradually increasing the gas pressure on one side. If the membrane is integral, then the gas will achieve breakthrough simultaneously in all pores (as indicated by the appearance of significant bubbles in the beaker in the figure) and only at the same pressure differential, the so-called bubble point pressure, which will depend on the liquid and the membrane pore size. However, if there is a defect in the membrane resulting a larger pore, then the pressure differential needed for the gas to escape through that pore will be smaller, and gas bubbles will appear at a lower pressure than the expected bubble point pressure (quantitative details on this are provided in Section 14.11).

The bubble point test is very commonly used to test membranes for liquid sterile filtration processes as shown in Figure 14.37. The test can be done before, during, or after filtration. However, the filtration process must be stopped, and an external pressurized gas must be admitted to the housing dome. If the pressure at which bubbles appear corresponds to the expected bubble point pressure, then the membrane is integral. If the bubbles appear at a lower pressure, the membrane is compromised. Other causes can also result in test failure, including using a wetting liquid with different surface tensions than recommended test liquid, integral filter but wrong pore size, high test temperature, or incompletely wetted membrane. Clearly, if the membrane is tested after the filtration process is completed and is found to be damaged, then all the liquid processed until that time must be reprocessed or discarded.

The bubble point test cannot be used with depth filters since these filters do not have uniform pore sizes. Even if appropriately wetted, the liquid trapped in larger pores will require a lower gas pressure than smaller pores, i.e. a single bubble point pressure for depth filters does not exist.

14.10.2.4 Diffusion Test This is a nondestructive test that can be used instead of the bubble point test. For membrane filters of large surface areas, it may be difficult to determine precisely when the bubble point pressure is reached. In such cases the diffusion test is especially appropriate to test the integrity of the membrane. The test is based on the principle shown graphically in Figure 14.38. If a membrane is dry, any small pressure differential across the membrane surface will enable the gas to flow minimally unimpeded across the membrane. However, if the pores in the membrane (schematically represented as cylindrical openings in the figure) are filled with a wetting liquid, the gas will no longer be able to move across the membrane unless the pressure differential exceeds the bubble point pressure, and the liquid in the pores is dislodged. However, although somewhat counterintuitively, a

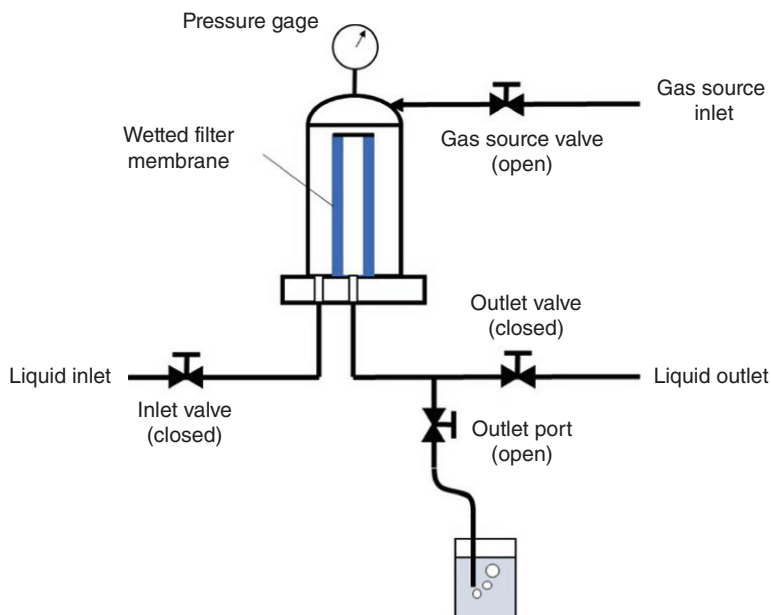


FIGURE 14.37 Bubble point test for membrane filter cartridges.

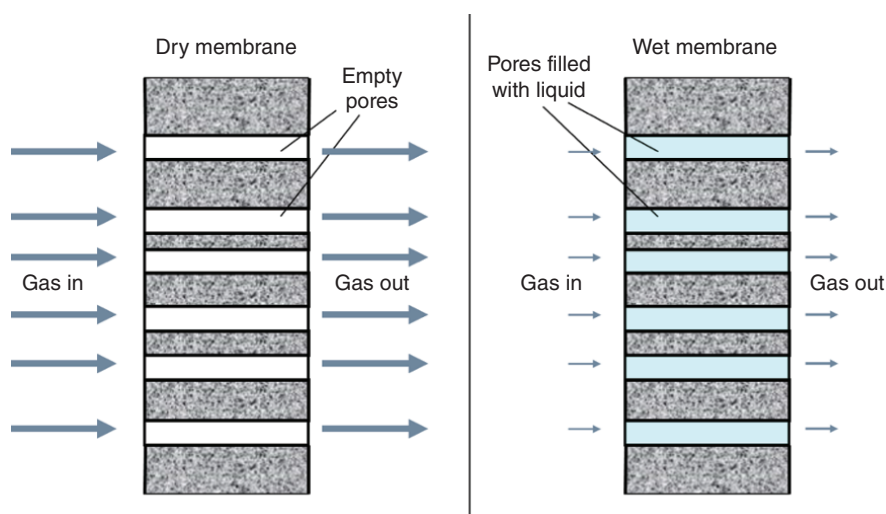


FIGURE 14.38 Diffusion test. Dry membrane (left). Wetted membrane (right).

small gas flow rate will always exist across the wetted membrane if a pressure differential exists, even if this pressure is lower than the bubble point pressure.

The reason for this is that the gas dissolves in the liquid in the pores at the high-pressure end of the membrane, diffuses (as dissolved gas) across the liquid in the pores toward the low-pressure end side, and then is released as a gas at the low-pressure end of the membrane. This small gas flow rate can be measured in integral membranes and can be even predicted. Therefore, the diffusion test consists in measuring

accurately this small gas flow rate across the membrane being used and comparing it to the expected value.

In an actual test, shown in Figure 14.39, the filter is wet with the appropriate test liquid (e.g. water for hydrophilic membranes, alcohol/water mixture for hydrophobic membranes), the pressure is slowly increased on the upstream side of the filter to the recommended test pressure (typically at least 80% of the minimum bubble point specification), the system is allowed to equilibrate, and the gas flow at the outlet port is measured using an accurate gas flow meter or even,

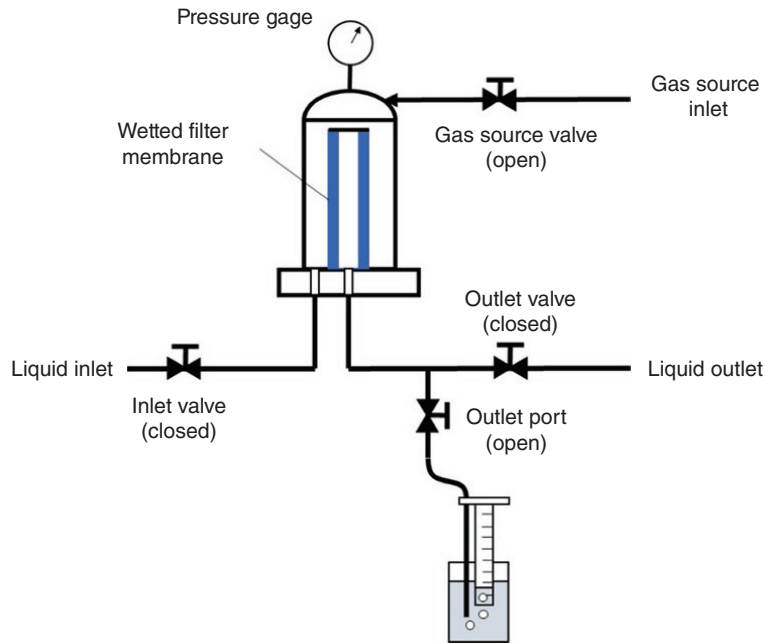


FIGURE 14.39 Diffusion test for membrane filter cartridges.

and very simply, by measuring for one minute the gas accumulated in an inverted graduated cylinder. If this flow is the same as that predicted, the membrane is integral. If not, any defect in the membrane will allow gas to flow at a higher rate, either as a result of a local breakthrough caused by the lower bubble point of that pore or because of the thinning of the liquid in that pore, leading to a high gas diffusion rate. Therefore, detecting a higher gas flow rate than expected could imply membrane failure.

14.10.2.5 Pressure Hold Test This test is also non-destructive and is just a variation of the diffusion test. Accordingly, it is conducted as shown in Figure 14.39 with one major difference: once the upstream side of the housing dome is pressurized with gas to the desired value (typically some 80% of the bubble point pressure), the gas source valve is turned off. Since gas diffusion across the membrane proceeds as before but there is no new gas source, the pressure in the housing dome decreases over time. If this pressure decreases with time as expected for an integral membrane, then the membrane is integral. If it proceeds more rapidly, membrane failure is suspected. This test has a real-life analog when one inflates a bicycle tire to a desired pressure and then detects the tire pressure over time (even qualitatively by detecting the tension on the tire). If the pressure decreases as expected, for example, over a one-month period, the tire is fine, but if the same pressure reduction occurs in a day, then the tire may be damaged.

14.10.2.6 Water Intrusion Test This test is only useful to test the integrity of hydrophobic air sterilizing membrane

filters or air vent. Because of their hydrophobicity, water will not wet these membranes and will not be able to penetrate the membrane pores. However, if enough pressure is applied, water can be forced into some pore structures and, if the membrane is not integral, even be able to cross the membrane if the defective pores are large enough.

The test consists of flooding with water the upstream side of a hydrophobic membrane (typically the outer side of a filter inside a cartridge) and applying pressure to it with a pressurized gas (Figure 14.40). As a result, the membrane pleats will bend, and, in addition, some water will enter the larger pore structure of the membrane, resulting in an initial pressure drop. After this equilibration period, the actual test can start during which the flow measuring apparatus measures the flow into the system required to maintain this initial pressure [71]. If the membrane is integral, this instrument will measure a small, steady flow of water through the membrane, corresponding to the flow recommended by the membrane manufacturer as a result of a similar standard test conducted on integral membranes (this flow is actually resulting from water on the high-pressure side of the membrane evaporating and escaping, through the pores, to the low-pressure side). However, if the membrane is defective, the measured flow will exceed this value since a larger amount of water will escape through the defect, and the membrane will have to be discarded.

14.10.2.7 Aerosol Test This test is used only for HEPA filters and other depth filters not to test them for sterility assurance, since these are not sterilizing filters, but to test their integrity and their ability to perform as expected. The

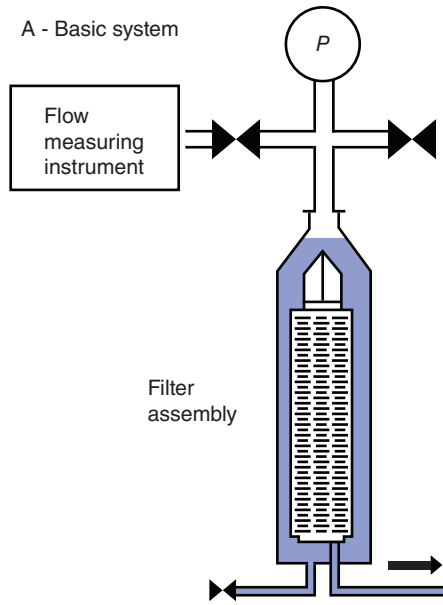


FIGURE 14.40 Water intrusion test apparatus. *Source:* From Jaenchen et al. [71].

test consists of generating a fog of droplets of poly-alpha-olefin (PAO) in the size range of 0.3–3 μm using an aerosol generator and introducing them into the air moving through the filter to be tested. The concentration of droplets appearing on the downstream side of the filter is measured using an aerosol photometer to ensure that they conform to the pre-defined standard for that filter.

14.11 ANALYSIS OF STERILE FILTRATION PROCESSES

14.11.1 Pressure Drop and Flow Rate Across Membrane Filters

The pressure drop, ΔP_f , experienced by a fluid that moves across a new, clean membrane is a function of the flow rate through the clean membrane, Q_0 , as well as a number of parameters pertaining to the characteristics of the membrane and the fluid. This relation can be described by different equations originally derived for flow through granular systems, such as Darcy's law:

$$\Delta P_f = \mu \frac{\delta Q_0}{\kappa A} = \frac{\mu R_0 Q_0}{A} \quad (14.86)$$

where

μ is the fluid viscosity (kg/(m·s)).

δ is the membrane thickness (m).

A is the membrane area (m^2).

κ is the hydraulic permeability (m^2).

R_0 is the hydraulic resistance of a clean membrane (m^{-1}).

If the membrane pores are assumed to be cylinders of diameter d_{pore} and assuming that the flow through them is laminar (Poiseuille flow), it is

$$\kappa = \frac{\varepsilon d_{\text{pore}}^2}{32\tau} \quad (14.87)$$

where

ε is the membrane void fraction.

d_{pore} is the pore diameter (m).

τ is the pore tortuosity (factor accounting for how much longer the actual average pore in the membrane is with respect to δ).

Another equation quantifying fluid flow through granular material is the Ergun equation, which, for the case in which the flow through the membrane is laminar (i.e. the typical case given the membrane small pore size), can be rewritten as the Blake–Kozeny equation, i.e.

$$\Delta P_f = 150 \frac{(1-\varepsilon)^2}{\varepsilon^3} \mu \frac{\delta Q_0}{D_p^2 A} \quad (14.88)$$

where

D_p is the effective particle diameter of particles making the solid granular bed, to be assumed to be proportional to the pore diameter d_{pore} for membranes.

In order to be used, these and similar equations rely on the accurate estimation of parameters such as the average pore diameter (which is squared in the equation and thus even more critical), the tortuosity factor, and the thickness of the membrane (rarely provided by manufacturers), all of which are difficult to determine precisely. Therefore, it is common practice to obtain parameters such as the hydraulic resistance for a clean membrane directly from simple experiments in which ΔP_f is obtained as a function of Q_0 . Figure 14.41 shows one of these charts, typically available from vendors. These empirical results show that the underlying relation between ΔP_f and $(\mu Q_0/A)$, as in both the Darcy's and Ergun equations, is appropriate to describe the process, although the proportionality constant between them should be determined experimentally for better quantification of the process.

The above equations apply to clean membranes. However, as the membrane is used, solids will be filtered out, and they will increase the hydraulic resistance of the membrane by obstructing the pores or accumulating and caking on its surface. The rate at which the filtration process will proceed will clearly depend on the type, size, and concentration of the solids contaminants, as well as the location where they will lodge (at the pore mouth, inside the pores, on the membrane surface forming a cake, etc.). Many different models exist,

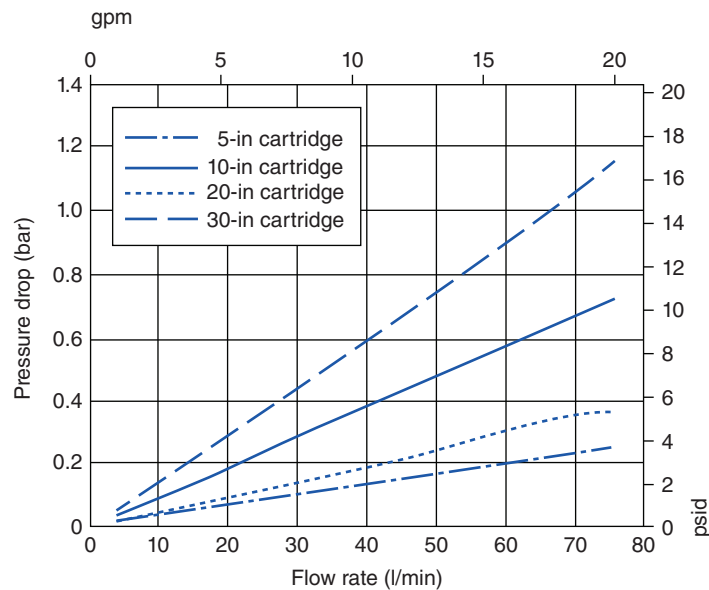


FIGURE 14.41 Typical ΔP_f vs. Q_0 chart for clean membranes (Millipore Express $0.2 \mu\text{m}$ SHF Hydrophilic Filter). *Source:* Courtesy of Millipore Sigma.

depending on the underlying blocking mechanisms, such as the standard blocking model (solids accumulate on the pore walls), complete blocking model (solids seal off pore entrances and prevent flow), cake filtration model (accumulation on surface), intermediate blocking model (sealing and accumulation), and others [72–74]. Each one of them results in a different prediction for the rate of flow reduction over time and may be more appropriate for different types of contaminants. Just as an example, assuming that the pressure differential ΔP_f is kept constant during filtration, the standard blocking model coupled with Darcy's law predicts that

$$\Delta P_f = \frac{\mu R Q}{A} \quad (14.89)$$

where the hydraulic resistance R changes with the cumulative volume, V , of filtrate generated at time t according to

$$R = \frac{R_0}{\left(1 - \frac{K_s V}{A}\right)^2} \quad (14.90)$$

where

K_s is the empirical plugging constant for the standard blocking model (in m^{-1}) that will depend on the type of solid particles and how they interact with the membrane.

Once this parameter has been obtained from experiments, it is possible to derive other expression for other variables of relevance including V as a function of time or Q as a function of time and be able to do scale-up of sterile filtration processes.

14.11.2 Calculation of the Bubble Point Pressure for Membranes of Uniform Pore Size

The prediction of the bubble point pressure for a membrane of uniform pore size is based on the physics of the capillary rise. If a liquid preferentially wets the solid surface of a capillary tube, it will rise in the capillary. At equilibrium, the upward force exerted by adhesion to the capillary tube (surface tension) will compensate the downward force caused by the column of liquid in the capillary (gravity force). At equilibrium in the capillary tube, it will be

$$g \rho_L h = \frac{4\gamma \cdot \cos \theta}{d_{\text{tube}}} \quad (14.91)$$

where

γ is the surface tension (dyne/cm).

θ is the contact angle, i.e. the angle formed by a drop of liquid in contact with the solid (Figure 14.42).

The term on the left-hand side in Eq. (14.91) has the dimension of a pressure, and it is in fact the pressure required to force the liquid out of the capillary and bring the liquid in it to its original liquid level.

In a porous solid matrix, the pores act as capillaries stacked next to each other. Therefore, if the pores are wetted with a wetting fluid (i.e. a fluid that adheres to the surface of the capillary and that has a small contact angle with the material making the matrix), the liquid will not be dislodged unless an extra pressure is applied by a gas on the pores to dislodge the fluid and break through the pores (bubble point).

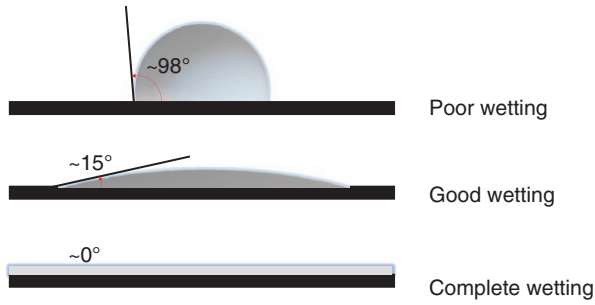


FIGURE 14.42 Contact angle of a liquid on a surface.

This pressure, the bubble point pressure Δp_{BP} , can be obtained from the following slightly modified version of Eq. (14.91):

$$\Delta p_{BP} = \frac{4k_{BP}\gamma \cdot \cos\theta}{d_{\text{pore}}} \quad (14.92)$$

where

$g\rho_L h$ has been replaced with the corresponding pressure term (Δp_{BP})

k_{BP} is a nondimensional shape correction factor accounting for the fact that the membrane is a tridimensional matrix with pores and imperfect cylindrical channels and it is not really made of ideally straight capillaries.

k_{BP} is typically determined experimentally, and it has been obtained for different porous materials, since the nature of the pore structure changes with the type of material and its fabrication. Common values are as follows:

- $k_{BP} = \sim 0.33$ for irregular shaped pores.
- $k_{BP} = \sim 0.25$ for more rounded pores.
- $k_{BP} = \sim 0.3$ for most membranes.

Values for k_{BP} are typically provided by the manufacturer or are incorporated in the manufacturer's bubble point specifications.

EXAMPLE PROBLEM 14.9

Let us calculate the bubble point pressure for a commercial hydrophilic sterilizing membrane cartridge filter (Millipore Express 0.2 mm SHF) and compare the result with literature data from the company. The wetting fluid is water.

Given:

$$d_{\text{pore}} = 0.2 \mu\text{m}$$

$$\text{Surface tension of water (at } 25^\circ\text{C)} = 72.8 \text{ dynes/cm}$$

We can use Eq. (14.92) to find Δp_{BP} , assuming that $k_{BP} \cong 0.3$. The result is 4.37 bar. The manufacturer specified that $\Delta p_{BP} \geq 4000$ mbar at 23°C , which is in agreement with the predicted value.

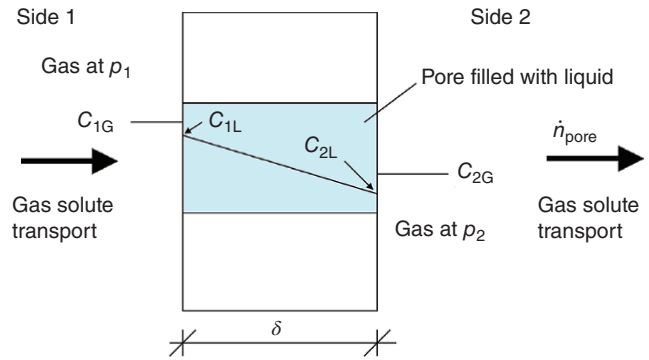


FIGURE 14.43 Gas diffusion in a pore.

If the same calculation is repeated for a similar membrane with a larger uniform pore size ($d_{\text{pore}} = 0.45 \mu\text{m}$), Eq. (14.92) predicts that $\Delta p_{BP} = 1.94$ bar. The value reported by the manufacturer is $\Delta p_{BP} \geq 1930$ mbar, i.e. in excellent agreement with the predicted bubble point pressure. Clearly, a larger uniform pore size results in a smaller bubble point pressure.

14.11.3 Calculation of the Gas Diffusion Rate in the Membrane Diffusion Test

In order to predict the rate of gas diffusion across a membrane filled with a wetting fluid, one can analyze the system using a simple schematic shown in Figure 14.43. This figure represents the cross section of a membrane subject to a differential pressure ($p_1 - p_2$), showing an idealized pore (not drawn to scale) filled with a wetting liquid.

In this figure, the dissolved gas on the higher-pressure side (side 1) has a concentration in liquid phase, C_{1L} , which is in equilibrium with the gas in the gas phase on side 1, C_{1G} . The corresponding concentrations on side 2 are C_{2L} and C_{2G} . Since $p_1 > p_2$, and therefore $C_{1G} > C_{2G}$, it follows that $C_{1L} > C_{2L}$ since the dissolved gas concentration is related to the gas concentration in the gas phase by the equilibrium relation $C_L = K_p C_G$ that applies to sparingly soluble gases. The difference in dissolved gas concentrations results in the transport of the dissolved gas between the two sides of the membrane that can be expressed using Fick's law of diffusion as

$$\dot{n}_{\text{pore}} = \frac{D_m}{\delta} A_{\text{pore}} (C_{1L} - C_{2L}) = \frac{D_m K_p}{\delta} A_{\text{pore}} (C_{1G} - C_{2G}) \quad (14.93)$$

where

A_{pore} is the cross-sectional area of a pore (m^2).

A is the cross-sectional area of membrane (m^2).

D_m is the diffusivity of gaseous solute in liquid inside the pores (m^2/s).

G_m is the volumetric flow rate of gaseous solute across the membrane (m^3/s).

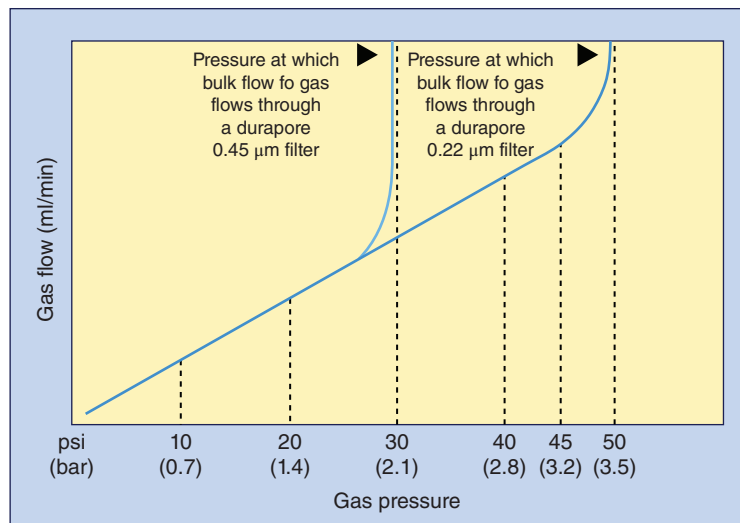


FIGURE 14.44 Gas flow rate as a function of pressure differential during a diffusion test. *Source:* Courtesy of Millipore Sigma.

\dot{n}_{pore} is the molar flow rate of gaseous solute across the membrane (mol/s).

\dot{n}_m is the molar flow rate of gaseous solute across a pore (mol/s).

K_p is the gas-liquid partition coefficient of gas solute.

n is the number of moles of gas solute (mol).

p is the gas pressure (Pa).

R_G is the ideal gas constant (J/(mol·K)).

T is the temperature (K).

V is the gas volume (m^3).

δ is the thickness of membrane (m).

ϵ is void fraction inside the membrane.

Equation (14.93) can be used to calculate the molar flow rate of dissolved gas, \dot{n}_m , across the whole membrane considering that \dot{n}_m is the sum of the flow rates for all the pores and that the area of a pore and the entire surface area of the membrane are related by $A = \epsilon A_{\text{pore}}$. Therefore

$$\dot{n}_m = \frac{D_m K_p}{\delta} A \epsilon (C_{1G} - C_{2G}) \quad (14.94)$$

Furthermore, this expression can be rewritten considering that the concentration of the gas in the gas phase can be expressed as $C_G = n_m/V = p/(R_G T)$, assuming that the ideal gas law ($pV = n_m R_G T$) applies. The final expression for the molar flow rate of dissolved gas, \dot{n}_m , across the whole membrane is then

$$\dot{n}_m = \left(\frac{D_m K_p}{\delta R_G T} \epsilon \right) A (p_1 - p_2) \quad (14.95)$$

In practice, this expression is more useful if the molar flow rate of dissolved gas is replaced with the volumetric flow rate

of dissolved gas, G_m , across the whole membrane using again the ideal gas law. If G_m is expressed in terms of the volumetric flow rate at a reference temperature and a reference pressure (for example, $T_{\text{ref}} = 25^\circ\text{C} = 298\text{K}$ and $p_{\text{ref}} = 1\text{atm}$), different than those of the test, then

$$G_m = \left(\frac{T_{\text{ref}}}{T} \cdot \frac{D_m K_p}{\delta} \epsilon \right) A \frac{(p_1 - p_2)}{p_{\text{ref}}} \quad (14.96)$$

This equation can be used to predict the gas flow rate during a diffusion test. A typical plot of G_m as a function of the pressure differential across the membrane is given in Figure 14.44.

EXAMPLE PROBLEM 14.10

Let us calculate the volumetric flow rate of air (reported at 25°C and 1 atm) for a commercial hydrophilic sterilizing membrane cartridge filter undergoing diffusion testing. The membrane is wetted with water, and the pressure difference across the membrane is kept at 2.8 atm. The test is conducted at 25°C .

Given:

$$d_{\text{pore}} = 0.2 \mu\text{m}.$$

$$A = 0.54 \text{m}^2.$$

$$\delta \cong 100 \mu\text{m}.$$

$$\text{Void fraction} \cong 80\%.$$

$$\text{Diffusion coefficient of air in water at } 25^\circ\text{C} = 2 \times 10^{-9} \text{m}^2/\text{s}.$$

$$\text{Henry's constant for air in water at } 25^\circ\text{C} = 1296 \text{atm}/(\text{mol/l}).$$

In this case, air is treated as a single gas since oxygen and nitrogen have similar properties and average diffusion coefficients, and the equilibrium constant is available. We can use Eq. (14.96) to solve this problem, but since the equilibrium between the air concentration in the gas phase and the dissolved air concentration in the liquid is expressed in terms of Henry's law, we have to slightly modify this equation. Henry's law for sparingly dissolvable gases is expressed as follows:

$$p = HC_L$$

This is an equilibrium relation where the air pressure corresponding to a given air concentration in the gas phase is used instead of C_G . However, $C_G = n_m/V = p/(R_G T)$ from the ideal gas law, and thus, at constant T , C_G and p are univocally related. This means that if the equilibrium is expressed as $C_L = K_p C_G$, then K_p can be expressed in terms of Henry's constant as follows:

$$K_p = \frac{R_G T}{H}$$

Then Eq. (14.96) becomes

$$G_m = \left(\frac{R_G T_{\text{ref}}}{p_{\text{ref}} H} \cdot \frac{D_m}{\delta} \varepsilon \right) A (p_1 - p_2)$$

Assuming that $T_{\text{ref}} = 25^\circ\text{C}$ and $p_{\text{ref}} = 1\text{ atm}$, i.e. reporting the volumetric air flow rate at this temperature and pressure, then $G_m = 27\text{ ml air/min}$. The manufacture's value is 30 ml/min that is in close agreement with the predicted value, especially considering that the void fraction and the membrane thickness were estimated.

14.11.4 Calculation of Pressure Decrease with Time in the Pressure Hold Test

In the pressure hold test, the rate of pressure decrease in the pressurized, sealed upstream side of the filter is monitored and related to the gas amount lost over time because of gas diffusion across the membrane, as shown in Figure 14.45.

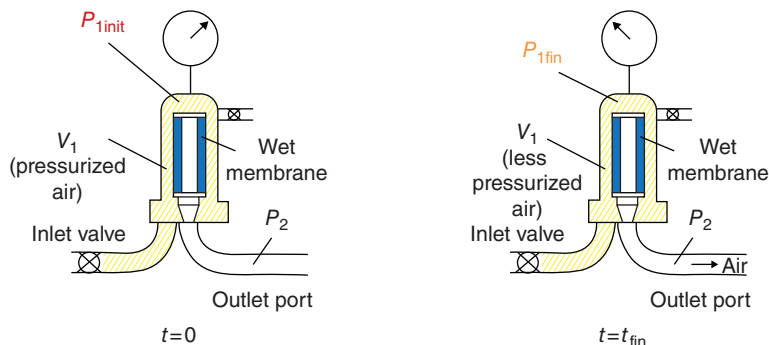


FIGURE 14.45 Pressure hold test. Initial status (left) and final status (right).

This system can be analyzed quantitatively considering what happens to the gas in the upstream sealed space, assumed to have a volume V_1 , using the simple schematic for the process depicted in Figure 14.46.

The unsteady-state mole balance for the gas in V_1 can be written as

$$\frac{dn_1}{dt} = -\dot{n}_m \quad (14.97)$$

Using the ideal gas law and recalling Eq. (14.95), this equation can be rewritten as

$$\frac{d}{dt} \left(\frac{p_1(t) V_1}{R_G T_1} \right) = - \left(\frac{D_m K_p \varepsilon A}{\delta R_G T} \right) (p_1(t) - p_2) \quad (14.98)$$

If the process is conducted at constant temperature, $T_1 = T_2 = T$, this expression can be integrated knowing that at $t=0$ the pressure upstream is $p_1 = p_{1\text{init}}$ to give

$$\ln \left(\frac{p_{1\text{fin}} - p_2}{p_{1\text{init}} - p_2} \right) = - \left(\frac{D_m K_p \varepsilon A}{\delta V_1} \right) t_{\text{fin}} \quad (14.99)$$

from which final pressure on the upstream side can be predicted to be

$$p_{1\text{fin}} = p_2 + (p_{1\text{init}} - p_2) \exp \left[- \left(\frac{D_m K_p \varepsilon A}{\delta V_1} \right) t_{\text{fin}} \right] \quad (14.100)$$

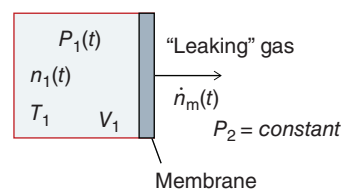


FIGURE 14.46 Schematic representation of gas leakage through the membrane filter during a pressure hold test.

An approximate version of this equation can be written recalling that for small values of x , it is that $\exp(-x) \cong 1 - x$, resulting in

$$p_{1\text{init}} = p_{1\text{fin}} \cong (p_{1\text{init}} - p_2) \left(\frac{D_m K_p \varepsilon A}{\delta V_1} \right) t_{\text{fin}} = \frac{R_G T}{V_1} \dot{n}_m t_{\text{fin}} \quad (14.101)$$

which can be rewritten as

$$\frac{p_{1\text{init}} - p_{1\text{fin}}}{p_2} \cong \frac{G_m t_{\text{fin}}}{V_1} \quad (14.102)$$

This is the expression most commonly used to analyze the results of the pressure hold test.

EXAMPLE PROBLEM 14.11

Let us calculate the decrease in pressure on the upstream side of a commercial hydrophilic sterilizing membrane cartridge filter at the end of a five minute period when the membrane undergoes a pressure hold test at 25 °C. The membrane is wetted with water. The initial upstream pressure is 3.8 atm (abs) = 3.813 bar, and the downstream pressure is atmospheric. The upstream volume is 3 l. The membrane characteristics and the physical parameters are as in the previous example.

The final upstream pressure can be calculated using Eq. (14.100), which, recalling that $K_p = R_G T/H$ from the previous example, can be written as

$$p_{1\text{fin}} = p_2 + (p_{1\text{init}} - p_2) \exp \left[- \left(\frac{D_m R_G T \varepsilon A}{\delta H V_1} \right) t_{\text{fin}} \right]$$

Substituting all numerical values results in $p_{1\text{fin}} = 3.804$ bar, which means that in five minutes the decrease in pressure upstream of the filter is $(p_{1\text{init}} - p_{1\text{fin}}) = 0.046$ bar, corresponding to 34.4 mmHg, which means that $((p_{1\text{init}} - p_{1\text{fin}})/p_2) = 4.5\%$.

Using instead the approximate equation (14.102) and inserting in it the value for G_m from the previous example (since the experimental conditions are the same) results in an identical value. Thus, the use of this second approach is thus well justified.

14.12 REGULATION OF STERILIZATION IN THE PHARMACEUTICAL INDUSTRY

14.12.1 Introduction

This section provides an overview on the regulation of the sterilization process. The two leading regulatory agencies are the US Food and Drug Administration (FDA) and the

European Union's European Medicines Agency (EMA). The regulatory guidance provided by these agencies serves as the framework for this section, although very similar guidance is provided by other regulatory agencies from around the world, such as the Japanese PMDA as well as from the World Health Organization (WHO). In general, the same principles are adopted worldwide. Regulation of sterile products mandates adherence to, first and foremost, cGMP principles, with additional specifics for ensuring sterility of the final product.

14.12.2 Sterilization and Regulation

As already mentioned, the sterility of a drug is defined as the complete absence of viable microorganisms. Testing for sterility of a preparation is evaluated statistically, creating a SAL, which was extensively discussed in previous sections. The SAL describes the probability of non-sterile units occurring in a single batch of sterile medicinal product. According to the pharmacopeia, the SAL for sterilized drugs in the final container is indicated as 1×10^{-6} .

Sterility is defined as a *critical quality attribute* for all sterile products. Sterility of the medicinal product cannot be assured by testing alone and needs to be assured by the use of suitable and validated manufacturing processes.

In order to minimize contamination during the manufacture of sterile products, it is important to ensure that certain general principles are complied with as follows:

- Engagement of qualified and well-trained personnel.
- The adoption of suitable premises for production.
- The use of suitable production equipment for manufacturing.
- Appropriate testing and controls to ensure sterilization.
- Documentation of manufacturing, environmental conditions, and associated controls.

Further to that, sterilization is achieved via two main methods:

- Terminal sterilization
- Aseptic processing

Terminal sterilization is defined as a process whereby a product is sterilized in its final container or packaging, which permits the measurement and evaluation of quantifiable microbial lethality [75].

Aseptic processing is defined as the processing and packaging process that are used to obtain a final product in a sterile state. Such processing requires careful control of all process steps to ensure that the product remains sterile throughout the production process.

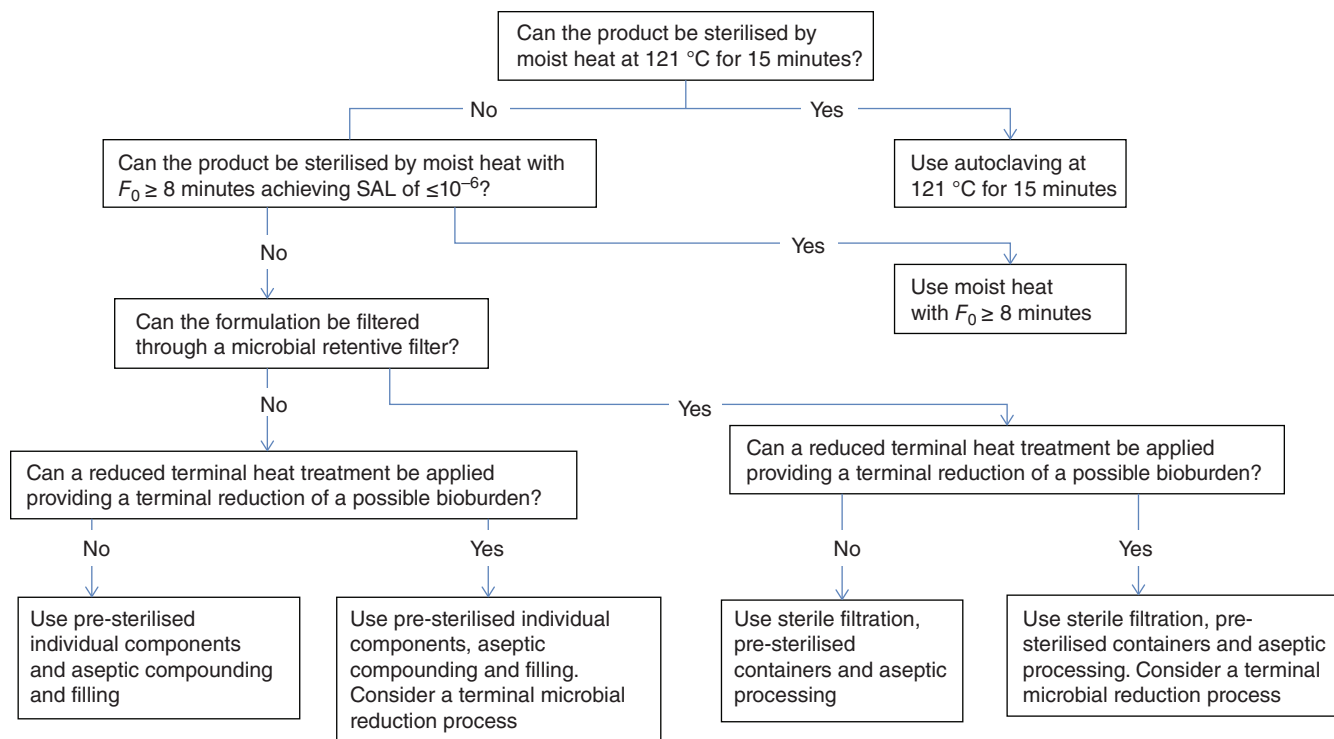


FIGURE 14.47 Decision tree for sterilization choices for aqueous products. *Source:* From http://www.ema.europa.eu/ema/index.jsp%3Fcurl%3Dpages/regulation/general/general_content_001431.jsp%26mid%3D.

In the *EU guideline on the sterilization of the medicinal product, active substance, excipient, and primary container* [76], the decision trees in Figures 14.47 and 14.48 are shown and are intended to assist in the selection of an appropriate sterilization method for a given product.

All regulatory agencies without exception note that terminal sterilization is the preferred method, and aseptic processing must be only used as a last resort if thoroughly justified. In the decision tree presented, aseptic processing is noted as an option of last resort. The reason is that aseptic processing involves the control of multiple steps, each of which bears some risk to contamination and thus must be carefully controlled and documented. With terminal sterilization, sterilization takes place on the final product with appropriate testing and controls, thus providing a higher degree of assurance that the sterilization process has been effective.

The focus of this section is the regulation of specifically terminal sterilization.

14.12.3 Sterility and Pharmaceutical Dosage Forms

For the EMA the following dosage forms must meet the requirements of sterility [77]:

- Injections
- Infusions

- Dosage forms that are applied to the eyes
- Dosage forms that are applied to mucous membranes.

The EMA does not enforce the sterility specification for aqueous-based oral inhalation solutions, suspensions, or sprays.

For the FDA, the requirements for sterility apply to the following, in addition to the EMA requirements (Table 14.9) [80–83]:

- Aqueous-based oral inhalation solutions, suspensions, or sprays must be sterile
- Inhalation solutions, suspensions, and sprays for oral delivery to the lungs.

14.12.4 Manufacture and Regulation of Terminally Sterilized Pharmaceutical Products

In the FDA guidance on the validation of sterilization [79], it is noted that: “The efficacy of a given sterilization process for a specific drug product is evaluated on the basis of a series of protocols and scientific experiments designed to demonstrate that the sterilization process and associated control procedures can reproducibly deliver a sterile product.” This means that for the FDA, in order for a sterilization manufacturing process to be valid, it must be validated. The principles for the validation must be carried out not only in accordance

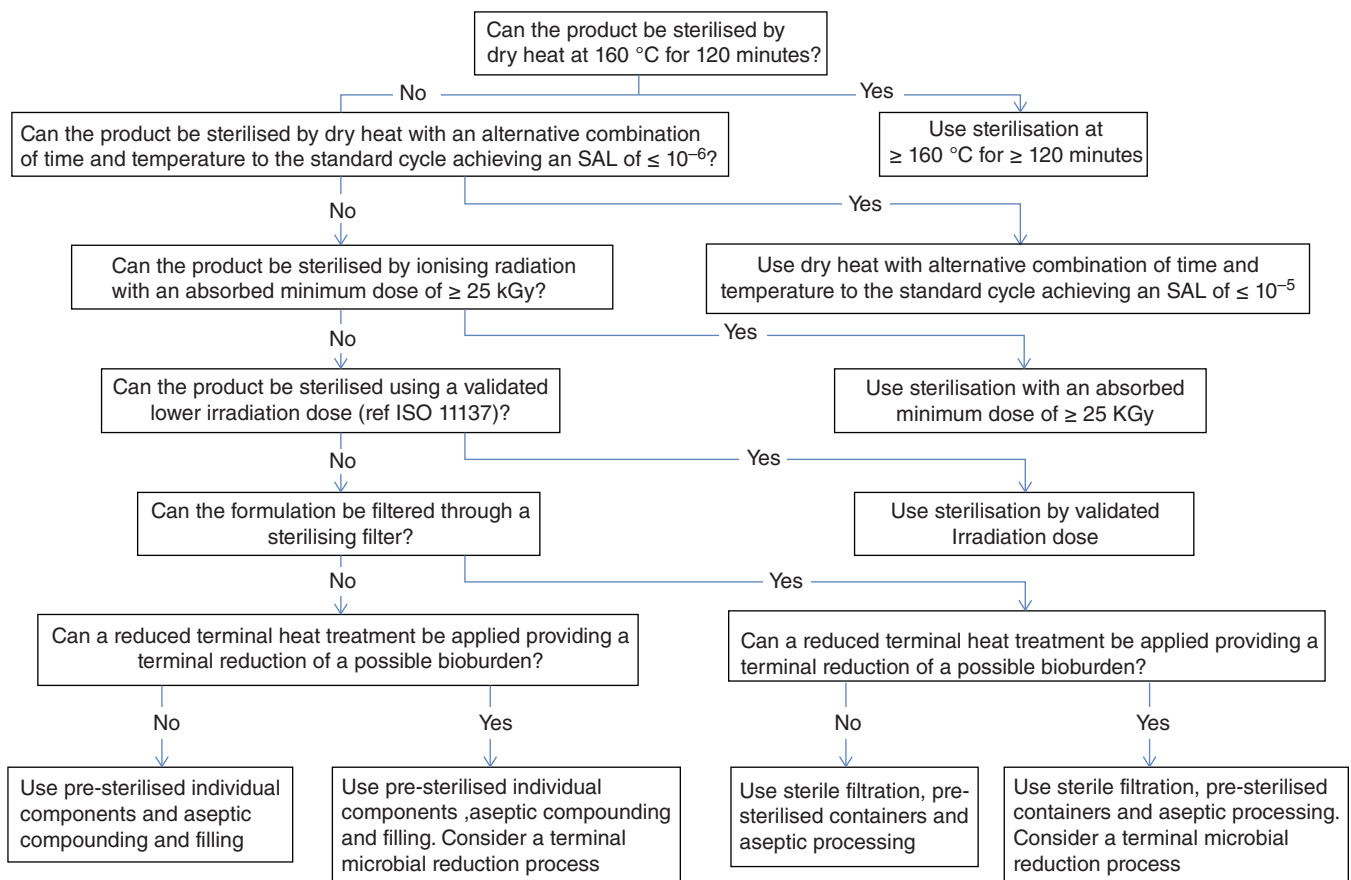


FIGURE 14.48 Decision tree for sterilization choices for nonaqueous liquid, semisolid, or dry powder products. *Source:* From http://www.ema.europa.eu/ema/index.jsp%3Fcurl%3Dpages/regulation/general/general_content_001431.jsp%26mid%3D.

TABLE 14.9 Summary of Relevant Regulations and Guidance

Author/Publisher	Title
European Pharmacopeia general chapter 2.6.1	Details of Methods and Sampling for Sterility Testing
European Pharmacopeia general chapter 5.1.1	Methods of Preparation of Sterile Products
European Pharmacopeia general chapter 5.1.2	Biological Indicators
European Pharmacopeia general chapter 5.1.5	Application of the F_0 Concept to Steam Sterilization of Aqueous Solutions
European Pharmacopeia general chapter 5.1.9	Guidance for Using the Test for Sterility
USP <1211>	Sterilization and Sterility Assurance of Compendial Articles (Manufacturing of Sterile Drug Products)
USP <1116>	Microbiological Monitoring of Clean Rooms and Other Controlled Environment
EMA/CHMP/CVMP/QWP/BWP/850374/2015	Guideline on the Sterilization of the Medicinal Product, Active Substance, Excipient and Primary Container (Draft)
Pharmaceutical Inspection Conventions (PIC/S)	Recommendation on the Validation of Aseptic Processes
Pharmaceutical Inspection Conventions (PIC/S)	Recommendation of Sterility Testing
Pharmaceutical Inspection Conventions (PIC/S)	Isolators Used for Aseptic Processing and Sterility Testing
FDA [78]	Guidance for Industry, Sterile Drug Products Produced by Aseptic Processing
FDA [79]	Guideline for the Submission of Documentation for Sterilization
ISO 13408-2: 2018	Aseptic Processing of Health Care Products: Part 2: Sterilizing Filtration
ISO 14644-1: 2015	Cleanrooms and Associated Controlled Environments: Part 1: Classification of Air Cleanliness by Particle Concentration
ISO 14644-2: 2015	Cleanrooms and Associated Controlled Environments: Part 2: Monitoring to Provide Evidence of Cleanroom Performance Related to Air Cleanliness by Particle Concentration
ISO 14644-4: 2015	Cleanrooms and Associated Controlled Environments: Part 4: Design, Construction and Start-up

to the guidance on sterilization (1994) but also with the updated guidance on validation [84] in consideration. Interestingly, it is noted in the 1994 guidance that it is, in principle, possible to validate a sterilization process without the need for three production batches.

As noted earlier, terminal sterilization may be affected by one of several modes:

- Steam (moist heat) and dry heat sterilization
- Radiation sterilization
- Chemical (liquid and gas) sterilization
- Sterile filtration.

The technical details of each mode of sterilization are covered in this chapter, and typical requirements for their control and acceptance are detailed in the EU guidance [76] and associated pharmacopeia monographs, as well as the FDA guidance on validation of sterilization [79], and summarized in the ensuing sections using the FDA 1994 validation guidance as a basis.

Once the mode of sterilization is adequately described, information on the controls and means of assessing effectiveness need to be provided.

14.12.5 Control of Sterilization Processes

14.12.5.1 Processing Environment Terminal sterilization cannot be accomplished without consideration of the environment and sequence of process steps, which lead to a product being sterilized. To that end regulatory authorities focus on four key areas in addition to the mode of sterilization:

- Equipment, processing, and facility design
- Air quality
- Water quality
- Personnel and training

A microbiological monitoring program for production areas along with a bioburden monitoring program for product components and process water should be established, including the autoclave cooling water. Details required include the frequency, methods used, and action levels, and data

summaries should be included. A description of the actions taken when specifications are exceeded is also required.

All of this must be carried in accordance with the relevant guidance and local cGMPs.

Specifications have been established for “clean rooms” where sterile manufacturing occurs. The established specifications are not harmonized between regulatory agencies, and differing terminology is used; therefore care must be exercised when designing a sterile manufacturing operation.

For the EMA the specifications are described in the EU GMP guide [85]. For the FDA this is spelled out in the USP guide (USP Chapter 1116) as well as the FDA guide on aseptic processing (Tables 14.10–14.12).

The applicability for each “clean room” type is provided in the EU GMP guide [86] and summarized in Table 14.13.

14.12.5.2 Container Closure and Package Integrity A description of the scientific validation studies (and data) in support of the microbial integrity of the drug packaging components must be described and include the following types of information (Table 14.14).

14.12.5.3 Sterility Testing Methods and Release Criteria Sterility test methods should be described and should include the protocol for the selection of representative units during production. When test methods differ significantly from compendial test methods, a demonstration of the equivalency to the compendial method should be provided. Testing performed within barrier systems should be described, and information concerning validation of the barrier system may be necessary.

The sterility test applied to the finished product should be validated for the product(s) concerned, and samples taken for sterility testing should be representative of the whole of the

TABLE 14.11 EU Classification: Maximum Permitted Airborne Particle Concentration During Classification

Grade	Maximum Permitted Number of Particles $\geq 0.5 \mu\text{m}$		ISO Classification in Operation/at Rest
	At Rest $\geq 0.5 \mu\text{m}/\text{m}^3$	In Operation $\geq 0.5 \mu\text{m}/\text{m}^3$	
A	3520	3520	5/5
B	3520	352000	5/7
C	352 000	3 520 000	7/8
D	3 520 000	Not defined	8

TABLE 14.10 FDA Clean Air Classification

Clean Area Classification (0.5 μm particles/ ft^2)	ISO Designation	$>0.5 \mu\text{m}$ Particles/ m^3	Microbiological Active Air Action Levels (cfu/ m^3)	Microbiological Settling Plates Action Levels (diam. 90 mm: cfu/4 h)
100	5	3520	1	1
1000	6	35 200	7	3
10 000	7	352 000	10	5
100 000	8	3 520 000	100	50

batch. Batch-processing records should be examined in conjunction with the results of the sterility tests.

To ensure an acceptable level of bacterial endotoxins in the finished drug product, the microbiological contamination of the components should be minimal. Specification limits for endotoxins and bioburden in components and bulk solution should be provided where relevant. It is not acceptable to rely on the sterilization process to solely reduce the microbial contamination.

For injectable products the water for injection and the intermediate, if appropriate, and finished products should be monitored for endotoxins, using an established compendia method that has been validated. In the event of sample failures, the cause of the failure should be investigated, and

necessary action should be taken. Alternative methods to those in the pharmacopeias may be used if they are validated, justified, and authorized.

14.12.5.4 Evidence of Formal, Written Procedures

Written procedures, designed to prevent microbiological contamination of drug products purporting to be sterile, must be established and followed. Such procedures should include validation of any sterilization process. Therefore, evidence should be provided that there are formal, written procedures describing the elements listed above and that these procedures are followed. Such evidence may consist of SOPs, listing of SOPs, and protocols submitted as part of these elements.

14.12.5.5 Parametric Release

Currently, FDA requires that sterile products meet certain sterility requirements before release to the market. In many cases, the requirements for batch release are fulfilled by conducting a sterility test on finished units drawn from the batch. Parametric release is defined as a sterility assurance release program where demonstrated control of the sterilization process enables the use defined critical process controls, in lieu of the sterility test, to fulfill the release requirements. Under this strategy, release of terminally sterilized products can be based upon meeting the defined sterilization parameters and not on

TABLE 14.12 EU Classification: Recommended Limits for Microbial Contamination in Operation

Grade	Air Sample (cfu/m ³)	Settle Places (Diameter 90 mm) (cfu/4 h)	Contact Places (Diameter 55 mm) (cfu/plate)
A	1	1	1
B	10	5	5
C	100	50	25
D	200	100	50

TABLE 14.13 Clean Room Classification and Application

Area Class	Requirements/Application
A	Sterile products – sterilized in the final container closure system – filling Sterile products – aseptic preparations aseptic preparation aseptic filling transfer of partially closed containers
B	Sterile products – aseptic preparations background areas for class A zones
C	Sterile preparations – sterilized in the final container closure system preparation of solutions (production step with high risk) filling Preparation of ointments, creams, suspensions, emulsions Sterile preparations – aseptic preparations preparation of solutions to be filled Background area for blow/fill/seal machines
D	Sterile preparations – sterilized in the final container closure system preparation of solutions and components for filling background area for blow/fill/seal machines Sterile preparations – aseptic preparations handling of components after the washing process background area for an isolator

TABLE 14.14 Container Closure and Package Integrity Requirements

Simulation of the stresses from processing	Experimental designs that simulate the stresses of the sterilization process, handling, and storage of the drug and their effects on the container closure system
Demonstrate integrity following the maximum exposure	Container closure integrity should be demonstrated on product units that have been exposed to the maximum sterilization cycle(s). If a product is exposed to more than one process, then exposure to the maximum cycle of all processes should be incorporated into the study design
Multiple barriers	Each barrier that separates areas of the drug product claimed to be sterile should be separately evaluated and validated
The sensitivity of the test	The sensitivity of the experimental method used for container closure integrity testing should be specified and provided
Integrity over the product shelf life	Microbial integrity of the container closure system should be demonstrated over the shelf life of the product

performing an approved sterility test. Meeting the requirements of the parametric release process can provide greater assurance that a batch meets the sterility requirement than can be achieved with a sterility test of finished units drawn from the batch.

14.12.5.6 Control Strategy for the Terminal Sterilization Program A control strategy is used to ensure that the acceptance criteria of the parametric release process and terminal sterilization cycle are met in order to ensure product sterility. The control strategy should include the following:

- The rationale for the methods implemented to monitor and control the terminal sterilization process used for the product release (the critical process parameters).
- The rationale for the selection of critical process parameter(s).
- A description of the acceptance criteria for parametric release.
- A description of the drug product and container closure system (including secondary packaging, as applicable) that will be part of the parametric release program.
- A description of the proposed production loading patterns and verification that they are within the validated limits for the terminal sterilization cycle or a statement that they have not changed since last approved and validated (as applicable).
- A description of the microbiological monitoring plan for the product and components prior to terminal sterilization or a statement that the plan has not changed since last validated.

14.12.5.7 Risk Assessment, Process Understanding, and Prior Knowledge Successful parametric release systems are based on the reliability of the control strategy of the sterility assurance program. The risk assessment should include the following:

- Consistency of performance of the terminal sterilization cycle within the validated limits.
- A discussion of risk to the sterility of the product relative to:
 - The production terminal sterilization cycle.
 - The production loading patterns.
 - The container closure system (including secondary packaging).
- Any potential contamination risks from the environment (as appropriate).
- Experience with the proposed or similar product (and container closure system) and proposed or similar sterilization process, the overall risks to sterility, and the steps have been taken to assess and control these risks.

For new products, prior knowledge from developmental and registration/exhibit batches may be sufficient.

14.12.5.8 Documentation for Parametric Release Process The following information specific to the proposed parametric release process should also be included in a description of a parametric release program:

- A complete and detailed description of the current relevant terminal sterilization cycle.
- Identification of the critical process parameters (process/cycle parameters and appropriate load monitors essential for product release) for the product(s) proposed for parametric release, including the minimum and maximum limits for these critical parameters.
- Acknowledgment that adherence to the critical parameters of the parametric release program will substitute for the performance of a sterility test as the primary release criterion for the product and that sterility test results from the finished product will not be used to overrule any failure to meet the acceptance criteria of the parametric release program.
- Acknowledgment that regardless of the batch release technique used, any specimen tested according to the reference test method for sterility will meet the criteria for sterility.
- A description of the sterilization load monitor that indicates:
 - The type of monitor being proposed.
 - How the load monitor will be used and analyzed.
 - The functions being measured by the monitor.
 - The rationale for the location of the monitor.
- Documentation of the control system to verify exposure of the load to the sterilization process.
- Revision of the certificates of analysis or batch release records for each product subject to parametric release to indicate that parametric release is now the method used to provide assurance of the requirement of sterility.

14.12.6 Regulation of Terminal Sterilization

The specific aspects to be controlled for terminal sterilization are described in the following sections. These in conjunction with the controls described above are necessary for the successful sterilization program.

14.12.6.1 Regulation of Moist Heat Sterilization The controls of moist heat sterilization begin with a description of the sterilization process. The elements to be sterilized must be described in sufficient detail and include descriptions of the drug product and the container closure system(s) to be sterilized (e.g. size(s), fill volume, and secondary packaging if used). This is followed by a description of the controls used

to assess the effectiveness of the operations and confirmation of the success of the sterilization operation.

The system under consideration needs to be first described in sufficient detail along with objectives for success (Table 14.15).

After describing the system under consideration, the means whereby the operations will be monitored are described (Table 14.16).

The controls used to measure the success (or lack thereof) and effectiveness of the operations must be provided (Table 14.17).

14.12.7 Regulation of Gas Sterilization with Ethylene Oxide

The specific elements necessary for gas sterilization are described in Table 14.18.

14.12.8 Regulation of Sterilization Using Radiation

The specific elements necessary for sterilization using radiation are described in Table 14.19.

14.13 CONCLUDING REMARKS

In summary, sterilization is regulated in accordance with cGMPs and local regulations. The choice of the sterilization method has to be justified, and the associated methods must be validated. In the EU, the method of sterilization described in the corresponding general monograph of the European Pharmacopoeia has to be used. All deviations have to be justified. The procedures for all sites (including outsourced activities) where sterilization is performed have to be documented.

TABLE 14.15 Description of Sterilization Process and Product

Item	Description
Container closure system	Descriptions of the drug product and the container-closure system(s) to be sterilized
The sterilization process	A description of the sterilization process used to sterilize the drug product in its final container-closure system, as well as a description of any other process(es) used to sterilize any necessary additional components, such as packaging, drug substance or drug product, and related items. The efficacy of these process must be justified
The autoclave process and performance specifications	A description of the autoclave process, including pertinent information such as cycle type (e.g., saturated steam, water immersion, and water spray), cycle parameters and performance specifications including temperature, pressure, time, and minimum and maximum F_0 . The type of autoclave used, as well manufacturer and model, must be described
Autoclave loading patterns	The autoclave loading patterns must be described
Methods and controls to monitor production cycles	Methods and controls used to monitor routine production cycles (e.g., thermocouples, pilot bottles, and biological indicators) should be described, including the number and location of each as well as acceptance and rejection specifications
Requalification of production autoclaves	A description of the program for routine and unscheduled requalification of production autoclaves, including frequency must be described
Reprocessing	A description and validation summary of any program that provides for reprocessing (e.g., additional thermal processing) of product should be provided. This could also affect the stability program

TABLE 14.16 Thermal Qualification of the Cycle

Heat distribution and penetration studies	Heat distribution and penetration study protocols and data summaries that demonstrate the uniformity, reproducibility, and conformance to specifications of the production sterilization cycle should be provided
Thermal monitors	The number of thermal monitors used and their location in the chamber should be described. A diagram of the layout is useful
The effects of loading on thermal input	Data showing the effects of loading on thermal input to product, including minimum and maximum loads
Information included in the batch record	The validated processes to be used for sterilization and for de-pyrogenation of any container-closure components should be described

TABLE 14.17 Microbiological Efficacy of the Cycle

Identification and characterization of bioburden organisms	The methods and results from studies used to identify and characterize bioburden organisms should be described
Specifications for bioburden	Specifications (alert and action levels) for bioburden should be provided. A program for routinely monitoring bio-burden levels and ensuring they remain within specifications must be provided
Identification, resistance, and stability of biological indicators	Information and data concerning the identification, resistance (<i>D</i> and <i>Z</i> values), and stability of biological indicators used in the biological validation of the cycle should be provided
Resistance of the biological indicator relative to that of bioburden	Studies characterizing the resistance of the biological indicator relative to that of bioburden may be necessary
Microbiological challenge studies	Microbiological validation studies that demonstrate the efficacy of the minimum cycle to provide a sterility assurance of 10^{-6} or better to the product under the most difficult to sterilize conditions must be provided

TABLE 14.18 Fumigation Requirements

Description of the sterilizer	The sterilizer(s) and controlled site(s) for prehumidification and aeration of the product load should be described
Cycle parameters	The parameters and limits for all phases of the cycle, e.g. prehumidification, gas concentration, vacuum and gas pressure cycles, exposure time and temperature, humidity, degassing, aeration, and determination of residuals should be specified. Specific procedures used to monitor and control routine production cycles to assure that performance is within validated limits should be provided
Microbiological methods	The microbiological methods (growth medium, incubation temperature, and time interval) for cultivating spores from inoculated samples during validation experiments should be described as well as the microbiological methods used as part of routine production cycles
Stability	The program for monitoring the stability of packaging and the integrity of the container closure system barrier over the claimed shelf life should be described

TABLE 14.19 Radiation Requirements

The facility and the process	The radiation facility should be identified. The radiation source, method of exposure (i.e. movement through the irradiator), and the type and location of dosimeters used to monitor routine production loads should be described. If the low dose site is not used for routine monitoring, data that show the dose relationship between the two sites should be provided
The packaging of the product	The packaging of the drug product within the shipping carton and within the carrier should be described
Multiple-dose mapping studies	Multiple-dose mapping studies for identification of low and high dose sites and demonstration of uniformity and reproducibility of the process should be described
Microbiological methods and controls	The microbiological methods and controls used to establish, validate, and audit the efficacy of the cycle should be described
Monitoring stability	The program for monitoring the stability of packaging and the integrity of the container closure system barrier over the claimed shelf life should be described

REFERENCES

- CFR 21: Section 600 (2017). Biological Products: General, Part 600.3.
- Errington, J. (2003). Regulation of endospore formation in *Bacillus subtilis*. *Nat. Rev. Microbiol.* 1: 117–126.
- World Health Organization (2004). Guidelines on viral inactivation and removal procedures intended to assure the viral safety of human blood plasma products. WHO Technical Report, Series No. 924.
- Armenante, P.M. and Kirpekar, A. (1997). *Sterilization in the pharmaceutical and biotechnology industry*. In: *Handbook of Downstream Processing* (ed. E. Goldberg), 261–308. New York: Chapman & Hall.
- Garfinckle, B.D. and Henley, M.W. (2000). Sterilization. In: *Remington: The Science and Practice of Pharmacy*, 20e (ed. A.R. Genaro). Philadelphia, PA: Lippincott, Williams & Wilkins.
- Marino, F. J., Industrial sterilization: a review of current principles and practices, in *Pharmaceutical Dosage Forms: Parenteral Medications*, Vol. 2 (1993), by Avis, K. E., Lieberman, H. A., and Lachman, L. (eds.), Marcel Dekker, New York, pp. 473–540.
- Schuler, M.L. and Kargi, F. (1992). *Bioprocess Engineering*, 297–306. Englewood Cliff, NJ: Prentice Hall.

8. Scott, E.M. and Gorman, S.P. (1991). *Disinfection, Sterilization and Preservation*, 4e (ed. S.S. Block), 596–614. Philadelphia, PA: Lea and Febiger.
9. United States Pharmacopeia (USP) (2017). *National Formulary, USP40-NF35. Microbiological Control and Monitoring of Aseptic Processing Environments (1116); Sterilization and sterility assurance of compendial articles (1211) and Sterilization of compendial articles (1229)*. Rockville, MD: The USP Convention, Inc.
10. Wittenburg, G., Lauer, G., Oswald, S. et al. (2014). Nanoscale topographic changes on sterilized glass surfaces affect cell adhesion and spreading. *Mat. Res. Part A* 102 (8): 2755–2766.
11. Richards, J.W. (1968). *Introduction to Industrial Sterilization*. New York: Academic Press.
12. Pflug, I.J. (1972). Heat sterilization. In: *Industrial Sterilization* (ed. G.B. Phillips and W.S. Miller). Durham, NC: Duke University Press.
13. Boca, B.M., Pretorius, E., Gochin, R. et al. (2002). An overview of the validation approach for moist heat sterilization, part I. *Pharm. Technol.* 26: 62–70.
14. Lin, S.H. (1975). A theoretical analysis of thermal sterilization in a continuous sterilizer. *J. Ferment. Technol.* 53: 92–98.
15. Lin, S.H. (1976). Continuous high temperature/short time sterilization of liquid foods with steam-injection heating. *Chem. Eng. Sci.* 31: 77–82.
16. Lin, S.H. (1979). Residence time distribution of flow in a continuous sterilization process. *Process Biochem.* 14 (7): 23–27.
17. Armenante, P.M. and Li, Y.-S. (1993). Complete design analysis of a continuous sterilizer for fermentation media containing suspended solids. *Biotechnol. Bioeng.* 41: 900–913.
18. Armenante, P.M. and Leskiewicz, M.A. (1990). Design of continuous sterilization systems for fermentation media containing suspended solids. *Biotech. Progress* 6 (4): 292–306.
19. de Ruyter, P.W. and Brunet, R. (1973). Estimation of process conditions for continuous sterilization of food containing particulates. *Food Technol.* 27: 44–51.
20. Manson, J.E. and Cullen, J.F. (1974). Thermal process simulation for aseptic processing of food containing discrete particulate matter. *J. Food Sci.* 39: 1084–1089.
21. Vogel, H.C. (ed.) (1983). *Fermentation and Biochemical Engineering Handbook*. Park Ridge, NJ: Noyes Publications.
22. Parrott, E.L. (1970). *Pharmaceutical Technology*. Minneapolis, MN: Burgess Publishing Co.
23. Martínez, J.E. (2002). Microbial bioburden on oral solid dosage forms. *Pharm. Technol.* 26: 58–70.
24. Agalloco, J. (2007). Understanding overkill sterilization: an end to the confusion. *Pharm. Technol.* 30: S18–S25.
25. Deindoerfer, F.H. (1957). Calculation of heat sterilization times for fermentation media. *Appl. Microbiol.* 5: 221–228.
26. Wang, D.I.C., Cooney, C.L., Demain, A.L. et al. (1979). *Fermentation and Enzyme Technology*. New York: Wiley.
27. Aiba, S., Humphrey, A.E., and Millis, N.F. (1973). *Biochemical Engineering*, 2e. New York: Academic Press.
28. Deindoerfer, F.H. and Humphrey, A.E. (1959b). Principles in the design of continuous sterilizers. *Appl. Microb.* 7: 265–270.
29. Moss, M.O. and Smith, J.E. (1977). *Industrial Applications of Microbiology*. London: Surrey University Press.
30. Blackebrough, N. (1968). *Biochemical and Biological Engineering Science*, vol. 2. London: Academic Press.
31. Charm, S.E. (1971). *The Fundamental of Food Engineering*, 2e. Westport, CT: AVI Publishing Company.
32. Wallhäuser, K.H. (1985). Sterilization in biotechnology. *J. Chem. Technol. Biotech.* 2: 699–723.
33. Kirk, R.E. and Othmer, D.F. (2000). *Encyclopedia of Chemical Technology*. New York: Wiley.
34. Armenante, P.M. (1994). Derivation of a rigorous equation for the calculation of the *F*-value in isothermal sterilization processes. *J. Pharm. Sci.* 83 (5): 668–673.
35. Bailey, J.E. and Ollis, D.F. (1986). *Biochemical Engineering Fundamentals*, 2e. New York: McGraw-Hill.
36. Blanch, H.W. and Clark, D.S. (1996). *Biochemical Engineering*. New York: Marcel Dekker Inc.
37. Feliciotti, E. and Esselen, W.B. (1957). Thermal destruction rates of thiamine pureed meats and vegetables. *Food Technol.* 11: 77–84.
38. Ashley, M.H.J. (1982). Continuous sterilization of media. *Chem. Eng. (London)* 1 (2): 54–58.
39. Deindoerfer, F.H. and Humphrey, A.E. (1959a). Analytical methods for calculating heat sterilization times. *Appl. Microb.* 7: 256–264.
40. Brenner, H. (1962). The diffusion model of longitudinal mixing in beds of finite length. Numerical values. *Chem. Eng. Sci.* 17: 229–244.
41. Levenspiel, O. (1958). Longitudinal mixing of fluids flowing in circular pipes. *Ind. Eng. Chem.* 50: 343–346.
42. Perry, R.H. and Green, D.W. (2008). *Perry's Chemical Engineers' Handbook*, 8e. New York: McGraw-Hill.
43. Levenspiel, O. (1971). *Chemical Reaction Engineering*, 2e. New York: Wiley.
44. Aiba, S. and Sonoyama, T. (1965). Residence time distribution of a microbial suspension in a straight pipe. *J. Ferment. Technol.* 43: 534–539.
45. Carslaw, H.S. and Jaeger, J.C. (1959). *Conduction of Heat in Solids*, 2e. Oxford: Clarendon Press.
46. Bird, R.B., Stewart, W.E., and Lightfoot, E.N. (2006). *Transport Phenomena*, 2e. New York: Wiley.
47. Geankoplis, C.J. (1983). *Transport Processes and Unit Operations*, 2e. Boston: Allin and Bacon.
48. Jeong, S.-G. and Kang, D.-H. (2017). Inactivation of *Escherichia coli* O157:H7, *Salmonella Typhimurium*, and *Listeria monocytogenes* in ready-to-bake cookie dough by gamma and electron beam irradiation. *Food Microbiol.* 64: 172–178.
49. Shehata, M.M.K., Gomaa, F.A.M., and Helal, Z.H. (2011). Effects of gamma and electron beam irradiation on viability and DNA elimination of *Staphylococcus aureus*. *Arch. Clin. Microbiol.* 2 (6:3): 1–9.
50. Trampuz, A., Piper, K.E., Steckelbergand, J.M., and Patel, R. (2006). Effect of gamma irradiation on viability and DNA of *Staphylococcus epidermidis* and *Escherichia coli*. *J. Medical Microbiol.* 55: 1271–1275.

51. Alexander, P. and Bacq, Z.M. (eds.) (1961). *Fundamentals of Radiology*, vol. 5. New York: Pergamon Press, MacMillan Co.
52. Gaughran, E.R.L. and Lloyd, R.S. (eds.) (1974). *Technical Developments and Prospects of Sterilization by Ionizing Radiation*, vol. 1. Montreal: MultiScience Publication, Ltd.
53. Matagne, D., Delbar, N., Hartmann, H.-J. et al. (2004). Development of a process using electron beam for a terminal sterilization for parenteral formulations of pharmaceuticals. *Radiat. Phys. Chem.* 71: 419–422.
54. Carlton, C.B. and Brown, B.J. (1981). Gene mutation. In: *Manual of Methods of General Bacteriology* (ed. P. Gerhardt). Washington, DC: American Society for Microbiology.
55. Aquino, K.A.S. (2012). *Sterilization by Gamma Irradiation, Gamma Radiation*. London: InTech.
56. Block, S.S. (ed.) (1991). *Disinfection, Sterilization and Preservation*, 4e, 167–181. Philadelphia: Lea and Febiger.
57. OSHA (2002). Ethylene Oxide Fact Sheet. U.S. Department of Labor, Occupational Safety and Health Administration (OSHA), Washington DC
58. Forstrom, R.J. and Wardle, M.D. (1979). Low temperature sterilization using H₂O₂ at reduced pressure, U.S. Patent, 4,169,124.
59. Moore, F.C. and Perkinson, L.R. (1979). Cold gas sterilization using H₂O₂. US Patent, 4,366,125.
60. Rice, R.G. and Netzwer, A. (1982). *Handbook of Ozone Technology and Applications*, vol. 1. Ann Arbor, MI: Ann Arbor Science.
61. Dufresne, S. and Richards, T. (2016). The first dual-sterilant low-temperature sterilization system. *Can. J. Infect. Control* 31 (3): 169–174.
62. Goulet, E., Mullins, S., Levine, R., and Cooper, J. (2014). Depyrogenation and sterilization of medical device materials with nitrogen dioxide gas. PDA 8th Annual Global Conference on Pharmaceutical Microbiology, Bethesda, MD (21–23 October 2013).
63. McDonnell, G., Ehrman, M., and Kiess, S. (2016). Effectiveness of the SYSTEM 1E liquid chemical Sterilant processing system for reprocessing duodenoscopes. *Am. J. Infect. Control* 44: 685–688.
64. Wang, J. and Toledo, R.T. (1986). Inactivation of microorganisms on polyethylene exposed to hydrogen peroxide vapors in air at various temperatures. In: *Current Technologies in Flexible Packaging* (ed. M.L. Troedel), 37–48. Philadelphia: American Society for Testing and Materials.
65. Wardle, M.D. and Renninger, G.M. (1975). Bacterial effect of hydrogen peroxide on spacecraft isolates. *Appl. Microbiol.* 30: 710–711.
66. Ito, K.A. (1973). Resistance of bacterial spores to hydrogen peroxide. *Food Technol.* 11: 59–66.
67. Mentel, R. and Schmidt, J. (1973). Investigation on rhinovirus inactivation by hydrogen peroxide. *Acta Virol.* 17: 351–354.
68. Angal, S. and Dean, P.D.G. (1989). *Protein Purification Methods: A Practical Approach* (ed. E.L.V. Harris and S. Angal), 254–261. Oxford: Oxford University Press.
69. Rodriguez, A.C., Young, B., Caulk, K. et al. (2001). Calculating accumulated lethality and survivorship in EtO sterilization processes. *Med. Device Diagn. Ind.* 23: 100–107.
70. Meltzer, T.H. (1989). Filtration: the practice of prefiltration and its related considerations. *Ultrapure Water* 6 (3): 24–34.
71. Jaenchen, R., Schubert, J., Jafari, S., and West, A. (1997). Studies on the theoretical basis of the water intrusion test (WIT). *Eur. J. Parenter. Sci.* 2 (2): 39–45.
72. Bolton, G., LaCasse, D., and Kuriyel, R. (2006). Combined models of membrane fouling: development and application to microfiltration and ultrafiltration of biological fluids. *J. Membr. Sci.* 277: 75–84.
73. Rajniak, P., Tsinontides, S.C., Pham, D. et al. (2008). Sterilizing filtration: principles and practice for successful scale-up to manufacturing. *J. Membr. Sci.* 325: 223–237.
74. Saleem, M., Alibardi, L., Cossu, R. et al. (2017). Analysis of fouling development under dynamic membrane filtration operation. *Chem. Eng. J.* 312: 136–143.
75. PMDA (2007). *Guidance on the Manufacture of Sterile Pharmaceutical Products Produced by Terminal Sterilization*. Tokyo, Japan: Pharmaceutical and Medical Devices Agency (PMDA).
76. EU (2016). Guideline on Sterilisation of the Medicinal Product, Active Substance, Excipient and Primary Container.
77. European Pharmacopoeia (EP) (2016). General Chapter 5.1.1 ‘Methods of preparation of sterile products’; General Chapter 5.1.2 ‘Biological indicators of sterilization,’ General Chapter 5.1.4 ‘Microbiological Quality of Pharmaceutical Forms’ and General Chapter 5.1.5 ‘Application of the Fo concept to steam sterilisation of aqueous preparations’.
78. Food and Drug Administration (FDA) (2004). *Guidance for Industry Sterile Drug Products Produced by Aseptic Processing: Current Good Manufacturing Practice*. Rockville, MD.; FDA.
79. Food and Drug Administration (FDA) (1994). *Guidance for Industry for the Submission Documentation for Sterilization Process Validation in Applications for Human and Veterinary Drug Products*. Rockville, MD.; FDA.
80. CDER (1994). Documentation for Sterilization Process Validation in Applications for Human and Veterinary Drug Products. Rockville, MD.,.
81. CDER (2002). *Guidance for Industry: Nasal Spray and Inhalation Solution, Suspension and Spray Drug Products – Chemistry, Manufacturing, and Control Documentation*. Rockville, MD.; FDA.
82. CDER (2001). *Guidance for Industry, Sterility Requirement for Aqueous-Based Drug Products for Oral Inhalation: Small Entity Compliance Guide*. Rockville, MD.; FDA.
83. CFR 21: Section 200.51 (2001). Aqueous-Based Drug Products for Oral Inhalation.
84. Food and Drug Administration (FDA) (2011). *Guidance for Industry Process Validation: General Principles and Practices*. Rockville, MD: FDA.
85. EU GMP Guide (2008). Annex 1 Manufacture of Sterile Medicinal Products, 2008; Annex 2 Biological Manufacture of Biological Active Substances and Medicinal Products for Human Use.
86. European Commission (EC) (2017). EudraLex: The Rules Governing Medicinal Products in the European Union, Volume 4, Good Manufacturing Practice - Guidelines on Good Manufacturing Practice Specific to Advanced Therapy Medicinal Products.

CONTROLLED RELEASE TECHNOLOGY AND DESIGN OF ORAL CONTROLLED RELEASE DOSAGE FORMS

AVINASH G. THOMBRE

Drug Product Design, Pfizer, Inc., Groton, CT, USA

XIAO YU (SHIRLEY) WU

Leslie Dan Faculty of Pharmacy, University of Toronto, Toronto, Ontario, Canada

MARY T. AM ENDE

*Process Modeling & Engineering Technology Group, Pfizer, Inc., Groton, CT, USA**

15.1 INTRODUCTION

An oral controlled release drug delivery system is designed to deliver a drug in a controlled and predictable manner over a period of time or at a predetermined position in the gastrointestinal tract. There are several other terms used interchangeably to describe controlled release dosage forms. The Food and Drug Administration defines modified release dosage forms as those whose drug release characteristics of time course and/or location are chosen to accomplish therapeutic or convenience objectives not offered by conventional dosage forms such as a solution or an immediate release dosage form [1]. Modified release oral dosage forms include extended release, i.e. dosage forms designed to make the drug available over an extended period of time after ingestion, and delayed release, i.e. dosage forms designed to provide a delay before drug release. Additionally, terms such as sustained release, prolonged release, pulsatile release, and targeted release have also been used in the literature. Orally disintegrating tablets that are designed to disintegrate more rapidly than an immediate release tablet are also considered to be controlled release dosage forms. They disintegrate on contact with saliva, thus eliminating the need to chew the tablet, swallow an intact tablet, or take the tablet with liquids [2].

Over the past five decades, oral drug delivery systems have matured and currently are a dominant segment of the pharmaceutical market. Oral dosage forms are preferred because of their convenience and cost-effectiveness. Although they were once considered quite exotic, oral controlled release systems have now become commonplace and their advantages accepted both in the development of new molecular entities and in product enhancement. The controlled release market was estimated to be worth over US\$31.68 billion in the United States and \$50.26 billion globally (data from QuintilesIMS MIDAS™, aggregate of four quarters ending 2Q2017). Some top-selling controlled release products in the US market are given in Table 15.1.

There are several reasons for pursuing the development of controlled release dosage forms. Controlled release formulations can reduce the dosing frequency and minimize side effects. Drugs with short biological half-lives (i.e. those where the drug is metabolized or rapidly eliminated from the bloodstream) have to be dosed frequently in order to maintain efficacious levels in the blood. By slowing the rate at which the drug is released, a controlled release dosage form can increase the apparent half-life and maintain efficacious levels for a longer duration, thereby reducing the need for frequent dosing. Reducing the dosing frequency to once daily assures patient convenience and compliance, and a reduction in the peak-to-trough blood concentrations of the drug results

*Current address: Lyndra Therapeutics, Watertown, MA, USA

TABLE 15.1 Some Top-Selling Controlled Release Products in the United States

Name	Drug	Indication	Company	Type of Controlled Release Formulation	US Sales, Aggregate of four Quarters Ending 2Q2017 (\$)
Tecfidera	Dimethyl fumarate	Multiple sclerosis	Biogen Inc.	Delayed release capsules	3.47 B
Oxycontin	Oxycodone	Opioid agonist for pain management	Purdue Pharma		1.80 B
Teva-Methylphenidate ER-C	Methylphenidate	Attention deficit hyperactivity disorder (ADHD)	Teva	Trilayer capsule-shaped tablet with two distinct drug layers and a push layer (osmotic technology) and a drug overcoat layer	1.32 B
Dexilant	Dexlansoprazole	Gastroesophageal reflux disease (GERD)	Takeda	Capsule containing two types of granules with two different release profiles (within an hour and 4–5 hours later)	1.22 B
Namenda	Memantine hydrochloride	Alzheimer's disease	Allergan	Capsule containing coated extended release beads	1.06 B
Metformin Lupin	Metformin hydrochloride	Type 2 diabetes mellitus	Lupin	Polymer-based oral drug delivery to the upper GI tract	1.00 B
Betanis	Mirabegron	Overactive bladder	Astellas	Extended release matrix tablet	0.93 B
Adderall XR	Amphetamine and dextroamphetamine	Attention deficit hyperactivity disorder (ADHD)	Shire Pharmaceuticals	Capsule containing two types of drug-containing beads designed to give a double-pulsed delivery	0.90 B
Ranexa	Ranolazine	Chronic angina	Gilead Sciences	Extended release matrix tablet	0.87 B
Seroquel	Quetiapine fumarate	Atypical antipsychotic	AstraZeneca	Extended release matrix tablet	0.74 B

in a more uniform therapeutic effect and can potentially lead to a lower total dose. Controlled release dosage forms can reduce undesirable side effects that are related to high and rapidly rising drug peak blood levels. In some cases, the undesirable side effects may be related to a local irritation of the upper part of the gastrointestinal tract by the drug. In such cases, a delayed release dosage form can help bypass the upper part of the gastrointestinal tract and reduce the frequency and intensity of these side effects.

Orally disintegrating tablets have the advantage that they can be taken without water. This can be very important to pediatric and geriatric patients and to patients who have difficulty swallowing tablets or capsules. A controlled release dosage form intended to avoid degradation of acid-labile drugs is typically an enteric-coated dosage form (delayed release). The enteric coat prevents drug release in the acidic environment of the stomach, and at a higher intestinal pH, the coating dissolves to enable drug release.

The rapid advance in the field of controlled release occurred because of two main reasons: (i) interdisciplinary teams worked together on novel concepts and designs for drug delivery devices and (ii) advances in many fields that could be related to controlled release. Chemical engineers and the science of chemical engineering played a major role by introducing concepts of mass transfer and drug diffusion through matrices and membranes, material properties of excipients, thermodynamics, and kinetics of drug release. The science of biopharmaceutics provided the understanding of gastrointestinal physiology and its relationship to controlled release dosage forms, both with respect to the transit of dosage forms and the absorption of drug as a function of position in the gastrointestinal tract. It also provided preclinical *in vivo* models such as beagle dogs that led to an increased understanding of the *in vivo* performance of controlled release dosage forms and their *in vitro*–*in vivo* relationships. Polymer science and engineering provided novel materials with a range of properties, which could be tailored to suit a particular application, e.g. polymers that eroded with time, thereby releasing the drug. Advances in the understanding of the pharmacokinetics and pharmacodynamics of drugs allowed controlled release dosage forms to be designed in a rational manner. Finally, advances in manufacturing science and engineering were important, e.g. advances in the ability to manufacture precise laser-drilled orifice in osmotic tablets at rates suitable for commercial production.

This chapter focuses on the design of oral controlled drug release dosage forms. However, the field of controlled release is much broader. It spans other pharmaceutical dosage forms such as long-acting injectables and implants, transdermal patches, ocular devices, and targeted drug delivery systems. Furthermore, controlled release is also used in veterinary applications [3] and diverse fields such as the sustained release of fertilizers, insecticides, herbicides, fragrances, and the food industry [4].

15.2 DEVELOPMENT OF CONTROLLED RELEASE FORMULATIONS IN AN INDUSTRIAL SETTING

The rational development of controlled release formulations in the setting of a large multinational pharmaceutical company with discovery and development operations typically starts with establishing the rationale for modifying the release rate and the desired product profile, i.e. defining the medical need. The next steps involve selection of the dose, delivery duration, and release kinetics based on known or assumed target blood levels. It is highly recommended that prior to initiating a development program, an assessment of the feasibility of developing a controlled release formulation based on the physicochemical and biopharmaceutical properties of the drug candidate [5] be conducted and the most appropriate technology be selected based on the attributes of the technology and manufacturing considerations such as availability of commercial-scale equipment, operator expertise, and prior experience with the technology.

Many compounds fail to become drugs because of their poor physicochemical and/or poor biopharmaceutical properties [6]. The physicochemical properties that have an impact on the feasibility of a controlled release formulation include molecular weight, partition coefficient, solubility, pH–solubility profile, potential for solubilization, salt forms, polymorphs, particle size distribution, and stability. The biopharmaceutical and pharmacokinetic properties that have an impact on the feasibility of a controlled release formulation include gastrointestinal transit of the dosage form, fed/fasted state, permeability, efflux, and extent of gut wall and first-pass metabolism. Good absorption throughout the length of the gastrointestinal tract is important in the successful development of controlled release formulations [7].

15.3 CONTROLLED RELEASE PROFILES AND MECHANISMS

15.3.1 Types of Controlled Release Profiles

Different drug release profiles and release rates may be required based on the pharmacokinetic and pharmacodynamic needs of the medication. Commonly used drug release profiles are illustrated in Figure 15.1 and include (i) zero-order release, (ii) first-order release, (iii) biphasic release, (iv) pulsatile release, and (v) delayed or timed release. The first four release profiles are based on the time dependence of their release rates, while the last is based on the onset of drug release. In the case of zero-order release, the release rate remains constant, reflected by a linear relation between the amount of drug released and time. In the case of first-order release, the release rate decreases exponentially with time. Biphasic, multiphasic, and pulsatile release are typified by

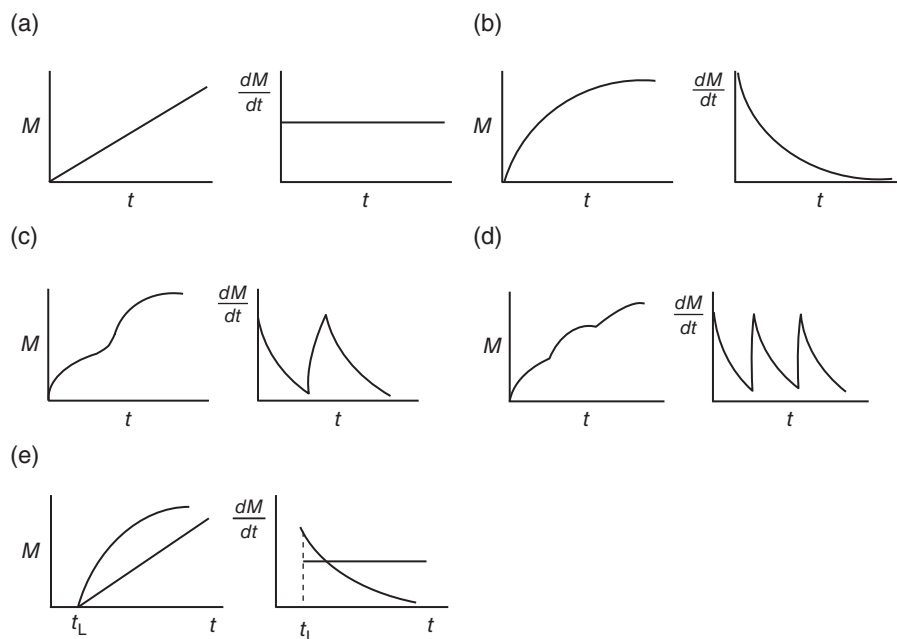


FIGURE 15.1 Schematic illustration of various types of release profiles commonly seen in oral controlled release dosage forms. (a) Zero-order release, (b) first-order release, (c) biphasic release, (d) Pulsatile release, and (e) delayed or timed release.

two or more modes of release. When there is no drug released until a lag time, t_L , the release profile is called delayed (or timed) release. After t_L , the release profile can be in any shape, such as zero-order or first-order release. It should be noted that there may be differences between the *in vitro* and *in vivo* controlled release profiles because conditions of pH, hydrodynamics, fluid volume, and the presence of enzymes and bacteria vary in the gastrointestinal tract.

15.3.2 Controlled Release Mechanisms and Structure of Controlled Release Systems

Various drug release profiles can be obtained by utilizing different drug release mechanisms, device geometry and structure, and materials. The following five major release mechanisms have been utilized alone or in combinations to design oral controlled release dosage forms: diffusion, erosion/ degradation, ion exchange, swelling, and osmotic pressure.

The drug delivery systems are frequently referred to by the mechanism that dominates the drug release rate. Corresponding mathematical models are then derived based on the dominating drug release mechanism, the geometry of the delivery system, and the boundary conditions that the delivery systems are exposed to. The following sections describe the major drug release mechanisms, delivery systems, and their associated release profiles and present essential mathematical equations of analytical or semi-analytical solutions derived from mechanistic models. Interested readers are referred to

specialized books [8–13] and original papers for the derivation of the equations.

15.3.3 Controlled Release via Diffusion

Drug release from a device is considered diffusion controlled when diffusion of drug molecules through the device is the rate-determining step. Depending on the structure of the delivery system, diffusion-controlled systems can be classified as membrane-reservoir (Figure 15.2a) or monolithic (matrix) systems (Figure 15.2b).

15.3.3.1 Membrane-Reservoir Systems In membrane-reservoir systems, there is a drug-rich core (drug reservoir) enclosed by a membrane, which may or may not contain drug initially. Drug diffusion from the reservoir through the membrane is the rate-limiting step. Each delivery system can be made into various geometries. Figure 15.3 shows membrane-reservoir systems of four basic geometries – slab, cylinder, sphere, and disk – which are commonly used for drug delivery. Irrespective of the geometry, membrane-reservoir systems should result in a zero-order release profile as long as the drug core provides a constant drug supply. This is true when excess amount of solid drug loaded in the core and drug dissolution is much faster than drug diffusion through the membrane. In this case a saturated drug solution at the inner side of the membrane is maintained at a constant concentration that normally equals the drug saturation solubility. Once the excess drug is dissolved, the drug core can no longer

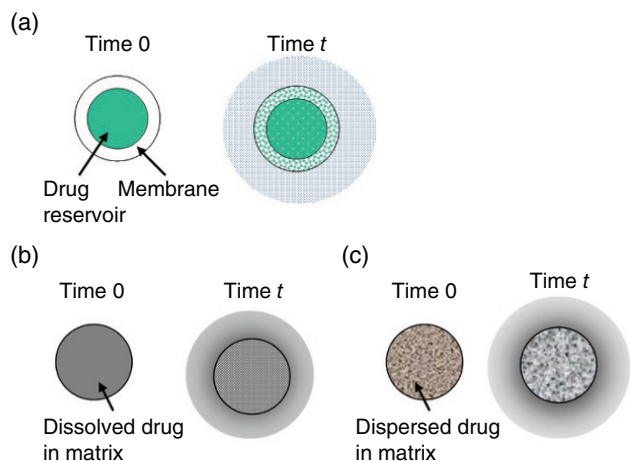


FIGURE 15.2 Schematic illustration of diffusion-controlled systems: (a) membrane-reservoir system, (b) matrix system, and (c) membrane-reservoir system of various geometries.

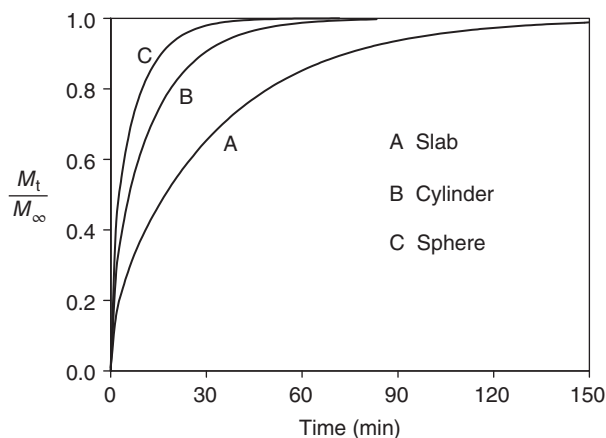


FIGURE 15.4 Comparison of release profiles of fractional drug release from one-dimensional slab, cylinder, and sphere with the same characteristic dimension ($a = 0.2$ cm) into a perfect sink, computed using AP-CAD[®] software.

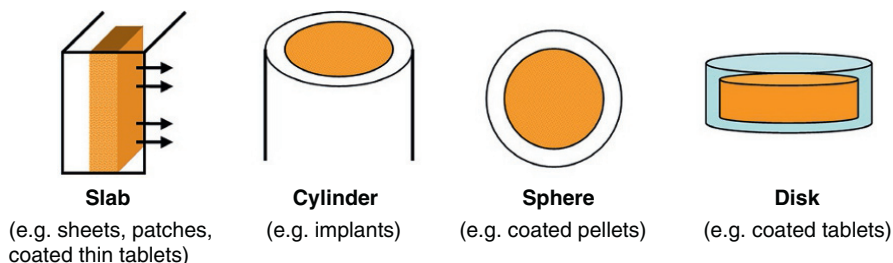


FIGURE 15.3 Membrane-reservoir systems of various geometries.

provide a constant supply, resulting in a decrease in the release rate.

15.3.3.2 Monolithic (Matrix) Systems In the monolithic (or matrix) systems, uniformly distributed drug is released by diffusion through the matrix. Depending on the loading level and drug solubility in the matrix, the drug may exist as a molecular solution (dissolved drug) or a particle dispersion (dispersed drug). The drug loading level (C_0) relative to the drug solubility (C_s) and initial drug distribution in the matrix can influence the release profiles, as do the volume of release medium and stirring condition.

A monolithic system usually provides first-order release profiles because the drug concentration within the matrix decrease with time and the diffusional distance increases with time. The nonlinearity of the release curve increases as the device is changed from slab to cylinder and from cylinder to sphere. Figure 15.4 compares release profiles of fractional drug release from one-dimensional (1-D) slab, cylinder, and sphere with the same characteristic dimension ($a = 0.2$ cm) into a perfect sink, computed using AP-CAD[®] software.

15.3.4 Controlled Release via Erosion or Degradation

Erosion- or degradation-controlled systems are a special case of matrix systems, in which matrix erosion or degradation is the rate-limiting step of drug release. Thus, while the rate of drug release for dissolution-based systems depend mostly on the drug solubility, erosion-based systems limit drug release by dissolution (erosion) or degradation of the materials that form the matrix. Pure erosion/degradation-controlled release is hard to find in oral controlled release dosage forms. Matrix erosion and degradation are often concurrent with other release mechanisms such as drug diffusion and dissolution. Biodegradable polymers typically used in erosion-controlled systems have been reviewed elsewhere [14].

15.3.4.1 Heterogeneous Erosion When the matrix is rigid and hydrophobic with minimal hydration in the release medium, entrapped drug is released mainly by matrix surface erosion, i.e. heterogeneous erosion (Figure 15.5a). If drug solubility in the medium is very low, drug release rate may still be dictated by matrix erosion (solution) even if the matrix

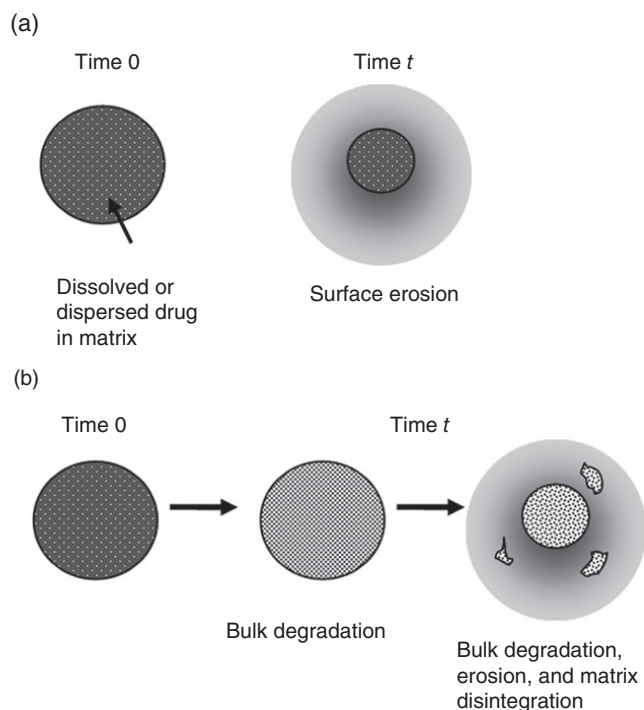


FIGURE 15.5 Schematic illustrations of erosion and degradation-controlled release. (a) Heterogeneous (surface) erosion. (b) Homogeneous (bulk) erosion.

is hydrophilic. The released drug particles may dissolve following release from the matrix. An ideal heterogeneous erosion-controlled system should give a zero-order release if it is a planar shape or a first-order release if it is a cylinder or a sphere. Again, the nonlinearity is higher for sphere than cylinder due to more dramatic reduction in the area toward the center of the device.

15.3.4.2 Homogeneous Degradation and Erosion When the matrix undergoes bulk degradation, the molecular weight of the matrix polymer decreases gradually, resulting in a higher drug diffusion coefficient in the matrix with time. Eventually the matrix may disintegrate and dissolve, releasing the remaining drug. This process is named homogeneous degradation and erosion (Figure 15.5b). Usually a first-order release curve is seen for a period of time, followed by an accelerated release when disintegration of the device occurs. The rate of degradation of conventional biodegradable polymers such as polylactides is too slow to be suitable for oral controlled release dosage forms that are retained in the gastrointestinal tract for a maximum of about 24 hours. Instead, microbially degradable polymers, especially azo-cross-linked polymers, which are degraded specifically by colonic bacteria, have been investigated for colonic drug delivery. The release profile of such delivery system is rather complex and can vary from near zero-order to first-order release.

15.3.5 Controlled Release via Ion Exchange

Ion-exchange resins, initially developed for water treatment, have been used in pharmaceutical dosage forms. Their applications include taste masking, sustained release, and gastric retention. They are designated either cationic or anionic based on the counterions. Cationic ion-exchange resins are composed of anionic groups such as $-\text{COO}^-$ and $-\text{SO}_3^-$ groups, while anionic exchange resins contain $-\text{NR}_2^+$ or $-\text{NR}_3^+$ groups.

When an ionic drug is loaded into a matrix, e.g. an ion-exchange resin or a polyelectrolyte with charges of opposite sign, its release from the matrix is normally controlled both by ion exchange and by diffusion because the release process involves several essential steps: (i) diffusion of counterions from release medium into the matrix, (ii) exchange of counterions with bound drug molecules in the matrix, and (iii) diffusion of free drug molecules out from the matrix into the medium.

These steps are depicted in Figure 15.6. In the case of hydrophobic polyelectrolytes, matrix swelling may also play a role in the release kinetics. In general, a first-order release profile is seen in ion-exchange-controlled delivery systems.

15.3.6 Controlled Release via Swelling

When a swellable glassy polymer matrix is placed in a thermodynamically compatible solvent, it undergoes an abrupt transition from the glassy state to the rubbery state. Because the polymer chains at the glassy state are rigid, drug diffusion in the glassy region is negligible as compared with that in the rubbery region. Pharmaceutical dosage forms are usually made from swellable hydrophilic polymers such as hydroxypropyl methylcellulose and polyethylene oxide. When a matrix tablet made from such polymer and loaded with drug is introduced into an aqueous medium, water penetrates into the matrix, wets the polymer and drug particles therein, and fills the pores. In the hydrated layer, drug particles start to dissolve and drug molecules diffuse out from the wetted zone, which has a boundary named the diffusion front (see Figure 15.7). The hydrated polymers chains gradually relax and disentangle, forming a gel layer. Drug diffusion in the gel layer is much faster than in the dry glassy core and in the slightly hydrated layer as well. In contrast to pure diffusion-controlled hydrophobic matrix drug system with little volume change during release, hydrophilic polymers undergo the glassy-rubbery transition and absorb large amounts of water due to osmotic pressure. As a result the volume of the device increases, so does the drug diffusion coefficient in the rubbery zone and matrix porosity if high quantities of water-soluble additives are added or in the case of high initial drug loadings. A matrix drug device is classified as being swelling controlled if the change in polymer morphology by interaction with the external release medium controls or alters

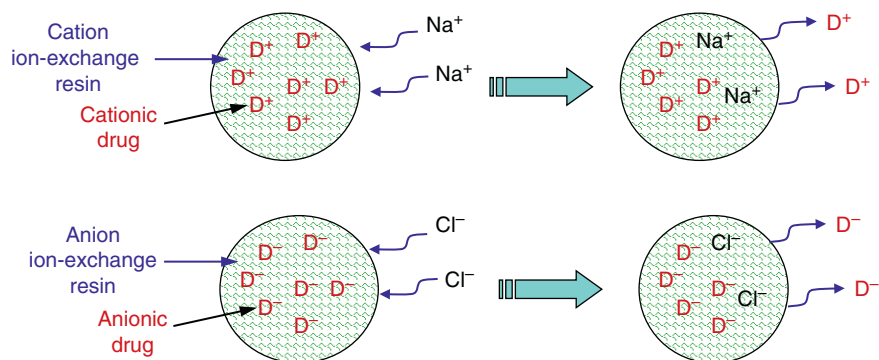


FIGURE 15.6 Schematic of ionic drug release from ion exchange resin that involves steps of counterion diffusion into the resin, exchange with bound drug, and diffusion of dissociated drug out of the resin.

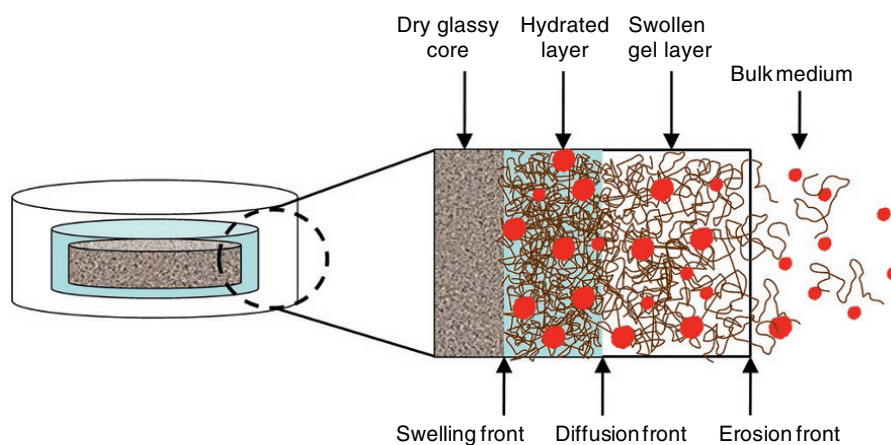


FIGURE 15.7 Illustration of drug release from a hydrophilic matrix tablet by swelling mechanism.

the drug release rate. Note that non-cross-linked hydrophilic polymers may dissolve before all payload is released, which is often seen in pharmaceutical hydrophilic matrices.

15.3.7 Controlled Release via Osmotic Pressure

Osmotic controlled release of drug molecules involves the regulation of osmotic permeation of water through the use of a semipermeable membrane. The diffusion of water across the semipermeable membrane induced by an existing chemical potential gradient between the dissolution medium and the tablet core creates a hydrostatic pressure. The hydrostatic pressure generated by the influx of water forces the release of a saturated solution of the drug through delivery ports in the device. In addition to the mechanism of osmotic pumping, drug release can also take place through the membrane as a result of the mechanism of solution–diffusion. Since the device volume is constant, the volume of drug solution delivered will be equal to the volume of osmotic water uptake within a given time interval. Therefore, the rate of drug delivery will be constant as long as a constant osmotic pressure

gradient is maintained across the membrane. Prolonged zero-order release can then be achieved with this system. However, as the reservoir concentration falls below saturation, the rate declines asymptotically. It is also conceivable for osmotic systems to achieve release rates much higher than systems that solely involve solution–diffusion mechanism.

Osmotic devices can be manually activated or self-activated. Manually activated devices have to be stored empty and loaded with water prior to use. Other versions have an impermeable seal between the semipermeable membrane and the water chamber, allowing the devices to be stored fully loaded with water. The osmotic pump then becomes activated when the seal is broken. Self-activated devices are activated by water imbibed from the gastrointestinal tract or the dissolution vessel medium driven by the device itself.

15.3.7.1 Rose–Nelson Pump The Rose–Nelson pump [15] shown in Figure 15.8 consists of a drug chamber, a salt chamber containing excess solid salt, and a water chamber. The drug and water chambers are separated by a rigid semipermeable membrane. Water moves from the water chamber

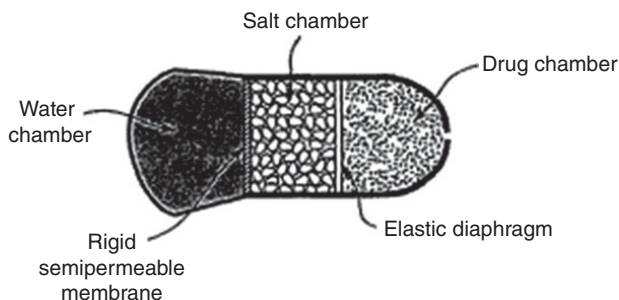


FIGURE 15.8 The three-chamber Rose–Nelson osmotic pump. *Source:* Reprinted from Santus and Baker [15], Copyright (1995), with permission from Elsevier.

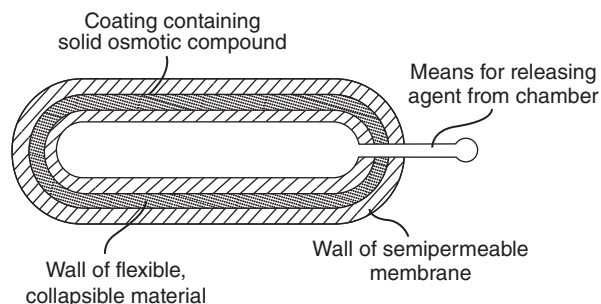


FIGURE 15.10 The Higuchi–Theuwes pump. *Source:* Reprinted from Santus and Baker [15], Copyright (1995), with permission from Elsevier.

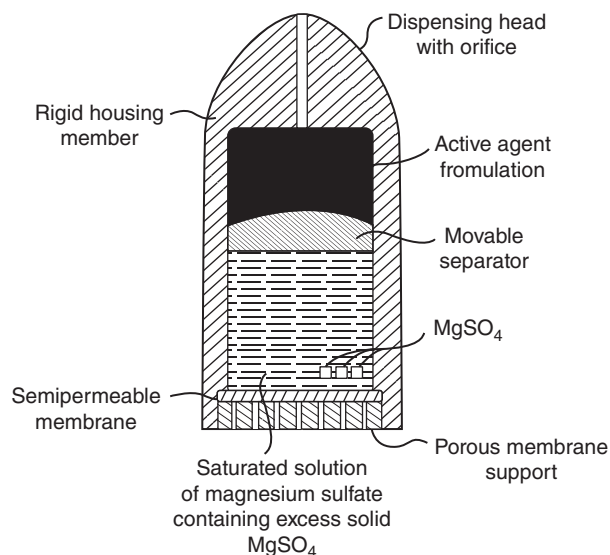


FIGURE 15.9 The Higuchi–Leeper pump. *Source:* Reprinted from Santus and Baker [15], Copyright (1995), with permission from Elsevier.

into the salt chamber as a result of the difference in osmotic pressure across the membrane. The increase in volume of the salt chamber as a result of water uptake moves the piston and causes drug to be pumped out of the device.

15.3.7.2 Higuchi–Leeper Pump The Higuchi–Leeper pump differs from the Rose–Nelson device in that the water chamber is absent (Figure 15.9). The Higuchi–Leeper pump usually consists of a salt chamber that contains a fluid solution with excess solid and a rigid housing with the semipermeable membrane supported on a perforated frame.

15.3.7.3 Higuchi–Theuwes Pump The semipermeable membrane in the Higuchi–Theuwes pump acts as the outer casing of the pump. As shown in Figure 15.10, the pump is composed of a rigid rate-controlling outer semipermeable membrane surrounding a solid layer of salt coated on the

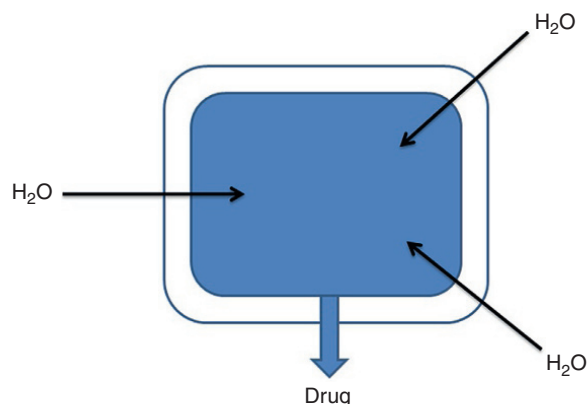


FIGURE 15.11 The elementary osmotic pump.

inside by an elastic diaphragm and on the outside by the membrane. During its operation, water is osmotically drawn by the salt through the semipermeable membrane. This water increases the volume of the salt chamber, forcing the drug release from the chamber.

15.3.7.4 Elementary Osmotic Pump In these systems, a semipermeable membrane with a delivery orifice surrounds an osmotic core that contains the drug. The delivery rate from these devices is regulated by the osmotic pressure of the osmotic agent of the core formulation and by the water permeability of the semipermeable membrane (Figure 15.11). For example, the OROS system developed by ALZA Corporation is used to deliver Acutrim, an over-the-counter appetite suppressant, at a controlled rate [16]. Similarly, Elan Corporation of Ireland has developed MODAS (multidirectional oral absorption system). This system differs from OROS in that it has a multitude of small pores through which the drug can exit.

15.3.7.5 Push–Pull Osmotic Pump The OROS system described in the previous section is somewhat limited because it can only deliver drugs with good aqueous

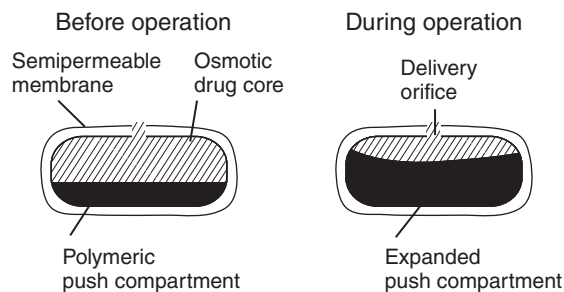


FIGURE 15.12 Illustration of the push-pull osmotic pump known as the gastrointestinal therapeutic systems (GITS). *Source:* Reprinted from Swanson et al. [18], Copyright (1987), with permission from Elsevier.

solubility. The push-pull osmotic pump, which delivers a suspension of drug, was an advancement over the elementary osmotic pump because it could be used for the delivery of low solubility drugs and it could be manufactured using conventional pharmaceutical equipment [17]. For this system, the core consisted of a bilayer tablet with one layer containing a swelling agent and the other layer containing the drug formulation. The swelling agent functioned to push a suspension of drug from the orifice (Figure 15.12). ALZA Corporation, which was acquired by Johnson & Johnson in 2001, developed the gastrointestinal therapeutic system (GITS) using this dosage form to deliver nifedipine to provide once-a-day dosing for hypertension [18].

15.3.7.6 Semipermeable Membranes Containing Micropores Innovations in the osmotic drug delivery field continued in the 1990s by the development of the controlled porosity osmotic pump tablet (CP-OPT) by Zentner and others [19–21]. The main advancement of the CP-OPT compared with the OROS[®] system was the new design of the semipermeable membrane to contain pores sufficient in size to eliminate the need for laser drilling an orifice. The CP-OPT membrane also contains a pore former and plasticizer. This osmotic dosage form is designed to deliver a drug solution by an osmotic mechanism, therefore limited in application to soluble compounds. In the late 1990s, Okimoto and Stella [22–26] advanced the CP-OPT technology to encompass poorly water-soluble compounds that could be solubilized by Captisol[™] (sulfobutyl ether- β -cyclodextrin or (SBE)_{7m}- β -CD) that serves as both a solubility-enhancing agent and an osmotic agent. The use of (SBE)_{7m}- β -CD enabled the osmotic release from CP-OPT of low solubility drug such as prednisolone, chlorpromazine, and testosterone.

The asymmetric membrane (AM) film-coated tablet is a unique embodiment within the field of osmotic drug delivery. The membrane is formed by a phase inversion process and is composed of several layers of polymer with a network of interconnecting pores [27]. The polymer acts as a

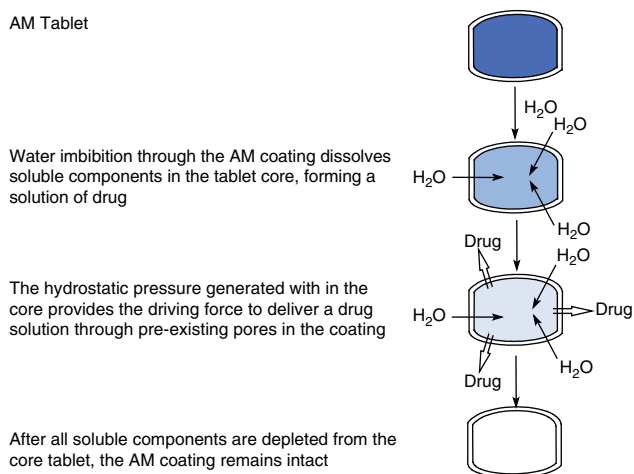


FIGURE 15.13 Semipermeable membrane containing micropores.

semipermeable barrier, while the interconnected pores provide a path for dissolved core components to exit. A laser-drilled orifice is not necessary in the AM system as required for the OROS technology and similar to the CP-OPT. In fact, the entire AM film coating acts as hundreds of preformed delivery orifices. Therefore, the drug release can be adjusted by varying the type and concentration of the pore former present in the semipermeable membrane as well as the membrane thickness [28]. Unlike the CP-OPT, in the AM tablet design, the porous semipermeable membrane contains polyethylene glycol in a dual role, serving as plasticizer and pore former. The holes through which drug is released are pores created in the tablet coating as a result of the method of coating and polymer solution used or occur when the water-soluble component of the tablet coating is leached out after the tablet is swallowed [15]. As with Theeuwes's elementary osmotic pump, a porous membrane surrounds an osmotic core containing the drug. It has been demonstrated that the mechanism of drug release from spherical beads consisting mainly of phenylpropanolamine hydrochloride and sucrose that were coated with a porous ethyl cellulose film is predominantly osmotic, irrespective of film porosity [29]. It has been shown that high water fluxes can be achieved with AM tablets [27, 28]. The asymmetric coating consists of a porous substrate with a thin outer skin. The high water fluxes from these asymmetric coatings permit the osmotic delivery of drugs with lower solubilities [28] (Figure 15.13).

15.3.7.7 Polymer Drug Matrix Systems Polymer drug matrix systems are composed of polymer-encapsulated drug particles dispersed within a polymer matrix (Figure 15.14). Several researchers have postulated different phenomena accounting for drug release. For example, Wright et al. [31] have postulated that drug release occurs as soon as water drawn osmotically in through the encapsulating polymer

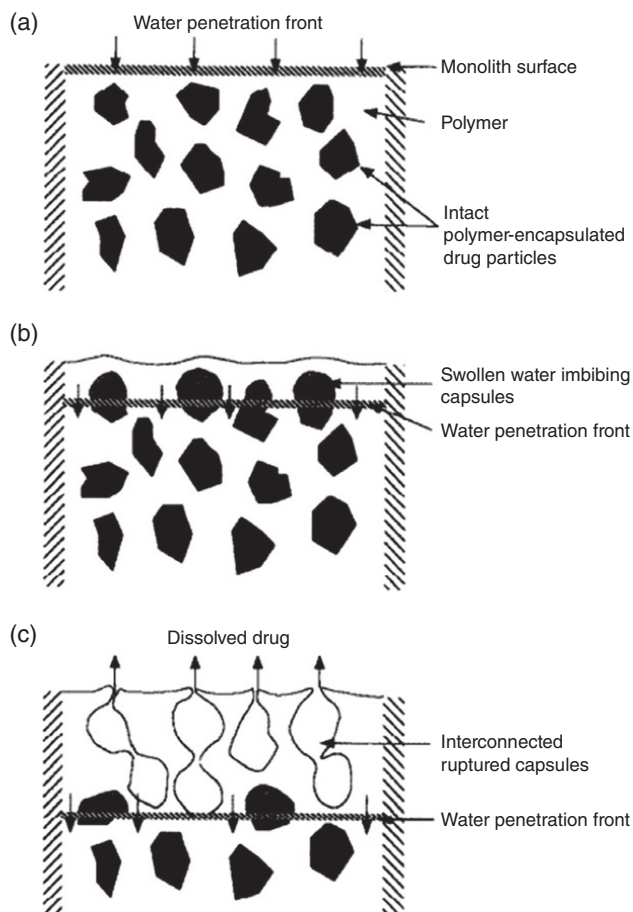


FIGURE 15.14 Polymer matrix system: (a) the diffusion of water to first layer of encapsulated particles, (b) water imbibition into encapsulated particles, and (c) zone of interconnected capsules, imbibing capsules, and intact capsules. *Source:* Reprinted from Amsden et al. [30], Copyright (1994), with permission from Elsevier.

causes the coating to rupture. An osmotic pressure gradient is then believed to pump the dissolved drug to the surface through fractures created via interconnected pores. In other words, after rupturing, osmotic pressure-driven convection is believed to be responsible for the release of the remaining solid material in the capsule. According to Amsden et al. [30], release by diffusion is the most likely phenomenon responsible for the release after the capsules rupture.

15.4 MATHEMATICAL EQUATIONS FOR DRUG RELEASE FROM CONTROLLED RELEASE DOSAGE FORMS

15.4.1 Diffusion-Controlled Systems

Diffusion-controlled systems can be described by Fick's second law. The general governing equation of release kinetics for 1-D release is

$$\frac{\partial C}{\partial t} = \frac{1}{x^\alpha} \left[\frac{\partial}{\partial x} \left(x^\alpha D \frac{\partial C}{\partial x} \right) \right] \quad (15.1)$$

where

$\alpha = 0$ for slab, $\alpha = 1$ for cylinder, and $\alpha = 2$ for sphere.

D is the drug diffusion coefficient in the device.

c is the drug concentration as a function of time t and distance x .

The 1-D release model is applicable to infinite large slab or infinite long cylinder, where drug release from the edge of the slab or the ends of the cylinder is negligible.

For general multidimensional problems, the governing equation is

$$\frac{\partial C}{\partial t} = D \left[\frac{\partial^2 C}{\partial x^2} + \frac{\partial^2 C}{\partial y^2} + \frac{\partial^2 C}{\partial z^2} \right] \quad (15.2)$$

With appropriate initial and boundary conditions, the partial differential equations can be solved analytically or numerically. Up to date, only handful explicit exact solutions, and approximate analytical solutions for simple geometries have been obtained. The final expressions are presented below without detailed derivation. Interested readers can find procedures of derivation in the cited references. For complex delivery systems with two- or three-dimensional (2-D or 3-D) release, numerical methods, such as finite element method [32–34], finite difference method [35, 36], and Monte Carlo method can be employed, which will not be elaborated here.

15.4.2 Membrane-Reservoir Systems

15.4.2.1 Exact Solution for 1-D Slab with Constant Reservoir in a Sink [9, 37]

$$M = \frac{DKC_r t}{\delta} + \frac{2KC_r \delta}{\pi^2} \sum_{n=1}^{\infty} \frac{(-1)^n}{n^2} \left[1 - \exp\left(\frac{-Dn^2 \pi^2 t}{\delta^2}\right) \right] + \frac{4C_m \delta}{\pi^2} \sum_{m=0}^{\infty} \frac{1 - \exp\left(\frac{-D(2m+1)^2 \pi^2 t}{\delta^2}\right)}{(2m+1)^2} \quad (15.3)$$

where

M is the cumulative amount of drug released from unit area.

δ is the membrane thickness.

C_r is the drug concentration in the reservoir, which is normally taken as the drug solubility in the presence of excess dispersed drug.

C_m is the initial drug concentration in the membrane.
 K is the partition coefficient of drug between the membrane and the reservoir.
 D is the drug diffusion coefficient in the membrane.

This equation predicts drug released from unit area until all dispersed drug is exhausted. By letting $C_m = 0$ and $t \rightarrow \infty$ in Eq. (15.3), drug released at steady state and time lag, t_l , can be obtained as

$$M = \frac{DKC_r}{\delta}(t - t_l) \tag{15.4}$$

Similarly by letting $C_m = C_r$ and $t \rightarrow \infty$, the initial burst time t_b is found from

$$M = \frac{DKC_r}{\delta}(t + t_b) \tag{15.5}$$

where

$$t_l = \delta^2/(6D). \\ t_b = \delta^2/(3D).$$

15.4.2.2 Analytical Solution for 1-D Slab with Nonconstant Reservoir in a Sink [37] After all solid drug is dissolved, the drug concentration in the reservoir decreases with time. The amount of drug released from unit area is then described by

$$M = C_s V \left[1 - \exp\left(-\frac{DK}{\delta V} t\right) \right] \tag{15.6}$$

where

V is the volume of the reservoir with unit area.

This equation is based on pseudo-steady-state assumption and mass balance.

Using Eq. (15.3) for time up to t^* , the time at which all dispersed drug is depleted, and using Eq. (15.6) after t^* , one can obtain a release profile covering the entire release course from constant reservoir to nonconstant reservoir. To find t^* let $M = (C_r - C_s)V$, substitute it into the left-hand side of Eq. (15.3), and then solve for t^* .

15.4.2.3 Exact Solution for 1-D Cylinder with Constant Reservoir in a Sink [37]

$$M = \frac{2\pi KC_r D t}{\ln\left(\frac{b}{a}\right)} + 4\pi \sum_{n=1}^{\infty} \left(\frac{KC_r J_0(b\alpha_n)}{J_0(a\alpha_n) - J_0(b\alpha_n)} + C_m \right) \frac{J_0(a\alpha_n) [1 - \exp(-D\alpha_n^2 t)]}{\alpha_n^2 [J_0(a\alpha_n) + J_0(b\alpha_n)]} \tag{15.7}$$

where

M is the cumulative amount of drug released from a cylinder of constant reservoir through a membrane of unit length.

a and b are the internal and external radius of the cylindrical membrane, respectively, which defines the membrane thickness $\delta = b - a$.

α_n 's are the positive roots of $J_0(a\alpha_n)Y_0(b\alpha_n) - J_0(b\alpha_n)Y_0(a\alpha_n)$.

J_0 and Y_0 are Bessel function of the first and the second kind of order zero.

Similar to slab case the corresponding steady-state drugs released with lag time or initial burst are

$$M = \frac{2\pi DKC_r}{\ln\left(\frac{b}{a}\right)}(t - t_l) \tag{15.8}$$

$$t_l = \frac{2\ln\left(\frac{b}{a}\right)}{D} \sum_{n=1}^{\infty} \frac{J_0(a\alpha_n)J_0(b\alpha_n)}{\alpha_n^2 [J_0^2(b\alpha_n) - J_0^2(a\alpha_n)]} \tag{15.9}$$

$$M = \frac{2\pi DKC_r}{\ln\left(\frac{b}{a}\right)}(t + t_b) \tag{15.10}$$

$$t_b = \frac{2\ln\left(\frac{b}{a}\right)}{D} \sum_{n=1}^{\infty} \frac{J_0^2(a\alpha_n)}{\alpha_n^2 [J_0^2(b\alpha_n) - J_0^2(a\alpha_n)]} \tag{15.11}$$

15.4.2.4 Analytical Solution for 1-D Cylinder with Nonconstant Reservoir in a Sink [37]

$$M = C_s V \left[1 - \exp\left(-\frac{2DK}{a^2 \ln\left(\frac{b}{a}\right)} t\right) \right] \tag{15.12}$$

where

V is the volume of the cylindrical reservoir with unit length.

15.4.2.5 Exact Solution for Sphere with Constant Reservoir in a Sink [37]

$$M = \frac{4\pi ab DK C_r t}{\delta} + \frac{8ab\delta(KC_r - C_m)}{\pi} \sum_{n=1}^{\infty} \frac{(-1)^n}{n^2} \left[1 - \exp\left(\frac{-Dn^2\pi^2 t}{\delta^2}\right) \right] + \frac{8b^2\delta C_m}{\pi} \sum_{n=0}^{\infty} \frac{1}{n^2} \left[1 - \exp\left(\frac{-Dn^2\pi^2 t}{\delta^2}\right) \right] \tag{15.13}$$

where

a and b are the internal and external radius of the spherical membrane, respectively, which defines the membrane thickness $\delta = b - a$.

This equation describes the drug release from constant reservoir through a spherical membrane. Similar to slab case the corresponding steady-state drugs released with lag time or initial burst are

$$M = \frac{4\pi abDKC_r}{\delta}(t - t_1) \quad (15.14)$$

$$M = \frac{4\pi abDKC_r}{\delta}(t + t_b) \quad (15.15)$$

where

$$t_1 = \delta^2/(6D).$$

$$t_b = b\delta^2/(3aD).$$

15.4.2.6 Analytical Solution for Sphere with Nonconstant Reservoir in a Sink [37]

$$M = C_s V \left[1 - \exp\left(-\frac{3bDK}{\delta a^2}t\right) \right] \quad (15.16)$$

where

V is the volume of the spherical reservoir.

15.4.2.7 Analytical Solution for 2-D Tablet with Constant Reservoir in a Sink [38]

$$M = 2 \left[\pi a^2 M_a + (H - \delta_a) M_r + M_c \right] \quad (15.17)$$

where M is the cumulative amount released from the axial direction and radial direction, M_a and M_r , are given in Eq. (15.3) and Eq. (15.7), respectively, and M_c is expressed below:

$$M_c = \frac{\pi K C_r t}{\ln\left(\frac{a}{b}\right)} \left\{ \frac{D_a}{\delta_a} \left[\frac{a^2 - b^2}{2} - a^2 \ln\left(\frac{a}{b}\right) \right] - D_r \delta_a \right\} \quad (15.18)$$

where

a and b are internal and external radius of the tablet.

δ_a is axial membrane thickness, the radial membrane thickness $\delta_r = b - a$.

H is the half-thickness of the tablet.

D_a and D_r are the drug diffusion coefficients in the axial and the radial directions, respectively.

The parameters for radial and axial directions can be identical for symmetric coating or different for asymmetric

coating. The amount of drug released at the steady state with time lag or initial burst is

$$M = 2\pi K C_r \left(\frac{a^2 D_a}{\delta_a} + \frac{2(H - \delta_a) D_r}{\ln\left(\frac{b}{a}\right)} \right) (t - t_1) \quad (15.19)$$

$$t_1 = \frac{\frac{a^2 h}{6} + 4(H - \delta_a) \sum_{n=1}^{\infty} \frac{J_0(a\alpha_n) J_0(b\alpha_n)}{\alpha_n^2 [J_0^2(b\alpha_n) - J_0^2(a\alpha_n)]}}{\frac{a^2 D_a}{\delta_a} + \frac{2(H - \delta_a) D_r}{\ln\left(\frac{b}{a}\right)}} \quad (15.20)$$

$$M = 2\pi K C_r \left(\frac{a^2 D_a}{\delta_a} + \frac{2(H - \delta_a) D_r}{\ln\left(\frac{b}{a}\right)} \right) (t + t_b) \quad (15.21)$$

$$t_b = \frac{\frac{a^2 h}{3} + 4(H - \delta_a) \sum_{n=1}^{\infty} \frac{J_0^2(a\alpha_n)}{\alpha_n^2 [J_0^2(a\alpha_n) - J_0^2(b\alpha_n)]}}{\frac{a^2 D_a}{\delta_a} + \frac{2(H - \delta_a) D_r}{\ln\left(\frac{b}{a}\right)}} \quad (15.22)$$

15.4.2.8 Analytical Solution for 2-D Tablet with Nonconstant Reservoir in a Sink [38]

$$M = C_s V \left\{ 1 - \exp \left[\frac{-2KD_r}{a^2 \ln\frac{b}{a}} - \frac{KD_a}{h(H - \delta_a)} - \frac{\pi K}{V} \left(\frac{D_a}{\delta_a} \left(\frac{a^2 - b^2}{2 \ln\left(\frac{a}{b}\right)} - a^2 \right) - \frac{D_r \delta_a}{\ln\left(\frac{a}{b}\right)} \right) \right] t \right\} \quad (15.23)$$

where

V is the volume of the tablet reservoir.

Although the equations for describing drug release kinetics for membrane-reservoir systems are presented above separately for constant and nonconstant reservoir, the total amount of drug released during the entire course can be combined seamlessly based on mass balance as outlined by Zhou et al. [38].

15.4.3 Monolithic (Matrix) Systems Containing Dissolved Drug ($C_0 \leq C_s$)

15.4.3.1 Exact Solution for 1-D Slab with Dissolved Drug ($C_0 \leq C_s$) in a Sink [9]

$$\frac{M_t}{M_0} = 1 - \sum_{n=0}^{\infty} \frac{8}{(2n+1)^2 \pi^2} \exp \left[\frac{-D(2n+1)^2 \pi^2 t}{4l^2} \right] \quad (15.24)$$

where

l is the half-thickness of the slab, $M_0 = 2C_0l$.
 C_0 is the initial drug loading.
 D is the drug diffusion coefficient.

If l is defined as the thickness of the slab, the equation is

$$\frac{M_t}{M_0} = 1 - \sum_{n=0}^{\infty} \frac{8}{(2n+1)^2 \pi^2} \exp\left[\frac{-D(2n+1)^2 \pi^2 t}{l^2}\right] \quad (15.25)$$

15.4.3.2 Exact Solution for 1-D Slab with Dissolved Drug ($C_0 \leq C_s$) in a Well-Stirred Finite Volume [9]

$$\frac{M_t}{M_\infty} = 1 - \sum_{n=1}^{\infty} \frac{2\alpha(1+\alpha)}{1+\alpha+\alpha^2 q_n^2} \exp\left(\frac{-D q_n^2 t}{l^2}\right) \quad (15.26)$$

where

q_n 's are the nonzero positive roots of $\tan q_n = -\alpha q_n$ and effective volume ratio, $\alpha = L/Kl$.
 K is the partition factor between solute in the slab in equilibrium with that in the solution. L is the thickness of the external volume, on one side of the slab, excluding the space occupied by the half-thickness of the slab.

Note that in a finite volume, the amount of drug released at equilibrium, M_∞ , may be smaller than the initial payload, M_0 , because the external medium may be saturated by the released drug before all payload is released. Should the saturation occur, the time for M_t/M_∞ to reach unity is shorter than M_t/M_0 , and the difference between these two increases as the effective volume ratio decreases. This phenomenon is seen in all geometries [32–34].

15.4.3.3 Exact Solution for 1-D Cylinder with Dissolved Drug ($C_0 \leq C_s$) in a Sink [9]

$$\frac{M_t}{M_0} = 1 - \sum_{n=1}^{\infty} \frac{4}{R^2 q_n^2} \exp(-Dq_n^2 t) \quad (15.27)$$

where

q_n 's are the roots of $J_0(Rq_n) = 0$.
 $J_0(x)$ is the Bessel function of the first kind of order zero.
 R is the radius of the cylinder.
 $M_0 = C_0\pi R^2$.

15.4.3.4 Exact Solution for 1-D Cylinder with Dissolved Drug ($C_0 \leq C_s$) in a Well-Stirred Finite Volume [9]

$$\frac{M_t}{M_\infty} = 1 - \sum_{n=1}^{\infty} \frac{4\alpha(1+\alpha)}{4+4\alpha+\alpha^2 q_n^2} \exp\left(\frac{-D q_n^2 t}{R^2}\right) \quad (15.28)$$

where

q_n 's are the nonzero positive roots of $\alpha q_n J_0(q_n) + 2J_1(q_n) = 0$.
 $J_1(x)$ is the Bessel function of the first order, $\alpha = \frac{V}{(\pi R^2 K)}$.
 K is the partition factor between solute in the cylinder and in the medium at equilibrium.
 V is the external volume excluding the space occupied by the cylinder.

15.4.3.5 Exact Solution for Sphere with Dissolved Drug ($C_0 \leq C_s$) in a Sink [9]

$$\frac{M_t}{M_0} = 1 - \frac{6}{\pi^2} \sum_{n=1}^{\infty} \frac{1}{n^2} \exp\left(\frac{-Dn^2 \pi^2 t}{R^2}\right) \quad (15.29)$$

where

R is the radius of the sphere.
 $M_0 = 4\pi R^3 C_0/3$.

15.4.3.6 Exact Solution for Sphere with Dissolved Drug ($C_0 \leq C_s$) in a Well-Stirred Finite Volume [9]

$$\frac{M_t}{M_\infty} = 1 - \sum_{n=1}^{\infty} \frac{6\alpha(1+\alpha)}{9+9\alpha+\alpha^2 q_n^2} \exp\left(\frac{-D q_n^2 t}{R^2}\right) \quad (15.30)$$

where

q_n 's are the nonzero positive roots of $\tan q_n = 3q_n/(3+\alpha q_n^2)$ and $\alpha = 3V/(4\pi R^3 K)$.
 K is the partition factor between solute in the sphere in equilibrium with that in the solution. V is the external volume excluding the space occupied by the sphere.

15.4.3.7 Exact Solution for 2-D Tablet with Dissolved Drug ($C_0 \leq C_s$) in a Sink [39] Considering drug release from all surfaces of a matrix tablet with symmetric properties, Fu et al. derived an exact solution:

$$\frac{M_t}{M_0} = 1 - \frac{8}{H^2 R^2} \sum_{m=1}^{\infty} \exp(-D\alpha_m^2 t) (\alpha_m^{-2}) \sum_{n=0}^{\infty} \exp(-D\beta_n^2 t) (\beta_n^{-2}) \quad (15.31)$$

where

α_m 's are the roots of $J_0(R\alpha) = 0$.

$\beta_n = (2n + 1)\pi/2H$.

R is the radius.

H is the half-thickness of a tablet.

15.4.4 Monolithic (Matrix) Systems Containing Dissolved Drug ($C_0 > C_s$)

15.4.4.1 Exact Solution for 1-D Slab with Dispersed Solute ($C_0 > C_s$) in a Sink [40]

$$M_t = \frac{2C_s}{\text{erf}(\beta^*)} \sqrt{\frac{Dt}{\pi}} \quad (15.32)$$

β^* can be found from the following equation:

$$\sqrt{\pi}\beta^* \exp(\beta^{*2}) \text{erf}(\beta^*) = \frac{C_s}{C_0 - C_s} \quad (15.33)$$

where

C_0 is the initial drug concentration.

C_s is drug solubility in the matrix.

This equation predicts a linear plot of M_t vs. \sqrt{t} , i.e. a square-root relationship. It is only applicable when the excess dispersed drug is present.

15.4.4.2 Analytical Solution for 1-D Slab with Dispersed Solute ($C_0 > C_s$) in a Sink [41]

$$M_t = \sqrt{DC_s(2C_0 - C_s)t} \quad (15.34)$$

This solution calculates drug released from unit area based on pseudo-steady-state assumption and is generally applicable for $C_0 > 3C_s$.

15.4.4.3 Analytical Solution for 1-D Cylinder with Dispersed Solute ($C_0 > C_s$) in a Sink [42]

$$M_t = \pi C_0 (R_0^2 - r^2) \quad (15.35)$$

$$\frac{r^2}{2} \ln\left(\frac{r}{R_0}\right) + \frac{R_0^2 - r^2}{4} = \frac{C_s Dt}{C_0} \quad (15.36)$$

where

R_0 is the radius of the cylinder.

r is the moving front of dispersed drug.

For a given series of r such as r_1, r_2, \dots, r_n , solve for $M_{t1}, M_{t2}, \dots, M_{tn}$ from Eq. (15.35) and t_1, t_2, \dots, t_n from Eq. (15.36), and then correlate M_t and t to get a release

profile. This solution is based on pseudo-steady-state assumption and is generally applicable for $C_0 > 3C_s$.

15.4.4.4 Analytical Solution for a Sphere with Dispersed Solute ($C_0 > C_s$) in a Sink [41]

$$M_t = \frac{4}{3}\pi R_0^3 C_0 - 4\pi \left[\frac{r^3 C_0}{3} + \frac{C_s r}{6} (R_0^2 + rR_0 - 2r^2) \right] \quad (15.37)$$

$$t = \frac{1}{6DC_s R_0} \left[C_0 (R_0^3 + 2r^3 - 3R_0 r^2) + C_s \left(4r^2 R_0 + R_0^3 \ln \frac{R_0}{r} - R_0^3 - R_0^2 r - 2r^3 \right) \right] \quad (15.38)$$

where

R_0 is the radius of the sphere.

r is dispersed drug moving front.

Using the same approach given in the case of 1-D cylinder, a correlation between M_t and t is obtained. This solution is based on pseudo-steady-state assumption and is generally applicable for $C_0 > 3C_s$.

15.4.4.5 Analytical Solution for a 2-D Tablet with Dispersed Solute ($C_0 > C_s$) in a Sink [43]

$$M_t = 2C_0\pi [H(R^2 - r^2) + zr^2] \quad (15.39)$$

where

R is radius.

H is half-thickness.

C_0 is initial drug loading.

r and z are the moving front of dispersed drug in the radial and the axial direction, respectively.

They are given as follows:

$$z = \sqrt{\frac{2D_a}{D_r} \left[\frac{r^2}{2} \ln \frac{r}{R} + \frac{R^2 - r^2}{4} \right]} \quad (15.40)$$

$$t(r) = \frac{C_0}{C_s D_r} \left[\frac{r^2}{2} \ln \frac{r}{R} + \frac{R^2 - r^2}{4} \right] \quad (15.41)$$

For a given series of r such as $r = 0, r_1, \dots, R$, solve for corresponding t and z from Eqs. (15.40) and (15.41), and then substitute r and z into Eq. (15.39) to calculate M_t and correlate M_t with t . This solution is based on pseudo-steady-state assumption and is generally applicable for $C_0 \gg C_s$.

15.4.5 Assumptions, Applications, and Implementations of Models for Diffusion-Controlled Systems

Certain important assumptions were used in the derivation of the models presented above for membrane-reservoir and matrix systems such as dissolution much faster than diffusion, constant material properties, and no dimensional change during the complete release process. If these assumptions can be justified for a given delivery system and release process, the mechanistic models for diffusion-controlled release can be applied for prediction of release kinetics, sensitivity tests of formulation variables, parameter identification of dosage forms, and *in vitro*–*in vivo* correlation. Compared with regression models such as $M_t = kt^n$, mechanistic models can reveal more information about the effects of important formulation variables on drug release kinetics, such as dimension (R , l , δ), geometry, material properties (D , C_s , K), and initial loading (C_0 , C_m). It is noticed that transcendental expression and nonlinear equations are involved in the mechanistic models, which is cumbersome for daily usage. However several computer software packages for dosage form design such as AP-CAD and Simulation Plus have been developed to implement the computation tasks.

Mathematical models aforementioned for diffusion-controlled systems describe the general trends of the drug release process. However, one major assumption made is that drug dissolution is much faster than drug diffusion, which means all or part of the drugs, depending on the drug solubility, have dissolved in the beginning of the release process. As a phenomenological approximation, it is acceptable for quick-dissolving drugs. While it may not be suitable for poorly water-soluble drugs, the drug dissolution process needs to be taken into account. Improved models have been proposed to embrace both drug diffusion and dissolution processes [44–49]. The governing equations describing diffusion- and dissolution-controlled drug release processes are presented as follows for a 1-D slab problem. The second term on the right-hand side of Eq. (15.42) depicts the change rate of concentration of dispersed drug due to drug dissolution that is described by Eq. (15.43) as an example:

$$\frac{\partial C_d}{\partial t} = D \frac{\partial^2 C_d}{\partial x^2} - \frac{\partial C_{sd}}{\partial t} \quad (15.42)$$

$$\frac{\partial C_{sd}}{\partial t} = -K_d(C_s - C_d) \quad (15.43)$$

where

C_d is the concentration of dissolved drug.

C_{sd} is the concentration of dispersed drug.

K_d is the dissolution rate coefficient of drug.

For the coupled partial differential equations of diffusion and dissolution, explicit exact or analytical solutions, like those for diffusion-controlled systems, have not been found yet. Hence numerical approaches such as finite element [46, 47] and finite difference [48] methods have been used to solve this mathematical problem.

15.4.6 Erosion-Controlled Systems

15.4.6.1 Surface (Heterogeneous) Erosion

15.4.6.1.1 Analytical Solutions for a 1-D Slab, Cylinder, and Sphere [49]

$$\frac{M_t}{M_\infty} = 1 - \left(1 - \frac{k_0 t}{a C_0}\right)^n \quad (15.44)$$

where

k_0 is surface erosion constant ($\text{mg}/(\text{h}\cdot\text{cm}^2)$).

a is the radius of a sphere or a cylinder or the half-thickness of a slab.

$n = 1, 2,$ and 3 for slab, cylinder, and sphere, respectively.

15.4.6.1.2 Analytical Solutions for a 2-D Tablet [13]

$$\frac{M_t}{M_\infty} = 1 - \left(1 - \frac{k_0 t}{r_0 C_0}\right)^2 \left(1 - \frac{2k_0 t}{l_0 C_0}\right) \quad (15.45)$$

where

r_0 and l_0 are the initial radius and thickness of the tablet, respectively.

15.4.6.2 Bulk (Homogeneous) Erosion There are few explicit analytical solutions available for bulk erosion problems. Lee developed a model for drug release from an erodible slab with consideration of simultaneous diffusion and erosion processes [50]. A more comprehensive model including erosion, diffusion, and chemical reaction was developed by Thombre and Himmelstein [51–53] for a slab in a sink. The model considered water, drug, acid generator, and acid with partial differential equations as follows:

$$\frac{\partial C_i}{\partial t} = \frac{\partial}{\partial x} \left(D_i(x, t) \frac{\partial C_i}{\partial x} \right) + v_i \quad i = A, B, C, E \quad (15.46)$$

where

C_i and D_i are the concentration and diffusion coefficient of the diffusing species.

v_i is the net sum of synthesis and degradation rate of species.

A, B, C, and E are water, acid generator, acid, and drug, respectively.

Concentration-dependent diffusion coefficient is expressed as

$$D_i = D_i^0 \exp \left[\frac{\mu(C_D^0 - C_D)}{C_D^0} \right], \quad i = A, B, C, E \quad (15.47)$$

Finite difference method was used to solve the equations with various initial and boundary conditions.

15.4.7 Swelling-Controlled Systems

Swelling-controlled release involving solvent penetration into and drug release out from a polymeric matrix system, such as 1-D planar and cylindrical devices, beads, or 2-D tablets, has been modeled and solved numerically [56–58]. The following presents governing equations for a 2-D tablet in a perfect sink. Equations (15.48)–(15.50) describe the rates of solvent penetration, drug diffusion, and dissolution, respectively. Dimensional change due to swelling and matrix erosion is expressed by Eq. (15.51), where the first term on the right-hand side accounts for the dimensional increase from swelling and the second term for polymer dissolution. Water concentration-dependent diffusion coefficients are expressed by Eqs. (15.52) and (15.53):

$$\frac{\partial C_w}{\partial t} = \frac{1}{r} \frac{\partial}{\partial r} \left(r D_w \frac{\partial C_w}{\partial r} \right) + \frac{\partial}{\partial z} \left(D_w \frac{\partial C_w}{\partial z} \right) \quad (15.48)$$

$$\frac{\partial C_d}{\partial t} = \frac{1}{r} \frac{\partial}{\partial r} \left(r D_d \frac{\partial C_d}{\partial r} \right) + \frac{\partial}{\partial z} \left(D_d \frac{\partial C_d}{\partial z} \right) - \frac{\partial C_{sd}}{\partial t} \quad (15.49)$$

$$\frac{\partial C_{sd}}{\partial t} = -K(C_s - C_d) \quad (15.50)$$

$$z_t r_t^2 = 2 \int_0^{z_t} \int_0^{r_t} \left[\frac{C_w(r, z, t)}{\rho_w} + \frac{C_d(r, z, t)}{\rho_d} \right] r dr dz + \frac{1}{2\pi\rho_p} \left(m_{p,0} - \int_0^t K_p A_s dt \right) \quad (15.51)$$

$$D_w = D_w^{eq} \exp \left[-\beta_w \left(1 - \frac{C_w}{C_w^{eq}} \right) \right] \quad (15.52)$$

$$D_d = D_d^{eq} \exp \left[-\beta_d \left(1 - \frac{C_w}{C_w^{eq}} \right) \right] \quad (15.53)$$

where

C_d is the concentration of dissolved drug.

C_{sd} is the concentration of dispersed drug.

C_s is the drug solubility.

K is the dissolution rate coefficient of drug.

C_w is the solvent concentration.

D_d and D_w are the diffusion coefficients of drug and solvent, respectively.

D_d^{eq} and D_w^{eq} are equivalent coefficients of drug and solvent at saturated solvent state.

β_d and β_w are characteristic constants for drug and solvent, respectively.

For initial drug loading below the drug solubility, the dissolution term in Eq. (15.49) can be omitted.

15.4.8 Ion-Exchange-Controlled Systems

15.4.8.1 Drug Loading onto Ion-Exchange Spheres [54]

$$F = \frac{3}{\lambda\theta_0} B_1 \left[\sqrt{\tau} \exp \left(-\frac{1}{4\tau} \right) + \sqrt{\pi\tau} \operatorname{erf} \left(\frac{1}{2\sqrt{\tau}} \right) - 2\sqrt{\tau} \right] \quad (15.54)$$

where

F is the fraction of drug loaded onto the sphere.

$\lambda = 3V/(4\pi R^3)$, $\theta_0 = C_0/C_{\max}$.

Here V is the external fluid volume, R is the radius of sphere, C_0 is the initial solute concentration in the external solution, and C_{\max} is the maximum solute binding capacity of the ion-exchange spheres. $\tau = Dt/R^2$, D is diffusion coefficient of polymer, and B_1 is obtained by solving the following equation:

$$\frac{B_1}{\sqrt{\tau}} = (1-\alpha)B_1\gamma - \beta[1 + B_1\gamma] \left\{ \theta_0 - \frac{3B_1}{\lambda} \left[\sqrt{\tau} \exp \left(-\frac{1}{4\tau} \right) + \tau\gamma - 2\sqrt{\tau} \right] \right\} \quad (15.55)$$

where

$\alpha = VRK_{\text{des}}/(DA)$.

$\beta = VRK_{\text{ads}}C_{\max}/(DA)$.

$\gamma = \sqrt{\pi} \operatorname{erf} [1/(2\sqrt{\tau})]$.

Here A is the surface area of sphere, K_{des} is the dissociation rate constant, and K_{ads} is the association rate constant.

15.4.8.2 Drug Release from Ion-Exchange Spheres [55]

$$\frac{M_\tau}{M_0} = 1 - \frac{3}{\theta_{\text{RS}}^0} \int_0^1 \left(\theta_{s^+} + \frac{K\theta_{s^+}}{\theta_{\text{Na}^+} + K\theta_{s^+}} \right) x^2 dx \quad (15.56)$$

$$\frac{M_\tau}{M_0} = 1 - \frac{3}{\theta_{RS}^0} \int_0^1 \left(\theta_{s^+} + \frac{\sqrt{K^2 \theta_{s^+}^4 + 8K \theta_{s^+}^2 \theta_{Ca^{2+}} - K \theta_{s^+}^2}}{4\theta_{Ca^{2+}}} \right) x^2 dx \quad (15.57)$$

where

$$M_0 = 4\pi R^3 C_{RS}^0 / 3.$$

C_{RS}^0 is the initial concentration of the drug in the sphere that is bound with binding sites of the ion-exchange polymer.

K is Langmuir isotherm constant, $x = r/R$.

C_m is maximum solute binding capacity of the ion-exchange sphere.

$$\theta_{s^+} = C_{s^+} / C_m, \theta_{Na^+} = C_{Na^+} / C_m, \text{ and } \theta_{Ca^{2+}} = C_{Ca^{2+}} / C_m.$$

Detailed numerical procedure can be found from the original reference.

15.4.9 Osmotic Pressure-Controlled Systems

The basic equations for the osmotic component of drug release versus time from osmotic pressure-controlled systems are obtained by expressing the mass delivery rate (dm/dt) from the dosage form as a product of the total volumetric flow rate (dV/dt) of water into the interior of the device and the concentration of drug (C) in the solution or suspension being released. For several osmotic dosage forms, the expression for the volumetric flow rate is derived from irreversible thermodynamics.

15.4.9.1 Miniosmotic Pump The pumping rate is given by the equation

$$\frac{dm_t}{dt} = \frac{A\theta\Delta\pi C}{h} \quad (15.58)$$

where

$\frac{dm_t}{dt}$ is the drug release rate.

C is the concentration of the drug in the chamber.

A is the surface area of the membrane.

θ is the osmotic permeability.

h is the membrane thickness.

$\Delta\pi$ is the osmotic pressure difference between the two solutions on either side of the membrane.

15.4.9.2 Elementary Osmotic Pump The release of drug from this system is controlled by the solvent influx (water) across the semipermeable membrane whereby this influx of water carries the drug to the outside via the orifice. According

to Theeuwes [59], the general expression for the solute delivery rate, dm/dt , obtained by pumping through the orifice can be described by

$$\frac{dm}{dt} = \frac{dV}{dt} C \quad (15.59)$$

where

C is the concentration of the compound in the dispensed fluid expressed per unit volume of the solution.

$\frac{dV}{dt}$ is the volume flux across the semipermeable membrane.

The volume flux is described as follow:

$$\frac{dV}{dt} = \frac{A}{h} L_p (\sigma\Delta\pi - \Delta P) \quad (15.60)$$

where

$\Delta\pi$ and ΔP are the osmotic and hydrostatic pressure differences, respectively, between the inside and outside of the device.

L_p is the hydraulic permeability.

σ is the reflection coefficient.

A is the membrane area.

h is the membrane thickness.

Since $\Delta\pi \gg \Delta P$ and the hydrostatic pressure inside the device is minimized as the delivery orifice increases, ΔP can be omitted from Eq. (15.56). Furthermore, when the osmotic pressure of the core, π , is significantly larger than the osmotic pressure of the dissolution fluid, Eq. (15.54) can be written as

$$\frac{dm}{dt} = \frac{A}{h} \theta \pi C \quad (15.61)$$

where

θ equals $L_p \sigma$.

Theeuwes [59] even characterized the mode of release mathematically, namely, zero-order delivery rate and non-zero-order release, over the entire life of the system:

$$\left(\frac{dm}{dt} \right)_z = \frac{A}{h} \theta \pi_s S \quad (15.62)$$

Equation (15.62) defines the zero-order release rate from $t = 0$ until a time t_z , where S is the solubility and π_s is the

osmotic pressure at saturation. The solubility in Eq. (15.62) replaces the concentration term, C , from time $t = 0$ to $t = t_z$ by assuming that the rate of dissolution of a single compound within the system is much larger than the rate of pumping.

The nonzero-order release rate, as defined by Theeuwes [59], as a function of time indicates a parabolic decline:

$$\frac{dm}{dt} = \frac{F_s S}{\left[1 + \frac{F_s}{V}(t - t_z)\right]^2} \quad (15.63)$$

where

F_s is the flux during the zero-order time and is related to the volume flux, F , into the device during nonzero release by

$$\frac{F_s}{F} = \frac{\pi_s}{\pi} = \frac{S}{C} \quad (15.64)$$

Moreover, the nonzero release rate can be further written as a fraction of the zero-order rate:

$$\frac{dm}{dt} = \frac{(dm/dt)_z}{\left[1 + (1/SV)(dm/dt)_z(t - t_z)\right]^2} \quad (15.65)$$

In order for the aforementioned equations defining the mode of drug release from an elementary osmotic device to be applicable, osmotic pumping has to be the sole mechanism of release. Therefore, the size of the orifice must be such that it is smaller than a maximum size to minimize the solute diffusion through the orifice. It is also imperative to have the orifice larger than a minimum size to reduce hydrostatic pressure inside the system. Hydrostatic pressure within the system will decrease the osmotic influx as well as it may cause an increase in the volume of the system.

The equations for the elementary osmotic pump represent the mass delivered per unit time due to the mechanism of osmotic pumping. In fact, the total mass delivered per unit time from such systems results from osmotic pumping, diffusion through the orifice, and diffusion through the membrane itself [59]. If diffusion through the orifice is negligible, then

$$\left(\frac{dm}{dt}\right)_t = \left(\frac{dm}{dt}\right)_0 + \left(\frac{dm}{dt}\right)_d \quad (15.66)$$

where

$\left(\frac{dm}{dt}\right)_p$ is the rate of release due to osmotic pumping.

$\left(\frac{dm}{dt}\right)_{dm}$ is the release rate resulting from diffusion, as demonstrated by Zentner et al. for KCl release rates from controlled porosity osmotic tablet [60].

The total zero-order release rate during the steady-state portion can then be expressed by

$$\left(\frac{dm}{dt}\right)_{t,z} = \frac{A}{h}(\theta\pi_s S + PS) \quad (15.67)$$

where

P is the permeability coefficient of the drug in the polymer.

In a similar fashion, the total nonzero-order rate can be given as

$$\frac{dm}{dt} = \frac{F_s}{S}C^2 + \frac{A}{h}PC \quad (15.68)$$

However, to express dm/dt as a function of time, the concentration, C , inside the system must be expressed as a function of time.

15.4.9.3 Semipermeable Membrane Containing Micropores

Many researchers have shown that besides simple diffusion, osmotic pumping mechanism contributes significantly to the release of drugs from film-coated preparations [29, 61]. The zero-order steady-state release for such systems under the influence of zero hydrostatic pressure can be expressed by Eq. (15.67). According to Lindstedt et al. [62], the equations presented by Theeuwes cannot describe osmotic pumping as contributing to the mechanism of drug release. They argue that since during zero-order release a steady state is maintained with no volume expansion of the tablets, the net bulk volume flux through the membrane must be zero. It therefore follows that release of drugs would be independent of osmotic pressure and be exclusively diffusive. Therefore, they presented the solute flux, F_s , as follows:

$$F_s = CF_v(1 - \sigma) + \frac{A}{h}P_s\Delta C \quad (15.69)$$

where

$$F_v = \frac{A}{h}L_p(\Delta P - \Delta\pi) \quad (15.70)$$

In order to remove the limitation of zero net bulk volume flux through the membrane, the membrane was considered to consist of two areas with different reflectivity. The release rate, Q , is the sum of the solute fluxes in areas 1 and 2:

$$Q = F_{s1} + F_{s2} = (1 - \sigma_1)F_{v1}C_1 + (1 - \sigma_2)F_{v2}C_2 + D_s \quad (15.71)$$

where $F_{v1} + F_{v2} = 0$ at steady state, since they are equal and in opposite direction, and

$$D_s = \left(\frac{A_1}{h_1} P_{s1} + \frac{A_2}{h_2} P_{s2} \right) \Delta C \quad (15.72)$$

is the diffusional release through areas 1 and 2. Equation (15.70) reduces to

$$Q = (1 - \sigma_2)(-F_{v1})C_s + D_s \quad (15.73)$$

The assumptions were that the bulk volume flux in area 1 is directed into the tablet and C_1 is zero and C_2 is equal to the concentration in the core, C_s , and $-F_{v1} = F_{v2}$.

Therefore,

$$Q = (1 - \sigma_2) \frac{A_1}{h_1} L_{p1} (\sigma_1 \Delta \pi - \Delta P) C_s + D_s \quad (15.74)$$

It was further inferred that the low-reflective area is very small compared with the total area. From this inference, it was concluded that the release rate may be calculated from

$$Q = (1 - \sigma_2) \frac{A}{h} L_{p1} (\sigma_1 \Delta \pi - \Delta P) C_s + D_s \quad (15.75)$$

If ΔP is assumed negligible and $\sigma_2 \left\langle \frac{\Delta P}{\Delta \pi} \right\rangle$ since the volume flux is directed out of the tablet and $F_{v2} > 0$, then Eq. (15.43) may be reduced to

$$Q = \frac{A}{h} L_{p1} \sigma_1 \Delta \pi C_s + D_s \quad (15.76)$$

15.4.9.4 Push–Pull Osmotic Pump Drug release from this system is controlled by the solvent influx (water) across the semipermeable membrane into the tablet core and resulting simultaneous push action from the swelling layer. According to Swanson et al. [18], the general expression for the solute delivery rate, dm/dt , obtained by pumping through the orifice can be simply modified from Eq. (15.28) to be

$$\frac{dm}{dt} = \frac{dV}{dt} C_s \quad (15.77)$$

where

C_s is the concentration of the drug in suspension in the dispensed fluid expressed per unit volume of the solution.

The osmotic volume flow into the osmotic compartment is described as

$$\left(\frac{dV}{dt} \right)_0 = \frac{k}{h} A_p(H) \pi_p(H) \quad (15.78)$$

where

k is the osmotic membrane permeability coefficient.

h is the membrane thickness.

A_p is the area of the push compartment.

π_p is the imbibition pressure of the push compartment.

An additional consideration for the push–pull system is the osmotic volume imbibition flow into the drug compartment described as

$$\left(\frac{dV}{dt} \right)_D = \frac{k}{h} (A - A_p(H)) \pi_D(H) \quad (15.79)$$

where

A is the total area of the dosage form.

π_D is the imbibition pressure in the drug compartment.

The total volume flow from the dosage form is the summation of the osmotic flow into the osmotic compartment and the osmotic imbibition flow into the drug compartment as described below:

$$\frac{dV}{dt} = \left(\frac{dV}{dt} \right)_0 + \left(\frac{dV}{dt} \right)_D \quad (15.80)$$

The concentration of dispensed drug from the dosage form can be expressed as

$$C_s = F_D C_0 \quad (15.81)$$

where

F_D is the fraction of drug in the drug compartment.

C_0 is the concentration of solids dispensed from the dosage form.

Substituting Eqs. (15.78) and (15.79) into (15.80) and Eqs. (15.80) and (15.81) into (15.77), the total drug release is given as

$$\frac{dm}{dt} = \left[\frac{k}{h} A_p(H) \pi_p(H) \right] + \left[\frac{k}{h} (A - A_p(H)) \pi_D(H) \right] F_D C_0 \quad (15.82)$$

15.4.9.5 Polymer Drug Matrix Systems The osmotic release mechanism from the polymer matrix type of device may be expressed by the model developed by Wright et al. [31]:

$$\frac{dm}{dt} = \frac{3\alpha\rho\phi\Delta\Pi S_0 L_p}{d \left(\lambda_b^3 - 1 \right) \left[1 - \left(\frac{6\phi}{\pi} \right)^{1/3} \right]} \quad (15.83)$$

where

$\frac{dm}{dt}$ is the zero-order release rate.

α is a constant of proportionality.

ρ is the solid density of the drug.

ϕ is the volumetric loading of the drug.

$\Delta\Pi$ is the osmotic pressure difference between the capsule solution and the external medium.

S_0 is the surface area of the device.

L_p is the polymer hydraulic permeability.

d is the particle size.

λ_b is the polymer extension ratio at rupture.

In deriving this model, it was assumed that the drug particles were spherical and were released by osmotic rupturing. Also, the matrix was considered to consist of two zones, a ruptured capsule zone separated from a zone of water imbibing capsules by a moving water front.

Schirrer et al. [63] developed the following model by assuming that the water flow into a capsule per unit time per unit area at the capsule-polymer interface is constant for a given osmotic agent and is directly proportional to the volumetric loading:

$$\frac{1}{Q_0} \frac{dQ}{dt} = \frac{S_0 \Phi}{V_0 \phi^{1/3} 3(\lambda_b - 1)} \quad (15.84)$$

where

Φ is the water flow into a capsule per unit time per unit area at the capsule-polymer interface. Q is the volume of salt released.

Q_0 is the initial volume of salt in the matrix.

V_0 is the initial volume of the device.

The rate of mass drug release is defined as

$$\frac{dm}{dt} = \rho \frac{dQ}{dt} \quad (15.85)$$

According to Amsden et al. [64], these models are limited because of their dependence on drug density and undefined constant of proportionality. Further, it is believed that at low drug loading a portion of the drug will be released by dissolution and diffusion. Based on these premises, a model was developed. Paramount to this model is the

assumption that the solutes that remain in a capsule after it ruptures were released by diffusion and not convection. Also, the model accounts for capsule swelling and that not all capsules in the monolith were ruptured. The model is as follows:

$$\frac{dm}{dt} = \frac{8\pi \{f(1-F_D)\phi\}^{2/3} S_0 L_p E \omega C_{\text{sat}} (\lambda_b^3 - 1)}{3d t_b^*} \quad (15.86)$$

where

f is the mass fraction of material remaining after the initial burst that is released by osmotic pressure-induced polymer rupturing.

$f(1-F_D)$ is the mass fraction of particles in the monolith released by rupturing.

F_D represents the mass fraction of particles released by dissolution and diffusion.

S_0 is as defined previously.

E is Young's modulus of elasticity.

$$f = 1 - \text{Exp} \left(-\frac{\pi}{4} \left(\frac{h_c}{h} \right)^2 \right) \quad (15.87)$$

where

h_c and h are the critical wall thickness for rupturing to occur and the average wall thickness, respectively.

$$C_s \omega = C_s \text{Exp} \left[\frac{-DA_0(t_b)}{V_f h} \right] \quad (15.88)$$

where

D is the drug diffusivity.

A_0 is the cross-sectional area of the channel.

C_s is the agent saturation concentration.

V_f is the capsule volume after rupture.

t_b is the time required for sufficient water to flow into a capsule to induce rupture.

Finally,

$$t_b^* = \frac{4L_p E t}{r_0^2} \quad (15.89)$$

where

r_0 is the initial particle radius.

CASE STUDY

Background for Example 15.1 A controlled release formulation was required for a drug to reduce the high dosing frequency related to its short half-life (2.5–3.8 hours) and to reduce C_{max} -related side effects. The projected dose strengths were 2 and 10 mg. The drug solubility was greater than 100 mg/ml at pH 4. The dose–solubility map indicated that an osmotic dosage form based on asymmetric membrane technology would be suitable for this drug [7].

EXAMPLE 15.1

An asymmetric membrane (AM) tablet core is composed of 10 mg of a highly soluble drug and mannitol ($\pi = 38$ atm) as osmogen. This core tablet is coated with a 15% w/w semipermeable coating composed of cellulose acetate, polyethylene glycol, acetone, and water. The drug release into various media (distilled water, sucrose solutions, and saturated drug solutions) is listed in Tables 15.2 and 15.3.

- What are the osmotic pressures of the sucrose solutions in Table 15.2?
- What osmotic pressure will shut down the osmotic release mechanism for this AM tablet?
- What percentage of the release mechanism is due to diffusion?

Solution

- The osmotic pressures for sucrose solutions can be located in the *Handbook of Chemistry and Physics* [65]. These literature standard values are plotted in Figure 15.15, similar to experimental results reported by am Ende and Miller [66], and demonstrate consistency with accepted standards. The sucrose solution concentrations of 0, 216, 364, 471, and 601 g/l have osmotic pressures of 0, 20, 40, 60, and 90 atm, respectively.
- The drug release profiles from Table 15.2 are plotted in Figure 15.16, and slopes for the initial 0–60% release are calculated using linear fits. The initial release rates as a function of sucrose media osmotic pressure are 5.1, 3.6, 2.5, and 1.0%/h for 0, 20, 40, and 60 atm, respectively. These resulting initial release rates are then plotted as a function of media osmotic pressure, as shown in Figure 15.17. The media osmotic pressure that would shut down drug release is determined by extrapolation to be 76 atm.
- The drug release profiles listed in Table 15.2 are plotted in Figure 15.18. The diffusional contribution is approximately 5% of the total drug release, as demonstrated when all osmotic release is shut off by saturating the media with drug.

EXAMPLE 15.2

Compressed tablets of a model drug were coated with ethyl cellulose and 20% diethyl phthalate, a plasticizer. The tablet

TABLE 15.2 Drug Release from Asymmetric Membrane Tablet into Media Containing Varying Sucrose Concentrations

Time (h)	% Drug Dissolved				
	Distilled Water	216.2 g/l Sucrose	363.7 g/l Sucrose	470.6 g/l Sucrose	601 g/l Sucrose
0	0	0	0	0	0
2	10.5	7.1	5.0	2.1	0.95
6	32.1	21.5	14.8	6.2	3.7
12	60.8	45.2	30.0	12.4	5.4
18	81.7	63.7	44.9	18.4	7.1
24	99.9	88.3	59.6	24.7	6.4

TABLE 15.3 Drug Release from Asymmetric Membrane Tablet into Saturated Drug Solution Media (% drug dissolved = mg released/total mg in tablet × 100%)

Time (h)	Π Release Shutoff	Diffusional Release Shutoff (D)	Cumulative Release ($\Pi + D$)	Actual Drug Release
0	0	0	0	0
2	5.1	13.1	18.2	12.8
6	3.7	40.9	44.6	—
8	—	—	—	52.2
16	—	—	—	84.5
18	7.1	86.1	93.2	—
24	6.4	91.8	98.2	93.9

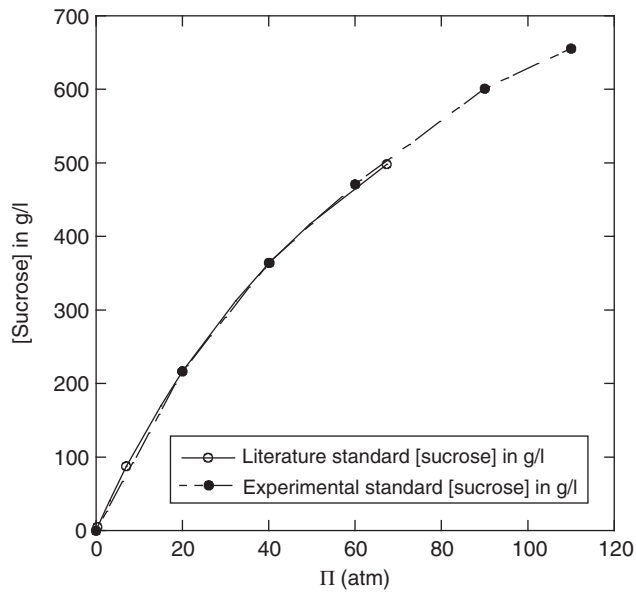


FIGURE 15.15 Calibration curve for sucrose solutions prepared to produce osmotic pressures ranging from 0 to 100 atm.

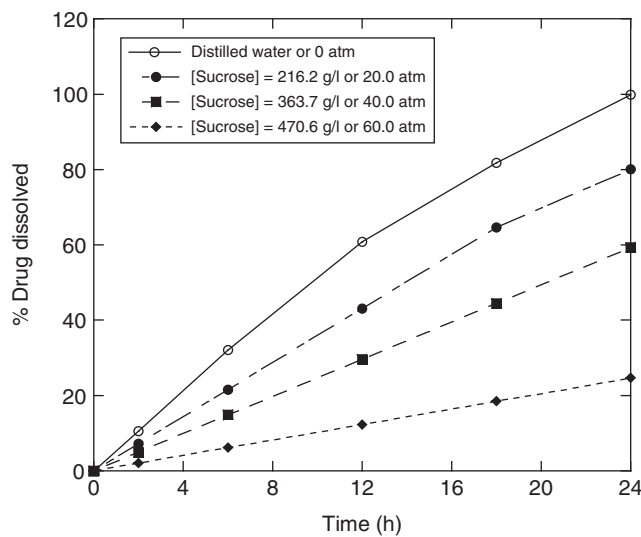
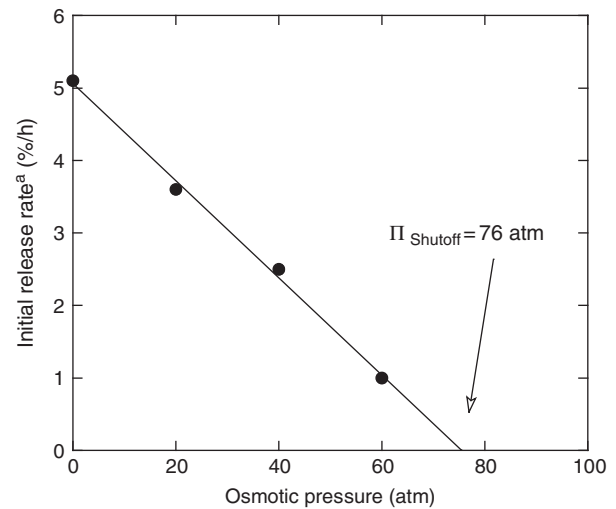


FIGURE 15.16 Drug release from AM tablets into varying concentrations of sucrose solutions and associated osmotic pressures determined in part (a).

0.6 cm in diameter and 0.3 cm in thickness contains 30% drug. The drug solubility in the tablet was estimated to be 0.1 g/cm^3 . The drug diffusion coefficient in the coating was evaluated previously to be $1 \times 10^{-8} \text{ cm}^2/\text{s}$ by fitting experimental release curves using a two-dimensional tablet model of membrane-reservoir system.

a. Calculate fractional release and release rate of a tablet and $20 \mu\text{m}$ coating thickness.



^aInitial release rate determined by linear regression over 0–60% release profile.

FIGURE 15.17 Initial drug release rates from AM tablets as a function of media osmotic pressure.

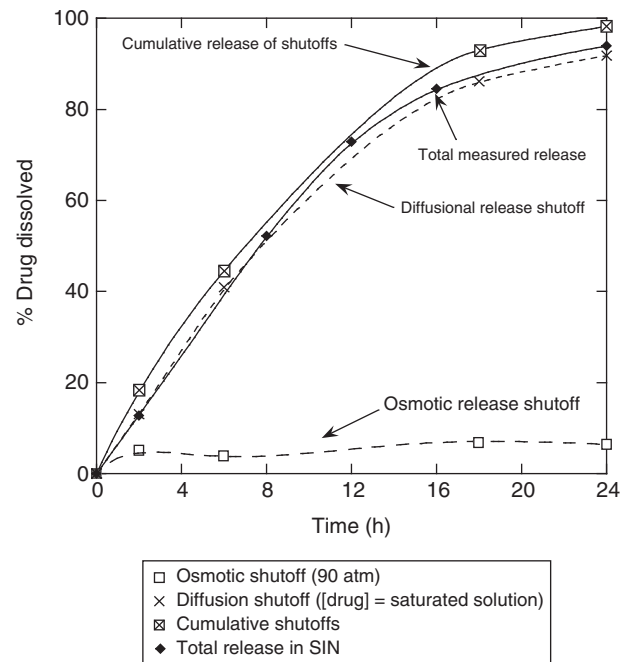


FIGURE 15.18 Drug release from AM tablets into media with osmotic pressure in excess of the determined shutoff, e.g. 90 atm, and into media saturated with drug to determine diffusional shutoff [54]. *Source:* Reprinted from am Ende and Miller [66], Copyright (2007), with permission from Springer Nature.

- b. If a zero-order release up to 12 hours is desirable, what coating thickness should be applied?
- c. What is the release rate at this coating thickness?

Solution

- The fractional release and release rate were computed based on Eqs. (15.17), (15.18), and (15.23) and plotted in Figure 15.19.
- By implementing computer simulation using various coating thickness, an optimal thickness of 31.5 μm was found to give zero-order release up to 12 hours.
- With this coating thickness, the tablet provides a release rate of 1.5 mg/h.

EXAMPLE 15.3

Inert matrix tablets of two model drugs of different solubilities were made by compression of drug–excipient matrix granules. The granules were prepared by using acrylic polymer dispersion (Eudragit[®] FS 30D) as the granulating agent. The tablet has a diameter of 0.6 cm and a thickness of 0.3 cm and contains 30% drug. The solubility of drug A in the tablet was estimated to be 0.1 g/cm³ and drug B 0.01 g/cm³. The drug diffusion coefficient in the coating was evaluated previously to be

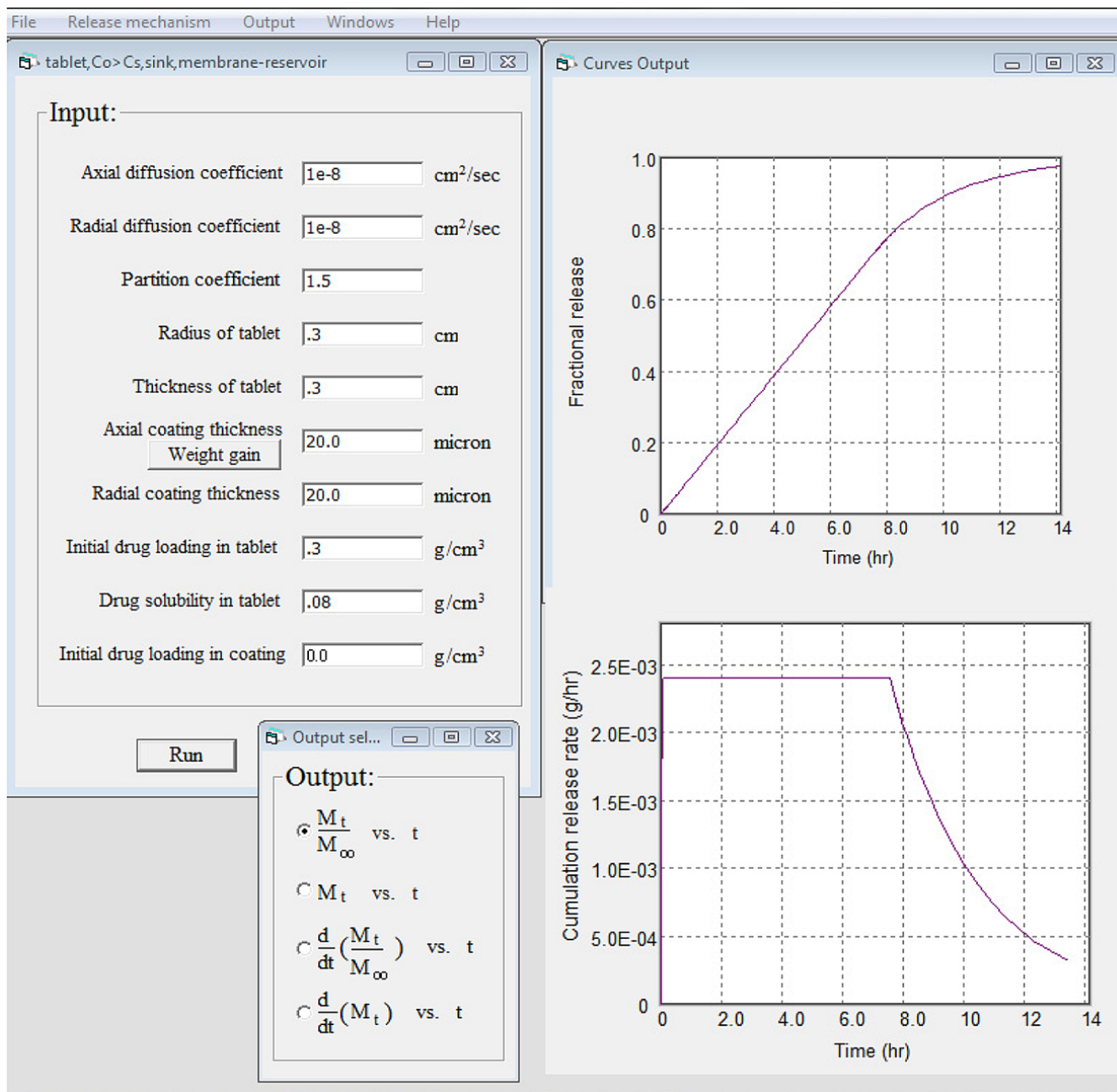


FIGURE 15.19 Fractional drug release profile (top right) and release rate (bottom right) of a 2-D membrane-coated tablet predicted by Eqs. (15.17) and (15.23) and computed by AP-CAD[®] software package. In this example identical axial and radial coating thickness and diffusion coefficients were used (see the input parameter values on the left panel).

$8 \times 10^{-7} \text{ cm}^2/\text{s}$ by fitting experimental release curves of tablets containing a low initial drug loading ($C_0 < C_s$) using the two-dimensional (2-D) tablet model (Eq. 15.31). How would drug solubility and dissolution influence the rate and release profiles?

Solution

Because the tablets contained initial drug loading greater than drug solubility, analytical solutions for 2-D matrix tablets with $C_0 > C_s$ (Eqs. 15.39–15.41) were used to compute drug

release profiles of drug A. It is reasonable to assume the dissolution of drug A is much faster than diffusion. The results are presented in Figure 15.21. For drug B, its solubility is one-tenth of that of drug A (Figure 15.20), and the model with a drug dissolution term (Eq. 15.49) was applied and solved numerically using AP-CAD® software package. The results are presented in Figure 15.20. Comparing the results in Figure 15.21 with those in Figure 15.20, it is seen that the release profile becomes more linear when drug dissolution becomes more significant.

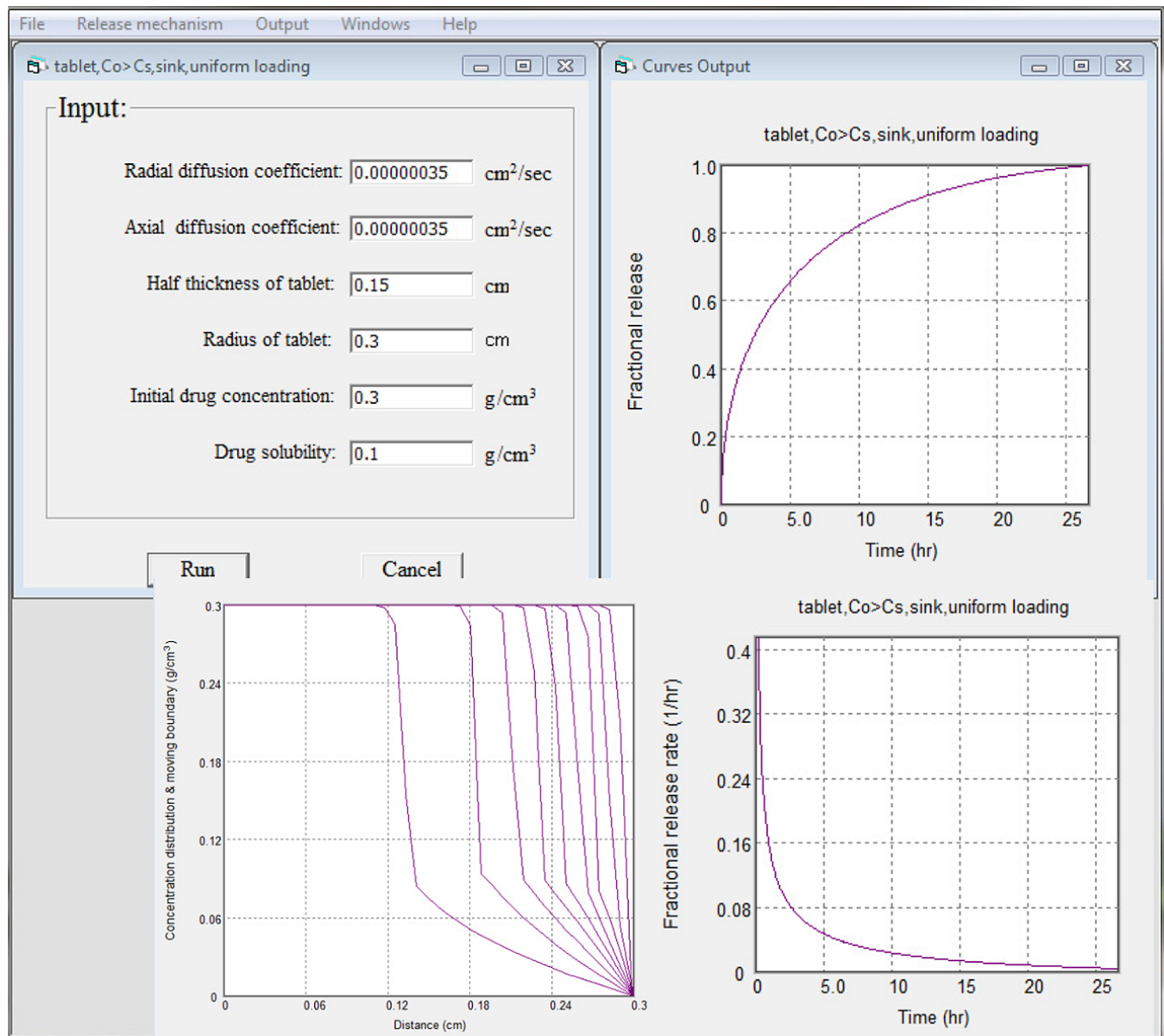


FIGURE 15.20 Fractional release (top right), release rate (bottom right), and concentration profiles at various times (bottom left) of a matrix tablet containing a water-soluble drug with the dimension, initial drug loading, and diffusion coefficient indicated in the left panel.

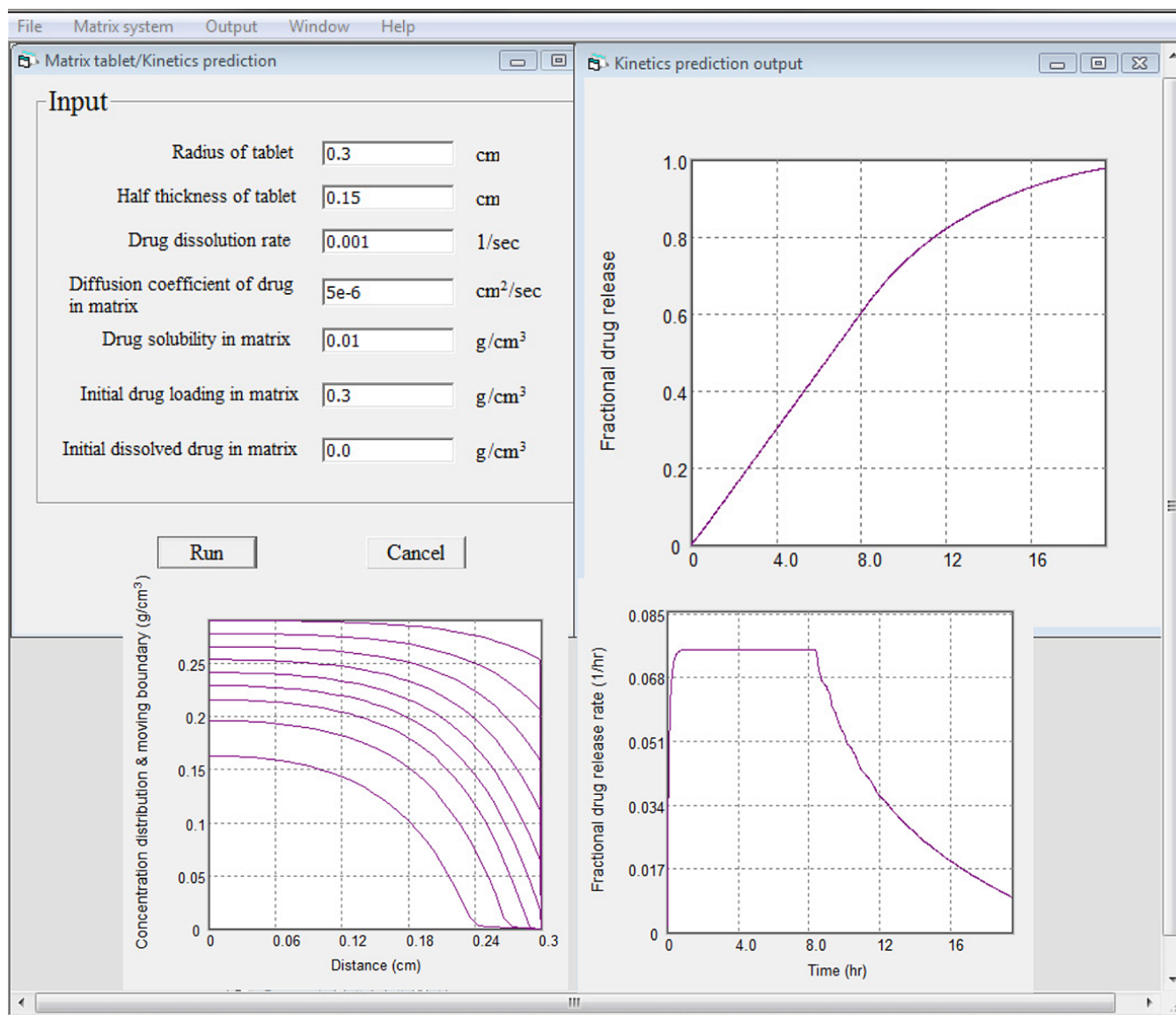


FIGURE 15.21 Fractional release (top right), release rate (bottom right), and concentration profiles at various times (bottom left) of a matrix tablet containing a poorly water-soluble drug with the dimension, initial drug loading, diffusion coefficient, and drug dissolution rate constant indicated in the left panel.

15.5 CONCLUSIONS

Controlled release technology and the design of controlled release dosage forms were discussed in this chapter. The equations governing the rate of drug release have been derived based on the dominant mechanism of drug release, e.g. Fick's second law of diffusion for non-erodible monolithic and reservoir devices and irreversible thermodynamics for certain osmotic systems. In addition to physicochemical factors, biopharmaceutical factors such as mechanism of drug absorption and gastrointestinal transit of the dosage form must be considered in the design of oral controlled release dosage forms.

REFERENCES

1. FDA (September 1997). Guidance for Industry SUPAC-MR: Modified Release Solid Oral Dosage Forms Scale-Up and Postapproval Changes: Chemistry, Manufacturing, and Controls; in vitro Dissolution Testing and in vivo Bioequivalence.
2. FDA (December 2008). Guidance for Industry Orally Disintegrating Tablets.
3. Rathbone, M.J. and Bowersock, T. (eds.) (2002). *Veterinary Drug Delivery*. Amsterdam: Elsevier p 284 pp.
4. Lakkis, J.M. (ed.) (2007). *Encapsulation Controlled Release Technologies in Food Systems*, 239. Chichester: Wiley.

5. Thombre, A.G. (2005). Assessment of the feasibility of oral controlled release in an exploratory development setting. *Drug Discovery Today* 10 (17): 1159–1166.
6. Lipper, R.A. (1999). How can we optimize selection of drug development candidates from many compounds at the discovery stage? *Modern Drug Discovery* 2: 55–60.
7. Thombre, A.G. (2000). Feasibility assessment and rapid development of oral controlled release prototypes. *ACS Symposium Series* 752: 69–77.
8. Carslaw, H. S.; Jaeger, J. C., *Conduction of Heat in Solids*. 2nd ed. Oxford University Press, Oxford, 1959; p 510 pp.
9. Crank, J., *The Mathematics of Diffusion*. 2d Ed. 1975; Oxford University Press, Oxford p 414 pp.
10. Baker, R. (1987). *Controlled Release of Biologically Active Agents*, 279. Wiley.
11. Fan, L.T. and Singh, S.K. (1989). *Controlled Release: A Quantitative Treatment (Polymers, Properties and Applications, 13)*. Springer-Verlag: New York.
12. Kydonieus, A. (ed.) (1992). *Treatise on Controlled Drug Delivery: Fundamentals, Optimization, Applications*, 553. New York: Marcel Dekker.
13. Kim, C.-J. (2000). *Controlled Release Dosage Form Design*, 310. Lancaster, PA: Technomic Pub. Co.
14. Carrier, R.L. and Waterman, K.C. (2006). Use of biodegradable polymers in oral drug delivery: challenges and opportunities. In: *Handbook of Biodegradable Polymeric Materials and Their Applications*, vol. 2 (ed. S.K. Mallapragada and B. Narasimhan), 33–56. Stevenson Ranch, CA: American Scientific Publishers.
15. Santus, G. and Baker, R.W. (1995). Osmotic drug delivery: a review of the patent literature. *Journal of Controlled Release* 35 (1): 1–21.
16. Ranade, V.V. and Hollinger, M.A. (1996). *Drug Delivery Systems*, 2e, 16–39. Boca Raton, FL: CRC Press.
17. Cardinal, J.R. (2000). Controlled release osmotic drug delivery systems for oral applications. In: *Drugs and the Pharmaceutical Sciences*, Transport Processes in Pharmaceutical Systems, vol. 102 (ed. G.L. Amidon, P.I. Lee and E.M. Topp), 411–444. Boca Raton, FL: CRC Press, Taylor and Francis Group.
18. Swanson, D.R., Barclay, B.L., Wong, P.S., and Theeuwes, F. (1987). Nifedipine gastrointestinal therapeutic system. *The American Journal of Medicine* 83 (6B): 3–9.
19. Zentner, G.M., Rork, G.S., and Himmelstein, K.J. (1985). The controlled porosity osmotic pump. *Journal of Controlled Release* 1 (4): 269–282.
20. Zentner, G.M., Rork, G.S., and Himmelstein, K.J. (1985). Osmotic flow through controlled porosity films: an approach to delivery of water soluble compounds. *Journal of Controlled Release* 2: 217–229.
21. Appel, L.E. and Zentner, G.M. (1991). Use of modified ethyl cellulose latexes for microporous coating of osmotic tablets. *Pharmaceutical Research* 8 (5): 600–604.
22. Okimoto, K., Miyake, M., Ohnishi, N. et al. (1998). Design and evaluation of an osmotic pump tablet (OPT) for prednisolone, a poorly water soluble drug, using (SBE)7m-b-CD. *Pharmaceutical Research* 15 (10): 1562–1568.
23. Okimoto, K., Ohike, A., Ibuki, R. et al. (1999). Design and evaluation of an osmotic pump tablet (OPT) for chlorpromazine using (SBE)7m-b-CD. *Pharmaceutical Research* 16 (4): 549–554.
24. Okimoto, K., Ohike, A., Ibuki, R. et al. (1999). Factors affecting membrane-controlled drug release for an osmotic pump tablet (OPT) utilizing (SBE)7m-b-CD as both a solubilizer and osmotic agent. *Journal of Controlled Release* 60 (2–3): 311–319.
25. Okimoto, K., Rajewski, R.A., and Stella, V.J. (1999). Release of testosterone from an osmotic pump tablet utilizing (SBE)7m-b-cyclodextrin as both a solubilizing and an osmotic pump agent. *Journal of Controlled Release* 58 (1): 29–38.
26. Stella, V.J., Rao, V.M., and Zannou, E.A. (2003). The pharmaceutical use of captisol: some surprising observations. *Journal of Inclusion Phenomena and Macrocyclic Chemistry* 44 (1–4): 29–33.
27. Herbig, S.M., Cardinal, J.R., Korsmeyer, R.W., and Smith, K. L. (1995). Asymmetric-membrane tablet coatings for osmotic drug delivery. *Journal of Controlled Release* 35 (2–3): 127–136.
28. am Ende, M.T., Herbig, S.M., Korsmeyer, R.W., and Chidlaw, M.B. (2000). Osmotic Drug Delivery from Asymmetric Membrane Film-Coated Dosage Forms. In: *Handbook of Pharmaceutical Controlled Release Technology* (ed. D.L. Wise), 751–785. New York: Marcel Dekker, Inc.
29. Narisawa, S., Nagata, M., Hirakawa, Y. et al. (1997). An organic acid-induced sigmoidal release system for oral controlled-release preparations. III. Elucidation of the anomalous drug release behavior through osmotic pumping mechanism. *International Journal of Pharmaceutics* 148 (1): 85–91.
30. Amsden, B.G., Cheng, Y.-L., and Goosen, M.F.A. (1994). A mechanistic study of the release of osmotic agents from polymeric monoliths. *Journal of Controlled Release* 30 (1): 45–56.
31. Wright, J., Chandrasekaran, S.K., Gale, R., and Swanson, D. (1981). A model for the release of osmotically active agents from monolithic polymeric matrixes. *AIChE Symposium Series* 77 (206): 62–68.
32. Zhou, Y. and Wu, X.Y. (1997). Finite element analysis of diffusional drug release from complex matrix systems. I. Complex geometries and composite structures. *Journal of Controlled Release* 49 (2,3): 277–288.
33. Wu, X.Y. and Zhou, Y. (1998). Finite element analysis of diffusional drug release from complex matrix systems. II. Factors influencing release kinetics. *Journal of Controlled Release* 51 (1): 57–71.
34. Wu, X.Y. and Zhou, Y. (1999). Studies of diffusional release of a dispersed solute from polymeric matrixes by finite element method. *Journal of Pharmaceutical Sciences* 88 (10): 1050–1057.
35. Siegel, R.A. (2000). Theoretical analysis of inward hemispheric release above and below drug solubility. *Journal of Controlled Release* 69 (1): 109–126.
36. Siepman, J., Streubel, A., and Peppas, N.A. (2002). Understanding and predicting drug delivery from hydrophilic matrix tablets using the “sequential layer” model. *Pharmaceutical Research* 19 (3): 306–314.

37. Good, W.R. and Lee, P.I. (1984). Membrane-controlled reservoir drug delivery systems. *Medical Applications of Controlled Release* 1: 1–39.
38. Zhou, Y., Chu, J.S., Li, J.X., and Wu, X.Y. (2010). Theoretical analysis of release kinetics of coated tablets containing constant and non-constant drug reservoirs. *International Journal of Pharmaceutics* 385: 98–103.
39. Fu, J.C., Hagemer, C., and Moyer, D.L. (1976). A unified mathematical model for diffusion from drug-polymer composite tablets. *Journal of Biomedical Materials Research* 10 (5): 743–758.
40. Paul, D.R. and McSpadden, S.K. (1976). Diffusional release of a solute from a polymer matrix. *Journal of Membrane Science* 1 (1): 33–48.
41. Higuchi, T. (1963). Mechanism of sustained-action medication. Theoretical analysis of rate of release of solid drugs dispersed in solid matrices. *Journal of Pharmaceutical Sciences* 52 (12): 1145–1149.
42. Roseman, T.J. and Higuchi, W.I. (1970). Release of medroxyprogesterone acetate from a silicone polymer. *Journal of Pharmaceutical Sciences* 59 (3): 353–357.
43. Zhou, Y., Chu, J.S., Zhou, T., and Wu, X.Y. (2005). Modeling of dispersed-drug release from two-dimensional matrix tablets. *Biomaterials* 26 (8): 945–952.
44. Chandrasekaran, S.K. and Paul, D.R. (1982). Dissolution-controlled transport from dispersed matrixes. *Journal of Pharmaceutical Sciences* 71 (12): 1399–1402.
45. Harland, R.S., Dubernet, C., Benoit, J.P., and Peppas, N.A. (1988). A model of dissolution-controlled, diffusional drug release from non-swelling polymeric microspheres. *Journal of Controlled Release* 7 (3): 207–215.
46. Frenning, G., Brohede, U., and Stromme, M. (2005). Finite element analysis of the release of slowly dissolving drugs from cylindrical matrix systems. *Journal of Controlled Release* 107 (2): 320–329.
47. Zhou, Y., Li, J.X., and Wu, X.Y. (2008). QbD of oral controlled release dosage forms by computational simulation. *Proceedings of the International Symposium on Controlled Release of Bioactive Materials*, 1144–1145. 12–16 July, New York, US.
48. Chang, N.J. and Himmelstein, K.J. (1990). Dissolution-diffusion controlled constant-rate release from heterogeneously loaded drug-containing materials. *Journal of Controlled Release* 12 (3): 201–212.
49. Hopfenberg, H.B. (1976). Controlled release from erodible slabs, cylinders, and spheres. Papers presented at [the] Meeting - American Chemical Society, Division of Organic Coatings and Plastics Chemistry, 36 (1), pp. 229–234.
50. Lee, P.I. (1980). Diffusional release of a solute from a polymeric matrix – approximate analytical solutions. *Journal of Membrane Science* 7 (3): 255–275.
51. Thombre, A.G. and Himmelstein, K.J. (1985). A simultaneous transport-reaction model for controlled drug delivery from catalyzed bioerodible polymer matrices. *AIChE Journal* 31 (5): 759–766.
52. Thombre, A.G. (1992). Theoretical aspects of polymer biodegradation: mathematical modeling of drug release and acid-catalyzed poly(ortho-ester) biodegradation. In: *Biodegradable Polymers and Plastics*, vol. 109 (ed. M. Vert, J. Feijen, A. Albertsson, et al.), 214–225. Royal Society of Chemistry.
53. Thombre, A.G. and Himmelstein, K.J. (1984). Modeling of drug release kinetics from a laminated device having an erodible drug reservoir. *Biomaterials* 5 (5): 250–254.
54. Abdekhodaie, M.J. and Wu, X.Y. (2006). Drug loading onto ion-exchange microspheres: modeling study and experimental verification. *Biomaterials* 27 (19): 3652–3662.
55. Abdekhodaie, M.J. and Wu, X.Y. (2008). Drug release from ion-exchange microspheres: mathematical modeling and experimental verification. *Biomaterials* 29 (11): 1654–1663.
56. Siepmann, J. and Peppas, N.A. (2000). Hydrophilic matrixes for controlled drug delivery: an improved mathematical model to predict the resulting drug release kinetics (the “sequential layer” model). *Pharmaceutical Research* 17 (10): 1290–1298.
57. Wu, N., Wang, L.-S., Tan, D.C.-W. et al. (2005). Mathematical modeling and in vitro study of controlled drug release via a highly swellable and dissoluble polymer matrix: polyethylene oxide with high molecular weights. *Journal of Controlled Release* 102 (3): 569–581.
58. Wu, X.Y. and Zhou, Y. (1997). Numerical simulation of controlled drug release from matrix tablets involving swelling, erosion and diffusion. *Pharmaceutical Research* 14: S716.
59. Theeuwes, F. (1975). Elementary osmotic pump. *Journal of Pharmaceutical Sciences* 64 (12): 1987–1991.
60. Zentner, G.M., Rork, G.S., and Himmelstein, K.J. (1986). Osmotic flow through controlled porosity films: an approach to delivery of water soluble compounds. In: *Advances in Drug Delivery Systems* (ed. J.M. Anderson and S.W. Kim), 217–229. New York: Elsevier.
61. Rekhi, G.S., Porter, S.C., and Jambhekar, S.S. (1995). Factors affecting the release of Propranolol Hydrochloride from beads coated with Aqueous Polymeric Dispersions. *Drug Development and Industrial Pharmacy* 21 (6): 709–729.
62. Lindstedt, B., Ragnarsson, G., and Hjaertstam, J. (1989). Osmotic pumping as a release mechanism for membrane-coated drug formulations. *International Journal of Pharmaceutics* 56 (3): 261–268.
63. Schirrer, R., Thepin, P., and Torres, G. (1992). Water absorption, swelling, rupture and salt release in salt-silicone rubber compounds. *Journal of Materials Science* 27 (13): 3424–3434.
64. Amsden, B.G., Cheng, Y.-L., and Goosen, M.F.A. (1996). A mechanistic study of the release of osmotic agents from polymeric monoliths. [Erratum to document cited in CA121:17834]. *Journal of Controlled Release* 38 (2,3): 275.
65. Weast, R. C.; Selby, S. M., *Handbook of Chemistry and Physics*, 55th Ed. 1974; Cleveland, OH: CRC Press, p 2304 pp.
66. am Ende, M.T. and Miller, L.A. (2007). Mechanistic investigation of drug release from asymmetric membrane tablets: effect of media gradients (osmotic pressure and concentration), and potential coating failures on in vitro release. *Pharmaceutical Research* 24 (2): 288–297.

16

PROCESS DESIGN AND DEVELOPMENT FOR NOVEL PHARMACEUTICAL DOSAGE FORMS

LEAH APPEL AND JOSHUA SHOCKEY
Green Ridge Consulting, Bend, OR, USA

MATTHEW SHAFFER
Lonza, Inc., Bend, OR, USA

JENNIFER CHU
FreeThink Technologies, Inc., Branford, CT, USA

16.1 INTRODUCTION

Strong fundamental knowledge of the formulation and process that is used for a pharmaceutical product is critical to ensuring efficacy, safety, and robust product quality. The design and development approach that is generally advocated in the pharmaceutical industry is called “quality by design” (QbD). This is the application of a scientifically logical approach in developing a formulation and process that is robust, well understood, and well characterized. Knowledge of science and engineering principles and how to apply them are imperative to this product development process.

This approach is especially important in novel dosage forms that are used to produce a drug product that may have even tighter tolerances for performance, stability, and/or manufacturability than a standard dosage form. For example, if the release rate of drug is governed by a functional coating, where the coating thickness and morphology impact the rate of release, it becomes critical to control the coating process such that it consistently provides the same coating quality. While this is important for cosmetic coatings too, the range of coating thickness that yields acceptable performance and appearance is much broader than that for a functional coating. This chapter demonstrates the application of energy and mass transport principles to both the dosage form mechanism of release and the manufacturing process for a novel pharmaceutical formulation.

The formulation and process utilized to make a drug product are coupled, and each must be examined in the context of the other. For example, in choosing the materials that are used in the formulation, it is important to understand the process implications of the selected materials. Similarly, understanding the mechanism of drug release is key to understanding which product attributes are most critical in achieving the target release profile. This knowledge can be used to help guide process design and development.

In this chapter, an osmotic rupturing multiparticulate formulation manufactured by fluid bed coating is used as a model to demonstrate the application of engineering principles to develop a formulation and process for a novel dosage form. The rupturing multiparticulate is designed to provide a burst of drug release at a specific point in time after dosing. The primary uses for this type of dosage form would be where a delayed release of drug is required; quite commonly this is in combination with an immediate release dose of drug in a single unit dose. This approach allows the combination of multiparticulates with different release profiles in a capsule to provide the overall target release for the product.

In order to identify a process suitable for manufacture of a formulation, it is important to first understand the critical attributes of the formulation. Secondly, it is important to have a general understanding of the process, in this case fluid bed coating, to understand the effect of the process equipment and key process parameters on the formulation. Finally, these

bodies of knowledge should be combined to design the most appropriate process for manufacture of a specific formulation. This chapter is organized accordingly. We will first introduce the formulation and its mechanism of release in order to understand the critical product attributes. Next we will discuss the fluid bed coating process from a general perspective. Finally, we will combine knowledge of the formulation and mechanism of release with this general process understanding to discuss specific fluid bed coating process considerations for the rupturing multiparticulate formulation.

16.2 ARCHITECTURE AND FORMULATION

The architecture of the multiparticulate system referred to in this chapter is modeled after formulations described by Ueda [1, 2] and Dashevsky [3]. The release of drug from a rupturing multiparticulate occurs when water passes through a delayed release functional coating into the multiparticulate, builds pressure, and eventually ruptures the delayed release coating, allowing the drug in the multiparticulate to be released. The rupturing multiparticulate is composed of a seed core, surrounded by drug and sweller layers. The final layer is composed of a semipermeable polymer that controls the rate of water ingress. This “delayed release” layer is considered a functional coating since it governs the drug release rate from the multiparticulate. Figure 16.1 illustrates the multiparticulate architecture.

The various layers of the multiparticulate must have certain attributes in order to achieve the target release profile. In

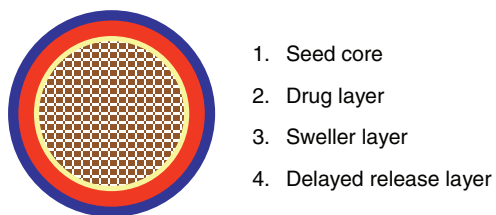


FIGURE 16.1 Osmotic rupturing multiparticulate architecture.

addition, all components that contact the drug must be chemically compatible with it. The general attributes of each component are described below, followed by the specific components used in the model system.

The drug in the model system is adiplon, a small molecule that has relatively high solubility over a physiologically relevant pH range (>2 mg/ml from pH 1–8). For the rupturing multiparticulate formulation, the desired dose was low (<10 mg). The desired release profile was an immediate release dose followed by a pulse of drug one to three hours after administration. Formulations for delayed release multiparticulates with one, two, and three hours lag times were identified. The osmotic rupturing multiparticulate was selected as the lead approach for pulsatile release. The exact composition of the adiplon multiparticulate formulation that was developed is shown in Table 16.1. The rationale for choosing the specific components for each layer is presented below.

The seed core material should provide an inert, durable, and smooth substrate for coating, and the size should be appropriate to achieve the target dose in the desired capsule size. Spherical microcrystalline cellulose (Celphere CP-708) with a mean particle size of 700 μm was used in the model system.

The drug layer should provide immediate drug release once the delayed release coating has ruptured. A water-soluble binder, HPMC E5 premium, was used in the model system with a high drug loading (75%A). The relatively high drug loading was selected to reduce processing time.

The sweller layer should provide sufficient driving force in the form of water activity at a sufficiently rapid rate to hydrate the core up to the point of rupture. The swelling component should be uniformly distributed to provide a smooth coating surface. Milled croscarmellose sodium (Ac-Di-Sol) with a water-soluble binder (HPC, Klucel EF) was used in the model system.

The delayed release layer has several properties that are critical to the performance of the system. The coating contains a semipermeable polymer that controls water ingress. This makes both the morphology and thickness of the coating critical to the

TABLE 16.1 Example Adiplon Multiparticulate Formulation Composition

Layer	Amount (mg/g Final Multiparticulates)	Component (% in Layer)	Name	Function
Seed core	391.5	700 μm Celphere	Microcrystalline cellulose (Celphere CP-708)	Substrate
Drug layer	14.6	HPMC (25%)	Hypromellose (E5 premium)	Binder
	43.9	API (75%)	API (milled)	Active
Sweller layer	128.6	HPC (28.6%)	HPC (Klucel EF)	Binder
	321.4	Ac-Di-Sol (71.4%)	Croscarmellose sodium (Ac-Di-Sol, milled)	Sweller
Delayed release layer	50	Talc (50%)	Talc (IMP-1889 L)	Coating strength modifier
	50	Ethyl cellulose (50%)	Ethylcellulose (Ethocel STD 10 cP)	Semipermeable polymer

performance. The coating in this layer must also fail via fracture or rupture; ideally the coating is brittle enough that it does not require a large degree of swelling prior to rupture. The semipermeable polymer selected was ethylcellulose. This polymer is commonly used in osmotic systems. A coating strength modifier, talc, was also used to increase the brittleness of the coating. Figure 16.2 shows the effect of two levels of talc on performance. Increasing the level of talc resulted in a much sharper burst, indicating more of the multiparticulates burst in a narrower time window. Figure 16.3 shows the performance of the one, two, and three hour delayed release formulations. All formulations had the same composition differing only in the thickness or coating weight of the delayed release layer. In order to further understand the mechanism of drug release from this

system, a more detailed explanation of the mechanism of release is given in the next section.

16.3 MECHANISM OF RELEASE

The physical model for the mechanism of release from rupturing multiparticulates is shown in Figure 16.4. The target release profile for a rupturing multiparticulate formulation is shown in Figure 16.5. The release profile has two primary components: the lag time, defined as the time between aqueous exposure of the multiparticulates and when rupture of the functional coating is initiated, and the duration of release, defined as the time between when rupturing begins and when drug release from the multiparticulates is substantially complete.

If a dissolution profile on a single multiparticulate were performed, it would show a lag time dependent on the thickness and composition of the delayed release coating and would have an immediate release profile once it ruptured (see Figure 16.6). However, since a dosage form consists of hundreds or thousands of multiparticulates, not one single multiparticulate, there is a distribution of final multiparticulate size, with small cores with thin coatings on one end of the distribution and large cores with thick coats on the other (see Figure 16.7). If this distribution is then translated to predicted performance, multiparticulates with different coating thicknesses would be expected to rupture at different times. If many of these multiparticulates are in a dosage form, then the overall dissolution profile is the composite of many individual dissolution profiles, and the overall dissolution profile will have a much broader duration of release (see Figure 16.8). The lag time for the dosage form is then defined as the amount of time prior to the first multiparticulate rupturing, and the duration of release is proportional to the breadth of the coating weight distribution.

The mechanism of release from rupturing multiparticulates can be presented mathematically. The water uptake by the multiparticulates can be represented by Equation 16.1, as demonstrated for osmotic systems by Theeuwes [4]:

$$\frac{dV_w}{dt} = \frac{A}{h} L_p (\sigma \Delta \Pi - \Delta P) \quad (16.1)$$

where

V_w is the volume of water in the multiparticulate.

t is time.

A is the cross-sectional area of the coating.

h is the coating thickness.

L_p is the mechanical permeability of the coating to water.

σ is the reflection coefficient.

$\Delta \Pi$ is the osmotic pressure difference across the coating.

ΔP is the hydrostatic pressure difference across the coating.

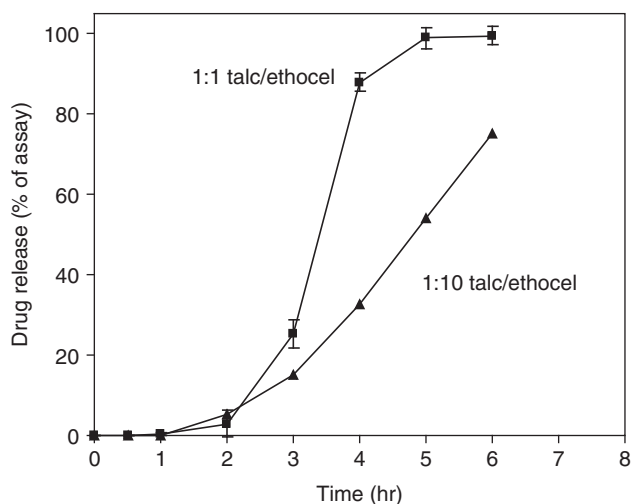


FIGURE 16.2 Effect of talc level in delayed release layer on performance of adipiplon multiparticulates.

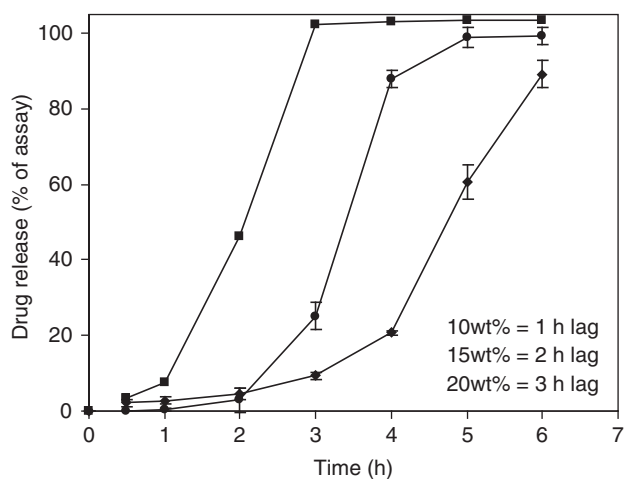


FIGURE 16.3 Effect of delayed release layer coating weight on performance of adipiplon multiparticulates.

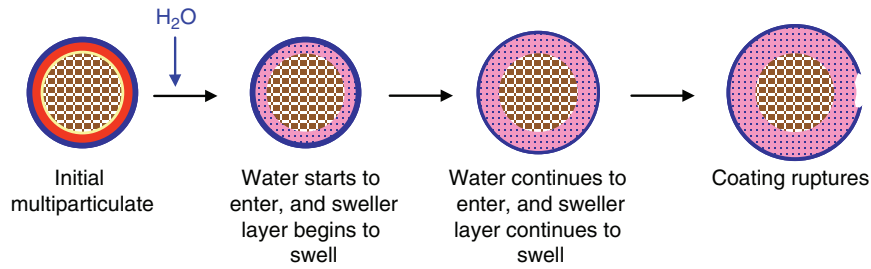


FIGURE 16.4 Mechanism of release for rupturing multiparticulates.

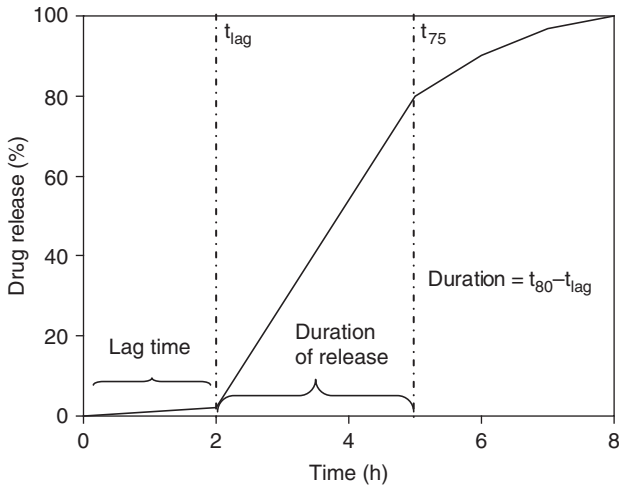


FIGURE 16.5 Example release profile of rupturing multiparticulates.

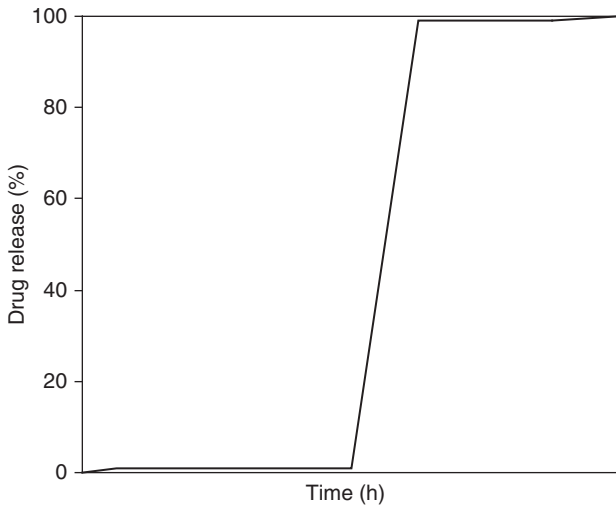


FIGURE 16.6 Single bead dissolution.

From a practical formulation approach, using this equation to guide formulation selection ensures the multiparticulate is formulated such that the osmotic pumping term is always significantly greater than the hydrostatic pressure

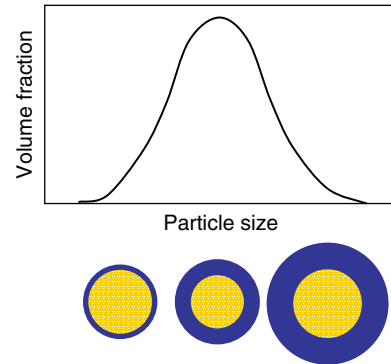


FIGURE 16.7 Coated particle size distribution and bead population.

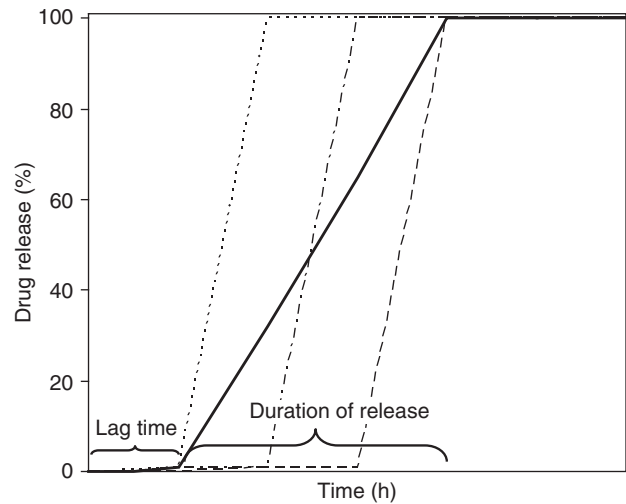


FIGURE 16.8 Composite dissolution profile.

resistance term. This will ensure that the direction of water flow is into the multiparticulates (e.g. $dV_w/dt > 0$) and that the multiparticulates will ultimately rupture.

To examine the hydrostatic pressure difference, ΔP , it is useful to use Laplace's law, as shown in Equation 16.2. This law describes the relationship of pressure of a sphere to wall tension or stress:

$$\Delta P = \frac{2E\epsilon_c h}{r} \quad (16.2)$$

where

E is the modulus of elasticity for the semipermeable coating.

ϵ_c is the engineering strain of the system (deformation of the system).

r is the radius of the multiparticulate.

Equation (16.2) assumes that the coating is perfectly elastic and does not yield (i.e. it displays Hookean behavior, where the strain is directly proportional to the stress) while it expands and that the multiparticulates are spherical with the coating thickness significantly smaller than the radius of the multiparticulate. For this system these are reasonable simplifying assumptions. For $dV_w^*/dt > 0$ (where * denotes the value at rupture), we can combine Equations (16.1) and (16.2) to obtain Equation (16.3):

$$\sigma \Delta \Pi > \frac{2E\epsilon_c^* h^*}{r^*} \quad (16.3)$$

Practically, this equation tells us that the pressure required to rupture the coating is directly proportional to the coating's modulus of elasticity, E ; the engineering strain, ϵ_c (a measure of how much the coating has changed in size); and the thickness of the semipermeable coating. Also, the pressure required to rupture the coating is inversely related to the radius of the multiparticulate, i.e. smaller multiparticulates are harder to rupture.

To summarize, from the preceding discussions of the formulation architecture and mechanism of release, critical product attributes include the size and distribution of seed cores, the uniformity of all coating layers, the potency of the drug and sweller layers, and the thickness and morphology of the delayed release layer. Now that the factors governing the release of the drug from the multiparticulates have been discussed, the next section covers understanding the fluid bed coating process both generally and specifically as it pertains to manufacturing this dosage form.

16.4 PROCESS

The primary process used to manufacture the rupturing multiparticulates is Wurster fluid bed coating. In this chapter, when we refer to fluid bed coating, it is always Wurster fluid bed coating. In order to obtain the target performance and have a robust and well-characterized process, it is important to have a good understanding of the fluid bed coating unit operation.

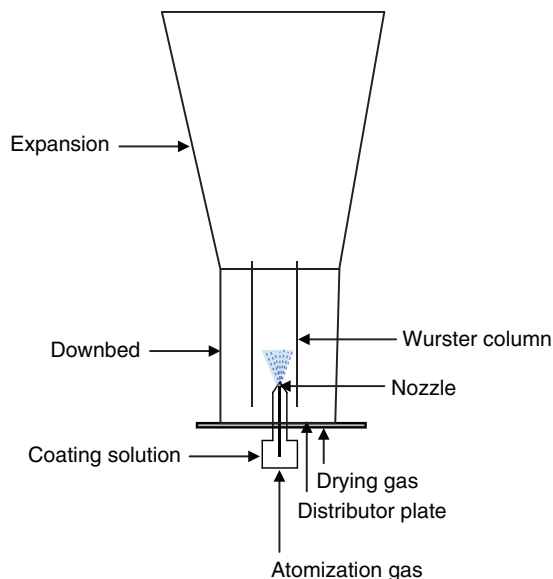


FIGURE 16.9 Fluid bed coater.

A schematic of a bottom spray fluid bed coater with Wurster column is shown in Figure 16.9. The fluid bed coater consists of an air distributor plate, a nozzle, a Wurster column, the expansion chamber, and the downbed, atomizing air and fluidizing airflow from the bottom to the top of the coater.

The fluid bed coating process can be envisioned as controlled circulation of particles through the Wurster column, where coating is applied via coating solution droplets from a two-fluid atomizer. Hot drying gas is introduced at the bottom of the fluid bed where the design of the distributor plate causes the majority of the gas to go up through the Wurster column. As a result, the velocity in the column is much higher than in the downbed, and there is a pressure differential between the column and downbed, causing particles to move from the downbed into the column. These entrained particles are transferred upward through the column where coating is deposited and partially dried. As the particles leave the column, they enter the expansion chamber where the particles continue to dry and the gas velocity drops (due to the equipment geometry) below the minimum entrainment velocity and the particles disengage from the gas stream and fall back to the downbed to start the cycle over again.

There are many subprocesses that occur in a Wurster fluid bed coating process. All of these subprocesses must be working appropriately to successfully coat particles. However, in the context of this chapter, the main phenomena that need to be considered in the fluid bed coating process are (i) the frequency of coating and circulation of individual particles as they affect the uniformity of the coating and (ii) coating deposition and drying as it directly affects the coating quality. The first phenomenon will be discussed qualitatively in this

chapter, while the second phenomenon will be addressed in more detail.

It is important to have a qualitative understanding of the frequency of coating individual particles as it can have a direct impact on the coating uniformity. First of all, it is important to understand that when particles pass through the Wurster column, only a small percentage are actually coated on each pass [5, 6]. The consequence is that in order to uniformly coat all particles, the coating process needs to be long enough to ensure a sufficient number of passes through the spray zone for each particle. In some cases, the coating solution solids content may need to be adjusted to ensure a sufficiently long coating time.

Particle circulation through the fluid bed must also be consistent for all particles in order to achieve uniform coating. Preferential entrainment due to static or particle size or uneven fluidization of the downbed can lead to nonuniform coating across particles during the process. Practically, particle circulation can be assessed through visual observations of fluidization and can be improved through the use of a narrower particle size distribution of cores and/or through increasing bed humidity to minimize static accumulation.

As the particles circulate through the Wurster column, coating solution droplets are deposited onto the particles. As the particles continue to circulate through the fluid bed, the coating droplets then dry. The properties of these droplets and the environmental conditions such as temperature and solvent concentration within the fluid bed directly affect the drying rate of both the droplets prior to deposition and the deposited coating. The drying rate can have a significant impact on the coating morphology and therefore the drug release rate if it is a functional coating. At faster drying rates the coating droplets contain less solvent and produce more porous coatings. These more porous coatings can be both more permeable and mechanically weaker than coatings applied under wetter conditions. Thus it is important to choose an appropriate drying rate for a particular coating process and ensure that it is maintained throughout the process.

Droplet size can also have a significant effect on the coating properties, since it impacts the drying rate. There are four variables that affect the droplet size: atomization gas flow rate (commonly controlled by pressure), atomizer design, spray rate, and solution properties. From a practical standpoint the last three variables are usually fixed for a given process and equipment train, leaving the atomization gas flow rate as the most common process variable. For most solutions, once there is sufficient atomizing gas flow to fully atomize the solution, there is generally a small effect of atomizing gas on droplet size (see Figure 16.10) [7].

The three process parameters that affect the droplet drying/coating formation conditions are drying gas flow rate, drying gas temperature, and spray rate. Practically, drying gas flow rate is constrained as it is coupled to particle

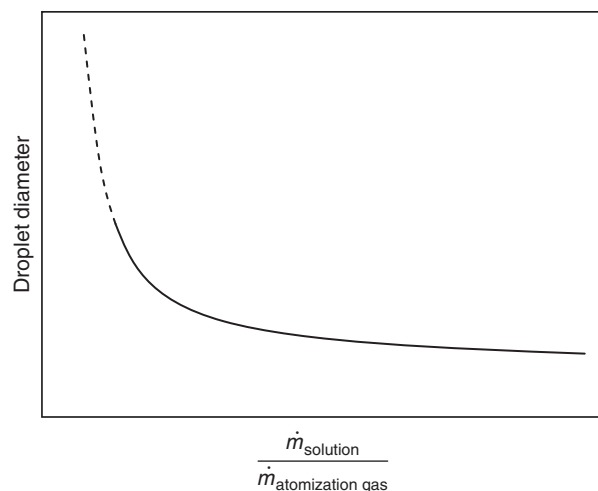


FIGURE 16.10 Droplet size as a function of $\dot{m}_{\text{solution}}/\dot{m}_{\text{atomization gas}}$.

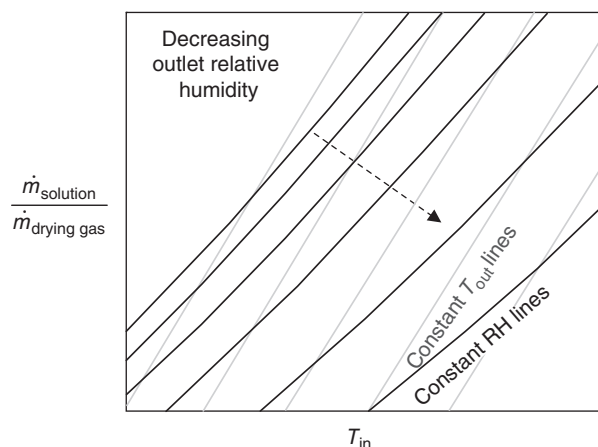


FIGURE 16.11 Outlet temperature and relative humidity as a function of $\dot{m}_{\text{solution}}/\dot{m}_{\text{drying gas}}$ versus T_{in} .

circulation, leaving spray rate and inlet temperature as variables. It is often helpful to evaluate the effect of these parameters on the driving forces for heat and mass transfer, namely, the dependent variables of bed temperature and solvent concentration. Figure 16.11 shows such a plot for an aqueous coating system with constant bed temperature and humidity lines. Such a plot can be constructed for any solvent/drying gas coating system using mass and energy balances and is demonstrated later in this chapter.

In the upper left-hand corner of the plot, there is a low driving force for mass transfer and thus slower drying rates. Moving to the lower right-hand corner, the solvent concentration decreases and temperature increases, resulting in much faster drying rates and thus decreasing coating efficiency and increasing coating porosity.

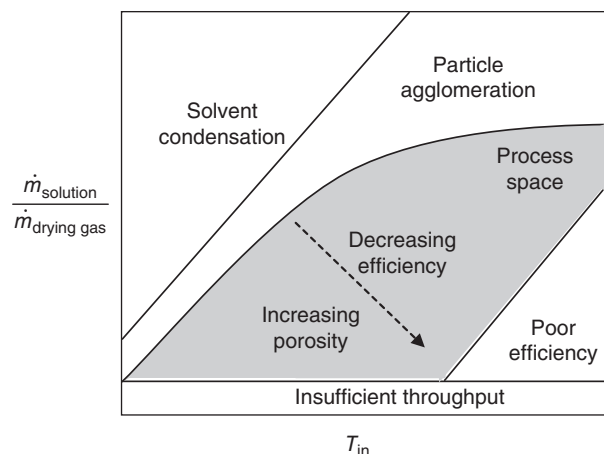


FIGURE 16.12 Conceptual process space.

This plot can be overlaid with the practical limitations of the coating process (some of which are determined experimentally) as shown in Figure 16.12 to define the acceptable processing space. This process map can be used as a guide to understand the effect of process variables on product properties within the process space.

16.4.1 Mass and Energy Balance

In order to understand the overall process on a macroscopic scale and draw the process plots above, it is necessary to evaluate the overall mass and energy balance for the unit operation. The objective is to evaluate the overall system when operating at steady state with a control volume drawn around the entire system, as shown in Figure 16.13. Using the principles of the conservation of mass and the first law of thermodynamics, the inputs of this control volume must be equal to the outputs plus any accumulation that occurs in the control volume.

The inputs into the control volume are:

- The coating solution’s mass flow rate, temperature, and solids content.
- The drying gas’s mass flow rate, inlet temperature, and relative humidity.
- The atomization gas’s mass flow rate, temperature, and relative humidity.

The outputs from the process are:

- The exiting gas stream’s mass flow rate, outlet temperature, and relative humidity.
- The heat loss from the system.

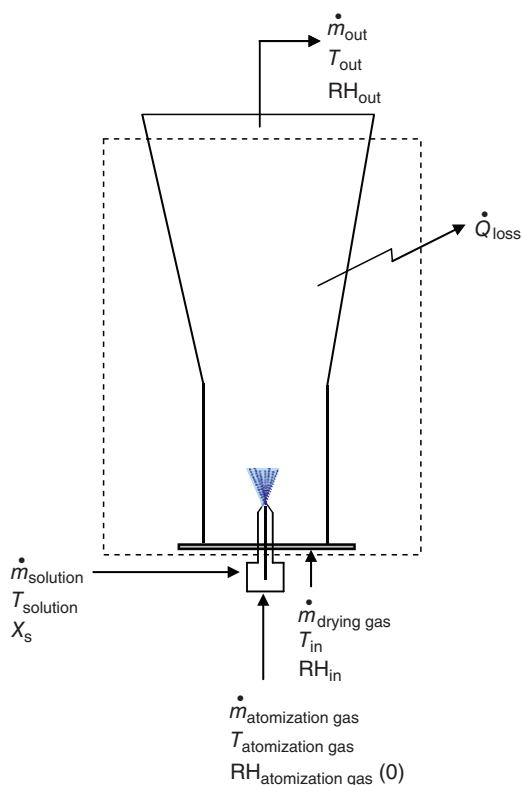


FIGURE 16.13 Control volume for fluid bed coater.

If an energy balance on the system at steady state is done and it is assumed that potential and kinetic energy changes are negligible, the following equation can be written:

$$\dot{m}\Delta H = \dot{Q} + \dot{W} \tag{16.4}$$

where

$\dot{m}\Delta H$ is the change in enthalpy for the system.

\dot{Q} is the rate of heat flow into the system.

\dot{W} is the rate of work done on the system.

It is assumed that there is negligible work done on the system and that the enthalpy contributions of the spray solids and atomizing gas are negligible. Therefore, only the enthalpy contributions of the drying gas and the solution are significant. Furthermore, it is assumed that the heat of vaporization is constant with respect to temperature and the spray solvent is completely evaporated. Under these conditions the energy balance can be written:

$$\begin{aligned} \dot{m}_{d\ gas} C_{p\ d\ gas} \Delta T_{d\ gas} + \dot{m}_{solution} (1 - x_s) C_{p\ solution} \Delta T_{solution} \\ + \dot{m}_{solution} (1 - x_s) \lambda_v = \dot{Q} \end{aligned} \tag{16.5}$$

where

- \dot{m} is the mass flow rate of the feed solution.
- C_p is the specific heat of the feed solution.
- x_s is the solid fraction in the coating solution.
- λ_v is the enthalpy of vaporization.

This equation can be rearranged to solve for the drying gas outlet temperature to construct the lines in Figure 16.11. Similarly, the lines of constant humidity (or solvent) can be constructed using psychrometric principles [8].

Now that the formulation, the mechanism of release, and the general fluid bed coating process have been discussed, we can use this information to understand critical process parameters for manufacture of rupturing multiparticulates.

16.4.2 Process Considerations Specific to Rupturing Multiparticulates

As was discussed in Section 16.3, critical product attributes for rupturing multiparticulates are particle size and distribution of the seed core, uniformity of the coating for all layers and the coating morphology, and thickness for the delayed release layer. Below specific processing considerations for rupturing multiparticulates that impact the critical product attributes are discussed.

16.4.2.1 Seed Core Many types of particles with a wide range of sizes can be coated in a fluid bed coater. In general with smaller particles, minimizing static, ensuring uniform coating thickness particle to particle, and running the process with high efficiency is more challenging than for large particles. Large particles present their own set of process challenges such as maintaining good fluidization while minimizing attrition.

The key properties of the seed core that are important to achieving a uniform coating for all layers are the particle size and size distribution. The seed core size and distribution serves as the basis for the overall final multiparticulate size and distribution. As discussed in the mechanism section, this is important because the thickness of the coating on each particle affects when that multiparticulate bursts.

Based on this understanding, it is very important in cases where a sharp delayed release pulse is desired to start the process with as narrow a particle size distribution of seed cores as possible. Furthermore, it is critical to run the process to minimize the breadth of the size distribution for the final coated multiparticulates. In some cases, sieving the starting seed cores to narrow the distribution may be appropriate.

16.4.2.2 Uniformity of Coatings: Drug Layer and Sweller Layer The critical product attributes for the drug and sweller layers are uniformity of the coating and achieving target potency of either drug or sweller. To ensure uniformity of the

coating, good atomization, sufficiently long process time, and good particle circulation are required during the process. To ensure that the potency of the drug and sweller is as desired, the coating efficiency should be high (i.e. efficiency as a function of process conditions should be understood and maximized).

Particle size can be a factor for materials applied as suspensions. A rule of thumb for coating multiparticulates is that particles in suspension should be at least one order of magnitude smaller in their longest dimension than the size of the core being coated. This aids in obtaining high coating efficiency.

16.4.2.3 Delayed Release Layer In addition to uniformity of the coating being critical for the delayed release layer, as we have seen in our discussion of mechanism of release, the morphology of this coating (e.g. porosity) is also critical to performance. As discussed, the drying rate of droplets has significant impact on coating porosity. Based on this, the acceptable process space for coating this layer is further constrained by not only efficiency but also product performance. Practically, once the process map introduced in Section 16.4 is established for this process, it can be used to define a range of processing parameters that results in acceptable product performance.

16.5 SUMMARY

The key to developing a successful novel dosage form from a processing perspective is understanding the underlying mechanism of release of the dosage form, the unit operations used in the process, and how the key process variables impact the product properties. While this can be more complex with novel dosage forms, the principles remain the same for all dosage form development.

16.6 PROBLEMS

1. An ethylcellulose-based delay-release coating, having an elastic modulus of 500 MPa and elongation at failure of 3%, is applied to a formulated core that is 1 mm in diameter. The coating is applied to achieve 25 wt% coating weight. Assume the coating and formulated core have similar densities and the final dosage form has a displaceable volume of 0.2 mm³.
 - a. What is the hydrostatic pressure required to rupture the coating?
 - b. What is the volume of water required to rupture the coating?
2. Given the information from the previous problem and assuming a mechanical permeability of the membrane

of $5 \times 10^{-7} \text{ cm}^2/\text{atm h}$, a reflection coefficient of 1, and a constant osmotic pressure difference of 50 atm between the media and the core, what is the approximate time to hydrate the core?

3. An aqueous suspension of ethylcellulose, TEC, and talc at 20% w/w total solids is being coated onto formulated multiparticulates in a fluid bed coater. The drying gas is conditioned to an inlet temperature of 55 °C and a measured dew point of 10 °C. The air flow rate is 600 cfm and the solution spray rate is 15 kg/h. Assume all the solvent is evaporated, the solution is at 20 °C, and the rate of heat loss is approximately 20 kJ/min.
 - a. Estimate the temperature of the bed.
 - b. What is the relative humidity of the exhaust gas?

16.7 PROBLEM SOLUTIONS

Solution

- a. Based on the applied coating, the coating thickness and the starting particle radius must first be determined. Assume that $h \ll r$ and the coating and core are similar density:

$$h = \frac{1 \rho_{\text{bead}}}{3 \rho_{\text{coat}}} \left(\frac{X_{\text{coat}}}{1 - X_{\text{coat}}} \right) \cdot r$$

$$h_0 = \frac{1}{3} (1) \left(\frac{0.25}{1 - 0.25} \right) \left(\frac{1.00 \text{ mm}}{2} \right) = 0.056 \text{ mm} = 56 \mu\text{m}$$

$$r_0 = \left(\frac{1.00 \text{ mm}}{2} + 0.056 \text{ mm} \right) = 0.556 \text{ mm} = 0.56 \text{ mm}$$

Plug these values into the equation for the hydrostatic pressure difference term at the point of failure indicated below:

$$\Delta P^* = 2E \frac{h_0}{r_0} \frac{\epsilon_e^*}{(1 + \epsilon_e^*)^3}$$

$$\Delta P^* = 2(500 \text{ MPa}) \left(\frac{0.056 \text{ mm}}{0.556 \text{ mm}} \right) \left(\frac{0.03}{(1.03)^3} \right) = 2.8 \text{ MPa} = 27 \text{ atm}$$

- b. The amount of water is equal to the displaceable volume plus the change in volume of the expanding particle (based on its strain). Use the definition of strain to define what the volume of the particle will be at the point of coating failure indicated below:

$$V_{\text{bead}}^* = V_0 (1 + \epsilon_e^*)^3$$

$$V_w^* = V_d + V_0 \left[(1 + \epsilon_e^*)^3 - 1 \right]$$

$$V_w^* = 0.2 \text{ mm}^3 + \frac{4\pi}{3} (0.556 \text{ mm})^3 (1.03^3 - 1) = 0.27 \text{ mm}^3$$

Solution 2

Use the differential equation for the volumetric flow rate of water through the semipermeable membrane, Equation (16.1) from the chapter. By assuming that the coating thickness and surface area do not change during the time of hydrating the core, the equation can be simplified since the hydrostatic pressure difference is negligible and integrated to result the following relationship:

$$\int_0^{V_d} dV_w = \frac{4\pi r_0^2}{h_0} L_p \sigma \Delta \Pi \cdot \int_0^t dt$$

Solving for time,

$$t_{\text{hydrate}} = \frac{V_d h_0}{4\pi r_0^2 L_p \sigma \Delta \Pi}$$

$$t_{\text{hydrate}} = \frac{2 \times 10^{-4} \text{ cm}^3 (0.0056 \text{ cm})}{4\pi (0.0556 \text{ cm})^2 \left(5 \times 10^{-7} \frac{\text{cm}^2}{\text{atm} \cdot \text{h}} \right) (1) (50 \text{ atm})} = 1.2 \text{ h}$$

Solution 3

- a. Based on the defined inputs of inlet temperature and dew point for the drying gas, a psychrometric chart [9] can be used to define the properties of the drying gas, namely, the specific volume and humidity, determined to be 0.94 m³/kg dry air (DA) and 0.0076 kg/kg DA, respectively. The heat capacity of water was taken to be 4.186 kJ/kg°C, and enthalpy of vaporization 2390 kJ/kg (for 310 K). The specific heat for the moist air was interpolated to be 1.026 kJ/kg°C based on the individual values for air and water.

First determine the mass flow rates of both dry and moist air:

$$\dot{m}_{\text{DA}} = \left(\frac{600 \text{ ft}^3}{\text{min}} \right) \left(\frac{\text{m}^3}{35.315 \text{ ft}^3} \right) \left(\frac{\text{kgDA}}{0.94 \text{ m}^3} \right) = 18.07 \frac{\text{kg}}{\text{min}}$$

$$\dot{m}_{\text{d gas}} = \frac{18.07 \text{ kg}}{\text{min}} + \frac{18.07 \text{ kgDA}}{\text{min}} \left(\frac{0.0076 \text{ kg}}{\text{kgDA}} \right) = 18.21 \frac{\text{kg}}{\text{min}}$$

Rearrange Equation (16.5) from the chapter, and solve for the outlet temperature using units on a per minute basis (i.e. 15 kg/h is 0.25 kg/min):

$$T_{\text{OUT}} = \frac{\dot{m}_{\text{d gas}} C_{p \text{ dgas}} T_{\text{IN}} + \dot{m}_{\text{soln}} (1 - x_S) C_{p \text{ soln}} T_{\text{soln}} + \dot{Q} - \dot{m}_{\text{soln}} (1 - x_S) \lambda_V}{\dot{m}_{\text{d gas}} C_{p \text{ dgas}} + \dot{m}_{\text{soln}} (1 - x_S) C_{p \text{ soln}}}$$

$$T_{\text{OUT}} = \frac{18.21 (1.026) (55) + 0.25 (1 - 0.20) (4.186) (20) + (-20) - 0.25 (1 - 0.20) (2390)}{18.21 (1.026) + 0.25 (1 - 0.20) (4.186)} = 28 \text{ }^\circ\text{C}$$

- b. Evaluate the mass balance for water in the system, and solve for the outlet humidity assuming complete evaporation of solvent:

$$H_{\text{OUT}} = H_{\text{IN}} + \frac{\dot{m}_{\text{soln}}(1-x_S)}{\dot{m}_{\text{DA}}}$$

$$H_{\text{OUT}} = 0.0076 + \frac{0.25(1-0.20)}{18.07} = 0.0187 \frac{\text{kg}}{\text{kgDA}}$$

The psychometric chart can be used to interpolate the relative humidity of the exhaust gas at the calculated outlet temperature of 28 °C, or alternatively the humidity can be converted into a partial pressure of water vapor and Antoine's equation used to calculate the saturation pressure for the solvent at the calculated outlet temperature:

$$\text{RH}_{\text{OUT}} = 78\%$$

ACKNOWLEDGMENT

The authors wish to acknowledge Neurogen Corporation for allowing us to use adiplon osmotic rupturing multiparticulate data in the presentation of the material in this chapter.

REFERENCES

1. Ueda, S., Hata, T., Asakura, S. et al. (1994). Development of a novel drug release system, time-controlled explosion system (TES). I. Concept and design. *Journal of Drug Targeting* 2: 35–44.
2. Ueda, S., Yamaguchi, H., Kotani, M. et al. (1994). Development of a novel drug release system, time-controlled explosion system (TES). II. Design of multiparticulate TES and in vitro drug release properties. *Chemical and Pharmaceutical Bulletin* 42 (2): 359–363.
3. Dashevsky, A. and Mohamad, A. (2006). Development of pulsatile multiparticulate drug delivery system coated with aqueous dispersion Aquacoat® ECD. *International Journal of Pharmaceutics* 318: 124–131.
4. Theeuwes, F. (1975). Elementary osmotic pump. *Journal of Pharmaceutical Sciences* 64 (12): 1987–1991.
5. Cheng, X.X. and Turton, R. (2000). The prediction of variability occurring in fluidized bed coating equipment. I. The measurement of particle circulation rates in a bottom-spray fluidized bed coater. *Pharmaceutical Development and Technology* 5 (3): 311–322.
6. Cheng, X.X. and Turton, R. (2000). The prediction of variability occurring in fluidized bed coating equipment. II. The role of non-uniform particle coverage as particles pass through the spray zone. *Pharmaceutical Development and Technology* 5 (3): 323–332.
7. Masters, K. (1985). *Spray Drying Handbook*, 4e, 236–252. New York: Wiley.
8. McCabe, W., Smith, J., and Harriott, P. (2001). *Unit Operations of Chemical Engineering*, 6e, 596–621. New York: McGraw-Hill.
9. Moyers, C. and Baldwin, G. (1997). Psychrometry, evaporative cooling, and solids drying. In: *Perry's Chemical Engineers' Handbook*, 7e (ed. R. Perry, D. Green and J. Maloney), 12–15. New York: McGraw-Hill.

MULTISCALE MODELING OF A PHARMACEUTICAL FLUID BED COATING PROCESS USING CFD/DEM AND POPULATION BALANCE MODELS TO PREDICT COATING UNIFORMITY

AVIK SARKAR, BRIAN SHOEMAKER[#], PANKAJ DOSHI, AND MARY T. AM ENDE*
Process Modeling & Engineering Technology Group, Pfizer Inc., Groton, CT, USA

DALIBOR JAJCEVIC, PETER BÖHLING, PETER TOSON, AND MATEJ ZADRAVEC
Research Center Pharmaceutical Engineering GmbH, Graz, Austria

JOHANNES G. KHINAST
Research Center Pharmaceutical Engineering GmbH, Graz, Austria
Institute for Process and Particle Engineering Graz University of Technology, Graz, Austria

17.1 INTRODUCTION

Fluidized bed devices are commonly used in pharmaceutical manufacturing operations to produce drug products. Examples of cases where fluidized beds are employed include granulation, drying, and coating. In this chapter, we will discuss fluid bed coating operations used to produce coated beads, specifically drug-layered beads. We will describe the fluid bed coating process using devices with a Wurster insert, which helps direct the fluidization flow patterns within the device. The objective of the coating process is to produce coated microspheres or beads (typically smaller than 1 mm) with the desired degree of coating uniformity. Such coated beads are often encountered, for example, inside filled capsules, such as commercially available Detrol[®]. These beads are intended for use by small children or elderly patients who have trouble swallowing standard-sized tablets. Multiparticulate beads can also be dosed in larger quantities, e.g. 2 g azithromycin single-dose pack, as a sachet.

The Wurster coating process will be the focus of this chapter. The manufacturing of the spherical microsphere cores that are to be coated are outside of the present scope. In particular, we will present the use of multiscale computational tools to help understand and design the fluid bed coating operation. Using these multiscale models, we will virtually quantify the bead-to-bead coating uniformity as well as the variability in the potency of the drug-layered beads. Certain simplifying assumptions are made regarding the coating spray, such as the flow or the droplets and their drying characteristics. Although these are important aspects of a coating process, it is imperative to start with a simpler approach to help understand bead coating in a more holistic manner; these assumptions may be relaxed at a later time.

A multiscale approach is developed to understand the fluid bed coating process: first, we set up and execute simulations where computational fluid dynamics (CFD) is coupled with discrete element method (DEM) to model the

*Current address: Lyndra Therapeutics, Watertown, MA, USA

[#]Current address: Department of Chemical & Environmental Engineering, Yale University, New Haven, CT, USA

flow of beads inside the coater. These simulations are computationally very expensive, and one can simulate only a few seconds or a few minutes of the flow, whereas a real coating operation may last for hours. To model the long-term coating behavior, reduced-order models are constructed – the detailed flow simulations supply the necessary inputs for these reduced-order models. These reduced-order models are compartment models that are computationally much faster and can be used to extrapolate the behavior of the system over several hours of operation. Examples of such reduced-order compartment models include Monte Carlo approaches and the Markov chain Monte Carlo method, both of which will be presently discussed.

17.2 FLUID BED COATING PROCESS

17.2.1 Wurster Coating Process: Equipment Geometry and Operation

The Wurster process has been widely utilized to coat particles and beads in the pharmaceutical industry. Beads are generally coated with some combination of the drug product (active pharmaceutical ingredient [API]), functional coatings that govern the release rate of the drug, and protective coatings that protect the inner layers from dissolving prematurely. There are two sources of airflow in this system – atomizing air to form the coating droplets and drying air to fluidize and dry the coated beads. The coating operation takes place in a fluidized bed (Figure 17.1), in which particles are fluidized by airflow supplied from a distributor plate at the bottom of the device. One or more nozzles, which are usually located at the bottom of the device for the Wurster tube configuration, supply an atomization airflow and the coating liquid in the form of fine droplets. The liquid droplets interact with the particles as they pass through the spray zone and deposit/coat the particles' surfaces. The coating liquid generally consists of several components, e.g. a mixture of solid particles and a liquid as either a solution or a suspension. The coated beads are dried by the hot airflow supplied from the distributor plate; the same airflow serves to fluidize the bed. During the drying process, the liquid components of the solution/suspension evaporate, and the remaining solid components form the coating. The coating process is run until the desired coating thickness is achieved. Other than the coating thickness, the important product quality criteria include the bead-to-bead coating uniformity and the coating quality, i.e. coat porosity, film continuity, etc. These criteria are influenced by process conditions, beads and suspension properties, and the design/size of the used fluid bed device; we will attempt to quantitatively link some of these process conditions to the desired quality criteria or attributes.

The fluid bed coating device can be divided into different regions through which the coated particles successively pass through (Figure 17.1). The particle coating cycle can be

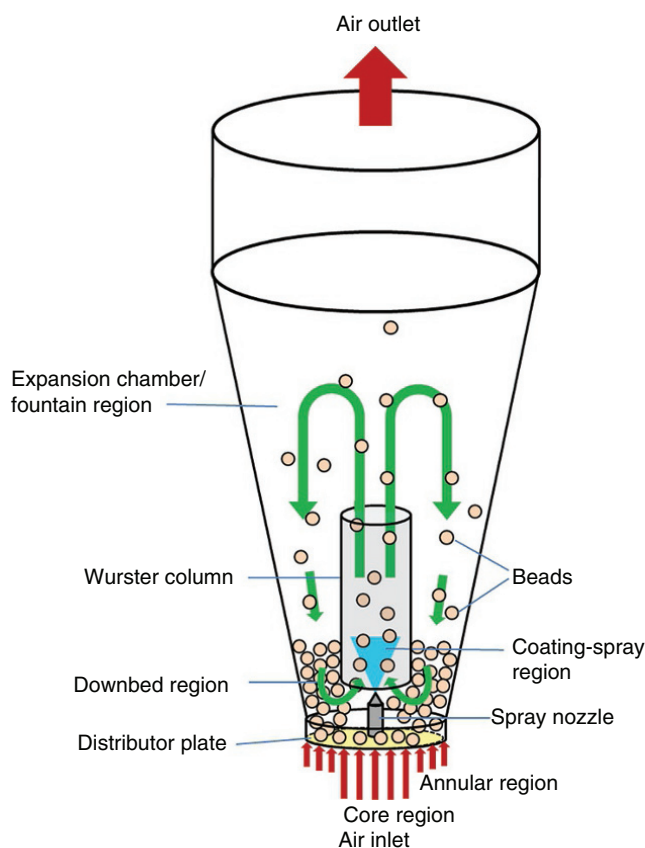


FIGURE 17.1 Schematic of a general Wurster fluidized bed coater, showing the bead flow patterns and general features, including bottom air inlet, spray nozzle, Wurster column, and top air outlet.

divided into two different parts: first, the sequence of coating and drying, wherein particles travel through the spray zone inside the Wurster tube, and, second, the subsequent drying wherein the particles move through the fountain and downbed regions before progressing to the next coating cycle.

Given this repeated flow cycling, the total amount of coating deposited onto a particle depends on the total number of particle passes through the spray zone and the amount of coating solution that the particle accumulates per pass. Therefore, it is evident that the residence time (RT) per pass in the Wurster zone and the cycle time (CT) are important factors in determining the final coating attributes; the RT per pass dictates the amount of coating accumulated per pass, whereas the CT provides the total number of passes over the duration of the coating process. In particular, quality attributes such as the film thickness and the bead-to-bead coating variability can be quantified by characterizing the Wurster-zone RT and CT. In the present chapter, the focus is on linking flow of the beads to the process parameters that influence bead trajectories to the bead-to-bead coating uniformity.

Besides these aspects, the spray and drying thermodynamics (evaporation/drying rate) play an important role in the product's quality. For instance, the atomization air flow rate

dictates the droplet size of the spray and, hence, the coating porosity and continuity. The thermodynamics of the spray and drying process are outside the scope of the current discussion, as is the intra-bead coating uniformity (i.e. the coating quality). These aspects are currently being studied and will be discussed in future publications.

Even though it is recognized that the Wurster-zone RT and CT are important aspects of the particle coating process, these quantities are difficult to obtain, and often impossible at commercial manufacturing scales, using experimental techniques. Fortunately, modern high-performance numerical simulations are available to address this gap, e.g. CFD coupled with DEM, referred to as CFD-DEM approaches, and/or CFD two-fluid model (CFD-TFM) simulations. These computational techniques allow detailed virtual interrogation of the effects of varying process parameters and/or geometry design/size on the desired flow information, viz. the RT and CT. Generally, it can be stated that the outcome of the computational investigation increases the level of process knowledge and understanding, where modern measurement techniques cannot be easily applied.

17.2.2 Experimental Measurement Techniques Applicable to Wurster Coaters

The US Food and Drug Administration (US FDA) has stressed the use of process analytical technology (PAT) as a mechanism to analyze/monitor and control pharmaceutical manufacturing processes through the measurement of critical process parameters (CPPs). Some of these CPPs are easy to measure, e.g. bed temperature, inlet/outlet air temperature, etc. For a fluidized bed, measurement of other process parameters is very difficult, and values that can be obtained are not representative of the entire batch, i.e. such measurements reflect only for some local conditions. Within a Wurster coater, the following process parameters are of interest:

- Fluidization (inlet) air flow rate.
- Inlet and outlet air temperature.
- Inlet and outlet air humidity.
- Atomization air flow rate and/or air pressure.
- Atomization air temperature.
- Spray rate.
- Bead size (online or offline).
- Coating thickness (online or offline).
- Air/gas temperature at different location within the Wurster coater.
- Static pressure at different locations inside the Wurster coater (usually before and after distributor plate and at outlet of Wurster coater).
- Loss of drying (LoD) (via probes or online measurement), which is a measure of the moisture content of the beads.
- Particle velocity, particle RT (within a zone), and CT.

Fluidization and atomization air flow rates are measured by an airflow meter (e.g. rotameter) that usually measures volumetric flow rate in a tube. The value can be estimated very precisely and represents an important process parameter.

For the measurement of the air/gas temperature, thermocouples are widely used. The fluidization and the atomization air temperature can be measured very precisely and represent the second set of important process parameters. Thermocouples are used to measure gas temperature, positioned inside the regions of interest. Note that the thermocouples do not directly measure the particle temperature. Instead, they actually measure the surrounding gas temperature, which can be assumed in equilibrium with the particle temperature.

Air humidity is the next CPP, which is usually measured for the inlet air as well as for the exhaust at exit of the Wurster coater. The air humidity can be measured, for instance, by a relative humidity sensor. The difference in humidity between the outlet and inlet to the coater depends on the amount of solvent that is evaporated from the coating spray.

The bead size distribution can be measured online or offline. The offline measurements are usually performed in such a way that the bead probes are taken out during the coating process. Applying, for instance, QICPIC, the particle size distribution can be measured by a high-speed image analyzer, which is using a pulsed light source with illumination times of less than 1 ns. The particles are optically frozen while at a high resolution and their size is evaluated.

Coating thickness can be measured online and/or offline using optical coherence tomography (OCT). OCT is an emerging optical imaging method that allows the acquisition of cross sections in a nondestructive and contactless manner. It is based on the phenomenon of white light interferometry and very sensitive to small changes in the refractive index of the sample. The outputs of the OCT online/offline measurements are, for instance, mean coating thickness and inter- and intraparticle coating variability. The results of the coating thickness measured by OCT technique are presented below.

17.3 COMPUTATIONAL METHODS TO MODEL MULTIPHASE FLOWS IN FLUIDIZED BEDS

To investigate a Wurster coater fluidized bed numerically, the simulation techniques and tools have to incorporate and model a multiphase system involving interactions between solids (particles), gas (air and vapor), and liquids (droplets and liquid deposition on the bead surfaces). There are two computational approaches that can be used, namely, coupled CFD-DEM and CFD-TFM, as introduced in Section 17.2.1.

The CFD-DEM model is based on an Euler–Lagrange approach. The gas phase is described as a continuum, i.e. as the Eulerian phase, via the Navier–Stokes conservation equations. The particle/bead phase is treated as a collection of discrete entities, i.e. the Lagrangian phase, whose

movement is tracked by applying Newton's second law of motion. The continuum and discrete phases are coupled via mass, momentum, and energy exchange. In CFD-DEM, particle contacts and collisions are explicitly resolved using the DEM approach (see Section 17.3.1).

In contrast to the CFD-DEM computational approach, CFD-TFM treats both phases as interpenetrating continua; the governing mass, momentum, and energy conservation equations are solved for both phases separately using the volume-averaged Navier–Stokes equations. Coupling between the two phases is through exchange terms in the conservation equations, most notably the gas–solid drag term.

The continuum closure laws for the gas phase are well established from single-phase literature. In contrast, continuum constitutive laws for the solid phase, especially stress relationships for frictional particles, are not yet well established. Therefore, CFD-TFM is generally less accurate than CFD-DEM but is computationally faster and, as such, well suited to large-scale simulations. In the following sections, the CFD-DEM and CFD-TFM approaches are discussed in greater detail.

17.3.1 CFD-DEM Method

17.3.1.1 Gas Phase The gas phase is modeled by solving the volume-averaged Navier–Stokes equations. All variables are considered to be locally volume-averaged quantities over a specific control volume, which is chosen to be at least one order of magnitude larger than the particle volume V_p [1]. The conservation of mass can be written as follows:

$$\frac{\partial}{\partial t}(\alpha_g \rho_g) + \nabla \cdot (\alpha_g \rho_g \mathbf{u}_g) = 0 \quad (17.1)$$

where

t is time.

α_g is the local volume fraction of the fluid.

ρ_g is the fluid (gas) density.

\mathbf{u}_g is the fluid velocity vector.

Similarly, the conservation of the momentum is

$$\frac{\partial}{\partial t}(\alpha_g \rho_g \mathbf{u}_g) + \nabla \cdot (\alpha_g \rho_g \mathbf{u}_g \mathbf{u}_g) = -\alpha_g \nabla p - \nabla \cdot (\alpha_g \boldsymbol{\tau}_g) + \alpha_g \rho_g \mathbf{g} - \mathbf{S}_M \quad (17.2)$$

where

p is the hydrodynamic fluid (gas) pressure.

$\boldsymbol{\tau}_g$ is the stress tensor.

\mathbf{g} is the gravitational acceleration.

\mathbf{S}_M is the interphase momentum transfer between particles and fluid, i.e. the gas-particle drag term.

The stress tensor is defined as

$$\boldsymbol{\tau}_g = -p\mathbf{I} + \boldsymbol{\sigma}_g \quad (17.3)$$

where

\mathbf{I} is the 3×3 unit/identity matrix.

$\boldsymbol{\sigma}_g$ is the viscous stress tensor, given by

$$\boldsymbol{\sigma}_g = \mu_g \left(\nabla \mathbf{u}_g + \nabla \mathbf{u}_g^T \right) - \frac{2}{3} \mu_g \nabla \cdot \mathbf{u}_g \mathbf{I} \quad (17.4)$$

with μ_g being the dynamic viscosity. The turbulent flow was modeled using standard k-epsilon turbulence model.

The coupling between the CFD and the DEM code is realized via the momentum source term S_M in Eq. (17.2), which represents the gas-particle drag force. The momentum source term, experienced by the fluid on the CFD side, is a sum of all drag forces of all particles in a DEM cell:

$$S_M = \sum_{N_p} \frac{V_p \beta}{\alpha_s} (\mathbf{u}_g - \mathbf{u}_p) \quad (17.5)$$

Several commonly applied drag models reported in the literature (e.g. Refs. [2–6]) describe the interphase momentum transfer coefficient β as a function of the solid α_s and fluid α_g volume fractions, the particle Reynolds number Re_p , the particle diameter d_p , and the particle density ρ_p . In our work, the Gidaspow [4] model was used, which combines the Ergun [7] equation for dense regimes and a correlation proposed by Wen and Yu [8] for more dilute regimes:

$$\beta = \begin{cases} \beta_{\text{Ergun}} = 150 \frac{\alpha_s^2 \mu_f}{\alpha_g d_p^2} + 1.75 \frac{\alpha_s \rho_f}{d_p} |\mathbf{u}_f - \mathbf{u}_p|, & \text{for } \alpha_g < 0.8 \\ \beta_{\text{Wen-Yu}} = \frac{3}{4} C_D \frac{\alpha_s \rho_g}{d_p} |\mathbf{u}_g - \mathbf{u}_p| \alpha_g^{-2.65}, & \text{for } \alpha_g \geq 0.8 \end{cases} \quad (17.6)$$

where the drag coefficient C_D and the particle Reynolds number Re_p are expressed as

$$C_D = \begin{cases} \frac{24}{\alpha_g Re_p} \left[1 + 0.15 (\alpha_g Re_p)^{0.687} \right], & \text{for } Re_p < 1000 \\ 0.44, & \text{for } Re_p \geq 1000 \end{cases} \quad (17.7)$$

$$Re_p = \frac{\alpha_g |\mathbf{u}_g - \mathbf{u}_p| d_p}{\mu_g} \quad (17.8)$$

To avoid discontinuity between the two equations, Gidaspow [4] introduced an additional function that describes the transition between the regimes:

$$\phi_{gs} = \frac{\arctan[150 \times 1.75 \cdot (0.2 - \alpha_s)]}{\pi} + 0.5 \quad (17.9)$$

Using Eq. (17.24), the interphase momentum transfer coefficient can be calculated as

$$\beta = (1 - \phi_{gs})\beta_{\text{Ergun}} + \phi_{gs}\beta_{\text{Wen-Yu}} \quad (17.10)$$

17.3.1.2 Solid Phase The movement of each particle (considered being a point source) is described by Newton's equation of motion [9]:

$$m_p \frac{d\mathbf{u}_p}{dt} = -V_p \nabla p + \frac{V_p \beta}{\alpha_s} (\mathbf{u}_g - \mathbf{u}_p) + \sum_{N_p} \mathbf{F}_{p \rightarrow p} + \sum_{N_w} \mathbf{F}_{p \rightarrow w} + m_p \mathbf{g} \quad (17.11)$$

where

- m_p is the particle mass.
- \mathbf{u}_p is the particle velocity.
- $-V_p \nabla p$ is the pressure gradient force (i.e. buoyancy force).
- β is the interphase momentum transfer coefficient (drag coefficient).

The terms $\mathbf{F}_{p \rightarrow p}$ and $\mathbf{F}_{p \rightarrow w}$ represent the interaction forces for particle–particle and particle–wall contacts, respectively, which are summed over all the particle contacts (N_p) and wall contacts (N_w) involving this specific particle. Finally, \mathbf{g} represents the acceleration due to gravity.

The angular momentum balance can be expressed as

$$I_p \frac{d\boldsymbol{\omega}_p}{dt} = \sum_{N_p} \mathbf{M}_p \quad (17.12)$$

where

- I_p is the moment of inertia.
- $\boldsymbol{\omega}_p$ is the angular velocity vector.
- \mathbf{M}_p is the torque vector summed over all the contacts N_p .

The moment of inertia, I_p , for spherical particles with radius r_p is given by

$$I_p = \frac{2}{5} m_p r_p^2 \quad (17.13)$$

Since fluidized beds often have regions operating near to the close-packing limit, particle–particle collisions are frequent. The literature offers several ways of describing particle–particle interactions: via continuum, hard-sphere, or soft-sphere models. The soft-sphere approach was applied in this work (Figure 17.2). When two particles are in contact with each other, a soft-sphere model assumes a slight overlap in the normal direction δ_n , resulting in a repulsive force. Its magnitude depends on the overlap between the two particles, their relative velocity, the spring stiffness, and the damping parameter:

$$\mathbf{F}_{p_A \rightarrow p_B, n} = [k\delta_n - \eta(\mathbf{u}_A - \mathbf{u}_B) \cdot \hat{\mathbf{n}}] \hat{\mathbf{n}} \quad (17.14)$$

where

k is the spring stiffness.

$(\mathbf{u}_A - \mathbf{u}_B)$ is the relative velocity vector of particles A and B.

$\hat{\mathbf{n}}$ is the unit vector along the centers of particle B and A.

η is the damping parameter provided in Jajcevic et al. [1] as

$$\eta = -2 \ln(e_{n, p \rightarrow p}) \frac{\sqrt{m_{AB} k}}{\sqrt{\pi^2 + (\ln(e_{n, p \rightarrow p}))^2}} \quad (17.15)$$

In this expression, $e_{n, p \rightarrow p}$ is the normal particle–particle coefficient of restitution, and m_{AB} is defined as $m_{AB} = (1/m_A + 1/m_B)^{-1}$, where m_A and m_B are the masses of particles A and B, respectively. The relative velocity vector is calculated as follows:

$$\mathbf{u}_{AB} = \mathbf{u}_A - \mathbf{u}_B - (r_A \boldsymbol{\omega}_A + r_B \boldsymbol{\omega}_B) \times \mathbf{n}_{AB} \quad (17.16)$$

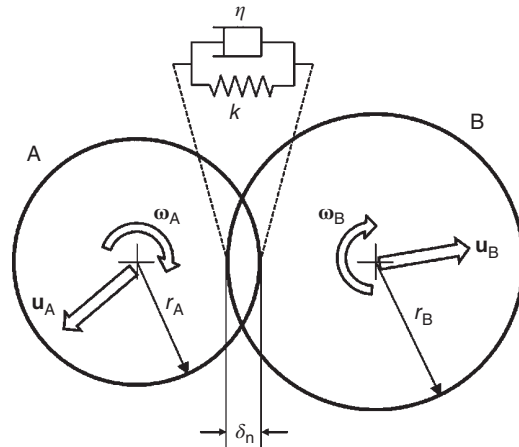


FIGURE 17.2 Modeling the contact dynamics in DEM using the soft-sphere model for particle–particle interactions. Source: Reprinted with permissions from Jajcevic et al. [1]. Copyright (2013) Elsevier.

and the normal relative velocity is

$$\mathbf{u}_{AB,n} = (\mathbf{u}_{AB} \cdot \mathbf{n}_{AB}) \mathbf{n}_{AB} \quad (17.17)$$

The unit vector between the center of particles A and B is given by

$$\mathbf{n}_{AB} = \frac{\mathbf{p}_B - \mathbf{p}_A}{|\mathbf{p}_B - \mathbf{p}_A|} \quad (17.18)$$

where

\mathbf{p} presents the position vector of the particle.

Note that we are using a linear spring and dashpot model for particle–particle contact; alternate models for normal contact can also be found in the literature.

For the tangential component of the contact force, a Coulomb-type friction law is used (see Refs. [1, 10]):

$$\mathbf{F}_{p_A \rightarrow p_B,t} = \begin{cases} -k_t \boldsymbol{\delta}_t - \eta_t \mathbf{u}_{AB,t}, & \text{for } |\mathbf{F}_{p_A \rightarrow p_B,t}| \leq |\mathbf{F}_{p_A \rightarrow p_B,n}| \\ -\mu_{fr,p \rightarrow p} |\mathbf{F}_{p_A \rightarrow p_B,n}| \mathbf{t}_{AB}, & \text{for } |\mathbf{F}_{p_A \rightarrow p_B,t}| > |\mathbf{F}_{p_A \rightarrow p_B,n}| \end{cases} \quad (17.19)$$

where

k_t , $\boldsymbol{\delta}_t$, η_t , $\mu_{fr,p \rightarrow p}$, and \mathbf{t}_{AB} are the tangential spring stiffness, tangential displacement, tangential damping coefficient, friction coefficient and tangential unit vector, respectively.

The tangential displacement over a small time step Δt is given by

$$\boldsymbol{\delta}_t = \mathbf{u}_{AB,t} \Delta t \quad (17.20)$$

The tangential velocity $\mathbf{u}_{AB,t}$ can be calculated as follows:

$$\mathbf{u}_{AB,t} = \mathbf{u}_{AB} - \mathbf{u}_{AB,n} \quad (17.21)$$

The tangential damping coefficient from Jajcevic et al. [1] is given by

$$\eta_t = -2 \ln(e_{t,p \rightarrow p}) \frac{\sqrt{\frac{2}{7} m_{AB} k_t}}{\sqrt{\pi^2 + (\ln(e_{t,p \rightarrow p}))^2}} \quad (17.22)$$

In this expression, $e_{t,p \rightarrow p}$ is the tangential coefficient of restitution. The tangential unit vector is calculated using the tangential velocity:

$$\mathbf{t}_{AB} = \frac{\mathbf{u}_{AB,t}}{|\mathbf{u}_{AB,t}|} \quad (17.23)$$

The particle–wall collision is expressed by treating the walls as a particle with infinite mass (i.e. $m_B \rightarrow \infty$; hence $m_{AB} = m_A$), infinitely large radius ($r_B \rightarrow \infty$), and no linear or angular velocities. The particle–wall interactions are computed by substituting the second particle’s properties with the wall properties.

17.3.1.3 Numerical Implementation of the Hybrid CFD-DEM Coupling Method For coupled CFD-DEM simulations, the code/software usually is composed of a CFD solver and DEM solver running in tandem. For the present work, the CFD solver uses the commercial software AVL FIRETM. The DEM code for the particles is based on eXtended Particle System (XPS) [11], developed in house at RCPE.

Interestingly, in this coupled approach, the DEM portion (i.e. XPS) is solved on the graphical processing unit (GPU) by Nvidia [12], while the CFD portion (AVL FIRETM) is solved using a traditional CPU approach, thus allowing an efficient coupling mechanism. This utilization of the GPU allows simulations with up to 100 million particles on a single GPU. Practically, we can perform CFD-DEM simulations with 10–20 million particles to generate results for the Wurster coating process in a reasonable time frame. Details on the complexities involved with coupling of these two codes running on two platforms can be found in Jajcevic et al. [1].

The CFD mesh is generated via an automated grid generation algorithm, resulting in an irregular spatial subdivision of the simulation volume inside the Wurster coater. When meshing for a CFD-DEM application, the cell sizes must be larger than the particle diameter; the ideal ratio of cell size to particle size is the subject of an ongoing debate (see Pepiot and Desjardins [13]). A general rule of thumb is that the cell volume should be greater than 10 times the particle volume, yet small enough to resolve the fine-scale features of the geometry – in our case, these would include the thickness of the Wurster tube (~1 mm) and the area of the atomization nozzle. From our experience, we find that a grid size Δx that is at least $2.5d_p$ is a good choice, but this is still coarser than the fine-scale features listed previously. To overcome this limitation of resolving fine geometric features using a coarser grid, Jajcevic et al. [1] developed a hybrid CFD-DEM coupling method involving a nontrivial mapping between an arbitrary-shaped CFD mesh and a regular DEM grid (termed linked cell space). Readers are urged to review Jajcevic et al. [1] for further details on this multigrid approach.

The data exchange between the CFD and DEM computational architectures, running on the CPU and GPU, respectively, are realized using the AVL Code Coupling Interface (ACCI) and a volume-weighted mapping algorithm (see Jajcevic et al. [1] for details). At the beginning of a time step, the fluid-flow data, such as velocity and pressure gradients, are transferred from the CPU to the GPU and mapped to the cell centers of the regular DEM grid. Next, the drag force and other source terms are calculated on the GPU and mapped

to DEM cells. The solid volume fraction is calculated based on the DEM cells and is also computed on the GPU. It is important to note that most of the data transfer occurs from the CPU to GPU, thus avoiding transfer of particle data from the GPU to the CPU. As the number of particles ($O(10^8)$) far exceeds the number of DEM cells ($O(10^6)$), transferring cell data is preferable to transferring particle data in terms of computational efficiency. Once the coupling source terms and the particle volume fraction are obtained on the DEM (GPU) side, the data is mapped back to the CFD code using the ACCI algorithm. Both codes can subsequently be executed and the calculation of the next time step can be performed.

17.3.2 CFD-TFM Method

The CFD-TFM approach is often preferred due to its computational efficiency; as particles are not explicitly modeled, the computational time does not depend on the number of particles simulated. In CFD-TFM, the conservation equations have similar structures for both gas and solid phases. The governing equations of both phases are formulated by using the volume-averaged Navier–Stokes equations. The gas-particle drag is the same as the expression used for the CFD-DEM approach (Eqs. 17.6 and 17.7). However, the particle collisions are not explicitly modeled, and the solid-phase stresses (granular pressure and viscosity) are derived from the kinetic theory of granular flows (KTGF) [14]. Generally speaking, the KTGF approach is more accurate for dilute, non-frictional particles, but it becomes less valid for denser flows with frictional particles.

Comparison of CFD-TFM predictions against other approaches, including CFD-DEM, has been a topic of ongoing study. In Wang et al. [15], a detailed comparison between results obtained from both CFD-TFM and the discrete particle model (DPM) is reported for a fluidized bed. In a later study, Chen and Wang [16] compared CFD-TFM, dense discrete particle model (DDPM), and CFD-DEM approaches for modeling impinging gas–solid flows. Although these studies demonstrate the applicability of the CFD-TFM method to denser granular flows, quantitative differences highlight the need to further improve the DDPM and CFD-TFM approaches.

Many CFD-TFM studies have been performed on spouted-bed devices with the purpose of assessing the validity of granular models and the influence of model parameters (e.g. Ref. [17]). Bao et al. [18] present a review of recent advances in spouted-bed devices. According to this review, Bao et al. conclude that the typical flow patterns of spouted beds can be well reproduced by both the CFD-TFM and CFD-DEM approaches, and the simulated characteristic properties such as spout diameter, minimum spouting velocity, and void fraction profile are in good agreement with experimental data. As such, it is worthwhile to investigate the applicability of the CFD-TFM model approach to our

current system of beads in a Wurster coater, before opting to choose the computationally expensive CFD-DEM model.

17.3.2.1 Governing Equations for CFD-TFM: Mass Conservation (Continuity Equation) In the two-fluid model, both the gas phase and the solid phase are described as fully interpenetrating continua using a generalized form of the Navier–Stokes equations for interacting fluids [19]. The continuity equations for the gas (subscript “g”) and solid (subscript “s”) phases are given by¹

$$\frac{\partial}{\partial t}(\alpha_g \rho_g) + \nabla \cdot (\alpha_g \rho_g \mathbf{u}_g) = 0 \quad (17.24)$$

$$\frac{\partial}{\partial t}(\alpha_s \rho_s) + \nabla \cdot (\alpha_s \rho_s \mathbf{u}_s) = 0 \quad (17.25)$$

$$\alpha_g + \alpha_s = 1 \quad (17.26)$$

where

t is time.

α_g and α_s are gas- and solid-phase volume fractions.

ρ_g and ρ_s are phase densities.

\mathbf{v}_g and \mathbf{v}_s are the velocities of the phases.

17.3.2.2 Governing Equations for CFD-TFM: Momentum Conservation Momentum conservation equations of gas and solid particulate phases are given in following form:

$$\begin{aligned} \frac{\partial}{\partial t}(\alpha_g \rho_g \mathbf{u}_g) + \nabla \cdot (\alpha_g \rho_g \mathbf{u}_g \mathbf{u}_g) = & -\alpha_g \nabla p + \nabla \cdot \bar{\tau}_g \\ & + \alpha_g \rho_g \mathbf{g} + \beta(\mathbf{u}_s - \mathbf{u}_g) \end{aligned} \quad (17.27)$$

$$\begin{aligned} \frac{\partial}{\partial t}(\alpha_s \rho_s \mathbf{u}_s) + \nabla \cdot (\alpha_s \rho_s \mathbf{u}_s \mathbf{u}_s) = & -\alpha_s \nabla p - \nabla p_s + \nabla \cdot \bar{\tau}_s \\ & + \alpha_s \rho_s \mathbf{g} + \beta(\mathbf{u}_g - \mathbf{u}_s) \end{aligned} \quad (17.28)$$

where

p is the pressure shared by all phases.

p_s is solids pressure.

\mathbf{g} is acceleration due to gravity.

β is interphase momentum exchange coefficient.

μ_g and μ_s are the shear viscosities of each phase.

λ_s is solids bulk viscosity.

¹In this chapter, we will use the subscript “s” for solids, representing the mean solid velocity averaged over an ensemble of particles. The subscript “p” will be used when referring to individual particles, e.g. when performing DEM calculations. The subscript “g” is used for the gas phase.

The terms $\bar{\tau}_g$ and $\bar{\tau}_s$ are gas and solid stress tensors, respectively, for which constitutive equations are needed to account for interactions within gas and solid phase:

$$\bar{\tau}_g = \alpha_g \mu_g \left\{ \left[\nabla \mathbf{u}_g + (\nabla \mathbf{u}_g)^T \right] - \frac{2}{3} \nabla \cdot \mathbf{u}_g \bar{\mathbf{I}} \right\} \quad (17.29)$$

$$\bar{\tau}_s = \alpha_s \mu_s \left[\nabla \mathbf{u}_s + (\nabla \mathbf{u}_s)^T \right] + \alpha_s \left(\lambda_s - \frac{2}{3} \mu_s \right) \nabla \cdot \mathbf{u}_s \bar{\mathbf{I}} \quad (17.30)$$

Since the concept of particles has disappeared completely in the two-fluid model formulation, the effect of particle–particle interactions can only be included indirectly, via the solid stress tensor and solids pressure. Closure expressions, i.e. correlations for the solids pressure and viscosity, are required to solve the CFD-TFM conservation equations. These closures for solid stresses are obtained from the KTGF [20]. The following constitutive closures are considered:

17.3.2.2.1 Solids Pressure (p_s)

$$p_s = \alpha_s \rho_s \Theta_s [1 + 2(1 + e)g_0 \alpha_s] \quad (17.31)$$

where

Θ_s is the granular temperature, which is the measure of particle velocity fluctuations.

e is the coefficient of restitution for solid–solid collisions.

g_0 denotes the radial distribution function.

The radial distribution function is a correction factor that modifies the probability of collision under close-packing conditions. Often used radial distribution function correlations are from Gidaspow [4] and Lun and Savage [21].

17.3.2.2.2 Solids Shear Viscosity (μ_s) Solids shear viscosity arises due to translational (kinetic) motion and collisional interaction of particles (see Refs. [22, 23]). A frictional component of viscosity can also be included to account for the viscous–plastic transition that occurs when particles of a solid phase reach the maximum solid volume fraction (see Schaefer [24]) for frictional viscosity and Johnson and Jackson [25] for frictional pressure):

$$\mu_s = \mu_{s,kin} + \mu_{s,coll} + \mu_{s,fric} \quad (17.32)$$

17.3.2.2.3 Solids Bulk Viscosity (λ_s) The solids bulk viscosity accounts for the resistance of the granular particles to compression and expansion [20]:

$$\lambda_s = \frac{4}{3} \alpha_s \rho_s d_p g_0 (1 + e) \left(\frac{\Theta_s}{\pi} \right)^{0.5} \quad (17.33)$$

where

d_p denotes the diameter of particles.

17.3.2.2.4 Interphase Momentum Exchange Coefficient (β)

The interphase forces between gas and solids include drag force, lift force, and virtual mass force among the main contributions. In coupling the equations of the two phases, due to the large difference in the densities ($\rho_s \gg \rho_g$), forces other than drag force are less significant and can usually be neglected. Momentum exchange between the phases is based on the gas–solid drag coefficient:

$$\beta = \frac{\alpha_s \rho_s f}{t_s} \quad (17.34)$$

where f denotes an exchange coefficient, related to the drag coefficient (C_D), and is usually a function of the particle Reynolds number (Re_s). The time scale t_s is the particulate relaxation time (related to the particle terminal speed), defined as

$$t_s = \frac{\alpha_s d_s^2}{18 \mu_g} \quad (17.35)$$

Various choices exist in the literature for the exchange coefficient f , depending on the specific conditions. Some of the existing drag models are summarized below:

- Syamlal–O’Brien model: Appropriate when the solids shear stresses are defined according to Syamlal and O’Brien [5].
- Parameterized Syamlal–O’Brien model [26]: Appropriate when the particle size lies between 40 and 500 μm and the particle density between 1.4 and 4.0 g/cm^3 ; bubbling typically forms directly at incipient fluidization (Geldart group B particles).
- Wen and Yu model [8]: Appropriate for dilute systems.
- Gidaspow model [23]: A combination of the Wen and Yu model [8] and the Ergun equation [7], recommended for dense fluidized beds.
- Huilin–Gidaspow model [27]: A combination of the Wen and Yu model and the Ergun equation.
- Gibilaro model [28]: Recommended for circulating fluidized beds.

Arguably, the choice of drag models plays the most important role in modeling gas–solid two-phase flows: different drag models significantly impact the solid-phase predictions [29]. Figure 17.3, adapted from Du et al. [29], plots some different drag functions (β) versus solid volume fraction. In the dilute region (solid volume fraction < 0.2), all drag models except for Richardson and Zaki [30] give similar results.

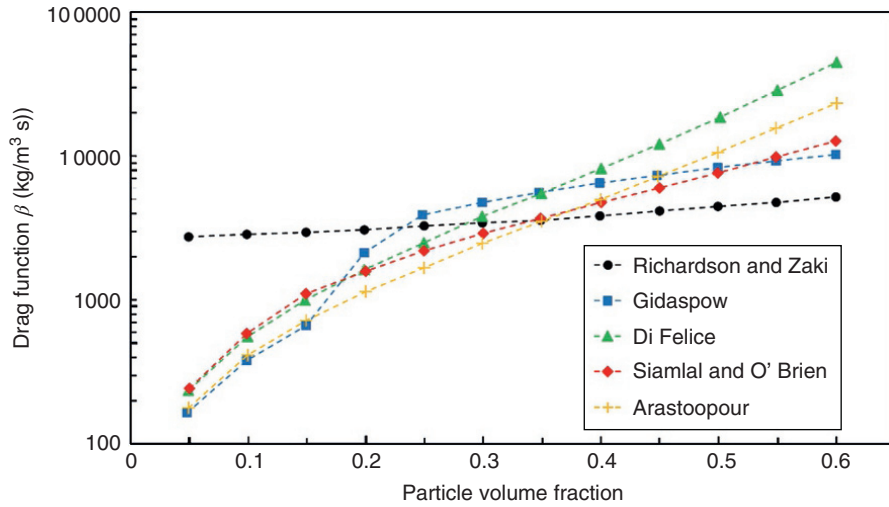


FIGURE 17.3 Comparison of different drag models used in CFD-TFM [29], shown as a function of particle solid volume fraction. Results are plotted for a representative particle size 1.44 mm and considering air as the fluid phase. *Source:* Reprinted with permissions from Du et al. [29]. Copyright (2006) Elsevier.

Beyond a solid volume fraction of 0.2, the differences become larger.

A comparison of these various drag laws suggests that an improper choice of drag models may yield inaccurate results or even lead to incorrect descriptions of gas–solid two-phase flow. Therefore, it is important to ensure that the chosen drag law is applicable to the system of interest. In the present work, we use validation, i.e. comparison of computational predictions against published experimental results, to ensure that our choice of drag model is indeed appropriate.

17.3.2.3 Governing Equations for CFD-TFM: Conservation of the Granular Fluctuating Energy (Granular Temperature Equation) In the kinetic theory for gases, the temperature is defined using the random fluctuating velocity of molecules. Analogous to the thermodynamic temperature of a gas, the granular temperature Θ_s is defined based on the random fluctuating velocity of the solid particles (solids fluctuating energy), given by

$$\Theta_s = \frac{1}{3} (u'_{s,x}u'_{s,x} + u'_{s,y}u'_{s,y} + u'_{s,z}u'_{s,z}) \quad (17.36)$$

The particle velocity can be expressed as the sum of the mean velocity and a superimposed fluctuating component representing the random vibratory motion of the particle: for example, $u_{s,x} = \langle u_{s,x} \rangle + u'_{s,x}$, where $\langle u_{s,x} \rangle$ is the mean velocity of all particles in a given volume (such as a CFD cell). In Eq. (17.36), $u'_{s,x}$, $u'_{s,y}$, and $u'_{s,z}$ represent this random fluctuating velocity of the solids along the coordinate axes. The source of the particle velocity fluctuations arises from random collisions between neighboring particles, which results in a random velocity field within the bed. Similar to

thermal temperature, the granular temperature Θ_s can be conducted (if a gradient of granular temperature exists), generated, and dissipated. Therefore, the conservation of granular temperature derived from the KTGF [14] takes the form

$$\frac{3}{2} \left[\frac{\partial}{\partial t} (\rho_s \alpha_s \Theta_s) + \nabla \cdot (\alpha_s \rho_s \mathbf{u}_s \Theta_s) \right] = (-p_s \bar{I} + \bar{\tau}_s) : \nabla \mathbf{u}_s + \nabla \cdot (\Gamma_{\Theta_s} \nabla \Theta_s) - \gamma_{\Theta_s} + \phi_{gs} \quad (17.37)$$

where the right-hand-side terms of equation represent the following: $(-p_s \bar{I} + \bar{\tau}_s) : \nabla \mathbf{u}_s$ is the generation of energy by the solid stress tensor, $\nabla \cdot (\Gamma_{\Theta_s} \nabla \Theta_s)$ is the diffusion of energy, Γ_{Θ_s} is granular temperature conductivity [22], γ_{Θ_s} is the collisional dissipation of energy [20], and ϕ_{gs} is the energy exchange between the gas phase and the solid phase [23].

17.3.2.4 Comparison of CFD-DEM and CFD-TFM Approaches It is difficult to generalize the relative advantages/disadvantages of the CFD-DEM and CFD-TFM approaches, as these are specific to the system of interest as well as the underlying physics of importance. However, we shall attempt to present a subjective comparison of the two as they apply to modeling fluidized beds, summarized in Table 17.1.

In our present work, we performed a comparison of the two approaches by validating against a published experimental data set by Li et al. [31, 32]. Both CFD-DEM and CFD-TFM techniques provided qualitatively similar results, but the CFD-DEM results showed slightly better agreement with the experimental data (results not presented as graphs). Therefore, we have chosen to use the CFD-DEM method

TABLE 17.1 Qualitative Summary of CFD-DEM vs. CFD-TFM Simulation Techniques as Relevant to Modeling Gas-Particle Fluidized Beds

	CFD-DEM	CFD-TFM
Preprocessing effort	<i>Normal</i>	<i>Normal</i>
Mesh requirements and sensitivity	<i>Mesh size should be larger than particle size; otherwise no special requirements</i>	<i>High quality mesh is required. Grid-independent solutions can be obtained for grid sizes smaller than $10d_p$</i>
Simulation stability	<i>Good numerical stability</i>	<i>Lower numerical stability as there are more coupled differential equations with nonlinear sources and closure terms</i>
Models and results reproducibility	<i>Tends to be more reproducible as particle collisions (solid stress) are resolved explicitly using DEM</i>	<i>Solid stresses are implicitly modeled via constitutive closures; there are a number of options for closures, and results depend strongly on the chosen closure relationship</i>
Scalability with respect to number of particles	<i>Computational costs for DEM increase with particle number; limited mainly to laboratory- or pilot-scale devices</i>	<i>Particles are not explicitly modeled; can handle larger systems with a relatively smaller computational overhead, compared with DEM (depending on the additional number of CFD cells)</i>

in the present work, and all future results will be based on CFD-DEM.

The primary motivation for using CFD-DEM over CFD-TFM in this work arises from the differences in the particle stress models in both approaches. In CFD-TFM, there are a variety of stress models derived for a number of particle types and conditions. However, most of them are focused on fine powders (generally $d_p < 250 \mu\text{m}$), whereas our present system is composed of larger beads ($d_p > 500 \mu\text{m}$). The application of the existing stress models, which were originally developed primarily for fine powders, to the larger beads in CFD-TFM simulations is questionable. Hence, the CFD-DEM approach, where particle collisions are explicitly resolved, is preferred as it makes fewer assumptions regarding particle collisions and stresses. Nonetheless, this section serves to inform the readers of the CFD-TFM approach as it may be more amenable to their particular system of interest. For example, beads produced by the melt spray congealed method are generally smaller ($\sim 250 \mu\text{m}$) and may be modeled well by CFD-TFM.

17.4 CFD-DEM MODEL OF THE WURSTER FLUID BED COATER

The primary focus of this chapter is to present a case study where a Wurster coating process is modeled using CFD-DEM. However, CFD-DEM can only simulate a few seconds to a few minutes of the process, whereas an actual coating operation can be several hours in duration. Therefore, we must find a way to extrapolate the short-term CFD-DEM data to predict the coating properties after many hours of operation. A multiscale framework is developed to predict the long-term coating behavior – results from the CFD-DEM simulations are incorporated into a reduced-order

compartment model. Section 17.4 discusses the setup of the CFD-DEM simulations and the results generated thereafter. Section 17.5 presents a simpler version of the a compartment model using Monte Carlo approach, and Section 17.6 presents a more detailed version using a Markov chain Monte Carlo approach.

In this section, we begin by recognizing the three important steps that are necessary to successfully perform CFD-DEM simulations. First, the geometry of the Wurster coater has to be described in detail, including determination of the airflow pattern through the distributor plate. Second, the particle properties that are inputs to the CFD-DEM model have to be estimated. Third and last, the fully coupled simulation needs to be set up and executed, followed by the analysis needed to generate the necessary flow information. The following sections (Sections 17.4.1–17.4.4) describe these three important steps in detail, before presenting some representative CFD-DEM results in Section 17.4.5.

17.4.1 Wurster Coater Geometry

In the first step, the geometry of the Wurster coater has to be defined. Wurster coatiers are, in general, frustum shaped (truncated cones) with one or more Wurster tubes positioned inside. The cross-sectional area of the coater increases with height, i.e. the upper region of the coater serves as the expansion chamber. This is to ensure that the gas velocity decreases with height and that the entrained particles can fall back down to the bed as the reduced gas velocity is no longer able to fluidize the beads. Figure 17.4 shows the design of some typical Wurster coater units, with small differences in the geometry. The Wurster coater unit of our interest is similar to the unit represented in Figure 17.4c.

For the present study, we will investigate the performance of a larger-scale Wurster coater (Glatt-15/30). This coater has

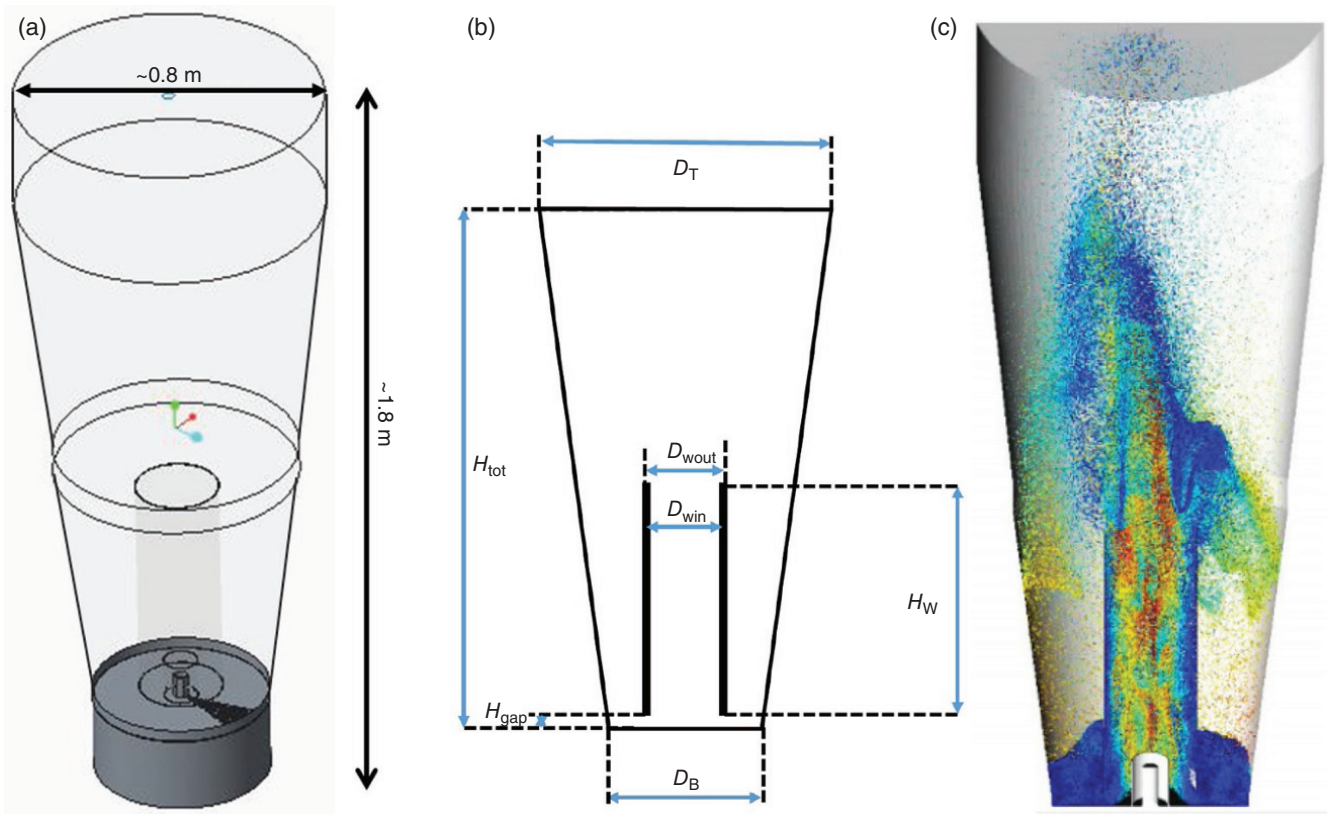


FIGURE 17.4 (a) Design of the Glatt-15/30 Wurster coater, showing some key dimensions. (b) Some of the dimensions/measurements needed from the actual coater to construct the CFD-DEM model. (c) Snapshot from a typical CFD-DEM simulation.

a height of approximately 2 m with a base diameter of approximately 46 cm (18 in) and is capable of processing up to 50 kg of beads. The Glatt-15/30 is generally used to manufacture clinical or commercial supplies of beads for pediatric drug programs. Details on the mesh used to discretize the simulation domain are discussed in Section 17.4.4.

17.4.2 Steady-State Simulation of Airflow Distribution Through the Distributor Plate

An important geometric detail of the Wurster coater is the bottom plate. The number and position of the bore holes, through which air is entering into the Wurster plenum region, has a significant influence on the particle fluidization. Figure 17.5a shows an actual distributor plate geometry. Several distinct zones with varying hole diameters are clearly visible. Close to the Wurster coater center, the size and number of holes is increased to allow more gas to pass through the central Wurster tube. Normally, a mesh is placed on top of the distributor plate to prevent particles from falling below the plate. Recall that for CFD-DEM simulations, the grid size must be larger than the particle size. The mesh required to explicitly resolve these air holes in the distributor plate would violate this requirement. Furthermore, it would be computationally too expensive to

fully resolve these fine holes in an already numerically demanding full-scale CFD-DEM simulation.

In our approach, we have first performed a steady-state CFD simulation of the airflow (no particles) through the distributor plate of the Wurster coater, with the holes explicitly resolved (Figure 17.5b). For this simulation, the geometry is fully discretized, including the inlet region before the distributor plate, all holes in the distributor plate, as well as the region inside the Wurster coater (Figure 17.6). The steady-state simulation determines the airflow distribution through the holes in bottom plate, specifically the distribution of air diverted through the core region (below the Wurster column) and the annular region (below the downbed region). We find that for this particular geometry and the chosen distributor plate, the flow through central core (regions “b” and “c” in Figure 17.5) is 64% and through the annular region (region “a” in Figure 17.5) is 36% of the total air flow rate. This relative air distribution is held constant and used to define the bottom-air-inlet boundary conditions in subsequent CFD-DEM simulations.

Almost all available CFD simulation software can be used to run steady-state single-phase turbulent flow simulation. As continuum media, air with constant material properties is normally used.

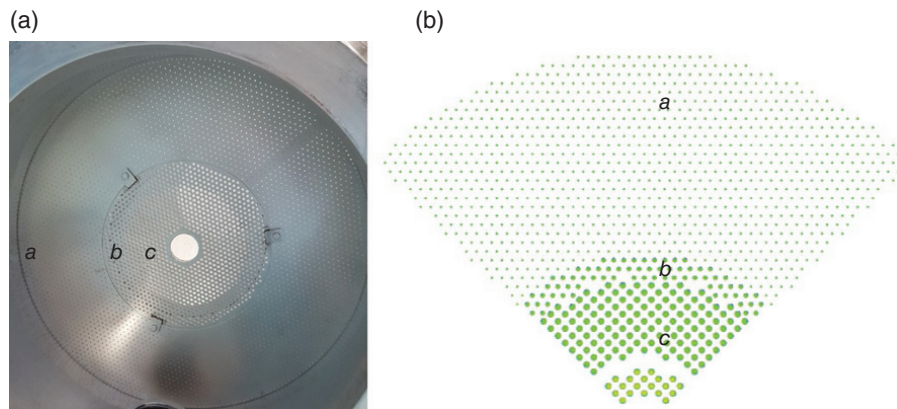


FIGURE 17.5 Distributor plate design: (a) actual geometry and (b) modeled distributor plate in CFD simulations. The regions “a,” “b,” and “c” are composed of holes with different diameters, with region “c” having the largest holes (4 mm) and region “a” having the smallest holes (1.5 mm).

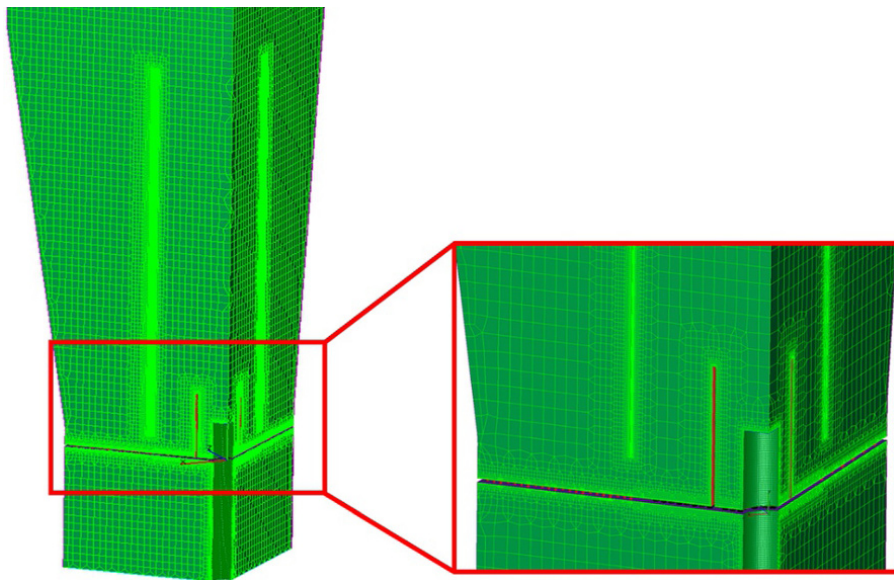


FIGURE 17.6 Determination of the airflow pattern through the bottom distributor plate via steady-state CFD simulations. The geometry along with the numerical mesh that explicitly resolves all the holes in the distributor plate is shown.

17.4.3 Determination of the Material Properties for DEM Simulations

To determine material properties needed as inputs to the coupled CFD-DEM simulations, several experimental characterization tests must be performed. An important input is the particle size distribution of the ingoing beads. The particle size distribution can be measured by a sieve analysis or by using automated size characterization test, such as QICPIC from Sympatec. The true density of the beads is also a required input, but it is relatively straightforward to obtain using a helium pycnometer.

Other particle properties of interest include the particle stiffness (i.e. Young’s modulus), particle–particle, and

particle–wall friction, as well as the particle–particle and particle–wall coefficient of restitution. Strictly speaking, each of these parameters should be calibrated using small-scale lab tests – the calibration framework would include experimental measurements such as angle of repose, powder rheometry using a Freeman FT4 rheometer, and other unit tests. For each of these experimental tests, corresponding DEM studies need to be performed with varying particle properties until the DEM parameters match the experimental results. However, for the present problem, we did not follow this rigorous approach and chose to use values for stiffness, friction, and restitution coefficient that are representative of the beads of interest, taken from published DEM studies for

pharmaceutical materials. For fluidized bed applications, the results are relatively insensitive to small changes in these particle–particle contact parameters. The dominant physics that governs flow behavior is the gas–particle drag, which depends on the diameter and density of the beads. Hence, the size and density of the beads is the more critical input, which we have characterized with greater rigor.

17.4.4 CFD-DEM Numerical Mesh

For the CFD portion of the coupled CFD-DEM simulation (solved on the CPU), a mesh with a reasonably small number of cells is adequate. Although the mesh used to determine the steady-state airflow may have several millions of cells (see Section 17.4.2), the number of cells for the coupled simulation should be kept as few as possible to keep computational costs low. The mesh used for the coupled CFD-DEM simulation is shown in Figure 17.7. The number of cells used for the coupled CFD-DEM setup (i.e. Figure 17.7) is at least an order of magnitude smaller than that used in Section 17.4.2 (i.e. Figure 17.6), i.e. a few hundred thousands of CFD cells are represented in Figure 17.7. This reduction in the number of cells is made possible by simplification of the distributor plate inlet, viz. the airflow distribution determined in Section 17.4.2 is used as the air inlet boundary condition of the coupled CFD-DEM simulation.

For the DEM portion (solved on the GPU), a mesh of hexahedral cells is used in the background for the particle–particle contact detection algorithms. Ideally, the size of these contact detection cells should be on the order of the particle size to optimize computational speed. However, the CFD-DEM coupling methodology requires the CFD cell sizes to be at least 2.5 times the particle diameter. In order to meet both requirements, a sub-grid methodology is used on the

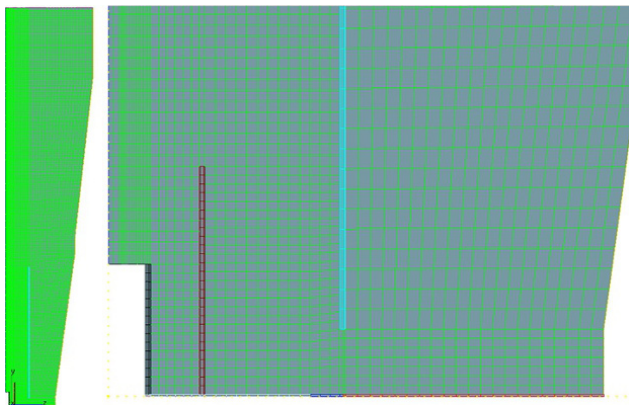


FIGURE 17.7 The computational mesh used for the CFD portion of the CFD-DEM simulations executed for the Glatt-15/30 coater. The holes in the distributor plate are not resolved explicitly (as in Figure 17.6), thereby reducing the number of computational cells.

DEM side, wherein one set of finer grids is used for contact detection algorithms and a second set with coarser cells is used for the purposes of coupling CFD with DEM.

There are several process parameters of interest when simulating a Wurster coater using CFD-DEM; each of these parameters can be adjusted to manipulate the process outcome. The purpose of the CFD-DEM studies is to link these process parameters to the process outcomes. The process parameters that we investigated are the particle size distribution, atomization air flow rate, Wurster gap height, and fluidization air flow rate (shown in Figure 17.8).

The process outcomes, i.e. output measurements, of interest include the residence time distribution (RTD) inside the Wurster region; the cycle time distribution (CTD) of the beads between successive passes through the Wurster column and downbed; the bead mass inside the Wurster column, i.e. number of particles that receive coating at any instant; and the mass flow rate through the Wurster tube, i.e. the circulation rate of beads (which is also the ratio of bead mass inside the Wurster column to the *mean* cycle time [MCT]). The following section (Section 17.4.5) presents some representative RTD and CTD results from the CFD-DEM parametric studies. Eventually, the RTD and CTD information will be used in reduced-order models to predict the long-term coating behavior.

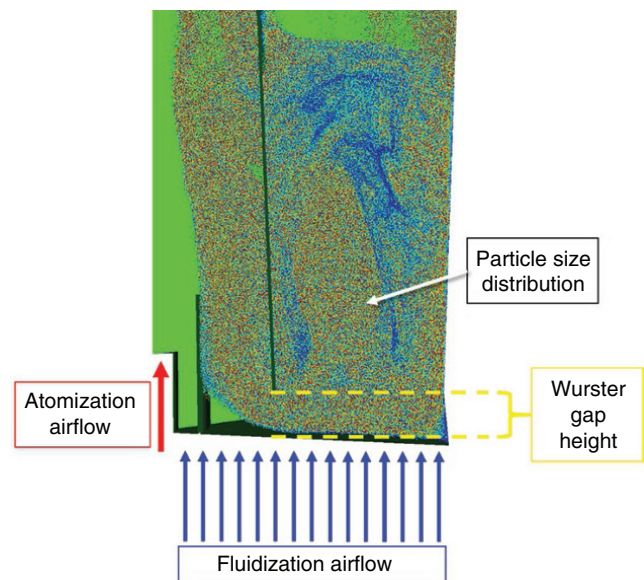


FIGURE 17.8 Snapshot of a CFD-DEM simulation showing the lower portion of the Wurster coater. The four parameters that were varied in this study are marked, viz. particle size distribution, atomization air flow rate, Wurster gap height, and fluidization air flow rate. The particles are colored by size, blue being the smallest and red being the largest particles.

17.4.5 Case Study: Influence of Fluidization Air Flow Rate

Fluidization air flow rate is a process parameter expected to have a significant influence on gas-particle flow dynamics inside the Wurster coater. In this section, we report the results from the parametric study of the fluidization air flow rate on the particle dynamics from the CFD-DEM simulations of the Glatt-15/30 Wurster coater. Through the results from this particular parametric study, i.e. effect of varying inlet air flow rate, we will demonstrate the typical data that may be generated from CFD-DEM simulations and present a framework where this data is propagated to reduced-order models to obtain long-term predictions.

The Wurster coater was loaded with 40 kg of uncoated beads, with the size distribution obtained directly from experimental Sympatec QICPIC measurements (Section 17.4.3). The bead sizes range from 500 to 710 μm ; the same PSD is used in the CFD-DEM simulations. The atomization air flow rate was set to 45 m^3/h , and the Wurster gap height was set to 30 mm – these were not varied for this particular parametric study.

Given the batch size of 40 kg, the total particles in the system were estimated to be about 300 million, which are too many to simulate using CFD-DEM. To reduce the computational costs, we exploit the azimuthal symmetry of the circular coater and model only a 30° segment full coater. We verified that this geometric simplification does not impact the flow predictions by demonstrating that the results from 90° and 30° segments were identical for a smaller 10 kg batch (presented at the AIChE 2016 Annual Meeting [33], and not shown separately as figures). With this simplified axisymmetric model, the number of CFD-DEM particles is now reduced to 18 million, which can be modeled using the existing hardware resources. The output measurements of interest, as described in the preceding section, are the RTD in the Wurster tube, CTD between successive Wurster tube visits, mass holdup inside the Wurster tube, and mass flow rate through the Wurster tube. In addition, we are also interested in the velocity profiles of the beads in order to better understand the flow dynamics inside the coater.

The RTD in the Wurster tube is a good indicator of the amount of the coating that particles receive. Figure 17.9 shows the RT inside the Wurster tube for three different fluidization air flow rates, spanning the typical range of operation for this device. Furthermore, the beads are divided into five size groups, each spanning approximately 40 μm , and for each simulation at a given air flow rate, the mean residence time (MRT) is reported individually for each size group. The intent is to better understand if there are any noticeable differences in the flow pattern based of the bead size – do particles spend more or less time in the spray zone if they are larger or smaller? The scatter bars for each binned data point represent one standard deviation of the full RTD for that particular bin.

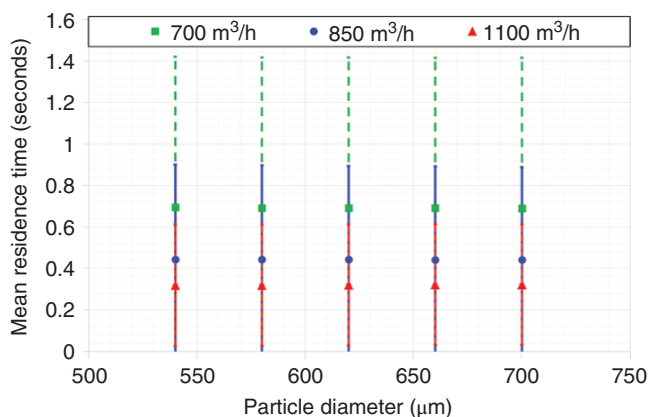


FIGURE 17.9 Mean residence time of particles for varying fluidization air flow rates in the Glatt-15/30 Wurster coater. The Wurster residence times are plotted for different size classes, with each size class having a width of approximately 40 μm . The scatter bars represent one standard deviation of the Wurster residence time distribution (RTD).

The results in Figure 17.9 show that, in general, beads spend a very short amount of time – a few seconds or less – per pass inside the Wurster column. There is little effect of the particle size on the Wurster RT; both the mean and standard deviation of the RT appear to be largely unaffected by bead diameter. (Upon closer inspection, there appears to be a minute decrease in the MRT with diameter, but this change is practically negligible.) However, there is a clear and significant impact of the fluidization air flow rate on the Wurster-zone RT. As the air flow rate is increased, the MRT in the Wurster column decreases, and the RTD becomes narrower (smaller scatter bars). It clearly shows that the particles spend less time in the spray zone at higher fluidization air flow rates.

Figure 17.10 shows the MCT between successive visits to the Wurster column shown, as before, for the five particle size groups and for the three fluidization air flow rates. The scatter bars represent one standard deviation of the CTD. The CT should preferably be sufficiently long for particles to dry completely before reentering the spray zone to avoid over-wetting and agglomeration, but short enough such that beads experience a large number of passes during the full process duration. A long CT for a subpopulation of the beads would suggest the presence of a poorly mixed or dead zone region in the coater, which is undesirable. Fortunately, the CT data shows that poor mixing is not an issue for the present Glatt-15/30 coater, also ascertained visually from animations of bead movement from the CFD-DEM simulations.

The CT decreases with increasing fluidization air flow rate and, to a smaller extent, with increasing particle size. A reduced CT indicates shorter time between visits to the spray zone. Hence, differences in the CT with bead size show that larger particles get coated more frequently compared

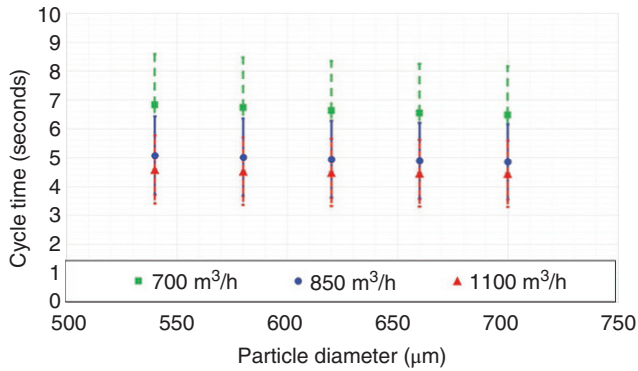


FIGURE 17.10 Mean cycle time of particles for varying fluidization air flow rates in the Glatt-15/30 Wurster coater. The cycle times are plotted for different size classes, with each size class having a width of approximately 40 μm. The scatter bars represent one standard deviation of the cycle time distribution (CTD).

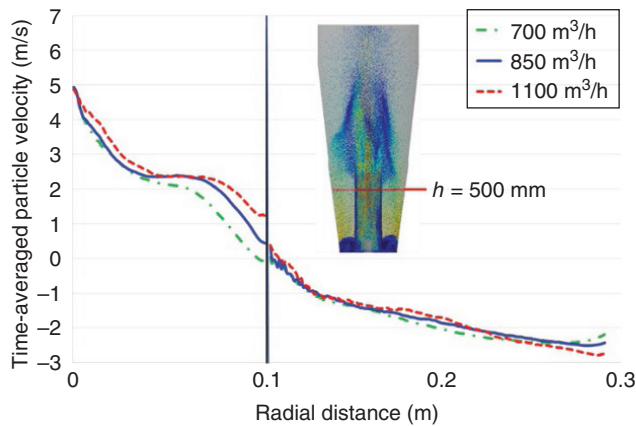


FIGURE 17.11 Radial distribution of particle velocities in the Glatt-15/30 Wurster coater at a height of 500 mm from the bottom distributor plate (the vertical height is marked in the inset). Results are shown for varying fluidization air flow rates.

with smaller ones. These size-based differences are expected to contribute to the heterogeneity in coating mass distributions at the end of process (which will be examined in Sections 17.5 and 17.6).

The time-averaged particle vertical velocity profile along the radial direction is shown in Figure 17.11. This profile is reported at a height of 500 mm above the distributor plate (compare with the Wurster gap height of 30 mm, total Wurster column height of 600 mm, and Wurster column diameter of 213 mm). The vertical gray line in the Figure 17.11 indicates the radial position of the Wurster column wall; radial locations smaller than 0.107 m lie inside the Wurster column and locations greater than 0.107 m represent the downbed region.

In the downbed region, particles generally have a negative velocity, indicative of the gradual downward motion of the bed as the beads circulate. There is little impact of the fluidization air flow rate on the flow velocities in the downbed region. Inside the Wurster tube, the particle velocity increases with increasing fluidization air flow rate, particularly close to the tube wall. Interestingly, for air flow rate of 700 m³/h, the time-averaged particle velocity near the Wurster tube wall is close to zero or less than zero. Of the particles near or at the wall, many have a negative vertical velocity, i.e. they are falling downward. This observation suggests backflow or back mixing of particles inside the Wurster tube, which also helps explain the wider RTD inside Wurster tube zone for the 700 m³/h operating conditions (larger scatter bars in Figure 17.9).

Finally, the mass flow rate through the Wurster tube and mass inside the Wurster tube are shown in Figure 17.12 for varying fluidization air flow rates. The mass flow rate increases, and the mass inside the Wurster tube decreases with increasing air flow rate, both observations being consistent with the RTD and CTD data. For a fixed spray rate of the coating solution/suspension, changes in the flow behavior inside the Wurster column, i.e. the number of particles as well as the speed of particle motion inside the spray zone, could potentially have an impact on the final coating uniformity of beads. The relationship between these flow properties – RTD and CTD – and the long-term coating uniformity will be examined in the following sections.

17.5 LONG-TERM COATING PREDICTIONS: SIMPLER MONTE CARLO APPROACH

CFD-DEM simulation results can be leveraged to make long-term coating uniformity predictions through a reduced-order Monte Carlo method. To this end, the fluid bed coater is subdivided into two zones: a spray region and a bed region. In this simpler version of the reduced-order model, the spray region is taken as the entirety of the Wurster column, and all other areas belong to the bed region.

The trajectory of a single particle consists of an alternating sequence of visits to each region. CFD-DEM simulations can be analyzed to obtain the durations of individual visits, i.e. the RT. A CFD-DEM simulation that is far shorter than a typical coating operation can still yield a large sample of individual RTs. Accordingly, single-visit RTDs are constructed for each region (Figure 17.13a and b), expressed as the respective probability distribution functions of the RT. Note that the RTD in Figure 17.13a is plotted for the fluidization air flow rate of 850 m³/h; hence, the mean and standard deviation of the distribution in Figure 17.9 are directly related to the data reported in Figure 17.13a. No particle size-based binning is performed in Figure 17.13, and the RTDs are computed for all the beads. The bed-region RTD in Figure 17.13b is strongly related to the CTD (Figure 17.10), but the

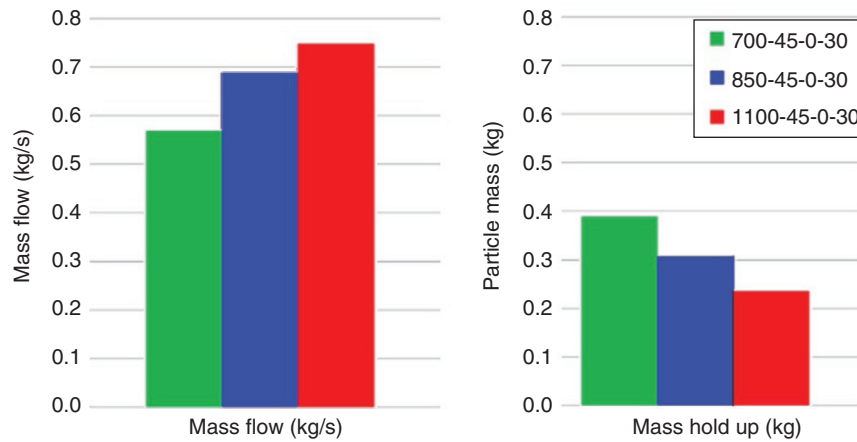


FIGURE 17.12 Mean mass flow rate through the Wurster column and bead mass inside the Wurster column in the Glatt-15/30 Wurster coater for varying fluidization air flow rates.

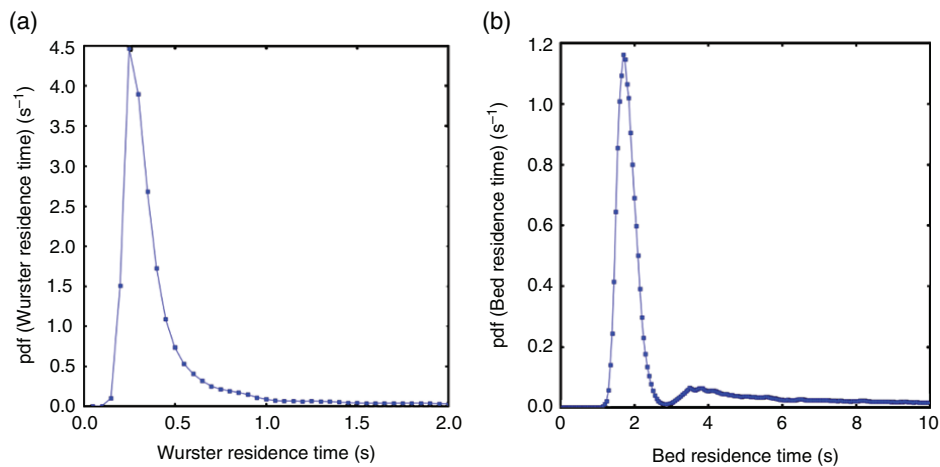


FIGURE 17.13 The single-visit residence time distribution (RTD) of beads in (a) spray region, i.e. inside the Wurster column and (b) in the bed region. The RTDs are shown for the operating conditions described in Section 17.4.5, but specifically for the fluidization air flow rate of 850 m³/h as a representative example.

distinction is that the CT is the sum of the Wurster-region and bed-region RT.

The flow of an individual particle can be considered as alternative successive visits to the Wurster (spray) region and bed region over the duration of the process. The duration of each visit to the Wurster and bed regions can be repeatedly sampled in an alternating fashion from the RTDs shown in Figure 17.13 to produce randomly generated RTs – this numerical approach of randomly sampling from predefined RTDs is called a Monte Carlo simulation. Throughout this sampling process, the cumulative amount of time spent in the Wurster column, i.e. the spray zone, is tracked. This procedure can be replicated for a large number of particles to generate a representative set of particle trajectories (albeit coarse grained to distinguish only between the spray and bed regions) and, hence, construct the cumulative RTD of

total time spent in the spray zone as a function of process time.

As a first approximation, the cumulative RTD in the spray zone can be taken as a surrogate for the total coating acquired by a particle. We can assume that over a small time interval, the delivered spray liquid is evenly distributed among all the particles currently inside the Wurster column. In reality, the rate of gain of coating mass for a particular particle in the Wurster column is dependent on several factors. Shielding may occur wherein the spray is blocked from reaching particles farther away from the spray nozzle by the particles that are closer to the nozzle. In addition, as the coating spray spreads radially outward from the nozzle, the volumetric density of spray droplets decreases; i.e. the spray droplets are not distributed equally among all beads in the spray zone. In this section, these complexities are overlooked, and particles in

the spray zone are assumed to accumulate coating at a constant, uniform rate. (This assumption will be relaxed in Section 17.6.) Therefore, the single-particle mass gain rate is

$$\dot{m} = \frac{Q s_{ml} \eta \left(\frac{1}{6} \pi d_p^3 \right) \rho_p}{M_w} \quad (17.38)$$

where

- Q is the total spray rate.
- s_{ml} is the solids loading in the coating solution/suspension.
- η is the fraction of coating spray that ends up on particles (coating efficiency).
- d_p is the particle diameter.
- ρ_p is the particle density.
- M_w is the average bead mass inside the Wurster column (see Figure 17.12).

The cumulative spray-zone RTD can be scaled by \dot{m} to determine the distribution of coating mass per particle.

Under the aforementioned assumptions, we can compute the coating mass distribution on all beads as a function of process time using the Monte Carlo coating model (Figure 17.14). The Monte Carlo predictions show that the distributions shift rightward with process time, i.e. the mean coating mass increases as more spray is applied to the beads.

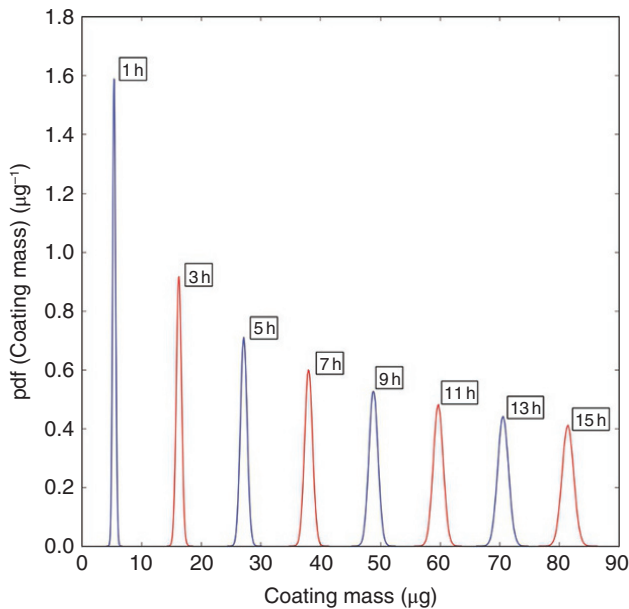


FIGURE 17.14 The evolution of coating mass distribution on beads as a function of process time. The span of the distribution, i.e. the variance of coating masses on beads, increases with process time. The distributions are normalized such that the area under each curve is unity.

Furthermore, the span of the coating mass distributions increases with process time, i.e. the variance of the absolute coating mass increases with process time. (The heights of the distributions appear to decrease as they are normalized to ensure that the area under each curve is unity.)

Once the coating levels on each individual particle are computed, the interparticle coating variability is measured via the coefficient of variation (CoV) of the bead coating mass, defined as

$$\text{CoV} = \frac{\sigma}{\mu} \quad (17.39)$$

where

- σ is the standard deviation of the coating mass.
- μ is the mean coating mass.

The CoV of coating mass is an important measure that defines the bead-to-bead coating uniformity – a smaller CoV value indicates a more uniformly coated batch. The CoV at any time point during the process can be obtained from the coating mass distributions shown in Figure 17.14; the CoV values, thus computed, are plotted as a function of time in Figure 17.15.

An important result that emerges is that the CoV is inversely proportional to the square root of process time ($\text{CoV} \propto t^{-1/2}$, Figure 17.15), regardless of the nature of the

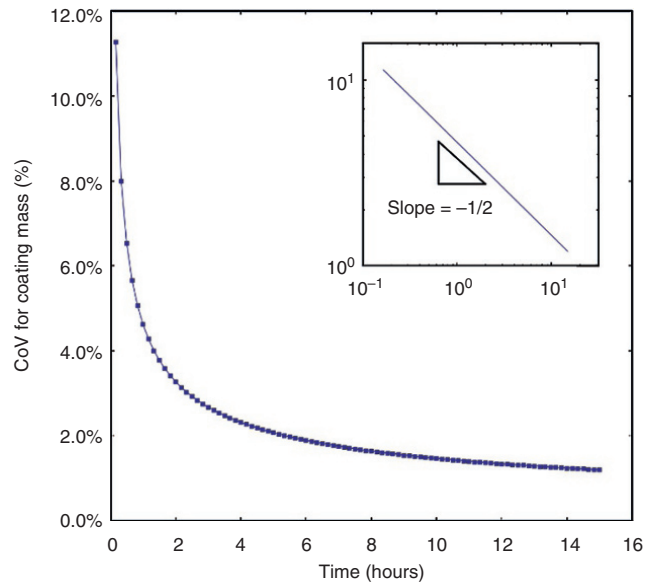


FIGURE 17.15 The coefficient of variation (CoV) of coating mass on the beads as a function of process time. The inset shows the same data plotted on a logarithmic scale, revealing the $\text{CoV} \propto t^{-1/2}$ relationship. The process conditions are same as those referred to in Figures 17.13 and 17.14.

underlying RTDs (other studies that arrive at the same result include, among others, Mann [34], Kalbag et al. [35], Kalbag and Wassgren [36], and Kumar et al. [37]). A practical implication of this finding is that for a given amount of spray to be applied, the coating uniformity may be improved by reducing the spray rate and extending the process time. Another use for CoV plots such as Figure 17.15 is to help optimize the spray rate. Although a longer-duration process improves inter-bead coating uniformity, the greatest improvement occurs near the start of the process. The gain toward the end of the process is marginal, and depending on the desired degree of uniformity, it may not be beneficial to have a very low spray rate, i.e. having an excessively long process duration from an industrial perspective. Low spray rates and long processing times present the risk of spray drying the spray droplets, which reduces the spray efficiency and can detrimentally affect the coating quality.

In the analysis presented thus far in this section, we have ignored the size distribution of the beads in the Wurster coater and assumed that all particles receive the same amount of coating, regardless of size. We shall relax this assumption because in a heterogeneous system, it is important to consider that not all particles in the spray zone will have the same coating mass gain rate. Depending on the bead size, we will now assume that the rate of coating mass accumulation is proportional to the cross-sectional area of the particle exposed to the spray. Accordingly, the single-particle mass gain rate is

$$\dot{m} = \frac{Qs_{ml}\eta\pi d^2}{S_w} \quad (17.40)$$

where

S_w is the average total surface area of all particles in the spray zone that can be determined directly from CFD-DEM.

The total coating mass acquired by any particle that visits the spray zone will then be the rate of mass gain (Eq. 17.40) multiplied by the RT in the spray zone. We can use the Monte Carlo approach to randomly sample the single-visit RT in the spray and bed zones. However, from the detailed CFD-DEM results presented in Section 17.4.5, we can now construct the RTDs for each size class (Figure 17.16). We have already seen that for the present Wurster coater, the bead size seems to have a negligible effect on the RT (see Figures 17.9 and 17.10), which is also corroborated when we examine the full RTD curves in Figure 17.16. Even though the changes with varying bead size are small, we will sample the single-visit RT from these new bead size-based RTDs (Figure 17.16) for our Monte Carlo simulation. For a single particle, the evolution of the cumulative spray-zone RT is determined by repeatedly sampling the spray-zone and bed-zone RTDs for the appropriate particle size class.

Now that we are considering bead polydispersity, the definition of the CoV needs to be refined. One may define the CoV for the coating mass on the beads, as we did in Figure 17.15 where we assumed that all particles gain coating at the same rate irrespective of size. However, given the fact that smaller size classes have smaller surface area per bead, these smaller particles will, in reality, accumulate less coating compared with larger particles. (Here, we assume that the droplet sizes are much smaller than the bead sizes and that the coating spray is uniformly distributed on the bead surface; in reality, there will be some intra-bead coating heterogeneities as spray droplets impact and spread unevenly on bead surfaces.) Therefore, the introduction of particle size polydispersity can lead to a coating mass CoV that approaches a nonzero asymptote for long process times (Figure 17.17a), compared with an asymptotic value of zero when all particles were assumed to have the same size-independent rate of mass gain (Figure 17.15). If one

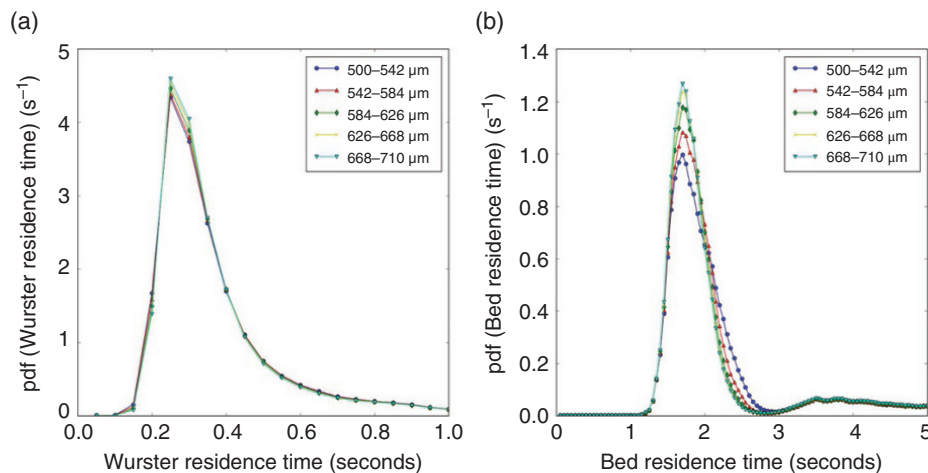


FIGURE 17.16 The residence time distributions (RTDs) of particles in the (a) spray region (inside Wurster tube) and (b) bed region of the Glatt-15/30 coater. The distributions are shown for a baseline air flow rate of $850 \text{ m}^3/\text{h}$ and using the same conditions as those used for Figures 17.13–17.15. The RTDs are shown for different particle size classes (refer to Figures 17.9 and 17.10).

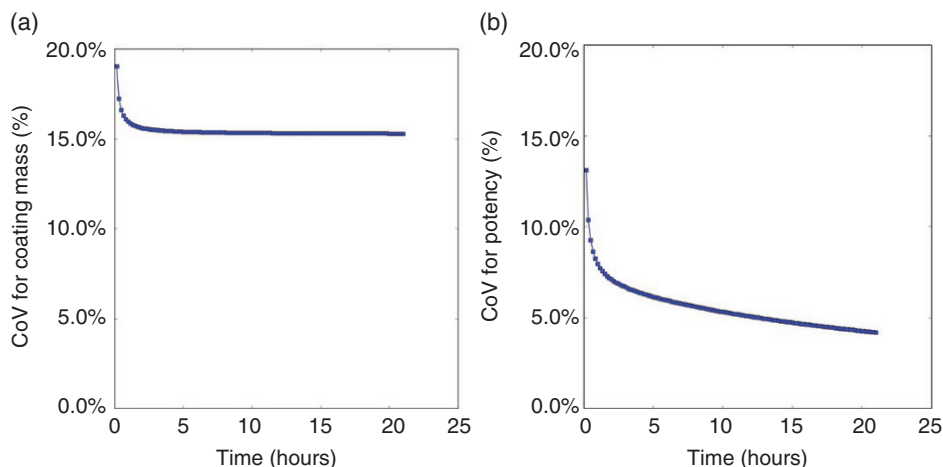


FIGURE 17.17 (a) CoV of coating mass for the system with nonuniform particle sizes. (b) CoV of potency for the system of nonuniform particle sizes.

considers a subpopulation consisting of same-sized particles, the associated coating mass CoV will be inversely proportional to the square root of time, approaching zero after very large process times; i.e. particles of the same size will accumulate nearly the same amount of coating. However, due to size-dependent small variations in flow behavior and differences in mass gain rate, subpopulations of different sizes may converge to different values of average coating mass per particle. Hence, the coating mass CoV may not converge to zero for the overall bead population.

In many applications, coating mass uniformity may not provide the most useful metric for product quality. For example, consider microspheres that are coated with a drug and then loaded into a capsule. Capsules are typically filled to a particular weight, so a consistent drug dosage can be achieved by maintaining a constant ratio of coating mass to total bead mass for each microsphere. (In instances where the coating solution is a mixture of drug and other components, the coating mass must be multiplied by the mass fraction of the API in the coating solution to obtain the actual amount of drug that is applied to the beads.) This ratio of API coating mass (after adjusting for the API mass fraction in the coating solution) to total bead mass is referred to as the potency, and the CoV of the bead potency would be of interest.

The CoV for potency is plotted in Figure 17.17b as a function of the process time. In this representative example, the potency computations take into account the amount of drug in the spray liquid, i.e. the nondrug and volatile components are excluded from the calculations. We see that although the CoV for coating mass is large (Figure 17.17a), the predicted CoV for potency is much smaller (<5%) at the end of the process. (Recall that the predictions in Figure 17.17 are under the more realistic assumptions where the coating mass gain is proportional to the bead surface area.) Such CoV predictions

can help in optimizing the process parameters, such as decreasing the spray rate to achieve a lower CoV, if needed. For pharmaceutical dosage forms, speaking very generally, the desired CoV for drug potency is on the order of 5% [38].

In other scenarios, dosage forms with a controlled release coating, a uniform coating thickness may be desired, which, in turn, dictates the release rate kinetics of the drug from the beads. For systems with nonuniform particle size distributions, uniformity in coating mass, potency, and coating thickness cannot be simultaneously satisfied. For instance, since smaller particles have a relatively lower core mass, they must necessarily receive less coating mass than larger particles in order to achieve the same potency. (Readers are reminded that we assume our spray droplets to be much smaller than the bead size; in reality, however, the coating thickness would also depend on the size of spray droplets and on how these spray droplets spread on the bead surfaces.) Therefore, optimization of fluid bed coating processes requires identification of the uniformity measurement most relevant to the performance of the final product.

17.6 LONG-TERM COATING PREDICTIONS: DETAILED STOCHASTIC MARKOV CHAIN OR COMPARTMENTAL APPROACH

In the previous section, we introduced a simpler stochastic (Monte Carlo) model where the bed was divided into two zones: spray and bed. In this section, we present a more detailed compartment model for predicting coating uniformity. This present model uses Markov chains to track the particle movement through different regions of the fluidized bed coater, i.e. the compartments. The coating mass for each particle inside the spray zone is calculated. From the spray mass per particle, we derive distributions of process-relevant

quantities, e.g. coating thickness, coating mass, and coating potency. In the Markov chain model, we can introduce multiple compartments and, therefore, incorporate some of the effects of spray heterogeneity (spray shielding) and expect to obtain better predictions of bead coating distributions. The particle trajectories obtained from CFD-DEM simulations (Section 17.4) act as inputs for the compartment model.

17.6.1 Markov Chains and Markov Process

A Markov chain is a stochastic process with a finite number of states and defined transition probabilities between these states. A Markov chain does not have a “memory,” i.e. the transition from one state to another only depends on the current state, not any past states [39]. Markov chains have been applied to a wide range of scientific fields, e.g. in mathematics, computer science, physics, astronomy, biology, game theory, and economy [40], and have several applications in chemical and pharmaceutical engineering, e.g. reaction kinematics [41–44], particle flow and mixing [45–48], process control and process management [49, 50], and fluidized beds [51–53].

An example of a Markov chain is given in Figure 17.18. The Markov chain has four states S_1 to S_4 and eight transitions with given probabilities. For example, the probability of an element (e.g. a particle) in the system to transition from state S_1 to S_2 is 90%. All transition probabilities can be summarized into a probability matrix \mathbf{P} , where P_{ij} is the probability of changing from state S_i to S_j . The value of P_{jj} is the probability of state S_j being unchanged within the time interval. The probability matrix corresponding to the system in Figure 17.17 is

$$\mathbf{P} = \begin{pmatrix} 0.1 & 0.9 & 0 & 0 \\ 0 & 0.2 & 0.8 & 0 \\ 0 & 0.7 & 0 & 0.3 \\ 0.4 & 0 & 0.6 & 0 \end{pmatrix} \quad (17.41)$$

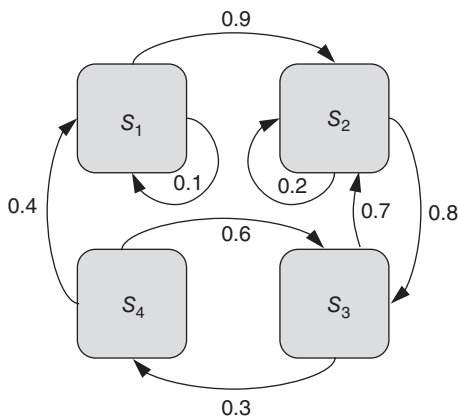


FIGURE 17.18 An example Markov chain with four states and eight transitions.

What the states represent depends on the application. In this chapter, the states represent compartments, which are distinct zones in the Wurster coater. The transition probabilities are related to the particle movement in the fluidized bed. In other applications (outside the scope of the present chapter), the states could also represent properties of particles, i.e. a range of particle size. The transition probabilities between particle sizes could then relate to agglomeration and breakage rates [51].

17.6.1.1 Formal Definition In the broadest sense, a Markov process is any stochastic process that satisfies the Markov property, i.e. the process can reliably be predicted with limited knowledge of the process history (the process is “memoryless”). A common restriction is that predictions of future states only depend on the current state, but this is not a necessary condition. The key property of Markov processes is that the quality of predictions cannot be improved even if the complete process history is known. Markov processes, in the broader sense, include both discrete and continuous processes [54].

In a narrower sense, the term Markov process is used for a time-continuous stochastic process on a finite or infinite countable state space, whereas the term Markov chains is used for time and state space discrete processes. In this chapter, we focus on discrete-time Markov chains (DTMC) on a finite state space.

17.6.1.2 Properties of Finite Discrete-Time Markov Chains (DTMC) Consider a Markov chain with N states and the state space $\mathcal{S} = \{S_1 \dots S_N\}$. The distribution over the state space can be given with a stochastic row vector \mathbf{x} of size N , where x_i defines the fraction of total elements (e.g. particles) that are present in the state S_i . A stochastic vector consists of only nonnegative probabilities, and the sum of all elements equals one:

$$\mathbf{x} = (x_1 \dots x_N) \quad (17.42)$$

with $\sum_{i=1}^N x_i = 1$ and $x_i \geq 0$

The DTMC advances in fixed time intervals Δt . If the distribution $\mathbf{x}^{(t)}$ at time step t is known, the distribution $\mathbf{x}^{(t+1)}$ in the next step of the process can be calculated by applying a probability matrix \mathbf{P} of size $N \times N$ to the stochastic vector. The probability matrix \mathbf{P} only contains nonnegative probabilities, and the probabilities in each row have a sum of one. This ensures that $\mathbf{x}^{(t+1)}$ is also a valid stochastic vector. The column sum is different from one in general, i.e. the different states have different probabilities to be occupied in the next time step:

$$\mathbf{x}^{(t+1)} = \mathbf{x}^{(t)} \cdot \mathbf{P} \quad (17.43)$$

with $\sum_{j=1}^N P_{ij} = 1 \forall i$ and $P_{ij} \geq 0$

If the matrix P is constant throughout the process, the Markov chain is homogeneous; if it changes over time, the Markov chain is inhomogeneous. If the probability matrix only depends on the time, i.e. $P^{(t)} = P(t)$, the Markov chain is **linear**, or if the probability matrix also depends on the current state, i.e. $P^{(t)} = P(t, \mathbf{x}^{(t)})$, it is nonlinear. A Markov chain of order m depends on the last m states, including the current state: $P^{(t)} = P(t, \mathbf{x}^{(t)}, \mathbf{x}^{(t-1)}, \dots, \mathbf{x}^{(t-m+1)})$.

In the homogeneous case, it is possible to calculate the n -step probability matrix P^n , because applying the same matrix to the stochastic row vector n times is equivalent to apply the matrix to the power of n once:

$$\mathbf{x}^{(t+n)} = \mathbf{x}^{(t)} \cdot \underbrace{P \cdot P \cdot \dots \cdot P}_{n \text{ times}} = \mathbf{x}^{(t)} \cdot P^n \quad (17.44)$$

A Markov chain is called *irreducible* if it cannot be divided into sub-chains and it is possible to reach all states from any state. A formal way to define the irreducibility in the homogeneous case uses the n -step probability matrix P^n : a homogeneous finite DTMC is with the probability matrix P irreducible, if there exists an integer value for n , such that all entries in P^n (P to the power of n) are strictly positive and therefore nonzero in particular.

In this chapter, we focus on *homogeneous* and *irreducible* Markov chains to describe the Wurster coating process. The process conditions of the Wurster coater are kept constant over time; thus the probability matrix P is constant, and the Markov chain is homogeneous. The irreducibility means that particles can move from any compartment to any other compartment in the Wurster coater over time. A reducible Markov chain would imply that there are two distinct closed particle flows in the Wurster coater that never exchange particles.

17.6.1.3 Exercise: Irreducible Markov Chain and Stationary Distribution Consider the Markov chain in Figure 17.18 and the corresponding probability matrix in Eq. (17.41). Check for irreducibility (a) intuitively by analyzing the graph and (b) by testing numerically the irreducibility criterion. (c) Start with a distribution $\mathbf{x}^{(0)} = (1, 0, 0, 0)$. Apply the probability matrix $n = 1, 2, 5, 10, 100, 1000$ times to see the distribution after n time steps $\mathbf{x}^{(n)}$. What are the corresponding n -step probability matrices P^n ? (d) Similar to (c) but start with a random initial distribution $\mathbf{x}^{(0)}$. What are the differences at low and high number of time steps?

$$P = \begin{pmatrix} 0.1 & 0.9 & 0 & 0 \\ 0 & 0.2 & 0.8 & 0 \\ 0 & 0.7 & 0 & 0.3 \\ 0.4 & 0 & 0.6 & 0 \end{pmatrix} \quad (17.45)$$

Solution

- a. Start at each state in the graph, follow the arrows, and check if all states can be reached from any state. Starting from S_1 , it is possible to reach S_1 and S_2 in the next time step, S_3 after two time steps, and S_4 after three time steps. Starting from S_3 , it is possible to reach S_2 and S_4 in the next time step and S_1 and S_3 in two time steps. Analogous for S_2 and S_4 , remember to check if it is possible to reach the starting step again.
- b. The criterion for irreducibility can be tested with the following piece of MATLAB® code. From the intuitive test in (a), we know that it takes at maximum three time steps to reach all states; therefore we expect that the smallest n that satisfies $P^n_{ij} > 0 \forall i, j$ is three.

```

%% Given: Probability Matrix

% less newlines in default Matlab output
format compact

P = [
0.1 0.9 0.0 0.0;
0.0 0.2 0.8 0.0;
0.0 0.7 0.0 0.3;
0.4 0.0 0.6 0.0;
];

%% (b) Irreducibility Test

NMAX=10;
found = 0;

Pn = P;
for n=2:NMAX
    Pn = Pn * P;
    if all(Pn(:) > 0)
        found=1;
        disp(['[PASS] Irreducibility Test
at n=' num2str(n)])
        break
    end
end
if not(found)
    disp(['[FAIL] Irreducibility Test for
n<=' num2str(NMAX)])
end

```

However, we see that the irreducibility test passes at $n = 4$. There is one nonzero entry in P^3 at position (4, 4), because it is not possible to start from state 4 and land on state 4 after *exactly* three steps. Check the graph in Figure 17.18!

```

%% Output:
[PASS] Irreducibility Test at n=4

```

c. MATLAB® supports the calculation of powers of matrices natively. We calculate P^n and $x^{(n)}$ for $n = 1, 2, 5, 10, 100, 1000$ and print the results:

```
%% (c) Distribution over time
```

```
x0 = [1 0 0 0]
```

```
for n=[1,2,5,10,100,1000]
```

```
    n = n           % force output of n
```

```
    Pn = P^n       % calculate n-step
```

```
           probability matrix
```

```
    xn = x0 * Pn   % apply to initial
```

```
           distribution
```

```
    disp(' ')     % newline
```

```
end
```

```
%% Output:
```

```
x0 =
     1     0     0     0
```

```
n = 1
```

```
Pn =
     0.1000     0.9000         0         0
         0     0.2000     0.8000         0
         0     0.7000         0     0.3000
     0.4000         0     0.6000         0
```

```
xn =
     0.1000     0.9000         0         0
```

```
n = 2
```

```
Pn =
     0.0100     0.2700     0.7200         0
         0     0.6000     0.1600     0.2400
     0.1200     0.1400     0.7400         0
     0.0400     0.7800         0     0.1800
```

```
xn =
     0.0100     0.2700     0.7200         0
```

```
n = 5
```

```
Pn =
     0.0346     0.5392     0.2513     0.1750
     0.0778     0.2899     0.5680     0.0643
     0.0224     0.5845     0.1962     0.1969
     0.0886     0.2382     0.6271     0.0461
```

```
xn =
     0.0346     0.5392     0.2513     0.1750
```

```
n = 10
```

```
Pn =
     0.0643     0.3635     0.4740     0.0983
     0.0437     0.4733     0.3360     0.1471
```

```
     0.0681     0.3431     0.4996     0.0892
     0.0397     0.4943     0.3095     0.1564
```

```
xn =
     0.0643     0.3635     0.4740     0.0983
```

```
n = 100
```

```
Pn =
     0.0542     0.4169     0.4068     0.1220
     0.0542     0.4169     0.4068     0.1220
     0.0542     0.4169     0.4068     0.1220
     0.0542     0.4169     0.4068     0.1220
```

```
xn =
     0.0542     0.4169     0.4068     0.1220
```

```
n = 1000
```

```
Pn =
     0.0542     0.4169     0.4068     0.1220
     0.0542     0.4169     0.4068     0.1220
     0.0542     0.4169     0.4068     0.1220
     0.0542     0.4169     0.4068     0.1220
```

```
xn =
     0.0542     0.4169     0.4068     0.1220
```

We see that the distributions $x^{(n)}$ converge to a stationary distribution for high values of n . High powers of the probability matrix P^n consist of copies of the converged distribution. The same is true for random initial distributions (d). While the early distributions (low n) are different from exercise (c), the late distributions are the same for both initial distributions. This is a fundamental property of all homogeneous irreducible Markov chains: they have a defined stationary distribution (often denoted as vector π) and the powers of the probability matrix P^n converge. The converged probability matrix P contains only copies of π . A rigorous proof of these properties is given in Ref. [54], but is out of scope of this chapter.

The Wurster coating process is a nice illustrative example of the stationary distribution. Although all particles are initially in the holdup region when the Wurster coater is filled, they will distribute over the process time. The overall mass fractions of particles in downbed, spray, and fountain region are constant throughout the process (stationary distribution), although individual particles are constantly cycling through these regions.

```
%% (d) Distribution over time, random
initial state
```

```
x0 = rand(1,4);
x0 = x0 / sum(x0)
```

```
for n=[1,2,5,10,100,1000]
```

```
    n = n           % force output of n
```

```
    Pn = P^n;       % n-step probability
```

```

matrices are the same      xn =
as in (c)                 0.0542    0.4169    0.4068    0.1220
% apply to initial
distribution
disp(' ')                 % newline
end                          n = 1000
%% Output                  xn =
                           0.0542    0.4169    0.4068    0.1220

x0 =
  0.4014    0.2035    0.3356    0.0595

n = 1
xn =
  0.0639    0.6369    0.1985    0.1007

n = 2
xn =
  0.0467    0.3239    0.5699    0.0596

n = 5
xn =
  0.0425    0.4857    0.3196    0.1521

n = 10
xn =
  0.0599    0.3868    0.4447    0.1086

n = 100

```

17.6.2 Compartment Model for Fluidized Bed Coating

17.6.2.1 Model Description CFD-DEM simulations are capable of producing highly accurate results; however, the high computational requirements (16 CPU hours per process second using the available hardware and software) make it impossible to simulate the complete coating process with durations on the order of 20 hours. In order to develop long-term predictions of quality attributes (averages and CoVs for coating thickness, coating mass, and potency), it is necessary to develop a simplified model.

In this section we introduce a Monte Carlo compartment model for long-term coating predictions (see Figure 17.19). Compartments are defined as nonoverlapping regions in the Wurster coater. Unlike the level of detail available in CFD-DEM simulations, the particle positions are not stored exactly in the compartment model; only the compartment number is tracked for each particle. The movement of particles between the compartments is modeled as a Markov chain. The transition probability matrix can be directly derived from the detailed particle trajectories in the CFD-DEM results.

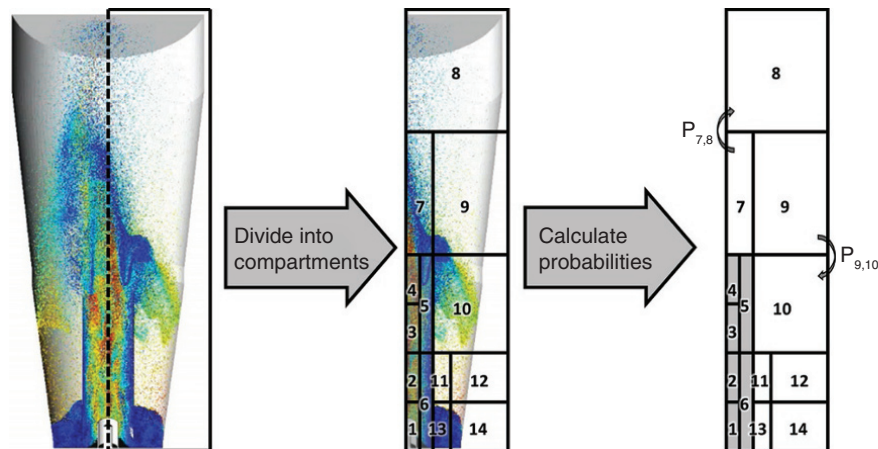


FIGURE 17.19 Subdivision of the Wurster coater into compartments. The movement of particles through compartments is modeled as a Markov chain process with a probability matrix P . The individual elements P_{ij} indicate the probability of a particle to move from compartment i to compartment j within a given time interval Δt . The spray zone is resolved into compartments 1–6 (gray).

The probability matrix indirectly incorporates information about all process settings that influence the particle movement, such as fluidization and atomization air flow rates, Wurster gap size, and batch size. Hence, the probability matrix will systematically vary as these process settings are changed. The other process parameters, for example, spray rate, spray formulation, or process duration, do not influence (or are assumed not to affect) the particle movement. This means that a change in the process conditions, say, the fluidization air flow rate, requires a new probability matrix (from the corresponding new CFD-DEM simulation), but a change in spray settings is possible without recalculating a probability matrix.

In order to obtain a measure for coating uniformity, a representative sample of particles (e.g. 1 000–10 000) is simulated in the compartment model. Each particle moves independent of the other particles through the compartments, described by the probability matrix. If one or more particles are in a *spray compartment* (a compartment that belongs to the spray region of the Wurster coater), the spray mass is distributed among the particles inside that particular compartment. Thus, at any instant, the initial sizes of particles and the spray rate within a particular spray compartment are known. Hence, it is possible to calculate and update the new coating thickness, particle size, and coating potency for each particle. The process-relevant quality attributes (e.g. coating uniformity) are obtained by calculating statistics of the per-particle coating thicknesses.

The following sections explain the model in detail: the calculation of the probability matrices from DEM data is covered in Section 17.6.2.2, with an example in Section 17.6.2.3 describing the interpretation of the probabilities and an exercise problem in Section 17.6.2.4. The calculation of coating mass per particle is shown in Section 17.6.2.5, and Section 17.6.2.6 derives process-relevant quantities from particle data.

17.6.2.2 Defining Compartments and Calculating Probability Matrices The probability matrix is the core of the Markov chain model. The probability matrix is directly computed from the particle movements observed in the detailed CFD-DEM simulations. To ensure a sufficiently good calculation for the probability matrix to be used in the Markov chain, the CFD-DEM simulations data should span at least a few cycles of particle movement through the Wurster column and bed regions.

The number and shape of compartments can be arbitrary; however, there are some performance and model quality considerations. Generally speaking, sections of interest within the Wurster coater should be resolved into finer compartments, e.g. the Wurster tube in Figure 17.19. However, finer compartments need smaller time step sizes. Ideally, the time step size Δt and compartment size Δc should be chosen in such a way that it is impossible for particles to skip

compartments in a single time step. Therefore, the limiting factor is the maximum particle velocity v_{\max} :

$$v_{\max} \Delta t \leq \Delta c. \quad (17.46)$$

If a finer compartment resolution is required, the time step Δt has to be reduced. As the probability matrix is only valid for a given time step Δt , reducing the time step would mean that the probability matrices would have to be recomputed from the CFD-DEM simulations for the new value of Δt . For our work, we have chosen a time step size of $\Delta t = 0.05$ s, which balances the desired minimum compartment size and the size of the output files generated by CFD-DEM (smaller time steps would require more frequent file output and, hence, larger file sizes).

Once a time step and compartment layout has been defined, the probability matrix has to be calculated. The particle positions are converted from Cartesian coordinates (x , y , z) in the Y-up convention (i.e. gravity along negative Y) to cylindrical coordinates (r , h , α):

$$\begin{aligned} r &= \sqrt{x^2 + z^2} \\ h &= y \\ \alpha &= \tan^{-1} \left(\frac{z}{x} \right) \end{aligned} \quad (17.47)$$

The values for r and h are then used to determine the compartment number of the particles. The compartment numbers over time per particle can then be stored in a simple table for further processing and calculation of probability matrices (see Section 17.6.2.4).

17.6.2.3 Example: Residence Time and Residence Probability This example presents the method to compute the RT quantities and probability matrices, given a representative set of particle trajectory data (for three particles). Table 17.2 shows example compartment data for 3 particles over 12 time steps. The highlighted data is the compartment number at each time step and used for the probability matrix calculation. The data highlights the connections between RT, CT, and RP. The RT is the time particles spend within a certain compartment. The CT is the time a particle needs to circulate through the Wurster coater, i.e. the time needed to visit all compartments and revisit the initial compartment. The RP is the probability that the particle is in the given compartment at a random point in time. The RP is related to the MCT, MRT in the spray zone, and the stationary distribution of the Markov chain (see Section 17.6.1.2):

$$RP(j) \approx P_{jj}^{\infty} \approx \frac{MRT(j)}{MCT} \quad (17.48)$$

Consider particle 1 in Table 17.2. It visits compartment 1 three times with durations $1\Delta t$, $2\Delta t$, and $3\Delta t$, where Δt is the

TABLE 17.2 Comparison of Residence Time and Residence Probability

	t_1	t_2	t_3	t_4	t_5	t_6	t_7	t_8	t_9	t_{10}	t_{11}	t_{12}	MRT(1)	MCT	RP(1) (%)
Particle 1	1	2	3	1	1	2	3	1	1	1	2	3	$2\Delta t$	$4\Delta t$	50
Particle 2	1	1	1	1	1	1	2	2	2	3	3	3	$6\Delta t$	$12\Delta t$	50
Particle 3	1	1	2	2	3	3	1	1	2	2	3	3	$2\Delta t$	$6\Delta t$	33

time step. This leads to an MRT for compartment 1 of $MRT(1) = 2\Delta t$. The cycles are of durations $3\Delta t$, $4\Delta t$, and $5\Delta t$, so the MCT is $4\Delta t$. The RP for compartment 1 is 50% because the particle is 50% of the time in compartment 1. The RP is also consistent with the MRT and MCT.

Particle 2 has the same RP as particle 1, but with a longer MRT and MCT. Particle 3 has the same MRT as particle 1, but the cycles are longer, leading to an overall lower RP. The distinction between RP and MRT is important to keep in mind when analyzing the results of the Markov chain, because the RP is the more direct output of the model.

17.6.2.4 Exercise: Calculating a Probability Matrix

Using the text file `example-compartments.txt` in the Appendix 17.A, (a) implement a script that analyzes

the compartment data to calculate the corresponding probability matrix. Write the probability matrix to a file. (b) Check the column and row sums of the calculated matrix. Is it a valid probability matrix?

Solution

- The file `example-compartments.txt` is a simple text file and can directly read in with MATLAB's `load` function. The number of compartments is directly obtained from the highest number in the compartment data.

The outer loop iterates over particles/lines and the inner loop iterates over time steps/columns. Each transition from one compartment to the next is counted in a matrix P . The row sums are then normed to one and the probability matrix can be written to a file.

```

%% (a) read compartment data

data = load('example-compartments.txt'); % compartments per particle over time
NC = max(data(:)); % number of compartments

%% calculate probability matrix

P = zeros(NC, NC); % probability matrix
for particle=1:size(data,1) % iteration over particles / lines
    for time=2:size(data,2) % iteration over time steps / columns
        src = data(particle,time-1); % source compartment
        dst = data(particle,time); % destination compartment
        P(src,dst) = P(src,dst) + 1; % count transitions from src to dst
    end
end

for i=1:NC
    P(i,:) = P(i,:) / sum(P(i,:)); % normalize row i
end

% storing probability matrix
save('example-probability-matrix.txt', 'P', '-ascii')

```

The resulting probability matrix is given as file `example-probability-matrix.txt` in the Appendix 17.A. For

reference, the first 4 rows and columns of the 8×8 matrix are printed here.

```

9.2165899e-01  6.9124424e-03  5.0691244e-02  2.0737327e-02  ...
1.2765957e-01  8.0851064e-01  6.3829787e-02  0.0000000e+00  ...
0.0000000e+00  0.0000000e+00  8.3435583e-01  1.6564417e-01  ...
0.0000000e+00  0.0000000e+00  2.1881838e-02  7.5492341e-01  ...
...           ...           ...           ...           ...

```

- b. The column and row sums of the probability matrix are calculated with the MATLAB built-in sum.

```
%% (b) column and row sums

% less newlines in default Matlab output
format compact

colSums = sum(P,1)
rowSums = sum(P,2)
```

The row sums must be equal to one – the column sums can be different from one. That means that every particle has to go somewhere in the next time step, but the different compartments have different probabilities to be chosen as a destination. Because all rows are normed to 1, it is a valid probability matrix.

```
%% Ouput

colSums = 1.1009    0.8154    0.9708
0.9413    0.9184    1.1942    0.9247
1.1344
rowSums =
 1.0000
 1.0000
 1.0000
 1.0000
 1.0000
 1.0000
 1.0000
 1.0000
 1.0000
```

17.6.2.5 Coating Mass Calculation After setting up the Markov chain to model the particle movement, the next step is to calculate the coating mass the particles obtained when they are inside the spray zone. If the spray zone is divided into multiple compartments, it is necessary to calculate the coating mass per compartment \dot{m}_j :

$$\dot{m}_j = \dot{m}_{\text{spray}} \cdot f_{\text{susp}} \cdot w_j \cdot \frac{N_{\text{CM}}}{N_{\text{real}}} \quad (17.49)$$

where

\dot{m}_{spray} is the total spray rate of the coating process.

f_{susp} is the solid fraction of the spray suspension, i.e. that part of the spray mass that remains as coating mass after the particles are dried.

w_j is the coating weight of compartment j , i.e. that fraction of the total spray that is distributed within compartment j .

The last term $N_{\text{CM}}/N_{\text{real}}$ scales the total spray mass from process scale with N_{real} particles (~300 million beads) down to the number of particles N_{CM} (~10 000 representative beads) in the compartment model.

Under perfect conditions the sum of the spray weights w_j is 1. If there are spray losses of, say, 3% in the process, the sum of weights equals 0.97. The results from CFD-DEM simulations (Section 17.4) have been used to obtain the coating weights w_j per compartment. A simple linear ray tracing algorithm was used to determine the (nonuniform) distribution of the spray among the compartments inside the Wurster coater. By tracking the termination points of a large number of rays emanating from the spray nozzle, i.e. the point where a ray encounters a particle, the relative distribution of the spray among the spray compartments could be calculated. As expected, the compartments closer to the spray nozzle received a greater amount of the coating liquid.

The spray mass per compartment is constant as long as the process settings do not change; therefore it can be calculated at the beginning of the simulation. However, the distribution of the spray mass between particles has to be performed in each time step Δt .

The coating mass added to compartment C_j within a time step is $\Delta t \cdot \dot{m}_j$. This coating mass is distributed among all particles i within compartment C_j . The amount of coating each particle receives is proportional to the particle surface area. Thus, the coating mass added to particle p at time step t is

$$m_{p,t} = \Delta t \cdot \dot{m}_j \cdot \frac{A_p}{\sum_{i \in C_j} A_i} \quad (17.50)$$

17.6.2.6 Critical Quality Attributes The most basic output of the compartment model is the coating mass for each particle. Based on the coating mass distribution and the initial particle size and masses, it is possible to calculate the critical quality attributes, such as the uniformity of coating thickness and coating potency.

The coating mass per particle is simply the sum of spray events until the current time step t . For the algorithm it is only necessary to track the cumulative coating mass $M_{p,t}$ per particle, not the mass $m_{p,\tau}$ of each single spray event per time step and particle:

$$M_{p,t} = \sum_{\tau=0}^t m_{p,\tau} \quad (17.51)$$

The current diameter $d_{p,t}$ of particle p at time step t can be calculated with

$$d_{p,t} = \left(d_{p,0}^3 + \frac{6}{\pi \cdot \rho_c} \cdot M_{p,t} \right)^{1/3} \quad (17.52)$$

where

$d_{p,0}$ is the initial particle diameter at $t = 0$.

ρ_c is the density of the coating mass.

$M_{p,t}$ is the total coating mass of particle p at time step t .

The coating thickness $T_{p,t}$ per particle is then simply

$$T_{p,t} = \frac{d_{p,t} - d_{p,0}}{2} \quad (17.53)$$

The coating potency $P_{p,t}$ of a particle is the ratio between active ingredient in the coating and the total particle mass. The mass of the active ingredient is calculated using the fraction of API in the spray solid, f_{API} , and the coating mass:

$$P_{p,t} = \frac{f_{API} \cdot M_{p,t}}{m_0 + M_{p,t}} \quad (17.54)$$

where

m_0 is the initial mass of the particle, i.e. the mass of the uncoated bead:

$$m_0 = \rho_p \cdot d_{p,0}^3 \cdot \frac{\pi}{6} \quad (17.55)$$

where the particle density is ρ_p .

It is possible to calculate and plot the particle size distribution, coating thickness distribution, coating mass distribution, and coating potency distribution at each time step. The distributions are characterized by arithmetic mean, standard deviation, and the CoV. The CoV is the standard deviation divided by the arithmetic mean, also known as relative standard deviation. The CoV is a dimensionless measure for the uniformity of the distributions (see Section 17.5).

17.6.3 Applications: Prediction of Coating Uniformity

This section discusses the output of the compartment model and coating uniformity predictions. Two measures for coating uniformity will be discussed: the CoV for coating thickness and the CoV for coating potency – the quantity of greater interest depends on the specific application (see Section 17.2 for a brief discussion). For protective coatings it is more important to optimize the coating thickness to ensure consistent drug release rate kinetics of all coated beads. However, for coatings composed of active drug substances, lower variation in coating potency is desired to keep the drug load consistent.

Our compartment model (see Figure 17.19) consists of 14 compartments of which 6 are spray compartments. The

probability matrices have been obtained from CFD-DEM simulations (Section 17.4). Unless otherwise noted, we use the representative baseline operating conditions for the results discussed in this section: the fluidization airflow is 850 m³/h, and the Wurster gap size is 30 mm. A spray formulation with 14% w/w solid fraction has been applied with a spray rate of 130 g/min for 18 hours.

17.6.3.1 Particle Size Distribution and Coating Thickness The compartment model enables us to directly calculate coating thickness and particle size distribution based on the initial particle sizes and coating mass per particle. Experimentally, the coating thickness is obtained by analyzing the shift of the particle size distributions. A comparison of predicted and measured particle size distributions is shown in Figure 17.20.

Figure 17.21 shows the predicted and measured particle average coating thickness. The experimental coating thickness has been obtained by taking samples during the second coating stage (PSD 1 to PSD 2, indicated by an arrow in Figure 17.20). The coating thickness has been measured with optical coherence technology, a technique that exploits the different refractive indices of coating beads and coating mass [55]. The compartment model predicts the shape of the particle size distribution and the particle growth rate with good accuracy, despite the underlying assumptions and simplifications embedded in the reduced-order compartment model.

17.6.3.2 Effect of Varying of Spray Rate and Coating Process Duration This section summarizes the effect of varying spray rates on coating uniformity. The total coating time (i.e. process duration) has been adjusted to keep the total coating mass constant. All compartment model runs have

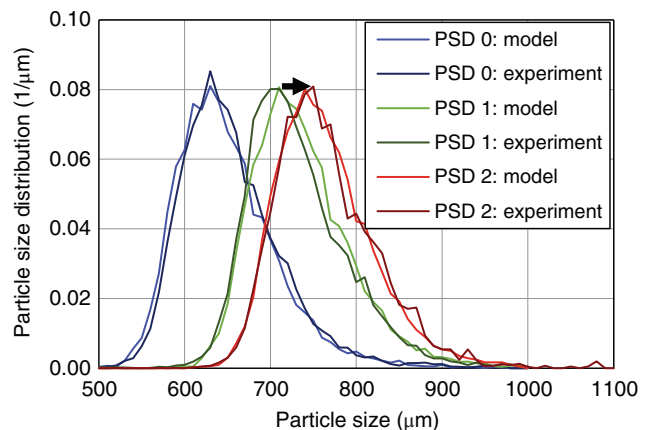


FIGURE 17.20 Comparison of experimental (dashed) and model (solid) particle size distributions. The results of two coating stages (going from PSD 0 to PSD 1 and from PSD 1 to PSD 2) are shown.

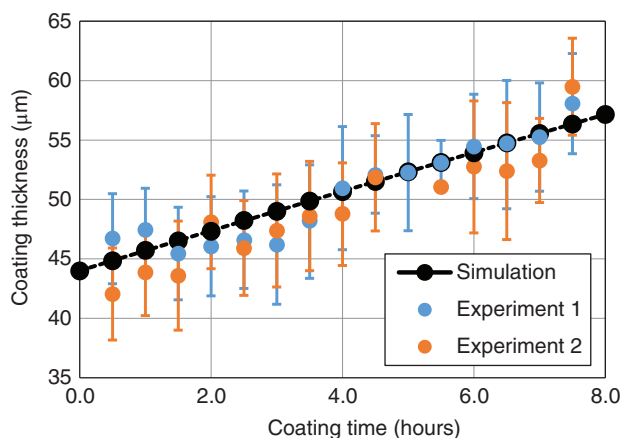


FIGURE 17.21 Comparison of the predicted particle growth, i.e. coating thickness, and OCT measurements of samples collected from coating experiments (two replicates). This coating step corresponds to the growth of particles from PSD 1 to PSD 2 in Figure 17.20.

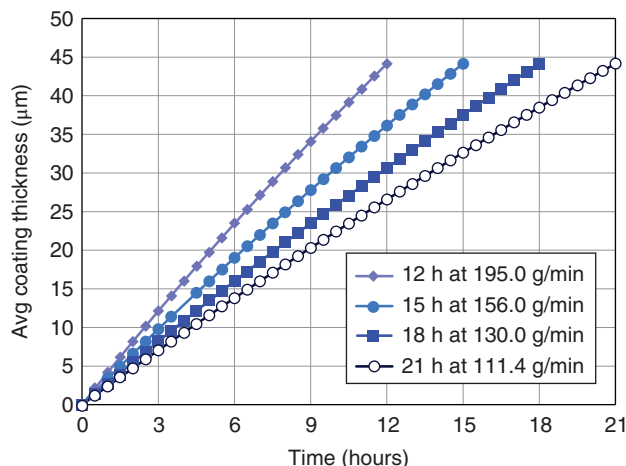


FIGURE 17.22 Coating thickness over time at different spray rates. The total coating time has been adjusted to keep the total spray mass constant across compartment model simulations.

been performed with the same probability matrix, because the process parameters that influence the particle movement (fluidization airflow, batch size, Wurster gap height) are kept constant; the spray rate does not affect the bead flow dynamics.

The average coating thickness at the end of the process is the same across the different compartment model runs (Figure 17.22) because the same amount of coating is applied to the same number of beads. However, the coating thickness uniformity is better for longer process times at lower spray rates (Figure 17.23), as reflected by a smaller CoV value. Although the coating thickness variability is higher (larger CoV) with increased spray rates, the different spray rates have a minor influence on the CoV for potency, i.e. bead-to-bead potency uniformity (Figure 17.24). Although the

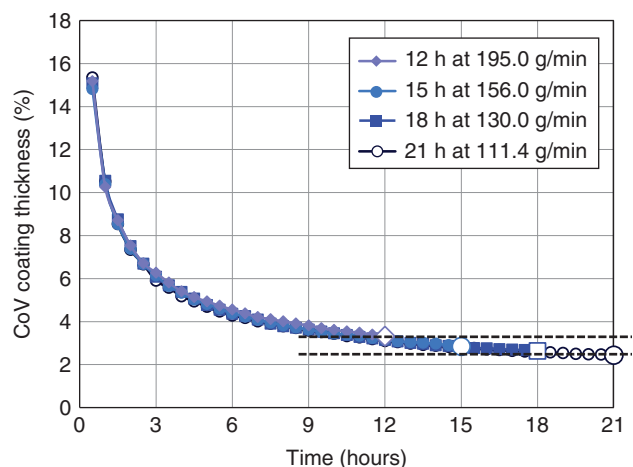


FIGURE 17.23 Time-evolution of the coefficient of variation of the coating thickness. Longer process times at lower spray rates improve the coating thickness uniformity.

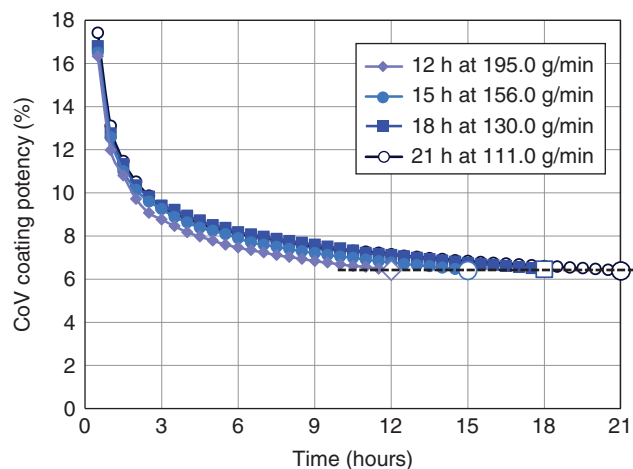


FIGURE 17.24 Time-evolution of the coefficient of variation of the coating potency. The different spray rates do not influence the potency uniformity of the coated beads.

CoV reduces further as the coating time is increased from 12 to 21 hours, all final CoV values are smaller than 4%, which is generally sufficient for the current product. These results suggest that it may be acceptable to have a shorter coating duration with a higher spray rate for the current product and process.

17.7 ALTERNATIVE APPROACHES TO PREDICT LONG-TERM COATING BEHAVIOR: POPULATION BALANCE MODELS (PBM)

The focus of this chapter has been on developing a multiscale modeling approach where particle-scale flow information,

obtained from detailed CFD-DEM simulations, are used to predict the long-term coating attributes using a reduced-order model. For our work, we chose to use stochastic particle-based methods: first, we presented a simpler Monte Carlo approach, followed by a more detailed Markov chain model. Both approaches entail tracking a finite number of discrete particles; thus, both approaches are stochastic in nature.

Stochastic approaches are not the only available method to describe the long-term growth of particles. Another technique to approach such problems is the population balance model (PBM) [56]. A number of studies have modeled coating processes using PBM (see, for example, Li et al. [57]), and some authors have specifically looked at Wurster coaters/granulators or other similar coating units (see Refs. [58–60]), which are but a few of the many studies available. The purpose of this section is to present a holistic overview of the PBM approach, without delving too deeply into the mathematical details.

The governing equation of PBM, in general, can be expressed as follows:

$$\frac{dn}{dt} + \frac{d(Gn)}{d\lambda} = \mathfrak{N}_{\text{birth}} - \mathfrak{N}_{\text{death}}, \quad (17.56)$$

where

λ is the quantity of interest that we wish to track.

n is the number density function (population) for the quantity of interest.

In our present example, we should track the coating mass on the beads. Therefore, we define $n(\lambda, t)d\lambda$ as the fraction of particles that have a coating mass that lies between λ and $\lambda + d\lambda$. Thus, the governing equation given by Eq. (17.56) tracks the population of particles that falls in each (infinitesimally small) size class given by the interval $[\lambda, \lambda + d\lambda]$.

The second term on the left-hand side describes the flux of particles between adjacent size classes; hence, G represents a local “velocity” with which the population moves. In the fluidized bed, beads gain coating mass in the spray zone and transition gradually from a smaller coating mass class to a larger coating mass class. The term G should then be a function of the spray rate (which we will not derive here, but readers can refer to any of the cited literature).

The terms $\mathfrak{N}_{\text{birth}}$ and $\mathfrak{N}_{\text{death}}$ on the right-hand side of Eq. (17.56) describe the “births” and “deaths,” respectively, in the population. An example would be in the case of, say, breakage of particles. The fracture of a larger particle into smaller daughter sizes would be reflected as $\mathfrak{N}_{\text{death}}$ for the larger size class and $\mathfrak{N}_{\text{birth}}$ events for the corresponding smaller size classes. In the present Wurster coating example, we can ignore $\mathfrak{N}_{\text{birth}}$ and $\mathfrak{N}_{\text{death}}$, so we do not expect partially coated particles to appear in any size class (unless they were deliberately introduced to the flow from outside). One scenario

where the birth and death terms may be relevant is twinning or agglomeration of wet beads, suddenly producing a composite bead with a larger amount of coating mass. We may choose to ignore this scenario for bead coating.

The next consideration is to incorporate the various regions in the Wurster coater, viz. the bed and spray zones, along with the RTDs in each zone. Li et al. [57] have incorporated the flow information from their spray coater via linking compartment models and PBM. In the simplest approximation, the coater was divided into two regions – bed and spray zones – and a system of two coupled PBM differential equations was developed. Each compartment was assumed to be a well-mixed continuous stirred tank reactors (CSTR), and the RTD was taken to be the exponential function $\text{RTD}(t) = \frac{1}{\tau} e^{-t/\tau}$, where τ is the mean RTD. To incorporate more complex flow behavior, one needs to consider a compartment model composed of a series of CSTRs connected in series or parallel that yields the desired RTD in each zone. The number of coupled PBM equations would also increase in this case.

Another complexity that we may want to consider is the size polydispersity of the beads: we will need to redefine our population to include the bead diameter, such that $n(\lambda, d_p, t) \cdot d\lambda \cdot dd_p$ is the population fraction of particles that have a coating mass between λ and $\lambda + d\lambda$ and have a bead diameter between d_p and $d_p + dd_p$; i.e. the population distribution now incorporates two internal coordinates λ and d_p . Although it is not difficult to formulate the new governing equations incorporating the bead sizes, solving the new set of coupled differential equations can be numerically challenging. Hence, most reported PBM studies implement simpler structures to represent their system.

In contrast to PBM, our chosen Monte Carlo and Markov chain approaches include the parameters of interest – viz. the flow and RTDs in the different zones and the effect of bead polydispersity – in a relatively easy manner. By defining a finite number of bead size classes and randomly sampling from RTDs to obtain particle trajectories, much of the numerical complexity of solving PBM could be avoided. Nonetheless, PBM is a more rigorous approach and is preferred in many scenarios.

17.8 SUMMARY AND OUTLOOK

In this chapter, we have presented methodology for predicting bead coating uniformity in a Wurster coating process. The multiscale approach is composed of two steps: high-resolution simulations to obtain particle trajectory data inside the fluidized bed and then extrapolating the flow results in a reduced-order model to compute the long-term coating attributes. The CFD-TFM and coupled CFD-DEM simulations were considered for generating the detailed particle

trajectories – eventually CFD-DEM was used for the present application.

Two coarse-grained approaches were used to extrapolate the particle trajectories to predict coating levels – a simpler Monte Carlo approach and a more detailed Markov chain model – both of which are stochastic techniques. The long-term coating attributes that were predicted by the reduced-order models include bead coating mass, potency, and thickness, with the coating CoV (which is a measure of uniformity) generally reported for these quantities. The predicted results show good agreement with experimental measurements.

From an industrial perspective, the purpose of these “virtual experiments” is to quantitatively understand the Wurster coating process and, in particular, the dependence of bead coating uniformity on process parameters and bead properties. Some key findings emerge from this study. From the CFD-DEM simulations, we ascertain that the bead flow inside the Wurster coating is robust; no segregation or dead zones are observed. Given the polydispersity of the beads, the fluidization behavior of different-sized beads was a concern – the question was whether the smaller beads have a significantly different flight time or CT and thus significantly different coating attributes compared with larger beads. The CFD-DEM results show that there are negligible differences in the Wurster (spray-zone) RTDs and CTD, demonstrating that there is practically little to no difference in fluidization behavior for different bead size classes.

The long-term coating models provided valuable information that help select and optimize the process parameters. For example, we have presented a scenario where the optimal operating conditions for minimizing coating thickness CoV are different from the optimal potency CoV. In practical terms, almost all the conditions explored yielded comparable metrics for coating uniformity – the CoV values predicted are already small (i.e. good inter-bead coating uniformity) and all conditions within a few percentage points. This serves to establish that the current equipment and process are robust within the ranges of operating conditions investigated. However, this generalization applies only to the quantity of the coating on the beads; no claims are yet being made about the quality of the coating, e.g. the coat porosity, coating defects, and coverage on the bead surface.

The methodology that was presented can be utilized in a variety of different situations. In addition to building a quantitative understanding of the current process, the methodology can be applied to virtually scale up the process. Generally, a coating process is first developed on a smaller scale (~1 kg of beads) during the early stages when drug product is scarce. Subsequently, the process is scaled up to clinical and commercial scales (~10–100 kg) while attempting to preserve bead coating attributes across scales. This process can be challenging as large-scale experiments are expensive (on the order of \$100 000 for a ~50 kg batch).

Therefore, it is preferable to eliminate experiments in favor of computational studies or “virtual experiments” that probe the parameters of interest.

To apply the multiscale approach for scale-up purposes, the CFD simulations need to be repeated for the different scales and at the different operating conditions of interest. Thenceforth, the compartment model needs to be modified: the underlying structure remains the same, but the compartments and the probability matrices need to be redefined. The effect of spray parameters can be explored with relative ease using the compartment model, with the goal of preserving the potency CoV or coating thickness CoV, whichever is of interest. Although we have not presented scale-up in the present chapter, it is certainly an area of active interest within our group and in the pharmaceutical industry in general.

Eventually, the goal is to achieve a desired dissolution profile for the beads – first *in vitro* and then *in vivo*. The dissolution profile, i.e. the rate of drug release from the beads, depends on the coating amount (coating thickness, coating mass, and potency) as well as the coating quality (coat porosity, droplet spreading, and coat continuity on bead surface). The present work addresses the first aspect, viz. coating amount on beads and bead-to-bead coat uniformity. The coating quality is hypothesized to depend on the drying thermodynamics of the beads within the beds; for example, rapid drying could potentially result in poorer coalescence of the droplets on the bead surface and, hence, a more porous coating. The effect of bead-drying thermodynamics is currently being studied and will eventually be used to develop a more predictive end-to-end modeling platform for the Wurster coating process.

17.A APPENDIX

List of files to be included as part of the exercise problems:

MATLAB files:

example_compartment.m
example_markov.m

Input text files:

example-compartment.txt
example-probability-matrix.txt
output-markov.txt
output-probability-matrix.txt

REFERENCES

1. Jajcevic, D., Siegmann, E., Radeke, C., and Khinast, J.G. (2013). Large-scale CFD-DEM simulations of fluidized granular systems. *Chem Eng Sci* 98: 298–310. <https://doi.org/10.1016/j.ces.2013.05.014>.

2. Beetstra, R., van der Hoef, M.A., and Kuipers, J.A.M. (2007). Numerical study of segregation using a new drag force correlation for polydisperse systems derived from lattice-Boltzmann simulations. *Chem Eng Sci* 62: 246–255. <https://doi.org/10.1016/J.CES.2006.08.054>.
3. Felice, R.D.I. (1994). The voidage function for fluid: particle interaction systems. *Int J Multiph Flow* 20: 153–159.
4. Gidaspow, D. (1994). *Multiphase Flow and Fluidization: Continuum and Kinetic Theory Descriptions*, 1e. San Diego, CA: Academic Press.
5. Syamlal, M. and O'Brien, T.J. (1989). Computer simulation of bubbles in a fluidized bed. *AIChE Symp. Ser.* 85: 22–31.
6. Koch, D.L. and Hill, R.J. (2001). Inertial effects in suspension and porous-media flows. *Anu Rev Fluid Mech* 619–647. <https://doi.org/10.1146/annurev.fluid.33.1.619>.
7. Ergun, S. (1952). Fluid flow through packed columns. *Chem Eng Prog* 48: 89–94.
8. Wen, C. and Yu, Y. (1966). Mechanics of fluidization. *Chem Eng Prog Symp Ser* 62: 100.
9. Cundall, P.A. and Strack, O.D.L. (1979). A discrete numerical model for granular assemblies. *Géotechnique* 29: 47–65. <https://doi.org/10.1680/geot.1979.29.1.47>.
10. Deen, N.G., Van Sint Annaland, M., Van der Hoef, M.A., and JAM, K. (2007). Review of discrete particle modeling of fluidized beds. *Chem Eng Sci* 62: 28–44. <https://doi.org/10.1016/j.ces.2006.08.014>.
11. Radeke, C.A., Glasser, B.J., and Khinast, J.G. (2010). Large-scale powder mixer simulations using massively parallel GPU architectures. *Chem Eng Sci* 65: 6435–6442. <https://doi.org/10.1016/j.ces.2010.09.035>.
12. Zeller, C. (2011). *CUDA C/C++ Basics What Is CUDA?* Santa Clara, CA: NVIDIA Corporation.
13. Pepiot, P. and Desjardins, O. (2012). Numerical analysis of the dynamics of two- and three-dimensional fluidized bed reactors using an Euler–Lagrange approach. *Powder Technol* 220: 104–121. <https://doi.org/10.1016/j.powtec.2011.09.021>.
14. Ding, J. and Gidaspow, D. (1990). A bubbling fluidization model using kinetic theory of granular flow. *AIChE J* 36: 523–538. <https://doi.org/10.1002/aic.690360404>.
15. Wang, J., van der Hoef, M.A., and Kuipers, J.A.M. (2013). Comparison of two-fluid and discrete particle modeling of dense gas-particle flows in gas-fluidized beds. *Chem Ing Tech* 85: 290–298. <https://doi.org/10.1002/cite.201200101>.
16. Chen, X. and Wang, J. (2014). A comparison of two-fluid model, dense discrete particle model and CFD-DEM method for modeling impinging gas: solid flows. *Powder Technol* 254: 94–102. <https://doi.org/10.1016/j.powtec.2013.12.056>.
17. Du, W., Bao, X., Xu, J., and Wei, W. (2006a). Computational fluid dynamics (CFD) modeling of spouted bed: influence of frictional stress, maximum packing limit and coefficient of restitution of particles. *Chem Eng Sci* 61: 4558–4570. <https://doi.org/10.1016/j.ces.2006.02.028>.
18. Bao, X., Wei, D., and Jian, X. (2013). An overview on the recent advances in computational fluid dynamics simulation of spouted beds. *The Canadian Journal of Chemical Engineering* September: <https://doi.org/10.1002/cjce.21917>.
19. Kuipers, J.A.M., Van Duin, K.J., Van Beckum, F.P.H., and Van Swaaij, W.P.M. (1992). A numerical model of gas-fluidized beds. *Chem Eng Sci* 47: 1913–1924. [https://doi.org/10.1016/0009-2509\(92\)80309-Z](https://doi.org/10.1016/0009-2509(92)80309-Z).
20. Lun, C.K.K., Savage, S.B., Jeffrey, D.J., and Chepurmiy, N. (1984). Kinetic theories for granular flow: inelastic particles in Couette flow and slightly inelastic particles in a general flow field. *J Fluid Mech* 140: 223–256. <https://doi.org/10.1017/S0022112084000586>.
21. Lun, C.K.K. and Savage, S.B. (1986). The effects of an impact velocity dependent coefficient of restitution on stresses developed by sheared granular materials. *Acta Mech* 63: 15–44. <https://doi.org/10.1007/BF01182538>.
22. Syamlal, M., Rogers, W., and O'Brien, T.J. (1993). *MFIX Documentation: Volume 1, Theory Guide*. Springfield, VA: National Technical Information Service.
23. Gidaspow D, Bezburuah R, and Ding J, (1992) Hydrodynamics of circulating fluidized beds: kinetic theory approach. Proceedings of the 7th International Conference on Fluidization, Gold Coast (Australia) (3–8 May 1992).
24. Schaefer, D. (1987). Instability in the evolution equations describing incompressible granular flow. *J Differ Equ* 66: 19–50.
25. Johnson, P.C. and Jackson, R. (1987). Frictional-collisional constitutive relations for granular materials, with application to plane shearing. *J Fluid Mech* 176: 67–93. <https://doi.org/10.1017/S0022112087000570>.
26. Syamlal, M., O'Brien, T.J. (1987). The derivation of a drag coefficient formula from velocity-voidage correlations. *Technical Note*, US Department of energy, Office of Fossil Energy, NETL, Morgantown, WV. https://www.researchgate.net/profile/Madhava_Syamlal/publication/242419434_The_Derivation_of_a_Drag_Coefficient_Formula_from_Velocity-Voidage_Correlations/links/00b7d52fa545283c82000000/The-Derivation-of-a-Drag-Coefficient-Formula-from-Velocity-Voidage-Correlations.pdf (accessed 9 November 2018).
27. Huilin, L. and Gidaspow, D. (2003). Hydrodynamics of binary fluidization in a riser: CFD simulation using two granular temperatures. *Chem Eng Sci* 58: 3777–3792. [https://doi.org/10.1016/S0009-2509\(03\)00238-0](https://doi.org/10.1016/S0009-2509(03)00238-0).
28. Gibilaro, L.G., Di Felice, R., Waldram, S.P., and Foscolo, P.U. (1985). Generalized friction factor and drag coefficient correlations for fluid-particle interactions. *Chem Eng Sci* 40: 1817–1823. [https://doi.org/10.1016/0009-2509\(85\)80116-0](https://doi.org/10.1016/0009-2509(85)80116-0).
29. Du, W., Bao, X., Xu, J., and Wei, W. (2006b). Computational fluid dynamics (CFD) modeling of spouted bed: assessment of drag coefficient correlations. *Chem Eng Sci* 61: 1401–1420. <https://doi.org/10.1016/j.ces.2005.08.013>.
30. Richardson, J.F. and Zaki, W.N. (1997). Sedimentation and fluidisation: Part I. *Chem Eng Res Des* 75: S82–S100. [https://doi.org/10.1016/S0263-8762\(97\)80006-8](https://doi.org/10.1016/S0263-8762(97)80006-8).
31. Li, L., Rasmuson, A., Ingram, A. et al. (2015a). PEPT study of particle cycle and residence time distributions in a Wurster fluid bed. *AIChE J* 61: 756–768. <https://doi.org/10.1002/aic.14692>.
32. Li, L., Rummelgas, J., van Wachem, B.G.M. et al. (2015b). Residence time distributions of different size particles in the

- spray zone of a Wurster fluid bed studied using DEM-CFD. *Powder Technol* 280: 124–134. <https://doi.org/10.1016/j.powtec.2015.04.031>.
33. Sarkar, A., Böhlring, P., Jajcevic, D., et al. (2016). Modeling an industrial-scale fluidized bed wurster coating process using CFD-DEM. AICHE Annual Meeting, San Francisco, CA (17 November 2016).
 34. Mann, U. (1983). Analysis of spouted-bed coating and granulation. 1. Batch operation. *Industrial & Engineering Chemistry Process Design and Development* 22: 288–292. <https://doi.org/10.1021/i200021a019>.
 35. Kalbag, A., Wassgren, C., Sumana Penumetcha, S., and Pérez-Ramos, J.D. (2008). Inter-tablet coating variability: residence times in a horizontal pan coater. *Chem Eng Sci* 63: 2881–2894. <https://doi.org/10.1016/j.ces.2008.03.009>.
 36. Kalbag, A. and Wassgren, C. (2009). Inter-tablet coating variability: tablet residence time variability. *Chem Eng Sci* 64: 2705–2717. <https://doi.org/10.1016/j.ces.2009.02.037>.
 37. Kumar, R., Freireich, B., and Wassgren, C. (2015). DEM–compartment–population balance model for particle coating in a horizontal rotating drum. *Chem Eng Sci* 125: 144–157. <https://doi.org/10.1016/j.ces.2014.06.046>.
 38. Food and Drug Administration (2014) Guidance for industry: Q4B evaluation and recommendation of pharmacopoeial texts for use in the ICH regions; Annex 6: uniformity of dosage units general chapter. <https://www.fda.gov/downloads/drugs/guidances/ucm085364.pdf> (accessed 9 November 2018).
 39. Markov, A.A. (1906). Extension of the law of large numbers to dependent events. *Bull Soc Phys Math* 2: 135–156.
 40. Berthiaux, H. and Mizonov, V. (2008). Applications of Markov Chains in particulate process engineering: a review. *Can J Chem Eng* 82: 1143–1168. <https://doi.org/10.1002/cjce.5450820602>.
 41. Antia, F.D. and Lee, S. (1985). The effect of stoichiometry on Markov chain models for chemical reaction kinetics. *Chem Eng Sci* 40: 1969–1971. [https://doi.org/10.1016/0009-2509\(85\)80135-4](https://doi.org/10.1016/0009-2509(85)80135-4).
 42. Irizarry, R. (2012). Fast compartmental Monte Carlo simulation of population balance models: application to nanoparticle formation in nonhomogeneous conditions. *Ind Eng Chem Res* 51: 15484–15496. <https://doi.org/10.1021/ie3011116>.
 43. Mizonov, V., Berthiaux, H., Arlabosse, P., and Djerroud, D. (2008). Application of the theory of Markov chains to model heat and mass transfer between stochastically moving particulate and gas flows. *Granul Matter* 10: 335–340. <https://doi.org/10.1007/s10035-008-0094-2>.
 44. Pippel, W. and Philipp, G. (1977). Utilization of markov chains for simulation of dynamics of chemical systems. *Chem Eng Sci* 32: 543–549. [https://doi.org/10.1016/0009-2509\(77\)87012-7](https://doi.org/10.1016/0009-2509(77)87012-7).
 45. Chen, S.J., Fan, L.T., and Watson, C.A. (1972). The mixing of solid particles in a motionless mixer: a stochastic approach. *AIChE J* 18: 984–989. <https://doi.org/10.1002/aic.690180515>.
 46. Gibilaro, L.G., Kropholler, H.W., and Spikins, D.J. (1967). Solution of a mixing model due to van de Vusse by a simple probability method. *Chem Eng Sci* 22: 517–523. [https://doi.org/10.1016/0009-2509\(67\)80034-4](https://doi.org/10.1016/0009-2509(67)80034-4).
 47. Inoue, I. and Yamaguchi, K. (1970). Particle motion in a mixer-mixing in a 2 dimensional V-type mixer. *Int Chem Eng* 10: 490.
 48. Rubinovitch, M. and Mann, U. (1983). Single-particle approach for analyzing flow systems. Part I: Visits to flow regions. *AIChE J* 29: 658–662. <https://doi.org/10.1002/aic.690290421>.
 49. Kamrunnihar, M., Huang, B., and Fisher, D.G. (2000). Estimation of Markov parameters and time-delay/interactor matrix. *Chem Eng Sci* 55: 3353–3363. [https://doi.org/10.1016/S0009-2509\(00\)00008-7](https://doi.org/10.1016/S0009-2509(00)00008-7).
 50. Yin, K.K., Liu, H., and NE, J. (2002). Markovian inventory policy with application to the paper industry. *Comput Chem Eng* 26: 1399–1413. [https://doi.org/10.1016/S0098-1354\(02\)00113-8](https://doi.org/10.1016/S0098-1354(02)00113-8).
 51. Catak, M., Cronin, K., and Medina-Tellez, D. (2011). Markov chain modeling of fluidized bed granulation incorporating simultaneous aggregation and breakage. *Ind Eng Chem Res* 50: 10811–10823. <https://doi.org/10.1021/ie102513v>.
 52. Dehling, H.G., Hoffmann, A.C., and Stuu, H.W. (1999). Stochastic models for transport in a fluidized bed. *SIAM J Appl Math* 60: 337–358. <https://doi.org/10.1137/S0036139996306316>.
 53. Mizonov, V., Mitrofanov, A., Ogurtzov, A., and Tannous, K. (2014). Modeling of particle concentration distribution in a fluidized bed by means of the theory of Markov chains. *Part Sci Technol* 32: 171–178. <https://doi.org/10.1080/02726351.2013.839016>.
 54. Rozanov, Y.A. (1982). *Markov Random Fields*. New York: Springer.
 55. Markl, D., Hanneschläger, G., Buchsbaum, A. et al. (2014). In-line quality control of moving objects by means of spectral-domain OCT. *Opt Lasers Eng* 59: 1–10. <https://doi.org/10.1016/j.optlaseng.2014.02.008>.
 56. Ramkrishna, D. (2000). *Population Balances: Theory and Applications to Particulate Systems in Engineering*. San Diego, CA: Academic Press.
 57. Li, J., Freireich, B., Wassgren, C., and Litster, J.D. (2012). A general compartment-based population balance model for particle coating and layered granulation. *AIChE J* 58: 1397–1408. <https://doi.org/10.1002/aic.12678>.
 58. Peglow, M., Kumar, J., Heinrich, S. et al. (2007). A generic population balance model for simultaneous agglomeration and drying in fluidized beds. *Chem Eng Sci* 62: 513–532. <https://doi.org/10.1016/j.ces.2006.09.042>.
 59. Ronsse, F., Pieters, J.G., and Dewettinck, K. (2007). Combined population balance and thermodynamic modelling of the batch top-spray fluidised bed coating process. Part I—Model development and validation. *J Food Eng* 78: 296–307. <https://doi.org/10.1016/j.jfoodeng.2005.09.030>.
 60. Saleh, K., Steinmetz, D., and Hemati, M. (2003). Experimental study and modeling of fluidized bed coating and agglomeration. *Powder Technol* 130: 116–123. [https://doi.org/10.1016/S0032-5910\(02\)00254-1](https://doi.org/10.1016/S0032-5910(02)00254-1).

18

PROCESS DESIGN OF TOPICAL SEMISOLIDS: APPLICATION OF FUNDAMENTAL CONCEPTS IN PHARMACEUTICAL ENGINEERING TO PEG OINTMENT DEVELOPMENT

AMANDA SAMUEL, THEAN YEOH, AND ROLF LARSEN
Formulation Process and Design, Pfizer Inc., Groton, CT, USA

AVIK SARKAR
Process Modeling & Engineering Technology, Pfizer Inc., Groton, CT, USA

18.1 INTRODUCTION

The topical administration of drugs is an important delivery route for achieving local effects in the skin, which is the tissue of interest for the treatment of dermatologic conditions. When developing topically administered drug products, the formulation, also referred to as the vehicle, is a key consideration because it can significantly influence delivery of the active pharmaceutical ingredient (API) into the skin, product efficacy, local tolerability of the product, and patient compliance.

The selection of a particular topical formulation type and the design of the formulation composition can be driven by many factors, including the physicochemical properties of the API, desired aesthetic properties, target indications, application sites, patient population, and competitive landscape. A wide variety of formulation types have been developed for topical administration to address these varied needs, including creams, lotions, gels, solutions, suspensions, ointments, and foams. For an overview of compositions, appearance, feel, and physical properties of these topical formulation types, the interested reader is referred to guidance published by the FDA on the classification of topical drugs [1].

Much has been written on the subject of designing formulations for topical drug delivery [2]. This chapter instead focuses on the unique considerations and challenges associated with the clinical and commercial-scale manufacture of topical semisolids. In particular, we will focus on one subset of topical ointments, polyethylene glycol (PEG)-based ointments.

18.2 PEG-BASED OINTMENTS

PEG (also referred to as macrogol)-based ointments typically consist of a solid PEG, such as PEG 3350 or 4000, combined with a lower molecular weight (MW) liquid PEG, such as PEG 200, 300, or 400. For example, PEG Ointment NF is mixture of PEG 400 and PEG 3350 in a 3 : 2 ratio [3].

PEG ointments can offer several advantages over other classes of topical formulations. Because PEGs are water miscible, PEG ointments are easily removed from the surface of the skin through washing with water, unlike hydrocarbon-based petrolatum ointments. Liquid PEGs are often good solubilizers of compounds that have poor solubility in water and/or oil, and as such, PEG ointments can provide an avenue for formulating hard-to-solubilize APIs. Additionally, PEG

ointments can accommodate a variety of polar cosolvents, penetration enhancers, and humectants including water, ethanol, propylene glycol, diethylene glycol monoethyl ether, and glycerin, which can be included in the formulation to tune the delivery, solubility, chemical and physical stability, viscosity, and organoleptic properties of the ointment.

PEG ointments can also have potential drawbacks for certain classes of APIs, however, and their use as vehicles for topical delivery should be predicated on suitable compatibility with the API of interest based on its physicochemical properties. While highly solubilizing, PEGs are not readily absorbed into the skin [4], unlike small molecule excipients such as propylene glycol and diethylene glycol monoethyl ether, which can act as both solubilizers and penetration enhancers [5–7]. PEG-based topical formulations have been associated with poor percutaneous delivery of drug substances including acyclovir and corticosteroids presumably due to reduced thermodynamic activity of the APIs as a result of persistence of unabsorbed PEG on the surface of the skin [8]. Another consideration when developing a PEG ointment is potential chemical reactivity of the API with the PEG itself and with impurities and degradation products, which can include organic peroxides, formaldehyde, and formic acid [9]. This chemical degradation, which can impact the shelf life of the drug product, can potentially be mitigated through the addition of antioxidants, including butylated hydroxytoluene (BHT) and butylated hydroxyanisole (BHA), ascorbic acid, propyl gallate, and tocopherol.

In this chapter we will focus on a PEG ointment composed primarily of an API in a mixture of PEG 3350, PEG 400, propylene glycol, and glycerin (Table 18.1).

18.3 MANUFACTURING PEG-BASED OINTMENTS

The previous sections have described the considerations involved in formulating a PEG ointment – in particular the advantages and disadvantages from skin delivery, solubility, and stability perspective. Once the composition of the formulation is finalized, we now need to devise a process to manufacture the ointment. This section of the chapter will discuss an example of how the manufacture of an ointment goes from

a benchtop lab scale (~100 g) to the commercial scale (~1000 kg).

The information provided so far in this chapter is for a representative PEG-based ointment, and we will continue to progress the same ointment to the manufacturing phase. A key feature of this ointment is that the API, which is a solid, is fully dissolved in the vehicle. We shall see that this will be important to the present choice of the manufacturing process and equipment. For the other types of formulations that we briefly touched upon in Section 18.1 – creams, lotions, suspensions, etc. – the manufacturing process may be slightly or greatly different.

It is important to understand the quality target product profile (QTPP) while developing the manufacturing process steps because the QTPP will guide the risk assessment and experimentation procedure used to define the processing ranges and parameter interactions. This will drive development of the formulation and process design space so that consistent quality product can be commercially manufactured. Additionally, the manufacturing processing constraints need to be understood so that the process developed for the clinical scale can be managed at the commercial scale. Some of these constraints include pumping and recirculation capabilities, heating and cooling capabilities, mixing and processing time, and equipment capabilities.

Additionally, it is important to evaluate API–excipient compatibility to determine critical exposure effects due to chemical reactions, degradation, and the impacts of possessing parameters. Based on this understanding, the order of addition can be developed in planning the process steps. In addition, understanding the API and excipient compatibility as well as the physical and chemical stability will guide the process development to understand the initial design space.

18.3.1 Manufacturing Scales

To conduct early research on the ointment stability and performance, small batches of the ointment, typically a few grams to a few hundred grams, are manufactured on a lab scale using benchtop equipment (including beakers, small overhead mixers, and handheld homogenizers). The process used at the lab scale may not be representative of the final manufacturing process.

It is very rare, and a large risk, to jump directly from making a few hundred grams at the lab scale to manufacturing hundreds or thousands of kilograms at the commercial scale. Moreover, smaller volumes are generally needed during the early stages of a clinical development program for toxicology and clinical studies before progressing to full-fledged commercial supply. The manufacturing process is almost always scaled up through several intermediate manufacturing scales (Figure 18.1). This helps build experience in the manufacturing process, helps identify and solve any issues with manufacturing, and therefore reduces the risk of having large failed

TABLE 18.1 Composition of a Generic PEG Ointment

Component	Function	Weight % Range
Active ingredient		Up to 2%
PEG 400	Cosolvent	30%
Glycerin	Cosolvent	15–20%
Propylene glycol	Cosolvent	15–20%
PEG 3350	Rheology modifier	30%

batches. During this scale-up process, it is recommended that the process be reviewed for scale-dependent process parameters and scale-independent process parameters. This will allow for smaller-scale data to be used for scale-independent parameters.

For our process, we staged the manufacturing scale-up by starting at the pilot scale (tens of kilograms), progressing to

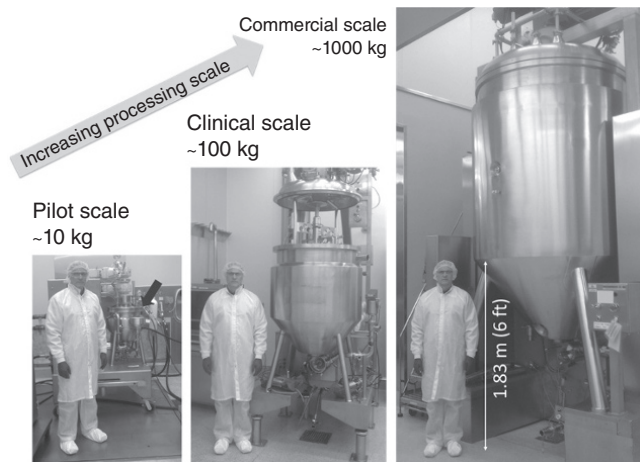


FIGURE 18.1 Scales of the manufacturing equipment for ointments, pictured with one of the coauthors to emphasize the relative sizes of the various units. The pilot-scale, clinical-scale, and commercial-scale units can make up to tens, hundreds, and thousands of kilograms of ointment, respectively. More details on these mixers will be discussed in the following section.

the clinical scale (hundreds of kilograms), and finally ending at the commercial scale (thousands of kilograms).

18.3.2 Process Equipment

The manufacturing of the ointment requires handling and manipulation of several phases: solids (drug powder, unmolten vehicle ingredients), viscous liquids (molten vehicle ingredients), and gas (air bubbles that get inadvertently trapped). Hence, the manufacturing equipment should be able to handle liquid to semisolid flows. A number of commercial vendors manufacture mixers capable of handling multiphase, viscous, and semisolid flows – such mixers are used not only to manufacture pharmaceutical products, but also for producing consumer goods and cosmetics (e.g. toothpaste, creams, gels), foods (pastes and emulsions, e.g. mayonnaise), and chemicals (e.g. paints).

Figure 18.2 shows mixing tanks at two different scales, as examples, made by FrymaKoruma. Such mixing vessels generally have the following features:

- Main mixing tank: Dictates the size of the ointment batch. Typical scales range from 1 kg up to approximately 3000 kg.
- Mixing elements – agitator and scraper: Typically counter-rotating mixing impellers to ensure uniform mixing throughout the mixing vessel. The scraper cleans material accumulated at the walls and helps improve heat transfer from the heated water jacket.

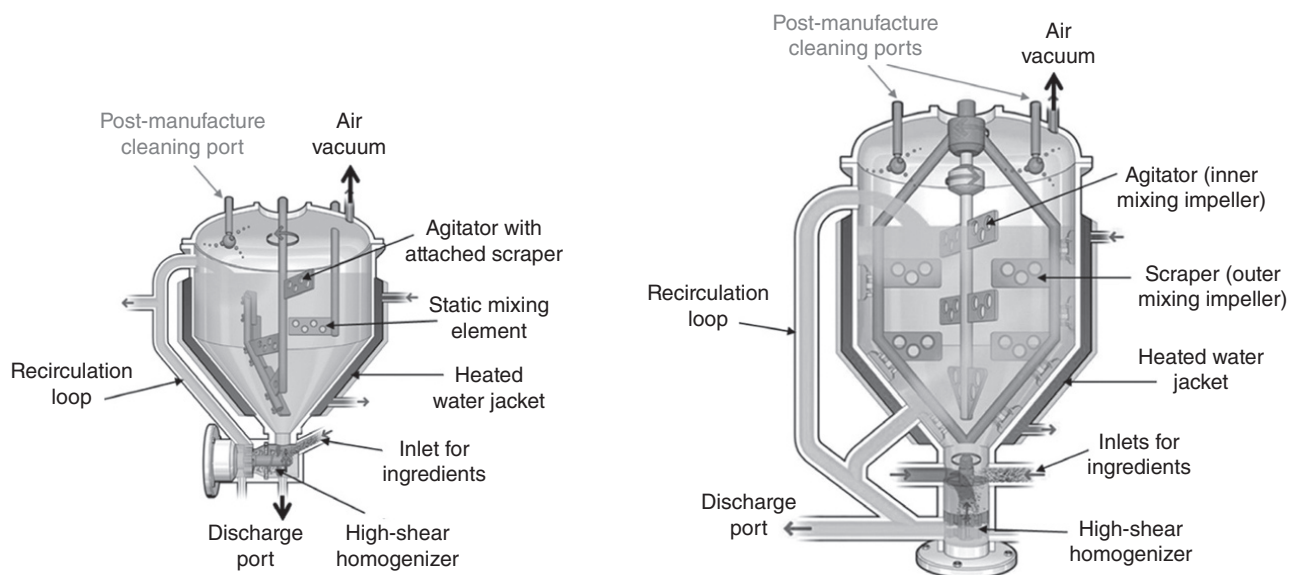


FIGURE 18.2 Manufacturing mixing vessels made by FrymaKoruma, showing a smaller-scale unit (left) and larger-scale unit (right), not drawn to scale. The smaller mixer on the left is used for smaller batch sizes, and the larger mixer on the right is more representative of the clinical- and commercial-scale units. Mixing vessels made by other equipment companies will differ in design, but the main operating principles remain similar. *Source:* Images adapted from FrymaKoruma (www.frymakoruma.com).

The mixing speeds, i.e. the rotation rates (RPM) of the agitator and scraper, are scale dependent. Typically, the agitator and scraper speeds that can be achieved gradually decrease with increasing scale.

- **Heated water jacket:** Provides heating/cooling by circulating hot/cold water along the outer walls of the tank. The temperature of the cooling water is generally either room temperature or chilled water whose temperature is specified by the manufacturing facility. Heating steam or superheated steam is used, as available in the facility (>100 °C). The heating or cooling rates are also scale dependent; larger batches take longer to heat or cool.
- **Inlet and discharge ports:** Used for adding ingredients during operation and emptying the tank at the end. The starting ingredients can be added through the open top lid before starting the mixing elements.
- **High-shear homogenizer:** A high-speed mixing element that generates a large amount of shear. Used for producing fine droplets for emulsions and breaking down powder agglomerates/aggregates. The operating speed (RPM) of the homogenizer should be governed by the shear strain rate that is required to emulsify the semisolid product. In the present case, our ointment is a solution, and, therefore, the high-shear homogenizer was not required; the high-shear elements (rotor and stator) were thus removed.
- **Recirculation loop:** Helps improve the product uniformity and distribution of heat through better mixing.
- **Air vacuum:** Reduces the amount of air bubbles that get entrapped in the ointment during manufacturing. Excessive air bubbles are undesirable, as we are seeking to make a continuous semisolid phase (and not an ointment that resembles whipped cream!).
- **Post-manufacture cleaning ports:** Most equipment will have provision to clean the unit using high-pressure water.

It is worthwhile to note that although the basic operating principles are similar across scales, there are notable differences in the size and geometry of the units. To be able to successfully and reliably scale up the manufacturing, a good understanding of the process and underlying physics is essential. The goal of scale-up is to define the new process parameters at the next scale that produce comparable product quality as the previous scale.

18.3.3 Process Flow

The process flow and order of excipient and API addition will be dependent on the API and formulation sensitivities, final product requirements, and the processing equipment that will be used. An understanding of the chemical stability, solubility, and miscibility of the API and the various excipients

across the required temperature range and mixing/hold times is needed to ensure that the drug product meets the target profile with regard to attributes such as potency, purity, content uniformity, microstructure, and rheological properties.

For example, a consideration when manufacturing a PEG ointment is that PEGs are susceptible to minor chemical degradation at temperatures as low as approximately 40 °C [10]. This may be of concern if the physicochemical properties of the API are such that it can chemically react with the resulting PEG degradation products, in turn leading to degradation of the API. In such cases, the manufacture process should be designed with this risk in mind. It may be advisable, for example, to limit the exposure of the drug product to high temperatures for extended periods of time during the manufacture process.

The process flow should also consider the equipment that will be used, which may depend on the equipment available at the development and production sites. As a manufacturing process is scaled up, the equipment will need to be evaluated to ensure consistent process conditions are maintained. Manufacture of the PEG ointment described in this chapter requires a vessel capable of mixing the material and maintaining a consistent temperature as the product cools and congeals. Additionally, the congealing process will increase loads to the mixing and scraping components as well and the ability to pump the material, and the mixing vessel must be capable of managing a range of product viscosities.

A general process flow diagram for the PEG ointment is shown in Figure 18.3. We will discuss the dissolution and congealing/annealing steps in greater detail in the next sections of this chapter.

18.3.4 Critical Process Parameters

There are two critical processing steps to manufacture this PEG ointment: dissolution of the API and congealing/annealing of the final product. For this formulation, the API is designed to be fully dissolved in the vehicle. Undissolved particles not only can be felt by the patient when the ointment is spread over the surface of the skin, but the presence of undissolved drug particles also can contribute to a lack of content uniformity of the bulk. As such, ensuring API dissolution in the bulk during processing is critical. There may be multiple ways to achieve the dissolution, including using a single vessel while mixing, including a dispersion/homogenizing unit, or coupling with a separate vessel with high-shear homogenization/emulsifier recirculation.

Congeaing/annealing of the final ointment is another critical step for this PEG ointment as it allows the liquid bulk to transform into a semisolid ointment with the desired microstructure and rheological properties. A unique characteristic of high MW PEG is the existence of multiple solid-state forms of PEG at ambient temperature [11]. As will be described in greater detail in Section 18.3.4.3, the relative

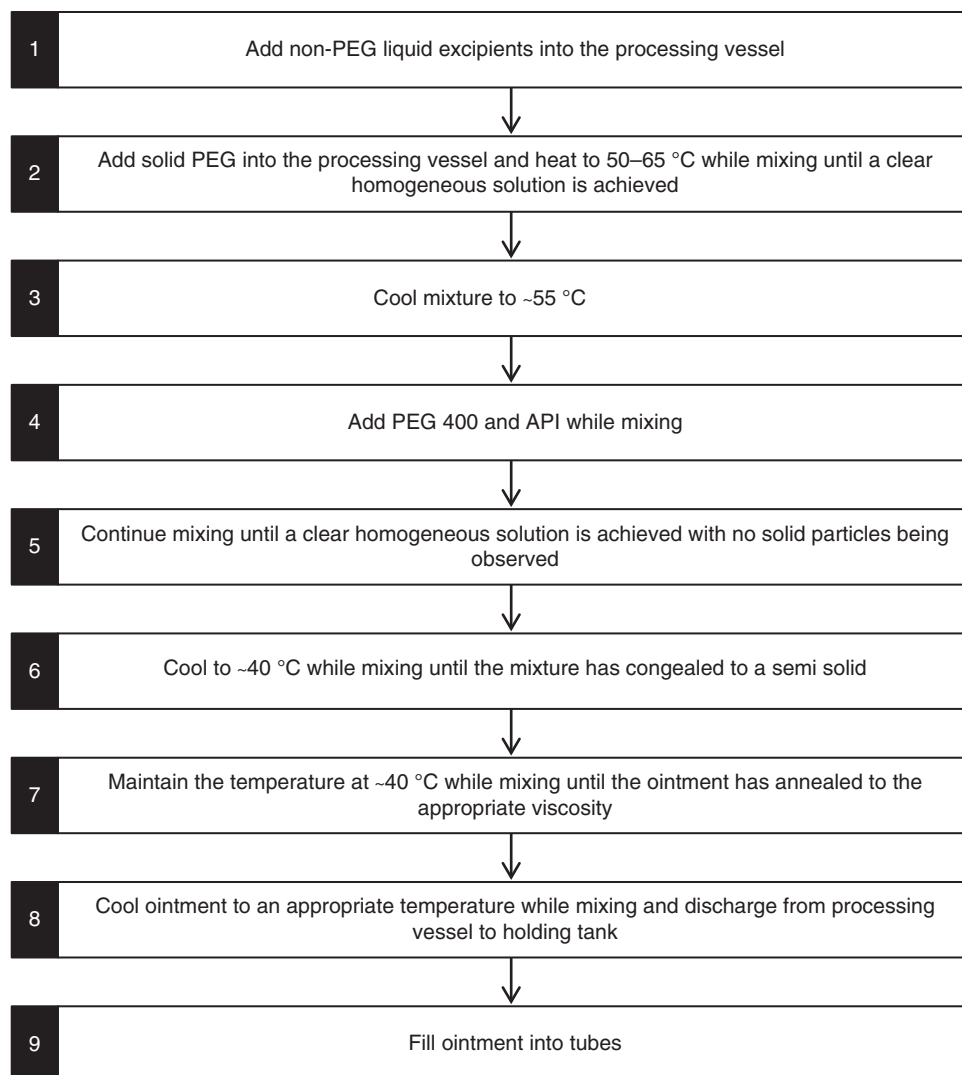


FIGURE 18.3 Typical process flow for PEG ointment manufacture.

fraction of the two solid-state forms is highly dependent on the processing conditions during congealing. As high MW PEG primarily functions as a viscosity agent/rheology modifier in the PEG ointment, the dominant solid-state form present in the ointment will have the most influence on the overall rheological characteristics of the ointment. Measured product viscosity, a rheological property, is typically a critical quality attribute for semisolid dosage forms. As such, proper specification of the congealing/annealing process parameters will be necessary to ensure manufacture of a PEG ointment that meets the desired viscosity specifications.

18.3.4.1 API Dissolution: Process Considerations

Dissolution of the API in the vehicle is a critical step in the manufacturing process of this PEG ointment. For many APIs the dissolution step is rapid and uneventful. However, some APIs take a long time to dissolve into the vehicle, and it

is important to design the process so that API particles are fully dissolved, and complete dissolution can be independently confirmed. As such, it is prudent to develop a rational order of addition of ingredients by taking into consideration API solubility in the liquid bulk and other contributing factors including the API and excipient physical and chemical attributes, temperature, and mixing efficiencies.

API particle size, distribution, cohesiveness, and wettability are important to understand as these attributes will affect the dissolution process. Smaller particle sizes will dissolve faster in the solvent vehicle, while larger particles may need to have longer contact times to dissolve. Highly cohesive API may tend to form large agglomerates or powder lumps while in solid bulk. If these lumps are not rapidly dispersed into the liquid phase through adequate mixing and good wetting with the solvent, dissolution time may increase. Micronizing the API may or may not help. While micronized particles may

have greater exposed surface, highly cohesive particles can also form agglomerates in the liquid phase, which result in an overall decrease in contact with the liquid bulk.

Analytical methods for the determining dissolution end point can be as simple as visual observation that a clear homogeneous solution is obtained. While this can be achieved by making visual observations of the bulk through taking samples, the removed samples need to be kept at temperature above the ointment congealing temperature in order to avoid rapid congealing. Alternatively, direct in-line processing monitoring techniques such as Raman spectroscopy and ATR-UV can be used to monitor dissolved API concentration profile [12].

Temperature experimentation at small scale is usually performed to develop the optimized temperature range for the dissolution of the API. Experiments should be performed to evaluate the degradation of the API that may determine the operating range for the process. Temperature variability and mixing efficiencies would need to be evaluated to determine the effects of scale.

Prior to adding the API to the solvent vehicle, it is important that this solvent vehicle is completely homogeneous and in a liquid state. Once this process step has been completed, the API can be added by using the vacuum capabilities of the equipment; as in the example using the FrymaKoruma, API is drawn in, while the dispersion homogenizer (DisHo) is in operation. This method uses the DisHo to break up agglomerations and disperses the API into the vehicle. The ideal process would be able to ensure that the particle is reduced to the primary particles as this will be the most efficient for the dissolution process. In some cases it may be advantageous to perform this more aggressively with an offline homogenizer fitted with appropriate rotor–stator gap width (micron sizing) that would break down particle agglomerates to their primary state. Within any unit operation this may take multiple passes depending on the efficiency of the unit being used.

18.3.4.2 API Dissolution: Computational Modeling to Aid Scale-Up In practice, full-scale experiments are not ideally suited to build fundamental process understanding of a process. In a full-scale process, a number of different, oftentimes competing, phenomena occur simultaneously; it is difficult to distinguish between or quantify/measure these different competing physical events. Moreover, our ability to instrument or see inside full-scale equipment is limited. Only small samples can be collected, or readings at a few points can be recorded, such as near the walls through viewing windows or sampling ports or at a few locations where probes can be inserted. In such situations, relying solely on experimental studies to design a process is not ideal, and it is recommended that scientists and engineers should use engineering and computational models to aid with process understanding.

We are interested in the dissolution step, which we have identified as one of our critical operations in the process.

Therefore, we will start with understanding dissolution of a single particle before proceeding to a collection of particles, i.e. a powder.

18.3.4.2.1 Single-Particle Dissolution A number of relationships exist that describe the dissolution of a particle of mass m , typically assumed to be spherical (the diameter will be denoted by s in the present chapter). For example, pharmaceutical scientists typically use the Noyes and Whitney [13] equation, where the rate of mass loss from a particle is given by

$$\frac{dm}{dt} = -\frac{\mathcal{D}}{l}A(c_{\text{sat}} - c_{\text{bulk}}(t)) \quad (18.1)$$

where

c_{bulk} is the concentration of the dissolved drug in the bulk of the solvent.

c_{sat} is the saturation solubility of the drug.

$A (= \pi s^2$, for spheres) is the surface area of the particle. (s is the particle diameter)

\mathcal{D} is the effective diffusion coefficient.

l is a length scale associated with the diffusion-boundary layer thickness around the particle that is on the order of the particle diameter.

It is well known in engineering literature that the mechanical stirring helps improve the mixing rates; however, there is no explicit dependence on the stirring rates in the Noyes and Whitney equation. Note that we describe \mathcal{D} as the *effective* diffusion rate, which implicitly lumps together all effects such as stirring and temperature. For our application, we want to decouple the pure diffusion effects from the flow (i.e. stirring) effects, as we wish to specifically examine the effects of stirring rates and manufacturing scales of the process.

The dissolution rate of a single particle is composed of two components (Figure 18.4). First is a pure diffusion contribution that represents, say, a single stationary particle in a large volume of stationary fluid. Second is an advective (sometimes referred to as convective) component that accounts for relative fluid flow over the particle surface; as the relative flow velocity increases, the dissolution rate increases. The Sherwood number is a dimensionless quantity that is used to describe the (nondimensional) rate of mass transport. Typically, the Sherwood number, Sh , is expressed as a function of the dimensionless Reynolds number ($Re = (\rho_f v s) / \mu_f$) and the Schmidt number ($Sc = \mu_f / (\rho_f \mathcal{D})$). For example, for dissolution of a spherical particle in a Newtonian fluid, the correlation for Sh is given as

$$Sh = \frac{h}{\mathcal{D}/s} = 2 + 0.552 Re^{1/2} Sc^{1/3} \quad (18.2)$$

originally by Frossling [14] but often referred to as the Ranz and Marshall [15] correlation. In this correlation, the

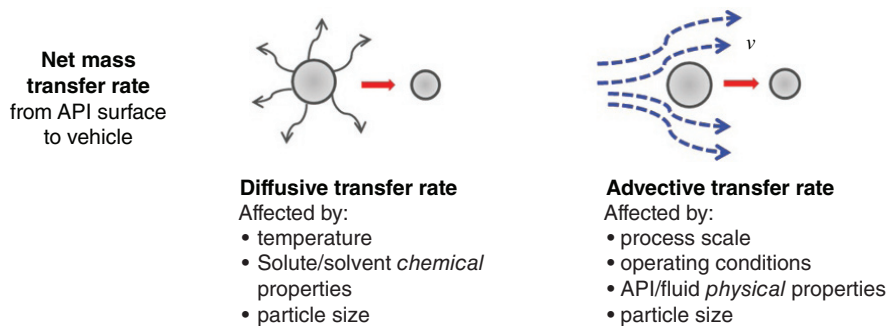


FIGURE 18.4 Schematic representation of the dissolution mechanisms for a single particle, showing mass transfer via diffusion and advection (or convection), along with the factors that affect both modes.

effect of the fluid-flow velocity over the particle surface, v , is explicitly accounted for via the Reynolds number. Here, h is the mass transfer coefficient from the bead to the fluid (SI units: m/s), which would replace the \mathcal{D}/l term in Eq. (18.1) as

$$\frac{dm}{dt} = -hA(c_{\text{sat}} - c_{\text{bulk}}(t)) \quad (18.3)$$

Such Sherwood number correlations are generally derived from experimental or numerical data, and the correlations change depending on the system of interest; for example, the functional relationship for $\text{Sh} = f(\text{Re}, \text{Sc})$ will change for flow and dissolution around cylinders, tubes, packed beds, etc. These correlations can be found in engineering handbooks, textbooks, and references.

For our application, we are interested in the same solute–solvent system at the same temperature. Therefore, the Schmidt number Sc , diffusion coefficient \mathcal{D} , and fluid properties (density and viscosity) do not change as we scale up. In addition, the Frossling/Ranz and Marshall correlation applies to Newtonian fluids; for a non-Newtonian shear thinning fluid such as the present ointment, a more generalized expression for the dimensionless mass transfer coefficient (Sh) can be written from Eq. (18.2) as

$$\text{Sh} = A + B(\text{Re})^C (\text{Sc})^D \quad (18.4)$$

where

A , B , C , and D are empirical constants that need to be determined from experiments and fitting.

Recognizing that the solute–solvent system of interest does not change, i.e. the viscosity (μ_f), density (ρ_f), diffusion coefficient (\mathcal{D}), and, hence, also the Schmidt number (Sc) are constant, we can rearrange the above equation and express the mass transfer coefficient in a dimensional form as

$$h = \frac{A_h}{s} + \frac{B_h}{s} (vs)^{C_h} \quad (18.5)$$

where

A_h , B_h , and C_h are parameters that need to be experimentally calibrated for the current solute–solvent system of interest.

Note that the Schmidt number, diffusion coefficient, fluid density, and fluid viscosity appear implicitly in these calibrated parameters. The Frossling/Ranz- and Marshall-type dissolution model, thus generalized, can help calculate the dissolution times for single particles or even powders composed of mono-disperse particles (albeit assumed to be spheres).

EXAMPLE PROBLEM

What is the predicted dissolution time for a single particle of size $150 \mu\text{m}$ with true density 1200 kg/m^3 in water (Newtonian fluid)? Start by assuming that there is no stirring and that there is an infinite expanse of fluid. Take saturation solubility as 120 kg/m^3 ($\sim 12 \text{ g/100 ml}$) in water and diffusion coefficient as $5 \times 10^{-10} \text{ m}^2/\text{s}$. What happens if you increase or decrease the particle size? Explore the diameter range of $100\text{--}300 \mu\text{m}$.

Solution

We will solve this problem in terms of the particle diameter, s , and track the evolution of the diameter starting from $s_0 = 150 \mu\text{m}$ until the diameter reduces to zero. The governing equation for dissolution is given by Eq. (18.3):

$$\frac{dm}{dt} = -hA(c_{\text{sat}} - c_{\text{bulk}}(t))$$

where

$$m = \rho_s \frac{1}{6} \pi s^3.$$

$$\rho_s = 1200 \text{ kg/m}^3.$$

$$A = \pi s^2.$$

$$c_{\text{sat}} = 120 \text{ kg/m}^3.$$

The problem assumes an infinite expanse of water (compared with the size of a single particle). Therefore, we can neglect the bulk concentration of the drug and simplify the mathematics as $c_{\text{bulk}}(t) \approx 0$.

The mass transfer coefficient can be calculated from the Frossling/Ranz and Marshall correlation (Eq. (18.2)):

$$\text{Sh} = \frac{h}{D/s} = 2 + 0.552\text{Re}^{1/2}\text{Sc}^{1/3}$$

Without stirring, $\text{Re} = 0$; hence,

$$h = 2 \frac{D}{s},$$

where

$$D = 5 \times 10^{-10} \text{ m}^2/\text{s}$$

Substituting in the dissolution equation

$$\frac{d}{dt} \left(\rho_s \frac{1}{6} \pi s^3 \right) = 2 \frac{D}{s} \cdot \pi s^2 \cdot \left(\underbrace{c_{\text{bulk}}(t)}_{\approx 0} - c_{\text{sat}} \right)$$

$$\text{or, } \frac{\rho_s ds}{2 dt} = -2 \frac{D}{s} \cdot c_{\text{sat}}$$

$$\text{or, } s \cdot ds = - \frac{4Dc_{\text{sat}}}{\rho_s} dt$$

Integrating both sides, with starting and ending limits,

$$\int_{s_0}^{s(t)} s \cdot ds = \int_0^t - \frac{4Dc_{\text{sat}}}{\rho_s} dt$$

$$\text{which yields } \frac{s(t)^2}{2} - \frac{s_0^2}{2} = - \frac{4Dc_{\text{sat}}}{\rho_s} t$$

$$\text{or, } s(t) = \sqrt{s_0^2 - \frac{8Dc_{\text{sat}}}{\rho_s} t}$$

The time taken for a 150 μm particle to dissolve fully, i.e. to reach a final size $s = 0$, is calculated as

$$s_{\text{final}} = 0 = \sqrt{s_0^2 - \frac{8Dc_{\text{sat}}}{\rho_s} t_{\text{final}}}$$

Substituting the values,

$$t_{\text{final}}|_{s_0 = 150 \mu\text{m}} = 56.25 \text{ s.}$$

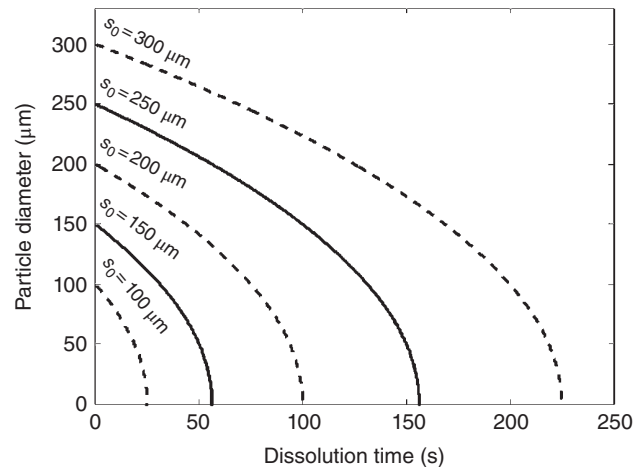


FIGURE 18.5 Evolution of the particle diameter with dissolution time for particles with varying initial size (see the example problem for full list of assumptions and material properties).

To examine the effect of initial particle size, simply vary the value of the initial particle size s_0 in the above equation. The effect of varying s_0 between 100 and 300 μm is plotted in Figure 18.5.

18.3.4.2.2 Population Balance Modeling (PBM) for Dissolution of Polydisperse Powders In this section, we will examine the main features of a population balance model, the model formulation, underlying assumptions, and how it was applied to solve the problem of powder dissolution. We will not delve very deeply into the underlying mathematics or derivation of the population balance equations; readers can find excellent textbooks [16] and a number of papers (to name a few examples, LeBlanc and Fogler [17], Bhaskarwar [18], and McCoy [19]) dedicated to this topic. In this section, we take the liberty of presenting equations without providing the full derivation and ask for the readers' indulgence in accepting the math as described. (Students are welcome to derive these equations on their own as a homework exercise.)

In the preceding section, we have described the dissolution of a single sphere. This single-particle model is applicable even if there are multiple particles in the system, but all of them are uniformly sized and are subjected to the same flow conditions, i.e. all particles have the same Sherwood number (Sh) or mass transfer rate (h). In reality, the ingoing solute will have a size distribution. Moreover, there will be differences in flow within the device, and there will be differences in the relative fluid-particle velocity, due to both differences in flow within the mixing tank and differences in particle size (finer particles will follow fluid streamlines more closely compared with more massive particles). Figure 18.6 (top part) shows a representative progression of the dissolution process for a typical powder, showing the particle size distribution as a function of time. As time progresses, the particle

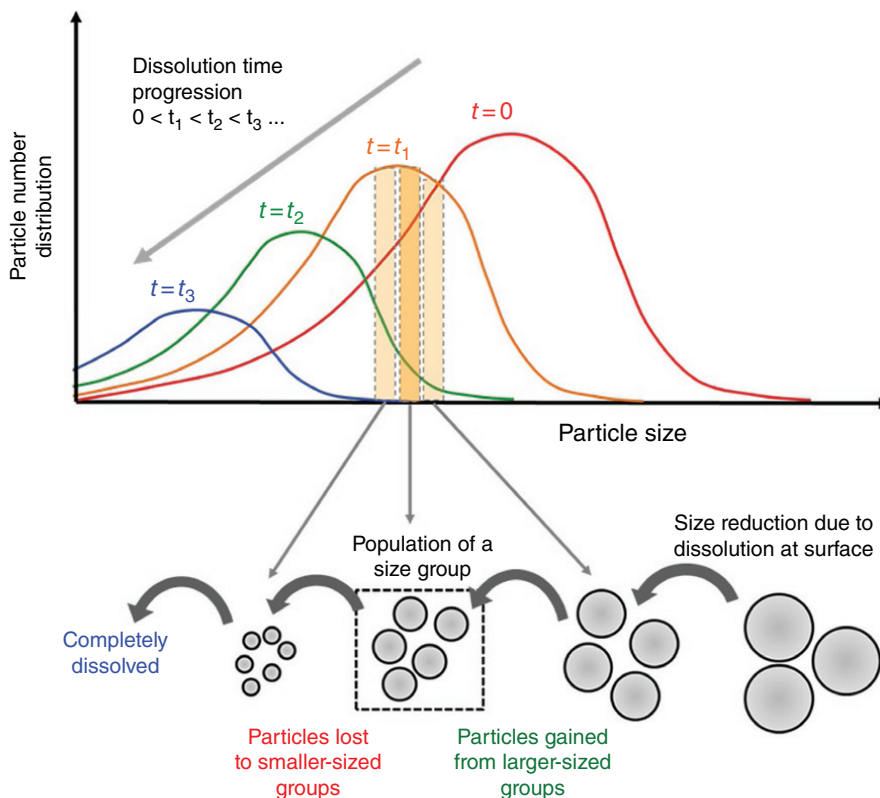


FIGURE 18.6 Schematic representation of the typical evolution of the particle number distribution during a dissolution process. Particle sizes as well as total number of particles decrease with dissolution time. The bottom part shows the mechanism by which particle diameters evolve and how particles move from larger-sized groups to smaller-sized groups.

size distribution shifts to the left as particles dissolve and shrink. Simultaneously, the total number of particles, reflected in the height of the particle number distribution function, decreases as smaller particles dissolve completely, leaving behind fewer undissolved particles.

We see that the population distribution of particle size evolves with time. A mathematical framework to describe such continuous (or even discontinuous) evolution of populations is given by the population balance modeling (PBM) approach. The PBM method has been utilized to study a variety of engineering problems such as coating (see Chapter 16), granulation, and milling. In the PBM approach, we track the population (via the number density function, f_s) of particles in an infinitesimally small particle size class. The number density function is defined such that $f_s(s, t)ds$ is the number of particles per unit suspension mass with diameters in the size range between s and $s + ds$. Note that the number density function depends on the diameter (s) and is a function of time (t).

Based on this definition for f_s , the governing PBM equation is written as

$$\frac{\partial}{\partial t} f_s(s, t) + \frac{\partial}{\partial s} [Gf_s(s, t)] = [\text{birth rate}] - [\text{death rate}] \quad (18.6)$$

Readers may immediately recognize Eq. (18.6) to be very similar to the general conservation equations that appear in heat/mass transport but with the spatial coordinate “ x ” replaced by the size coordinate “ s .” Indeed, the PBM equations follow the same approach used to derive mass or momentum conservation equations. As such, the term $\partial/\partial s[Gf_s(s, t)]$ represents an advective transport of the population. In momentum conservation, the term analogous to G represents the velocity, $v = dx/dt$. Similarly, the transport “velocity” for the PBM is given by replacing “ x ” with “ s ,” as $G = ds/dt$.

From our single-particle dissolution law, we have

$$\frac{dm}{dt} = -hA(c_{\text{sat}} - c(t)) \quad (18.7)$$

where

h is our mass transfer coefficient (see Eqs. (18.2), (18.4), and (18.5))

$c(t)$ is the dissolved bulk concentration of the drug in the solvent (same as $c_{\text{bulk}}(t)$ in Eq. (18.3), but we have dropped the “bulk” subscript).

Substituting the particle mass as $m = \rho_s \frac{1}{6} \pi s^3$ and particle area as $A = \pi s^2$ in Eq. (18.3), and then rearranging, we get

$$G = \frac{ds}{dt} = \frac{-2h(c_{\text{sat}} - c(t))}{\rho_s}. \quad (18.8)$$

In PBM, we track the populations of particles in each size class and update the population of each class as particles leave this particular size group (“deaths”) or enter the size group from other groups (“births”). Thus, the birth and death terms in the governing PBM equation are analogous to source and sink terms, respectively, appearing in fluid-flow conservation equations. For a dissolution process, there are no such abrupt birth or death events; particle distributions evolve smoothly through gradual mass dissolution at the surface (i.e. via the advection term G). Therefore, the birth and death terms in Eq. (18.6) for our dissolution problem are zeros. Examples of death events would be cases that may have particle breakage or fracture, causing a sudden “death” of particles from the particular larger-sized class. The daughter particles that are formed as a result of this breakage event are examples of “births” in the corresponding smaller-sized classes. This does not occur in the present case, or at least not within the confines of our present model, but could be included if formation or disintegration of agglomerated granules is considered.

Substituting Eq. (18.8) in the governing PBM equations, and setting the birth and death terms to zero, we get

$$\frac{\partial}{\partial t} f_s(s, t) + \frac{\partial}{\partial s} \left[\left\{ \frac{-2h(c_{\text{sat}} - c(t))}{\rho_s} \right\} f_s(s, t) \right] = 0 \quad (18.9)$$

This is a simple description of the dissolution process where $f_s(s, t)$ and $c(t)$ are the main unknown for which solutions are sought. The initial and boundary conditions need to be specified. The initial condition is given as $f_s(s, t = 0) = f_0(s)$, where $f_0(s)$ is the initial particle size distribution of the powder to be dissolved. The boundary condition is given at the lower limit $s = 0$; at this limit, there is no flux, i.e. $G(s = 0, t) = 0$. At the lower boundary $s = 0$, particles do not reduce any further in size but rather go completely into solution. This equation describes the evolution of the solid particulate (solute) phase.

As there are two unknowns – $f_s(s, t)$ and $c(t)$ – in Eq. (18.9), a second equation is necessary. A complementary equation for the liquid solvent phase can be derived to describe the concentration increase as particles dissolve, given by

$$\frac{d}{dt} c(t) = \int_{s=0}^{s=\infty} h(\pi s^2) (c_{\text{sat}} - c(t)) f_s(s, t) ds, \quad (18.10)$$

with the initial condition $c(t = 0) = 0$, i.e. the starting point is a clean solvent medium. The integral in the above equation essentially is the summation of mass loss by a particle in each size class mass (rate of mass loss is $h(\pi s^2)(c_{\text{sat}} - c(t))$), weighted by the number of particles (per unit volume) in that size class (the weightage, i.e. number of particles is given by $f_s(s, t) ds$).

So far, we have described the PBM equations for a simple dissolution problem but have not discussed how the effects of fluid flow are incorporated. Recall that the mass transfer coefficient is a function of the fluid-flow velocity, as well as particle size, i.e. $h = h(s, v)$. The expression for $h(s, v)$ may be obtained from the literature (e.g. the Frossling or Ranz and Marshall correlation) or may be obtained through calibration. In this work, we followed an experimental calibration approach to determine fitting parameters A_h , B_h , and C_h in Eq. (18.5). We will not describe the calibration procedure in great detail, but just focus on the key concepts of setting up a PBM system and interpretation of the results.

A key point to note is that $h(s, v)$ depends on the relative fluid-particle slip velocity, v , which is most likely to vary as we scale up. As a simple approximation, one could consider the extreme limits of $v = 0$ and $v = v_t$ to obtain the longest and shortest possible dissolution times, where v_t is the particle’s terminal velocity. However, in reality, different particles will have different slip velocities, depending on the level of agitation that the particles experience that, in turn, depends on the scale of operation and the operating conditions (impeller and scraper rotation speeds).

A more precise measure of the distribution of slip velocities may be obtained from multiphase computational fluid dynamics (CFD) simulations of the system of interest. Figure 18.7a shows a representative snapshot from CFD simulations of the pilot-scale equipment (smallest unit in Figures 18.1 and 18.2). These Eulerian–Eulerian CFD simulations were performed with a non-Newtonian fluid phase and particle phase. Once the simulations achieve steady state, we find that the mixer is very well mixed, and therefore we can ignore any spatial variation in the fluid-particle speeds. Each mixer has an agitator and a scraper, the rotational speeds for which can be specified independently. The simulations were performed at three different agitation conditions, representing low, medium, and high impeller speeds, which we express as the dimensionless Froude number given as

$$\text{Fr} = \frac{\omega^2 D_{\text{mixer}}}{2g} \quad (18.11)$$

where

ω is the rotational speed of the scraper or agitator.
 D_{mixer} is the characteristic diameter of the mixer.
 g is the acceleration due to gravity.

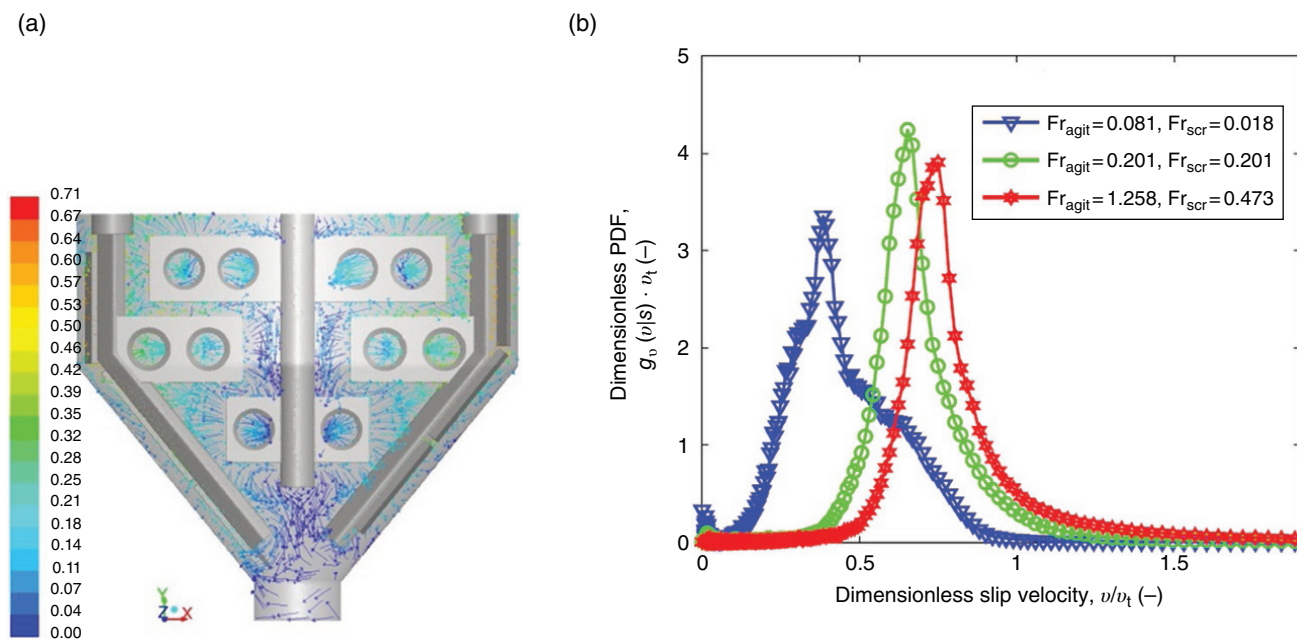


FIGURE 18.7 (a) Computational fluid dynamics (CFD) simulation snapshot for the pilot-scale equipment. Fluid-flow vectors shown colored by magnitude. (b) The probability distribution function for fluid-particle slip velocity, $g_v(v|s)$, scaled by terminal velocity v_t , shown for low, medium, and high agitation conditions.

The CFD results were analyzed to construct the probability distributions for the fluid-particle slip speeds (Figure 18.7b). The slip-velocity results were scaled with the particle terminal velocity; scaled thus, the distributions obtained for different particle sizes collapsed, thereby allowing us to use a master curve to determine the slip-velocity distribution for any particle diameter (valid within the range of 1–1000 μm). Interestingly, we find that the slip velocity for some particles exceeds v_t , especially at the higher agitation conditions. For each multiphase CFD simulations performed at varying scales and operating speeds, the corresponding slip-velocity distributions were generated.

Once we have determined the slip-velocity distribution, $g_v(v|s)$, we can use this to compute the effective mass transfer coefficient $h_{\text{eff}}(s)$, given by

$$h_{\text{eff}}(s) = \int_{v=0}^{v=\infty} h(v,s) \cdot g_v(v|s) dv \quad (18.12)$$

where the integral considers all possible slip velocities between 0 and ∞ , but weighted via the distribution function $g_v(v|s)$.*

*We borrow notation used in Bayesian probability: $g_v(v|s)$ is to be read as the conditional probability density function of the slip velocity (v , the continuous random variable), with given particle size (s). Hence, $g_v(v|s)$ can be thought of as a family of probability distributions when the given value of s is varied.

We will not discuss the full derivation, but ask the readers to accept it at present (will be presented in a separate publication in greater detail). When using Eq. (18.12), we implicitly assume that all particles of a given size exist uniformly mixed throughout the dissolution vessel and that there is no size-based segregation; this assumption is true in general for larger, well-mixed vessels such as the FrymaKoruma units.

We can now substitute Eq. (18.12) back in Eqs. (18.9) and (18.10) and rewrite our governing PBM equations as

$$\frac{\partial}{\partial t} f_s(s,t) + \frac{\partial}{\partial s} \left[\left\{ \frac{-2h_{\text{eff}}(s)[c_{\text{sat}} - c(t)]}{\rho_s} \right\} f_s(s,t) \right] = 0, \quad (18.13)$$

$$\frac{d}{dt} c(t) = \int_{s=0}^{s=\infty} h_{\text{eff}}(\pi s^2) (c_{\text{sat}} - c(t)) f_s(s,t) ds, \quad (18.14)$$

with the same previously listed boundary conditions. Therefore, we now have a model for our dissolution process that takes into account the full particle size distribution, as well as the effects of scale and agitation (implicitly via h_{eff}). The predicted outputs are the number density function of undissolved particles, $f_s(s,t)$, and the concentration of the dissolved drug, $c(t)$.

18.3.4.2.3 Predictions Using the Population Balance Model Now, we shall present the numerical solutions to the governing PBM equations developed in Section 18.3.4.2.2 and use the model to address an example of scale-up. A solver for the system of PBM equations was

developed in-house by the authors (in C++); we choose to spare the readers from the lurid details of implementing a numerical solver and present only the final results. It should be noted that the model implementation was validated successfully against an analytical solution (provided in section 2.11.1 of Ramkrishna [16]).

Before discussing the simulation results, we will first examine the model assumptions and inputs required to the model. Below are some of the key assumptions:

1. The PBM framework considers the entire mixing tank as a lumped system; there are no spatial variations or sub-compartments within the tank. This assumption is valid for very well-mixed tanks, confirmed from the flow and mixing times predicted by the CFD simulations.
2. Following the above well-mixed assumption, particles of any given size class are assumed to be uniformly distributed throughout the tank and conform to the slip-velocity distribution shown in Figure 18.7b. This is especially true when there are a large number of particles in the mixing tank, such as for powders with fine particle sizes.
3. There are no particle–particle interactions, valid when the particle suspension is dilute. In the present case, particle loading is only approximately 2% by mass.
4. Particle shapes are assumed to be spherical, although in reality drug crystals are somewhat oblong (needle shaped). In future, there is scope to further improve the correlation for h by incorporating shape effects more accurately, but it is outside the scope of the current work.

There are three required inputs to the PBM model:

1. Initial size distribution of particles ($f_0(s)$): This is obtained easily from direct experimental particle size measurements of the powder using laser diffraction (Malvern, Sympatec) or image processing (QICPIC).
2. Dissolution rate parameters: The fitting parameters in h (Eq. (18.5)) need to be determined before using the PBM model. A suitable correlation for $h = h(s, v)$ can often be found in the literature or engineering handbooks. Although we chose to calibrate these parameters, for now we will assume that a suitable constitutive relationship for $h(s, v)$ is known a priori.
3. Flow conditions, given by the relative-flow-speed distribution function $g_v(v | s)$: This distribution will vary based on scale and operating conditions, obtained from a series of CFD simulations in the present work.

Now that all required inputs to the model have been gathered, we can start generating predictions. As a representative example, let us start with the model results for the pilot-scale unit with medium agitation conditions. Recall that there are two coupled differential equations: Eqs. (18.13) and (18.14) represent the number distribution of undissolved particles and concentration of the dissolved drug, respectively.

Figure 18.8a shows the solution to the first equation (Eq. (18.13)). The numerical solution obtained for the number distribution function is converted to a mass distribution function for convenience; most papers/reports and experimental techniques generally report the mass distribution rather than the number distribution. Starting with the initial

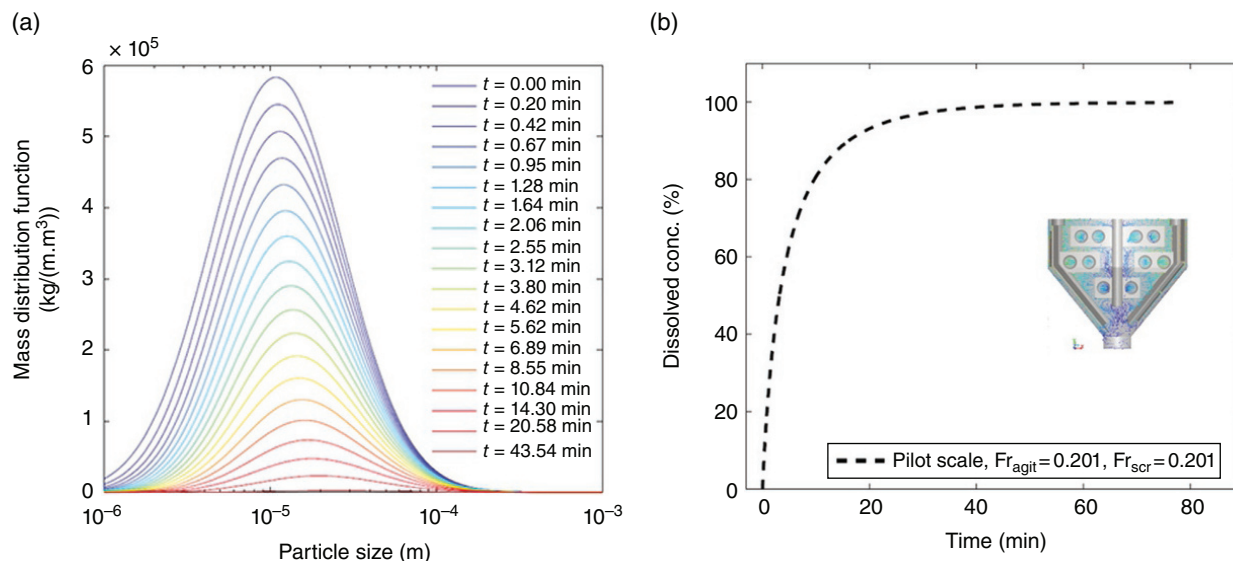


FIGURE 18.8 The PBM predictions presented, as a representative example, for the pilot-scale unit at medium agitator and scraper mixing speeds ($Fr_{\text{agit}} = 0.201$, $Fr_{\text{scr}} = 0.201$). (a) The mass distribution function of undissolved drug particles as a function of time. (b) The complementary dissolved drug concentration as a function of time.

size distribution of particles, which follows a lognormal distribution, the evolution of the mass distribution for various size classes is shown as a function of time. As time progresses, the number density magnitudes diminish and fewer particles remain in suspension. Although all size classes show decrease in numbers with time, the smaller particles disappear faster – these smaller particles have larger surface-to-volume ratios, and the dissolution rate is fastest for these particles (also reaffirmed by Eq. (18.5), where h varies inversely with particle diameter s). After long times, only a few large particles remain that continue to shrink and eventually dissolve completely.

Figure 18.8b presents the solution to the second governing equation, Eq. (18.14), which provides the concentration of the dissolved drug with time. The concentration versus time curve is typical of most dissolution processes: the dissolution is fast initially, but slows down with increasing time. We should mention that the drug loading is slightly higher than 50% of the saturation solubility. Hence, as more drug goes into solution, the driving force $\Delta c = c_{\text{sat}} - c$ reduces, further slowing down the dissolution rate. Had the drug loading been much lower than the saturation solubility (i.e. if $c \ll c_{\text{sat}}$, refer to example problem in Section 18.3.4.2.1), then the dissolution would have been much faster, and the process would have carried a lower risk. It is always worthwhile to examine the amount of solute relative to the saturation solubility before attempting to design any dissolution unit operation.

Now, we shall present an example of using the PBM framework to virtually scale up the unit operation. The dissolution process will be scaled up from the pilot scale to the clinical scale (Figure 18.9a). The dissolution performance obtained with medium agitation at the pilot scale will be the target condition; the goal is to obtain a dissolution profile that is very similar to the profile obtained at the target condition (i.e. match Figure 18.8b).

Many of the factors that affect dissolution do not change as we scale up, including process temperature, the solvent, and the solute. The changes that do occur include the size and scale of the mixing tanks, as well as the agitation conditions. Particles are agnostic to these device-scale changes: microscopically, individual particles are not sensitive to these macroscopic properties. However, particles are sensitive to the localized microscopic flow around particles that, in turn, depends on the scale and agitation conditions. The key to scale-up is to preserve this localized flow around particles, viz. the fluid-particle slip-velocity distribution, between the pilot- and clinical-scale mixing tanks. Figure 18.9b shows these slip-velocity distributions: the target clinical-scale slip-velocity distribution that we wish to preserve (shown as a black dotted line), and the slip-velocity distributions at the larger clinical scale for low, medium, and high agitation conditions are presented.

Based on Figure 18.9b, we can instantly make an educated guess that the ideal scaled-up, clinical-scale operating

conditions are likely to lie between the low and medium agitation speeds. The PBM predictions for the dissolution times are shown in Figure 18.9c; only the final few percentage points are shown as the objective is to match the pilot-scale dissolution time – 99.9% dissolved in 78 minutes – as closely as possible. The clinical-scale process at low agitation yields a 99.9% dissolution time of 108 minutes, which is much longer than our desired target of 78 minutes. At high and medium agitation, the process takes 74 and 83 minutes, respectively. Practically, both conditions are acceptable as the processing times are within ± 5 minutes of the target. From an operational safety perspective, however, it is unadvisable to run the processing equipment at the highest speed settings. Therefore, we propose the medium agitation conditions as the recommended scaled-up conditions.

As noted earlier, the ideal scaled-up conditions are between the medium and high agitation speeds – such an intermediate speed may have precisely yielded a 74 minute dissolution process. Recall that the slip-velocity distribution for each set of agitation conditions is generated from a corresponding CFD simulation – these CFD simulations incur a moderate to high computational cost; each run can take several hours to a few days to execute (compared with a few minutes for the PBM solver). Although we could have explored the clinical-scale speed range between medium and high agitation more thoroughly, such a detailed study was not warranted as our recommended scaled-up conditions were sufficient from an industrial perspective.

18.3.4.3 Annealing: Thermally Induced Irreversible Viscosity Change of PEG Ointments Compositionally, a PEG ointment comprises a solid PEG phase and a liquid phase. As mentioned in previous sections, manufacturing process plays a critical role in determining the rheological attributes of a PEG ointment. Our experience with this composition showed that incorporating an annealing step during manufacturing impacts the final product quality by way of improving ease of spreading of the ointment during application.

For the PEG ointment composition as described in this chapter, a post-congealing hold or annealing at 40 °C was found to cause an irreversible decrease in product viscosity, which turned out to be desirable. This behavior was first observed when ointment samples were studied using rheometry. It was observed that the viscosity of initial or fresh ointment samples can be reduced if stored at 40 °C for a period of time or if the samples undergo temperature cycling between 25 °C and 40 °C, as shown in Figure 18.10. Once this change occurred, no further change in viscosity was observed upon long-term storage at 40 °C [20].

To better understand the rheological properties of the ointment, viscosity-temperature sweep profiles were generated on similar ointment samples (Figure 18.11). It was found that fresh ointment samples showed two viscosity transitions: one in the range of 38–42 °C and one in the range of 44–48 °C

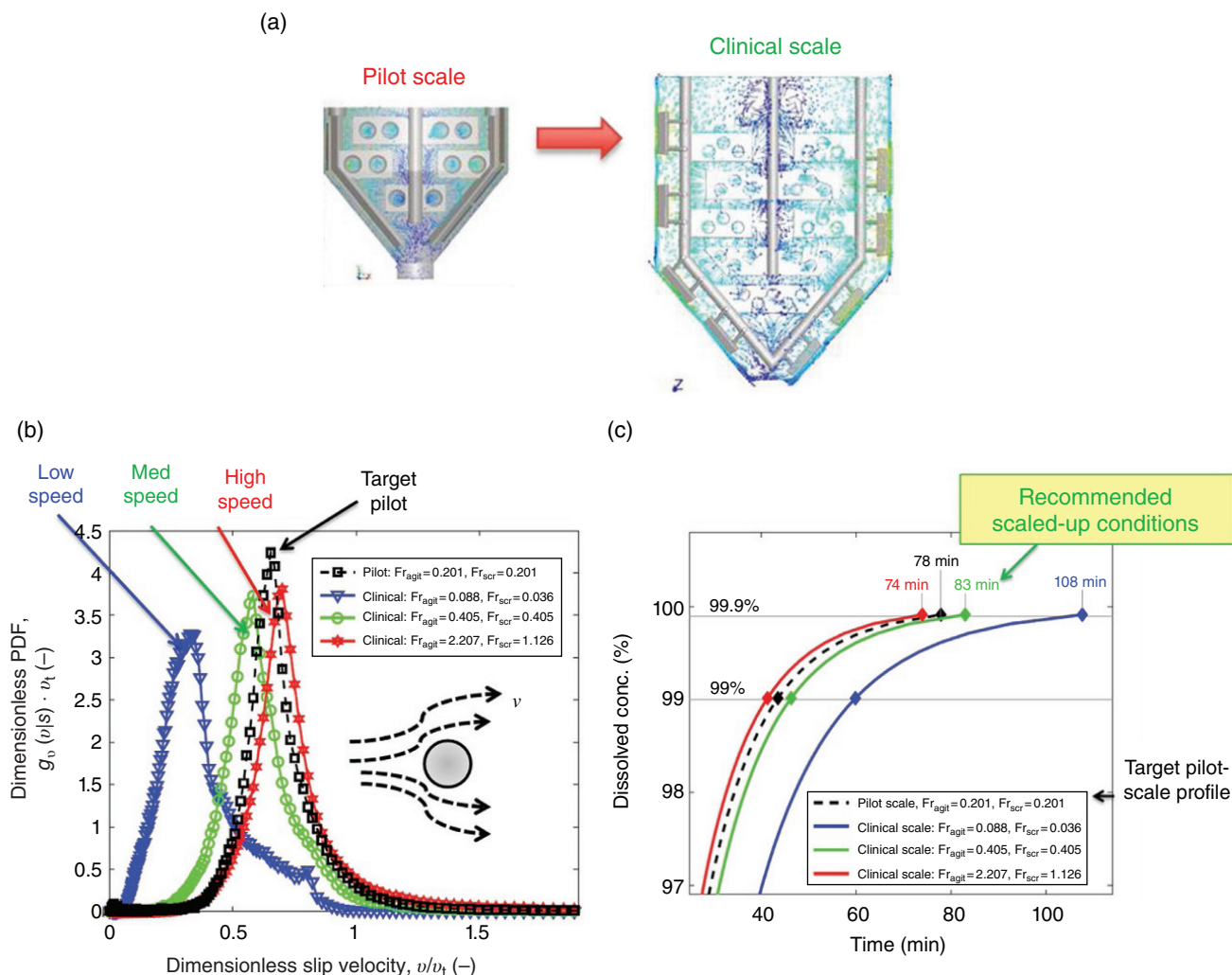


FIGURE 18.9 An example of using the PBM framework for scale-up. (a) Schematics of the units of interest (not drawn to scale); the goal is to scale up from the pilot scale to the clinical scale. (b) Slip-velocity distributions, obtained from CFD simulations, of the pilot-scale unit – target to be matched after scale-up (dotted black line), and distributions for low, medium, and high agitation speeds at the clinical scale. (c) Dissolution curves showing the pilot-scale target to be matched (dotted black line) and the predicted dissolution profiles for low, medium, and high agitation speeds at the larger clinical scale; zoomed view showing the final few dissolution percentage points.

that correspond to melting to form liquid. On the other hand, ointment samples that had previously been stored at 40 °C for either 2 hours or 51 hours showed only a single melting transition in the range of 44–49 °C.

To further investigate the transitions, DSC scans were generated on the ointment samples before and after three temperature cycles between 25 and 40 °C, Figure 18.6. The initial or fresh ointment sample showed two endothermic transitions centering around 41 and 46 °C, respectively. The temperature ranges at which the two thermal transitions occurred also coincided with the viscosity transitions observed previously. On the other hand, the ointment sample exposed to three cycles of temperatures between 25 and 40 °C only showed a single thermal transition centering

around 47 °C with lower viscosity in the range between 25 and approximately 37 °C.

Craig and Newton have suggested that high molecular PEGs (3000–20 000 MW) may exist in two physical forms – an extended form and a folded form [11]. The extended form is considered the more stable form, while the folded form is considered metastable and exhibits a lower T_m than the extended form. Buckley and Kovacs further suggested that chain unfolding may occur during heating, resulting in the transition to a more stable extended form [21]. These reported observations, although based on characterization of solid high molecular PEGs, are still relevant to the PEG ointment system as PEG 3350 exists as dispersed solid particles at room temperature in the ointment.

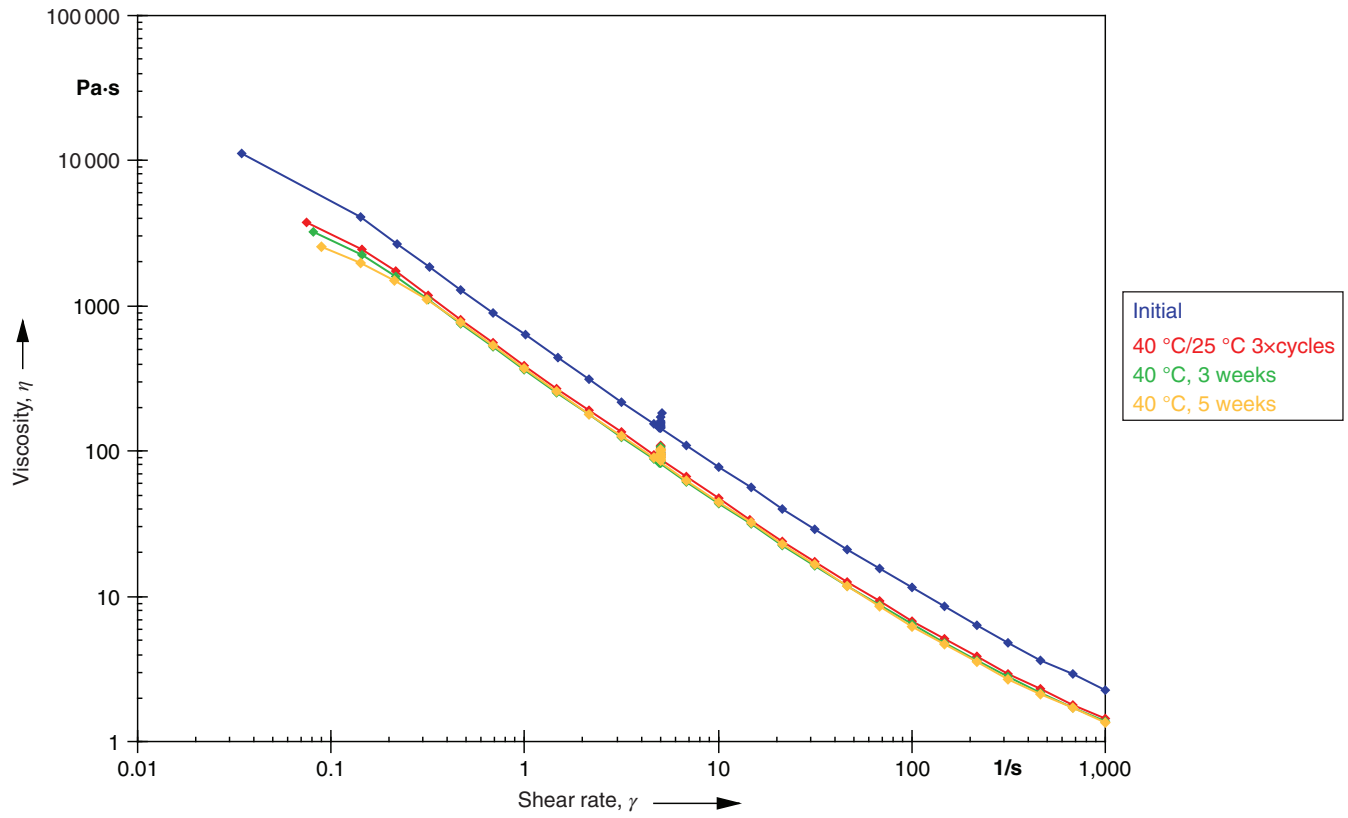


FIGURE 18.10 Viscosity–shear rate profiles of PEG ointment samples after exposure to 40 °C.

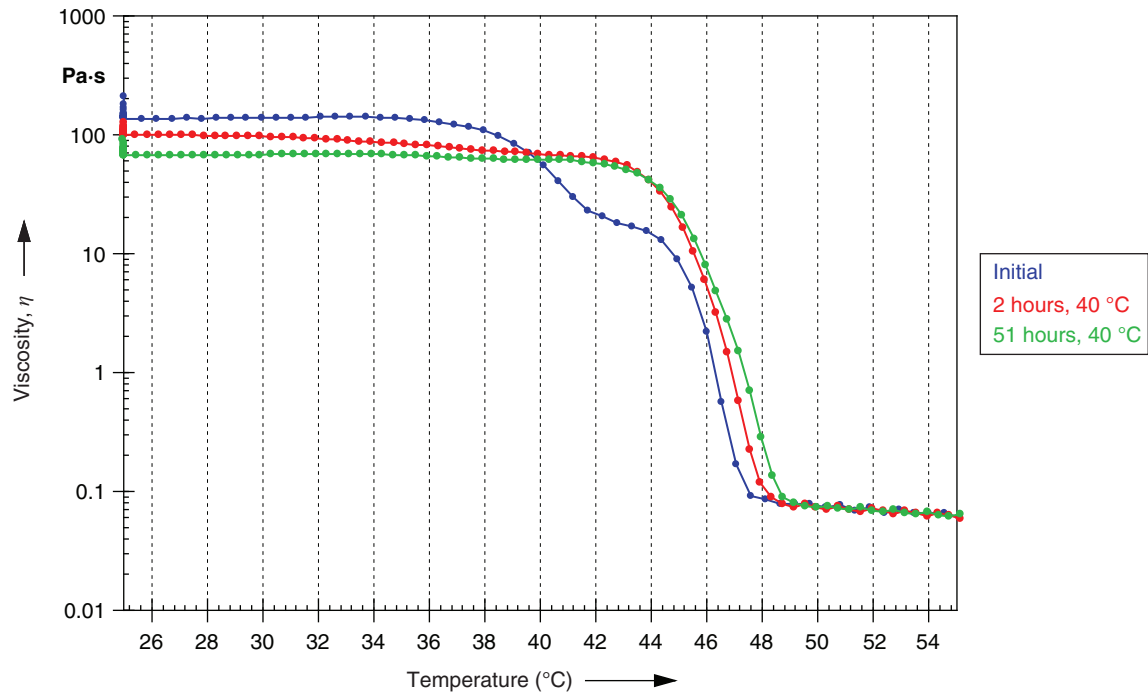


FIGURE 18.11 Viscosity–temperature profiles of PEG ointment samples after exposure to 40 °C.

Based on the observed transitions in Figure 18.12, the two DSC endothermic events (~41 and ~46 °C) would appear to correspond to the melting of folded and extended forms of PEG 3350, respectively. Furthermore, exposure to 40 °C likely caused the fraction of PEG 3350 in the metastable folded form to melt and recrystallize to form the stable extended configuration. As a result, we identified a way to control for consistency of product viscosity by enriching the stable polymorph form of PEG 3350 in the product. It also became apparent that such control can be achieved by annealing the ointment during manufacturing.

To demonstrate the relevance and criticality of this material attribute on product quality, a small sensory panel study was conducted to assess firmness and ease of spreading of four blinded ointment samples that were processed differently.

Figure 18.13 shows the evaluation result of four placebo ointments of the same composition. Samples were (A) ointment without annealing, (B) ointment with post-manufacturing annealing at 40 °C for eight days, (C) ointment incorporating 2 h in-process annealing at 40 °C, and (D) ointment incorporating 2 h in-process annealing at 40 °C with post-manufacturing annealing at 40 °C for eight days.

Most panelists were able to identify ointment sample A as having different firmness and ease-of-spreading properties compared with samples B, C, and D. Sample A was the only sample that did not undergo any form of annealing and had the highest measured viscosity. This physical property was correctly noticed by most panelists who collectively assigned a higher firmness score and a lower ease-of-spreading score for sample A.

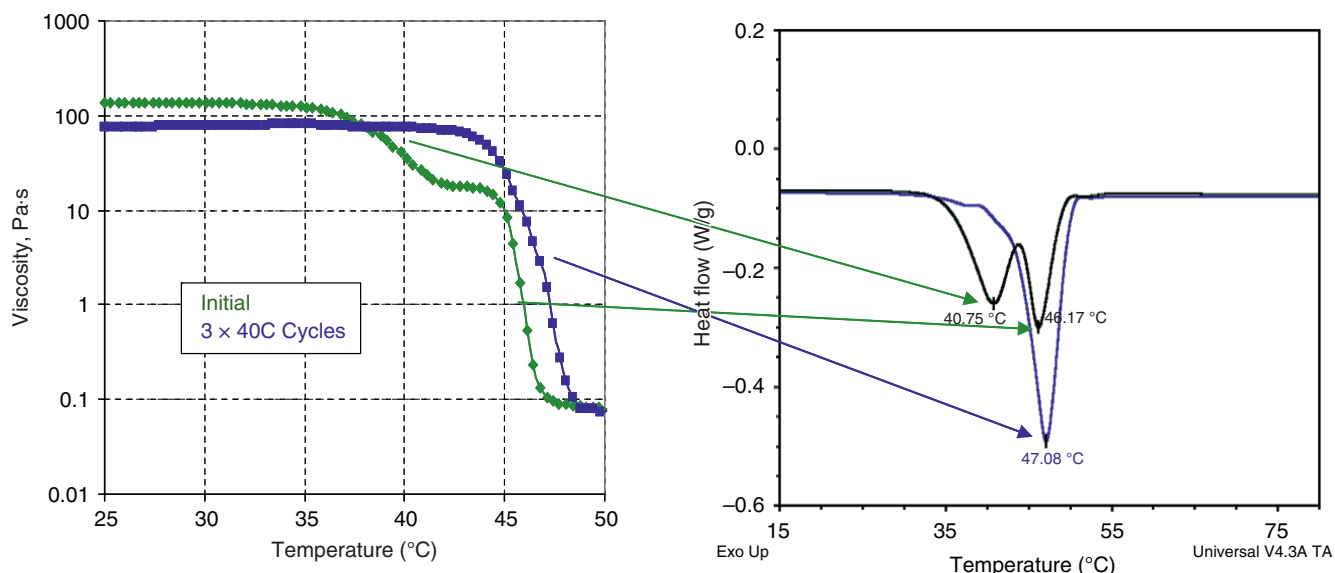


FIGURE 18.12 DSC profiles (right) of PEG ointment samples before and after exposure to 40 °C and comparison with viscosity–temperature profiles (left).

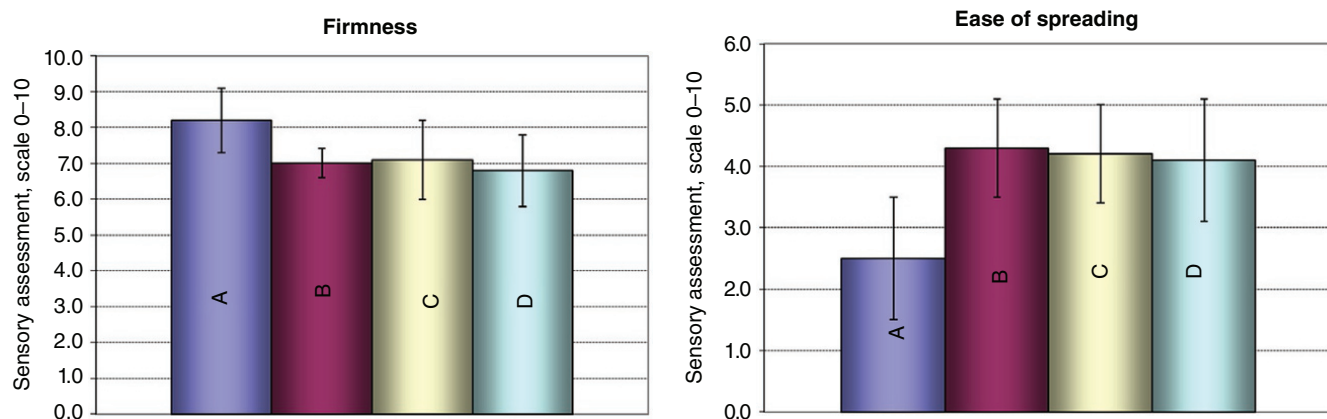


FIGURE 18.13 Sensory evaluation of ointment properties: firmness (left) and ease of spreading (right).

The overall result of this finding illustrates the impact on product quality by a material attribute that can be controlled through the manufacturing process. Based on this insight, it was recommended that the manufacturing process includes an annealing step as a means to control the viscosity of the final ointment product.

18.4 SUMMARY

This chapter described a scale-up approach for a topical PEG ointment. While some aspects, such as the importance of annealing, are specific to compositions like the one described, which is composed of a high MW PEG, others, such as API dissolution and process modeling, are applicable to any type of semisolid formulation. Overall, when designing a scale-up approach for a semisolid formulation, one should consider the physicochemical properties of the API and the excipients being used (e.g. solubility, miscibility, chemical reactivity, melting/boiling points, solid crystalline form), the target quality attributes of the final product (e.g. viscosity, chemical degradation profile), and the required processing steps and available equipment (e.g. mixing at constant temperature, high shear to emulsify and/or aid dissolution of formulation components).

Additionally, we have demonstrated the capabilities of engineering computational models to aid process design and scale-up. By developing and utilizing a PBM framework, we have helped build a fundamental understanding of the manufacture process and enabled a robust scale-up strategy for moving from the pilot scale through to clinical scale, which can be further extended to the commercial scale by following the same methodology. The general PBM framework described in this chapter can be followed to model other dissolution processes, extending beyond the manufacture of topical products.

REFERENCES

- Buhse, L., Kolinski, R., Westenberger, B. et al. (2005). Topical drug classification. *International Journal of Pharmaceutics* 295 (1–2): 101–112.
- Kuswahyuning, R., Grice, J.E., Moghimi, H.R., and Roberts, M. S. (2015). Formulation effects in percutaneous absorption. In: *Percutaneous Penetration Enhancers - Chemical Methods in Penetration Enhancement* (ed. N. Dragicevic-Curic and H.I. Maibach), 109–134. Springer.
- Andrews IV, A.T. (2007). Filtered models for gas-particle flow hydrodynamics. PhD thesis. Princeton University.
- D'souza, A.A. and Shegokar, R. (2016). Polyethylene glycol (PEG): a versatile polymer for pharmaceutical applications. *Expert Opinion on Drug Delivery* 13 (9): 1257–1275.
- Osborne, D.W. (2011). Diethylene glycol monoethyl ether: an emerging solvent in topical dermatology products. *Journal of Cosmetic Dermatology* 10 (4): 324–329.
- Oliveira, G., Hadgraft, J., and Lane, M.E. (2012). The role of vehicle interactions on permeation of an active through model membranes and human skin. *International Journal of Cosmetic Science* 34 (6): 536–545.
- Fasano, J.W., ten Berge, W.F., Banton, M.I. et al. (2011). Dermal penetration of propylene glycols: measured absorption across human abdominal skin in vitro and comparison with a QSAR model. *Toxicology in Vitro* 25 (8): 1664–1670.
- Freeman, D.J., Sheth, N.V., and Spruance, S.L. (1986). Failure of topical acyclovir in ointment to penetrate skin. *Antimicrobial Agents and Chemotherapy* 29 (5): 730–732.
- Barnes, K. (2005). Croda presents new high purity excipients. <https://www.in-pharmatechnologist.com/Article/2005/11/11/Croda-presents-new-high-purity-excipients> (accessed 8 November 2018).
- At what temperature will CARBOWAX™ polyethylene glycol products degrade? (2018). https://dowac.custhelp.com/app/answers/detail/a_id/3821 (accessed 8 November 2018).
- Craig, D.Q.M. and Newton, J.M. (1991). Characterisation of polyethylene glycols using differential scanning calorimetry. *International Journal of Pharmaceutics* 74 (1): 33–41.
- Wan, B., Zordan, C.A., Lu, X., and McGeorge, G. (2016). In-line ATR-UV and Raman spectroscopy for monitoring API dissolution process during liquid-filled soft-gelatin capsule manufacturing. *AAPS Pharm SciTech* 17 (5): 1173–1181.
- Noyes, A.A. and Whitney, W.R. (1897). The rate of solution of solid substances in their own solutions. *Journal of the American Chemical Society* 19 (12): 930–934.
- Frossling, N. (1938). Über die verdunstung fallender tropfen. *Gerlands Beiträge zur Geophysik* 52: 170–215.
- Ranz, W. and Marshall, W. (1952). Evaporation from drops. *Chemical Engineering Progress* 48 (3): 141–146.
- Ramkrishna, D. (2000). *Population Balances: Theory and Applications to Particulate Systems in Engineering*. Academic Press.
- LeBlanc, S. and Fogler, H.S. (1987). Population balance modeling of the dissolution of polydisperse solids: rate limiting regimes. *AIChE Journal* 33 (1): 54–63.
- Bhaskarwar, A.N. (1989). General population balance model of dissolution of polydisperse particles. *AIChE Journal* 35 (4): 658–661.
- McCoy, B.J. (2001). A new population balance model for crystal size distributions: reversible, size-dependent growth and dissolution. *Journal of Colloid and Interface Science* 240 (1): 139–149.
- Kasli, S.I., Yuhas, L.M., and Margulis, L.A. (2010). Investigation of thermal transition of a PEG-based ointment by rheology and DSC. American Association of Pharmaceutical Sciences Annual Meeting, New Orleans, LA (14–18 November 2010), p. W4121.
- Buckley, C.P. and Kovacs, A.J. (1976). Melting behaviour of low molecular weight poly (ethylene-oxide) fractions 2. Folded chain crystals. *Colloid and Polymer Science* 254 (8): 695–715.

19

ACHIEVING A HOT MELT EXTRUSION DESIGN SPACE FOR THE PRODUCTION OF SOLID SOLUTIONS

LUKE SCHENCK, MIKE LOWINGER
Merck & Co., Inc., Rahway, NJ, USA

GREGORY M. TROUP, LI LI, CRAIG MCKELVEY
Merck & Co., Inc., West Point, PA, USA

19.1 INTRODUCTION

Industrial applications of hot melt extrusion date back to the early 1900s. This process has since grown in use, and today is one of the most widely applied unit operations in the polymer industry. Recent increases in both patents and publications indicate that it is rapidly becoming a key processing route for pharmaceutical dosage units as well [1].

Pharmaceutical extrusion has been used to process pastes (e.g. wet granulation) and polymer melts in which an active pharmaceutical ingredient (API) is present in an amorphous and/or crystalline state. The extrusion process has enabled a wide range of product applications including (i) solid solutions for the oral delivery of insoluble poorly soluble APIs, (ii) implants, (iii) intraoral delivery, (iv) ophthalmic delivery, (v) controlled release (via matrix or multiparticulates), (vi) conventional (i.e. API is in crystalline state) tablets by continuous wet granulation, and (vii) nanocrystalline formulations.

Though many uses exist for hot melt extrusion, its application to produce solid solutions meeting pharmaceutical quality requirements remains limited. This chapter outlines design for Six Sigma (DFSS)-led development activities used to define the design space for an extrusion process, generating a solid solution. The use of DFSS methodology is a natural fit given its long-standing application to new technologies and added benefit of having significant overlap with the FDA's quality by design (QbD) initiative. The benefits of QbD development include added regulatory flexibility enabling continuous process improvement post filing, more prioritized development,

and an overall more effective management of risk. This text outlines how QbD methodology was reduced to a development road map using the DFSS tool set.

This chapter begins with an overview of solid solutions to provide a motivation for implementing hot melt extrusion in the pharmaceutical industry followed by a more detailed review of the extrusion process itself. The risk assessment activities that set the stage for more focused development efforts are reviewed, and the final section summarizes the use of process analytical technology (PAT) to facilitate definition of a multifactor design space. The chapter ends with a forward-looking vision of how this design space ultimately might be translated into a control strategy.

19.2 INTRODUCTION TO SOLID SOLUTIONS

Solid solutions have been developed largely to modulate undesirable drug properties – particularly the poor aqueous solubility and/or wetting behavior of development candidates. Product pipelines in major pharmaceutical companies are increasingly composed of insoluble drug candidates driving the need for alternative oral delivery strategies including nanocrystalline, lipid, and solid solution-based formulation technologies [2]. By some estimates the fraction of development candidates considered very soluble has dropped below 10% [3]. A primary driver for deploying solid solution approaches in pharmaceutical development is to increase the exposure of orally administered poorly soluble active

compounds, and several reviews on this subject have been written [4–7].

Solid solutions are solutions of API in a glassy polymer often prepared by melt, solvent, and/or mechanical means by processes such as extrusion, spray drying, or mechanical activation, respectively. Increased oral absorption from solid solution formulations is achieved by supersaturation and/or *in situ* formation of nanoparticles [8, 9]. Increasing the apparent solubility of active compounds drives both potentially faster dissolution and permeability rates [10], making it possible to achieve dose proportional increases in exposure at higher doses and reduced potential for formulation-related food effect. In addition to these benefits, solid solution formulations enable combination products in a solid format and have been used to bridge from liquid-filled capsule-type self-emulsifying and/or self-microemulsifying formulation approaches (Figure 19.1). Solid solutions have also found

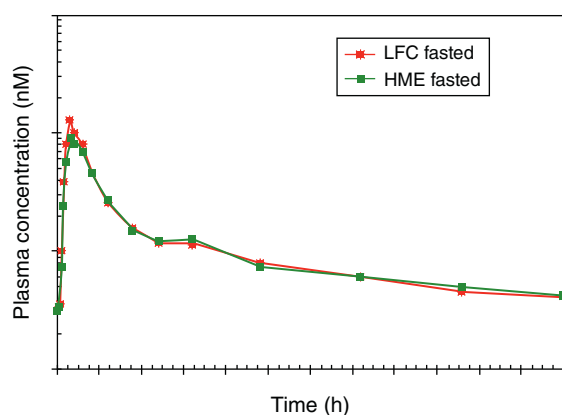
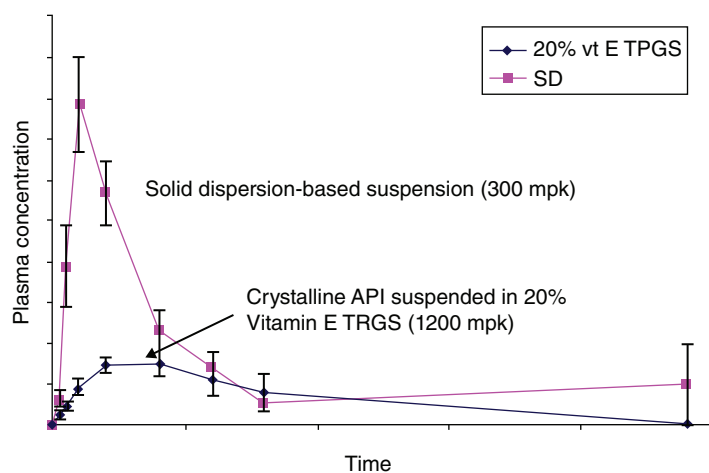


FIGURE 19.1 Plasma concentration following oral administration in fasted healthy adults of liquid-filled capsule (circles) and a solid solution intermediate-based tablet (squares) (compound A; $n = 24$; 200 mg).



use in the preclinical setting, often in the form of suspensions [11]. The exposure increase compared to common alternative formulations can be dramatic (Figure 19.2). The benefit in exposure does come with a commensurate physical instability risk posed by these stabilized amorphous systems, though several compounds have been successfully launched.

Any process that reliably produces homogeneous glasses with consistent properties can be used to make solid solutions (Figure 19.3); however, there are relatively few papers that have discussed the relationship of process selection on product performance [12–16]. Those authors that have broached this subject have typically compared the performance of identical or unique formulations prepared at a single set of operating conditions using multiple preparation processes. It is not possible to make generalizations about superiority of a particular process for solid solution manufacture from the current body of published work. This is largely because an understanding of the operating space explored and associated characterization data used to determine if a homogeneous glass was produced in each case was not presented.

Each preparation process possesses advantages and disadvantages. Solvent-based processes are generally easier to scale down to mg-scale while extrusion approaches generally provide greater production rates per equipment volume. Access to multiple approaches is likely required to broadly enable a diverse portfolio of compounds with solid solution technology. The identification of drug candidate polymorphs may cause development challenges for spray drying process development in the identification of suitable solvent systems or extrusion processes in the case of thermal degradation or temperature-induced polymorphic transitions.

A particularly desirable aspect of extrusion in today's pharmaceutical development environment is the ability to continuously process in a direct-to-drug-product manner (Figure 19.4). Extrusion naturally lends itself to continuous

Cpd	SD susp. dose	Reference formulation	Exposure increase
A	300 mpk	1200 mpk 20% Vitamin E TPGS	2–8X
B	100 mpk	100 mpk Imwitor 742:PS 80	67X
C	100 mpk	300 mpk Methocel Suspension	2–6X
D	200 mpk	750 mpk 10% PS80	4–16X

FIGURE 19.2 Preclinical impact of solid solution suspensions. (Left) Plasma concentration profile following oral administration in male Sprague-Dawley rats (compound B; $n = 4$). (Right) Cross-project preclinical formulation comparison of solid solution-based suspensions and a variety of reference formulations (mpk, milligrams per kilogram).

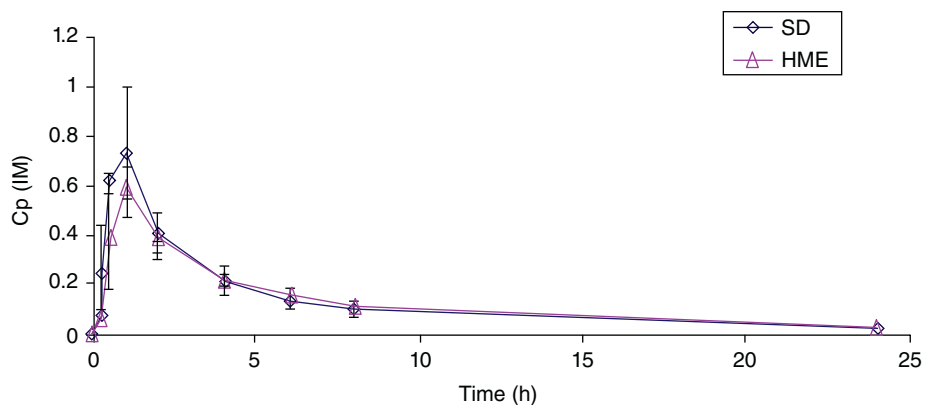


FIGURE 19.3 Comparison of identical solid solution formulations prepared by spray drying (SD) and extrusion (HME). Plasma concentration profile following oral administration of compound B in HPMCAS-LF as tablets (male beagle dogs; $n = 6$; crossover; 50 mg dose).



FIGURE 19.4 Direct-to-drug-product extrusion process train schematic, highlighting bulk feeders, precision loss in weight feeders, liquid injection, and calendaring of tablet between two chilled rolls.

processing from raw materials pneumatically conveyed into individual feed hoppers supplied by bulk containers to molded drug product fed into bulk product containers or even blister packages. This process has a small specific volume, making it amenable to real-time quality control and keeping the equipment footprint on a manufacturing floor exceptionally small. The product in an extrusion process naturally flows through a relatively narrow cross section, making continuous and direct process analytic interrogation straightforward (e.g. reducing issues of sampling).

19.2.1 Systematic Development Strategy

Prior to starting the design space definition process, efforts began by attempting to translate existing knowledge (i.e. experimental information, first principles understanding, models and best practices gleaned from peer-reviewed literature) into a comprehensive view of the extrusion process. Process input parameters were summarized to ensure none of the process parameters were overlooked. This was facilitated with process mapping exercises following fishbone or Ishikawa diagram methodology.

Subsequent risk assessment activities focused the parameters from the process map to a subset of potential critical process parameters (CPPs). These potential CPPs were identified as having a higher probability of impacting potential critical quality attributes (CQAs) as defined from an understanding of solid solution product requirements relevant to the patient. The risk-based evaluation of all parameters against the potential CQAs employed a quality function deployment (QFD) grid consistent with house of quality methodology [17]. This risk assessment exercise was not one of the tools explicitly outlined in the ICH Q9; however, it successfully managed the complexity of the hot melt extrusion process and delivered on the spirit of providing “transparent and reproducible methods to accomplish steps of the quality risk management process based on current knowledge about assessing the probability, severity, and, sometimes, detectability of the risk.” [18] The outcomes of this exercise were a summary of clearly prioritized development targets focused on a manageable set of potential CPPs.

Development efforts aimed to generate a fundamental understanding of the system, recognizing that the degree of regulatory flexibility attainable was predicated by depth of knowledge. Enhancing this fundamental understanding was a focus on scale-independent parameters rather than scale-dependent parameters. An example of achieving this more fundamental understanding of the extrusion process by studying scale-independent parameters is demonstrated here for the case of shear stress in the extruder. Shear stress is a key parameter thought to influence quality attributes including degradation and the capacity to achieve a molecularly dispersed product. Shear stress (the scale-independent parameter) is primarily manipulated via screw speed or rpm (the

scale-dependent parameter). While rpm is an easily accessible parameter to incorporate into design of experiments (DOEs), only studying the influence of rpm on product quality would have resulted in at best a correlative understanding of the system. This type of understanding does not capture the added impact of the degree of fill, scale, equipment manufacturer, or the result of wear in the extruder, all of which can change the shear stress at a given rpm. Focusing on scale-dependent parameters (i.e. rpm) rather than scale-independent parameters (i.e. shear stress) would have limited the potential to apply findings broadly to both anticipated and unanticipated process changes.

Initial small-scale experiments were conducted via DOE and successfully identified one scale-independent parameter that had a disproportionate impact on the key product quality attributes. This also yielded supportive data for more definitive CQA and CPP definitions. The next stage of development sought to build empirical models for the relationship between key scale-independent and scale-dependent parameters. This was achieved through response surface mapping, which was needed due to the multidimensional and quadratic means by which scale-dependent parameters (e.g. rpm, degree of fill, clearance) influenced scale-independent parameters (e.g. shear stress). This DOE was conducted on commercial-scale equipment incorporating inline PAT that facilitated evaluation of both product quality and process robustness. Robustness was assessed by introducing known perturbations at all processing conditions and monitoring the system’s capacity to dampen the upset. The combined understanding of quality attributes as a function of scale-independent parameters, the empirical models from response surface mapping, and the PAT data on process robustness allowed definition of a design space that achieved balance between achieving target quality attributes and optimizing operating conditions including throughput and process stability.

19.2.2 HME Process Overview and Mapping

The intent of process mapping is to create an objective view of the extrusion process and generate the set of potential process inputs to serve as the basis for risk assessment activities. Consider extrusion as a series of suboperations: (i) material feeding, (ii) powder conveying and pre-venting, (iii) melting and mixing, (iv) melt conveying and venting, and (v) pumping, shaping, and cooling. Each of these suboperations was mapped independently. The following sections include the map of process inputs and discuss some of the features of each suboperation.

19.2.3 Material Feeding

Extruder feed systems in many ways ultimately control the content uniformity of product. While extruders offer some

back mixing to dampen out high-frequency feed rate perturbations, low-frequency disturbances in material feed rates can result in compositional variations in the extrudate and, hence, compromised product quality [19]. One route to decouple feeder performance from the extrudate compositional uniformity would be to pre-blend all of the feed streams. However, pre-blending requires an extra unit operation, presents a risk for segregation, and is generally complex when one or more components are liquids. Figure 19.5 is a process map outlining a subset of the potential process inputs for the material feeding process.

Solids feed to the extruder can be achieved via volumetric or gravimetric feeders. Volumetric feeders are best suited for applications where the materials flow well and control of the composition of the feed stream will not vary (i.e. for use with powder pre-blends). The applications presented here focus on gravimetric feeders. These feeders are often single or twin-screw feeders coupled to a load cell(s). Constant flow rates into the extruder are attained via loss in weight feedback control.

A key consideration when evaluating solids feeders is that flow rates from these units are never fully continuous. Intrinsic to powder conveying screws is some level of pulsing into the extruder, since the feed material is not a continuous media. For very poorly flowing material exhibiting strong propensity for avalanching, the frequency of the pulsing for a twin-screw feeder could be expressed by

$$\frac{\text{rpm}}{(2n-1) \cdot 60} \tag{19.1}$$

where

n is the number of flights in the feed screws.

For example, for a twin-screw feeder with single-flighted conveying screws feeding poorly flowing material, running at 60 rpm would result in a pulse every second. Similarly the concept of continual versus pulsed addition applies for liquids. At low liquid flow rates, any inconsistent flow or dripping should be evaluated to ensure it is not occurring at a frequency undampened by the extruder’s back mixing. This can be overcome by trying to achieve back pressure to deliver the liquid as a continuous stream.

Powder feeders can be coupled to the extruders such that material drops from the end of the feed screws into an open barrel section. A limitation of this approach is that incorporation of the powders into the extruder is limited to a fraction of the open barrel feed port where the extruder’s down turning screw conveys material into the extruder [20]. The feeding of low bulk density powders can be particularly problematic and can significantly constrain the maximum attainable throughput. A common means to overcome flood feed limitations of low bulk density materials is via a side stuffer. Here, the loss in weight feeder delivers powder

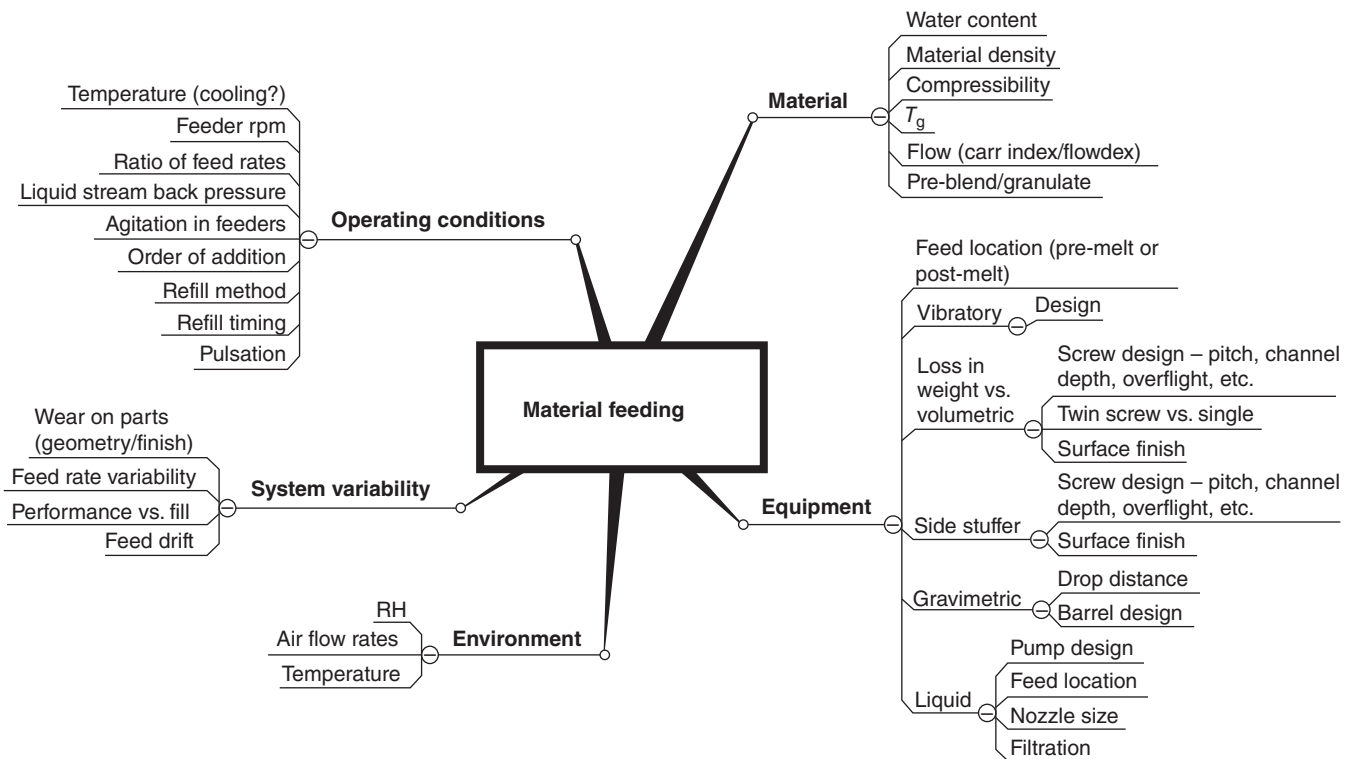


FIGURE 19.5 A summary of input parameters of the material feeding process.

gravimetrically to volumetric (constant rpm) conveyors coupled to the side of the extruder. These conveying screws produce some pre-densification and force the powder into the extruder, enabling higher throughputs for low bulk density powders. Successful densification in the side stuffer requires sufficient venting of entrapped air.

A consideration when designing feed locations is the disparities in shear stress, temperature, and mixing histories that materials experience. The highest viscosity that polymers transition through occurs at the melt onset, which is also the point of maximum shear stress. Active ingredients or liquids could be added upstream or downstream from this point. Adding the active downstream limits the total time at temperature and reduces the shear stress; however it also potentially reduces the extent of mixing achieved. An additional consideration when determining the feed location is the difficulties in mixing materials, having large viscosity differences. Lower viscosity materials can act as lubricants, reducing shear rates and mixing intensity [21]. In extreme cases, streaming of the low viscosity material through the extruder may be observed. Special attention should be paid to the mixing sections of the screw profile when there is a need to incorporate materials, having large viscosity differences.

19.2.3.1 Powder Conveying and Venting The next task that must be accomplished in the extruder is conveying the bulk powder to the melt zone. A list of process inputs to consider when feeding and conveying low bulk density powders is listed in Figure 19.6. In conveying sections, material passes through the extruder via drag flow with very little pressure generation. In an ideal conveying scenario, material would demonstrate perfect slip with the screw and perfect friction with the barrel.

Figure 19.7 highlights aspects of screws with particular relevance to conveying. Increasing flight width results in greater pumping efficiency of the screws. Increasing flight pitch, or angle of the helix, generally results in faster material conveying per revolution, although low bulk density materials may not follow this trend. The internal (D_i) and external (D_o) diameters of the screw have numerous implications. The channel depth ($D_o - D_i$) sets the free volume of the extruder and largely determines the maximum feed rate (without the use of mechanical pre-densification via a side stuffer). The D_i generally limits available torque at small scales, though at larger scales the torque constraints of the drive motor may be limiting. Though a larger D_i enables greater torque, this generally reduces throughput by sacrificing free volume.

Conveying involves added complications for pharmaceutical applications because many of the feedstocks possess

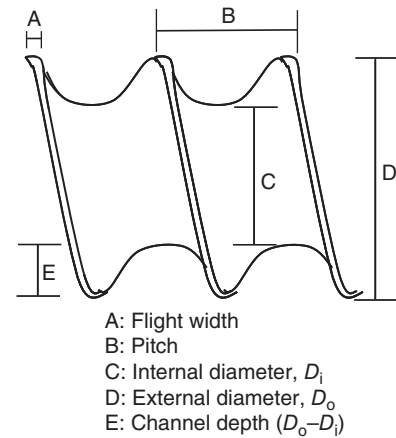


FIGURE 19.7 A subset of the critical design components of conveying screw elements.

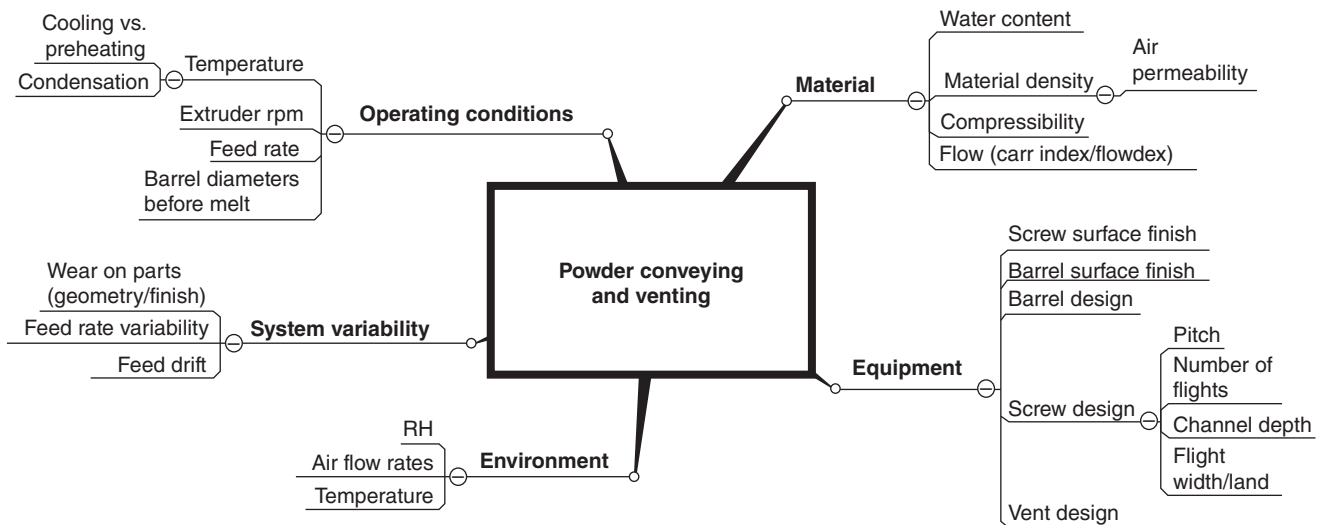


FIGURE 19.6 A summary of input parameters of the powder conveying and venting suboperations.

substantially lower bulk densities than commodity polymer extrusion operations based on 3 mm pellets. It is not uncommon for an API to have 0.2–0.4 g/cc bulk density. Many of the pharmaceutical polymers and surfactants are also only available from manufacturers as powders (not pellets), having bulk densities of 0.4–0.6 g/cc. As noted in the previous section, this low bulk density largely impacts the maximum feed rate achieved in the extruder due to flooding of the feed port. Increasing screw speed (in some cases) [22], employing a side stuffer, or increasing the free volume of the screws can enable higher feed rates. One way to achieve a higher free volume in the feed section is to use undercut, square channel screws in the feed zone [23]. While these are not self-wiping, this may not be a concern since there is no molten material in the feed section, and this would not extend the heat history of the material.

A further challenge in feeding low bulk density powders involves the entrained air. Low bulk density feed materials undergo significant densification during extrusion because up to 80% of powder feed streams are entrained air. Effective venting mechanisms are essential to remove this air and maximize throughput [22]. In addition to venting air, moisture removed from the feedstock by heating in the extruder may condense on colder upstream barrel sections or feed powder in the feed zones [20]. Given these two considerations, proper venting in the feed section is critical.

19.2.3.2 Melting and Mixing Achieving a melt is generally accomplished through the input of energy by the extruder into the formulation. A generic free energy diagram (Figure 19.8) illustrates this principle. The formulation components, whether crystalline API or amorphous polymer, undergo a transformation to a more mobile and deformable state at higher temperatures. The melt extrusion process enables energy input through both thermal and mechanical means. Thermal energy input is typically achieved through electric or oil heating of the barrel, which is transferred to

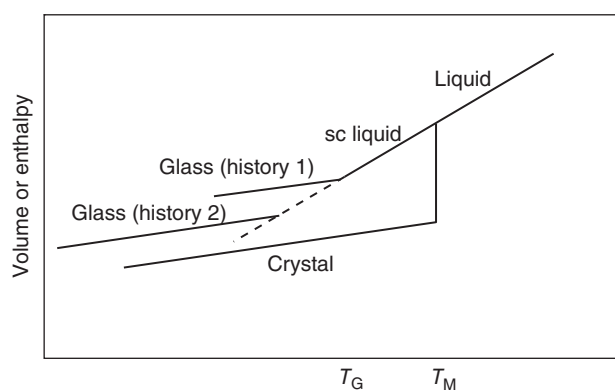


FIGURE 19.8 Behavior of glass-forming single component material.

the formulation via conduction. This method can be efficient in small-scale extruders, though conductive heating alone is generally not a sufficient source of energy to achieve a melt due to the poor thermal diffusivity of polymers [24]. Relying heavily on conductive heating also poses scale-up challenges since heat transfer is a function of surface area, while scale-up to preserve key material properties generally occurs on a volumetric basis. Moreover, greater product temperature uniformity could be expected when minimal conductive heat is transferred from the barrel [25].

While conductive heat can contribute to the total energy input at smaller scales, melting is largely achieved by viscous dissipation via frictional forces (including interparticle, material/wall, and material/screw friction) [24, 26]. Some estimates suggest at least 80–90% of the energy to achieve a melt is supplied by the extruder's screws [27]. This energy is generally referred to as the specific energy, which is the ratio of mechanical energy (as measured by a wattmeter on the drive motor) to feed rate in units of kW·h/kg. It is important to note that the wattmeter reading does not account for losses to the thrust bearing (which can be significant if high pressures are generated in the extrusion process) or drive motor efficiency at the given rpm, though these adjustments can be made. Mechanical energy input is achieved through deliberate design of the extruder screws to impart mechanical stress on the formulation, with specific energy often serving as a target to ensure process consistency upon scale-up [28].

Extruder screws are generally modular and consequently allow for a number of different configurations, which have a direct impact on the specific energy, residence time distribution, and maximum shear stress, among other process responses. The same formulation extruded under different conditions, therefore, may exhibit disparate levels of quality [29].

Pharmaceutical extrusion often features the use of double-flighted screws, creating three distinctly separate channels down the length of the extruder barrel. A configuration using only conveying elements would largely move material through the extruder in plug flow with minimal back mixing or material transfer between these three channels. The only mixing achieved in a pure conveying system would be laminar in nature and would potentially be less than expected due to viscous polymers not following no-slip boundary layer conditions [30]. As such, under these conditions there is almost no high-frequency disturbance dampening, potentially leading to poor compositional uniformity. Compositional uniformity requires back mixing and is achieved via screws designed to allow pressure flow to cause material movement between the channels. Here the screw flights are opened, often taking the form of mixing blocks. Figure 19.9 attempts to show conceptually how conveying elements are discretized in a way that allows exchange of material via mixing blocks. While mixing blocks are the conventional means to achieve a melt and sufficient back mixing,

other routes include blister rings [24], gear or turbine type elements [31], or simply staggering or offsetting conveying elements [20].

Mixing screw elements have been designed to ensure compositional uniformity. The axial width of a mixing paddle is an indicator of the magnitude of global pool capture (a region of the screw channel known for its high shear) and determines the extent to which mechanical energy will be imparted by a single mixing element. In plastics extrusion, wide mixing paddles reduce particle size through attrition (dispersive mixing), whereas narrow paddles and similar lower energy elements are utilized primarily to achieve compositional and thermal homogeneity (distributive mixing) [32]. Formulations consisting of miscible components benefit from both dispersive and distributive mixings to impart

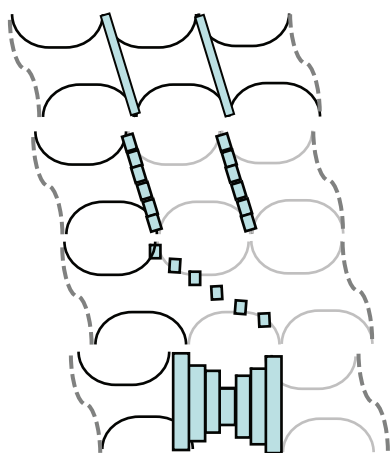


FIGURE 19.9 Illustration of the transition from conveying elements with a closed channel (top) to mixing elements with an open channel (bottom) to produce back mixing.

energy for melting and create surface area for diffusion, thereby yielding uniformity at a molecular level [33, 34]. The number of mixing cycles (cycles of volume expansion and compression during screw rotation) may also serve as a measure of mixing intensity. This mechanism is analogous to kneading pizza dough as a method to incorporate and mix ingredients (Figure 19.10) [20, 32].

19.2.3.3 Melt Conveying and Venting Downstream of the mixing paddles, the aggregate material is in a molten state and any additional conveying before the die continues to constitute a formulation’s time at temperature, which is the most relevant component of the residence time distribution as it relates to product quality (i.e. degradation vs. homogeneity). See Figure 19.11 for a list of process input parameters related to this stage. The conveying screws are generally less than completely filled, and residence time is a function of screw speed. This is in contrast to residence time in full barrel sections (e.g. reverse conveying elements, neutral mixing sections) where residence time is a function of throughput, with minimal contributions from screw speed [23].

Venting is an important aspect of the melt conveying step. The melting of API and excipients liberates solvents (e.g. water), and without proper venting, even small quantities of water or other volatiles could result in bubbling/foaming as the melt exits the die and/or result in undesirable residual solvents in the extrudate that could impact product properties. Devolatilization is generally a mass transfer limited process, as volatile materials must diffuse through the melt. Key influencing parameters include temperature (with elevated temperatures increasing the diffusion coefficients), feed rates (which impact residence time), and the screw profile. The screw profile affects both the residence time in venting sections and diffusion distances. Screw sections under the vent

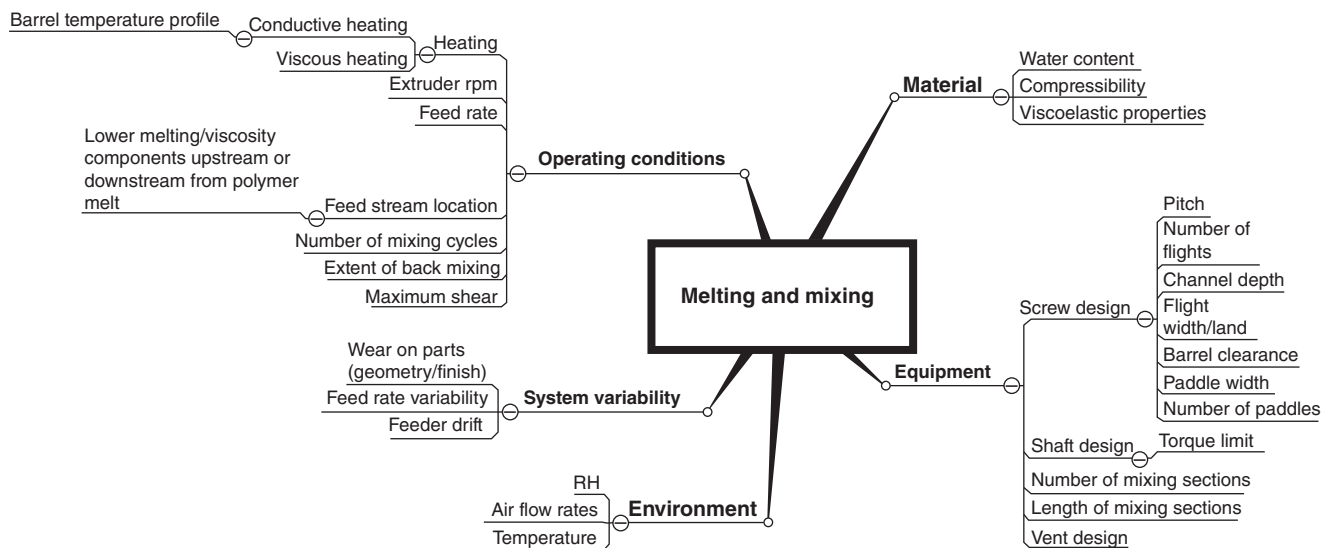


FIGURE 19.10 A summary of input parameters of the melting and mixing sub-operations.

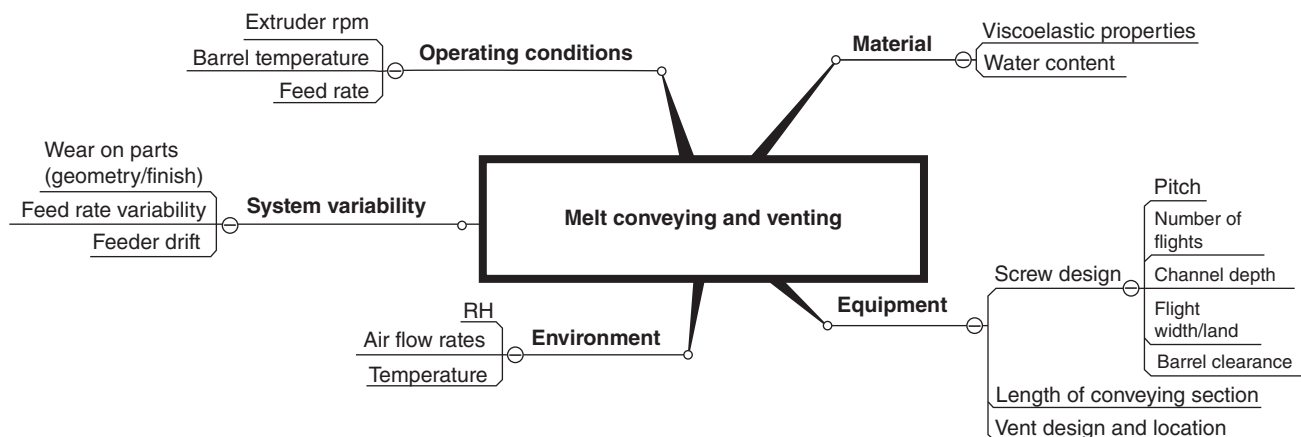


FIGURE 19.11 A summary of input parameters of the melt conveying and venting.

are typically designed to minimize local degree of fill using multi-flight large pitch screws to maximize surface area and minimize diffusion distance [27]. Screw profiles can also be designed to enhance the surface renewal phenomena to further reduce the diffusion distances.

The extrusion of formulations containing API solvates is an intriguing application of venting to enable what has not been possible with conventional pharmaceutical processing. A given drug molecule can exist as any number of solvates, depending on its chemical synthesis route. Safety considerations prohibit the vast majority of solvates from moving into drug development despite their sometimes favorable physical properties (i.e. improved flow, enabling better control of feeding to the extruder, and improved content uniformity of the extrudate). Using the melt extrusion process, an API solvate may be fed into the extruder, with the solvent removed upon melting. The final drug product should contain low levels of residual solvent, meeting ICH [35] specifications.

19.2.3.4 Pumping/Shaping and Cooling The stage of extrusion most closely associated with its namesake is the pumping of molten extrudate through a die. Die geometry may play a role in the final product, such as in the production of transdermal films, which would require a slit die. The molten extrudate may also be processed downstream via conventional unit operations (i.e. milling and compression), and in this case precise die geometry is not critical. A common shape for pharmaceutical extrusion application is a circular die, with multiple strands facilitating more rapid quench cooling. In some cases, particularly for multiparticulate controlled release applications, the extrudate strand may be passed through a die face cutter, resulting in small pellets [36]. Here strand diameter plays a more central role in product quality and is determined by the die geometry, viscoelastic properties of the formulation, process conditions (e.g. pressure), and material flow rates.

The extruder die can be roughly characterized as a resistance to material flow. Sufficient pressure is generated in the screw channels prior to the die to overcome this resistance. The pressure resistance can again be considered a function of the die geometry and the melt's viscoelastic properties. Significant pressure increases near the die may result in melt temperature increases due to viscous dissipation [37], resulting in the product's maximum melt temperature. For inviscid materials, additional heating during pumping out of the die may not be significant. In this case, the maximum product temperature may be achieved upstream of the die at some point after achieving a melt.

The extrudate can take many paths to a finished pharmaceutical dosage form following extrusion. The formulation is often quenched using such methods as passing along a conveyor belt with compressed air or feeding through chilled stainless steel rolls. Once cooled, the extrudate may be sized using conventional pharmaceutical mills and then compressed into tablets or filled into capsules. Alternative options for manufacturing finished pharmaceutical dosage forms include direct shaping methods of the extrudate. Directly formed tablets may be created by calendaring [38] or injection molding [39]. Figure 19.12 shows an example of a directly shaped dosage form. Molding enables production of complex shapes with features like embossing to improve patient compliance, enable branding opportunities, and prevent counterfeiting. Molding also has the potential to reducing the need for fillers and compression aids. See Figure 19.13 for a list of process input parameters related to this stage.

19.3 RISK ASSESSMENT

19.3.1 Quality Attribute Definition

Multiple processes and process conditions can produce material with similar *in vivo* responses (as demonstrated in

Figure 19.3), provided the process achieves a true solution. In the case of extrusion, heat generally drives miscibility unless the enthalpy of mixing is very unfavorable at high temperatures. Miscible formulations produced by extrusion will be homogeneous provided sufficient mixing, time, and heat.

While the bulk of extrudate properties can be explained by the heat history and energy input, relaxation state and particle physical attributes (e.g. particle size) may also impact



FIGURE 19.12 Example of a directly shaped tablet following a melt extrusion process illustrating the ability to form complex shapes, such as the Merck corporate logo (image weight ca. 200 mg).

formulation performance. The relaxation state can be varied systematically by changing quench rate. Figure 19.8 is a common illustration of how different quench rates of a single component system can lead to materials of different relaxation states and properties.

HME processing conditions are generally constrained at one end by the maximum throughput. This operating limit is usually characterized by low specific energy, short residence time, low product temperature, and/or limited mixing (Figure 19.14). An extreme case for a low energy limit would be to employ conveying elements, no barrel heat and no die restriction [40]. With more conventional compounding screw

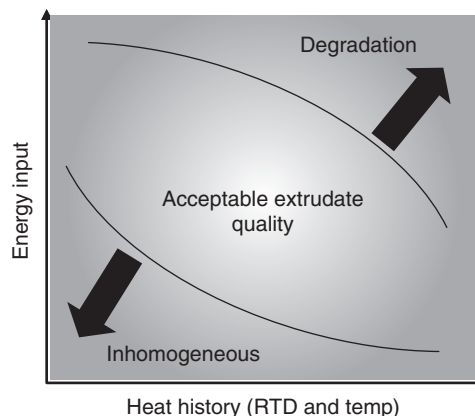


FIGURE 19.14 A schematic showing the balance of energy input and heat history, with the extreme combinations of these resulting in either degradation or inhomogeneous material.

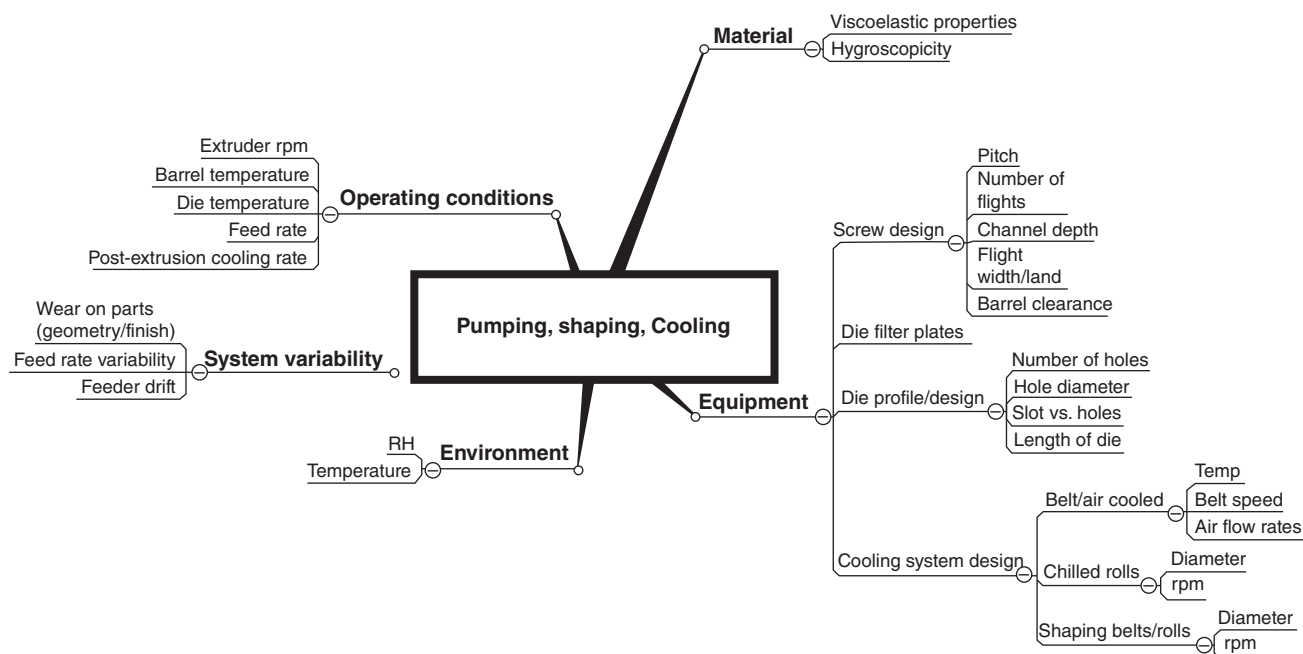


FIGURE 19.13 A summary of input parameters of melt pumping, shaping, and cooling.

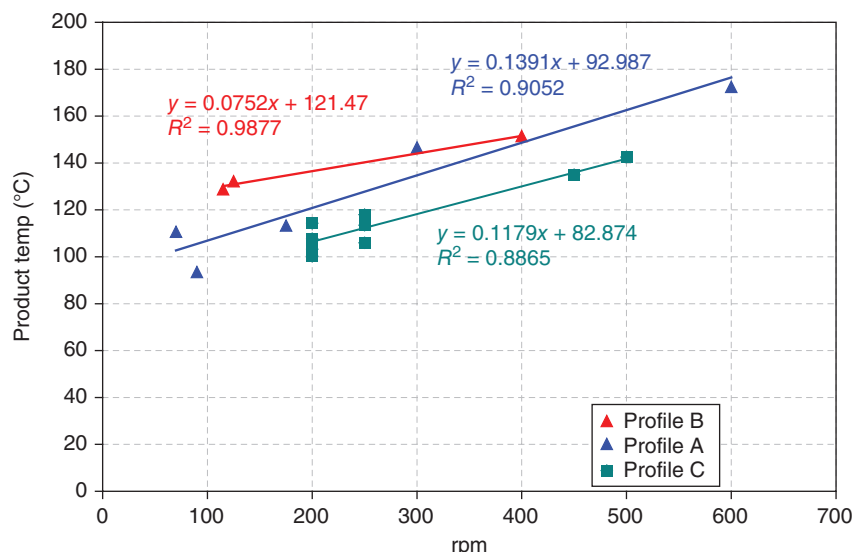


FIGURE 19.15 Impact of RPM and screw profile on the product temperature. Screw profiles varied from modest to aggressive: B → C → A.

profiles, it is possible that the low energy limit resulting in unacceptable product quality will lie outside of the accessible operating space (e.g. a torque limit may be encountered before inhomogeneous product is produced).

The upper end of HME processing conditions as shown in Figure 19.15 may be constrained by thermal degradation. Efforts to circumvent thermal degradation in HME include the use of plasticizers with the goal to process at low barrel temperatures and/or with a less aggressive screw profile [41]. Incorporating antioxidants during extrusion [42] and nitrogen blanketing can effectively stabilize oxidatively sensitive drugs. Elucidating degradation mechanisms for the polymers and developing analytical characterization methodologies is more complex compared with thermal degradation of small drug molecules for which stability indication assay method and LC-MS are commonly used.

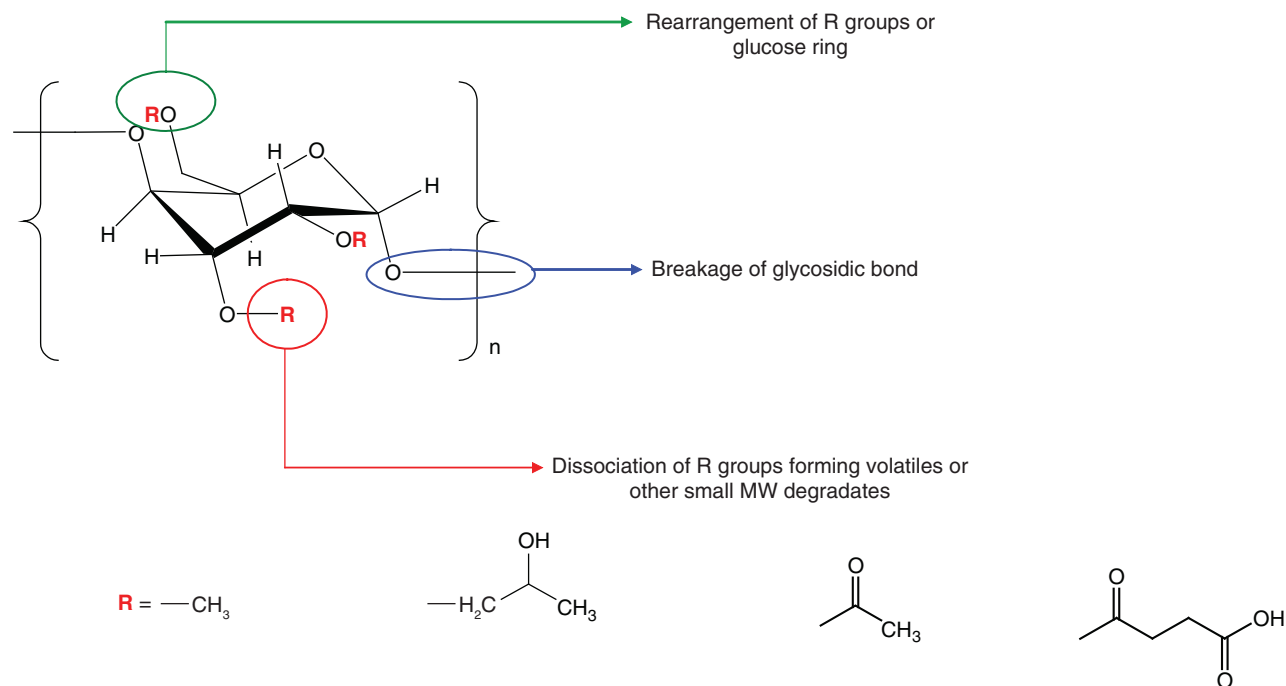
The implication of polymer degradation is twofold: (i) potential safety concerns of the degradation products at levels exceeding the ICH qualification threshold and (ii) potential impact on polymer functionality (e.g. stabilizing high energy amorphous drug in solid state and/or enhancing solubilization of poorly soluble compounds during *in vitro* drug release).

Hydroxypropylmethylcellulose acetate succinate (HPMC-AS) is a widely used polymer for making solid solutions owing to its unique physical and chemical properties including high T_G , ample solubility in organic solvents, and potential for diversified molecular interactions with a broad range of poorly soluble drugs. HPMC-AS is a linear polymer consisting of $\beta(1\rightarrow4)$ linked substituted D-glucose units. Three possible reactions could occur after thermal stress: (i) the dissociation of substitution groups, (ii) the breakage of glycosidic linkages, and (iii) rearrangement of the polymer backbone via intramolecular reactions (Scheme 19.1). A detailed understanding in

degradation chemistry involves polymer characterization using an array of analytical techniques such as TGA, GC-MS, size exclusion chromatography (SEC), NMR, IR, or mass spectrometry.

TGA is an effective tool for monitoring polymer breakdown products as a function of temperature. It has been routinely used to derive activation energy associated with polymer decomposition [43]. TGA analysis of HPMC-AS showed a modest volatile formation (2%) at temperatures ranging from 175 to 250 °C (linear ramp in 5 minutes). Headspace GC-MS analysis showed the volatiles mostly consist of CO₂, formic acid, acetic acid, succinic acid, and other unidentified small molecular species. The volatiles are likely the by-products of dissociation of side chains from the polymer backbone. The TGA/GC-MS results are consistent with those from variable temperature IR, where an increase in OH signal was observed, indicative of the loss of R groups. At representative HME processing temperature (e.g. <230 °C), the loss of R groups is believed to dominate the reaction pathway, leading to polymer degradation. TGA was also conducted for a series of other polymers including HPMC-phthalate, HPMC-trimellitate polyvinyl pyrrolidone, and polyvinyl pyrrolidone-polyvinyl acetate copolymer (PVP-PVAc). The polymer decomposition rate appears to increase in the order of PVP<PVP-PVAc<HPMC-AS<HPMC-T and HPMC-P.

Thermally induced glycosidic bond breakage could lead to depolymerization or cross-linking of the polymer, which could potentially compromise polymer functionality. SEC results of HPMC-AS processed at several extrusion conditions suggest that polymer breakdown or cross-linking is not likely to occur at typical operating temperatures (e.g. 170–230 °C). This is consistent with biorelevant dissolution of HME extrudate samples in which typical process



SCHEME 19.1

conditions had minimal impact on dissolution rate and apparent drug concentration. Under extreme processing conditions (e.g. 220 °C barrel temperature coupled with an aggressive screw profile), polymer cross-linking is evident, as shown by a significant drop in polymer aqueous solubility. The nature of the cross-linked polymer is not clear and remains the subject of further investigation. The polymer cross-linking could significantly compromise *in vitro* performance of solid dispersions due to the entrapment of drug molecules by cross-linked polymer networks.

19.3.2 Risk Assessment Tool

The foundation of risk assessment began with existing knowledge of how the extrusion process conditions would likely influence product quality. The quality attributes of the extrudate include achieving a molecularly dispersed glass having micro- and macroscale compositional uniformity and no degradate products. The CPPs to achieve this molecular dispersion include delivering appropriate thermal driving force without degrading any of the components and coupling this thermal driving force with sufficient time and adequate surface area for molecular diffusion to occur. Ensuring macroscale compositional uniformity required process conditions optimized to minimize sensitivity to feed rate disturbances.

A series of scale-independent parameters were understood to influence each CQA. Formulation-dependent parameters such as viscosity and glass transition temperature are also important; however due to the limited ability to adjust these,

their discussion is not included in detail here. Thermal driving force was influenced by the specific energy, maximum product temperature, and product temperature distributions. Sufficient time at temperature was influenced by the residence time and cooling rate. The surface area for diffusion was influenced by effective expansion/compression mixing cycles, mixing intensity, and the shear stress profile. The process robustness, or the ability to dampen feed input perturbations, was influenced by the extent of back mixing and the residence time distribution.

The risk assessment efforts sought to prioritize process parameters and identify those that should be carefully evaluated. There are complex interactions between many of the scale-dependent parameters, scale-independent parameters, and quality attributes. The QFD tool was thought to better address these complexities compared with more traditional, linear risk assessment tools like fault trees or failure mode effects and analysis.

19.4 ACHIEVING A DESIGN SPACE

Initial efforts focused on determining how scale-independent parameters impact quality attributes. Integral to this was identifying characterization methods indicative of patient-relevant quality attributes. A battery of solid state and thermal characterization methods coupled with a DFSS-driven measurement system analysis were used to identify a technique believed to be a good indicator of molecularly dispersed extrudate.

A quantitative assessment of the correlation between scale-independent parameters and quality attributes was evaluated within the framework of a DOE, with the scale-independent parameters set as the design factors. Constitutive equations and commercially available software based on one-dimensional solutions to heat, mass, and momentum balances were an integral part of this design. These models were used to determine how to adjust scale-independent parameters (screw profile, rpm, throughput, barrel temp, etc.) to achieve low and high factor level set points for scale-independent parameters (residence time distribution, specific energy, shear stress, mixing, etc.). Analysis of the result focused on quantitatively linking variations in the quality attributes to the scale-independent parameters.

While scale-up of extrusion processes is well understood [23, 24, 27] and achieving comparable product quality attributes upon scale-up is certainly feasible, it is unrealistic to expect all scale-independent parameters will be equivalent upon scale-up. For this reason, it was particularly important to have a deep understanding of which scale-independent parameters most influence quality attributes and where to focus scale-up efforts.

Results from these experiments indeed suggested that variations in the key measurements of quality could be attributed to a single scale-independent parameter. This greatly simplified which scaling rules to apply, how to approach scale-up, and how to define scale-up success.

Initial experiments also highlighted that while constitutive equations and the modeling software provided a reasonable first approximation, they did not sufficiently explain the relationship between scale-dependent parameters and scale-independent parameters. Statistically designed experiments were used generate empirical models for this purpose.

Figure 19.15 illustrates one case of observed multifactor interactions between dependent variables used to describe independent variables. During these experiments, feed rate, barrel temperature, screw speed, and screw profile were manipulated to achieve targeted responses in the scale-independent parameters. Figure 19.15 shows how the response for the product temperature (scale-independent parameter) can be largely explained just by the screw profile and screw speed (scale dependent). Attempts were made to generate this correlation in one comprehensive empirical equation via DOE; however, the screw profile had too large of an impact and required a multiple linear regression approach to effectively describe the observed behavior. The product temperature dependence is a stronger function of screw speed as the aggressiveness of the screw profile is increased (Figure 19.15).

Further response surface mapping experiments sought to capture factor interactions and curvature. Figure 19.16 shows contour lines for the mean residence time as a function of barrel temperature and feed rate. This illustrates the nature of the barrel temperature–feed rate interaction and that a quadratic model was necessary. Here, lower feed rates resulted in

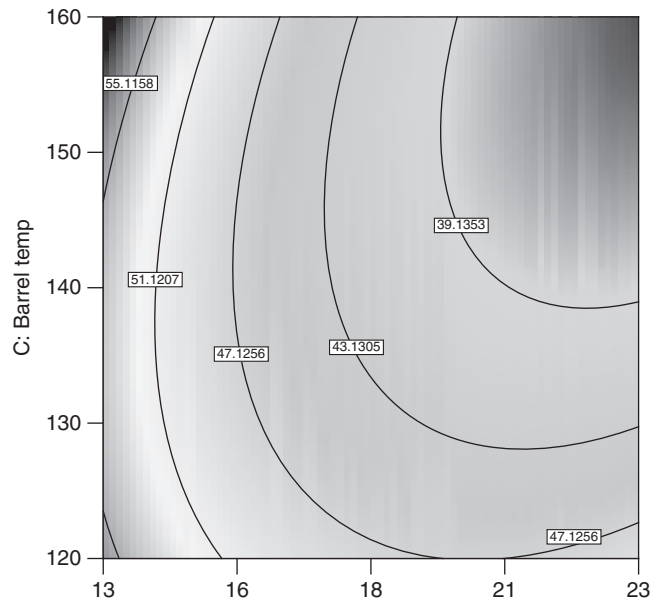


FIGURE 19.16 Impact of feed rate (kg/h) and barrel temperature (°C) on mean residence time.

longer mean residence times. This is expected since lower feed rates would equate to more back mixing per unit mass. Longer residence times were observed at higher feed rates when the barrel temperature was decreased from 160 to 120 °C. This could be due to the impact of lower barrel temperatures, achieving lower product temperatures and hence higher viscosity. This would increase backup lengths in filled sections, effectively extending residence time. Similar figures and empirical models were generated for the other key scale-independent parameters.

Understanding curvature and parameter interactions enabled a multifactor optimization of processing conditions. The resulting equations also made a detailed definition of the design space possible. The models provided insight into how quality attributes could be influenced in regions outside the space explored by the DOE. Several points were included outside the space of the DOE where a specific combination of operating conditions could achieve higher throughputs. This high throughput could not be universally achieved such that it could not have been a high factor setting in the DOE. Using the empirical model built via the DOE, the quality responses at these extreme points were predicted accurately.

19.5 PROCESS ROBUSTNESS

19.5.1 Introduction

The extrusion process is the integration of the mass feeding systems with the extruder itself, and when operated with multiple feed streams, an understanding of the overall system

dynamics is vital to ensuring product quality. The compositional homogeneity of the extrudate is dependent on performance and stability of each independent mass feeding system. Process upsets and perturbations as well as unstable mass feeder operation and degrading mass feeder performance threaten product quality.

In this work, PAT was used as a key enabler for holistic process understanding and quality control. An in-line transmission near-infrared measurement system was developed and implemented to monitor the composition of the multi-component melt stream exiting the extruder in real time. Real-time measurement of product composition during operation enables an operator to verify that the process is stable, identify process upsets, and ensure the targeted set point is realized. This PAT tool was also used to develop a dynamic model of the process to both pulsed and stepped inputs. The process model was then used to understand the disturbance dampening capabilities of the process and to inform performance requirements for the mass feeding systems.

19.5.2 System Identification

A series of system identification tests were conducted, which consisted of both pulsed [44] and step change inputs. Pulsed inputs were achieved by rapidly introducing pre-weighed amounts of API into the feed throat of the extruder. Step change inputs were achieved by changing the mass flow set points on the feeder controllers. The dynamic response of the process to the input signal was measured in each test case. The process responses to the compositional step change identification tests were fit to a first-order plus dead-time (FOPDT) model (Eq. 19.2), where $I(t)$ is the component concentration as a function of time, K is the process gain ($K = 1$, mass flow changes are fully realized), Δx is the magnitude and the direction of the step change, $u(t - t_0)$ is the unit step change with dead time, t is time, t_0 is dead time, τ is the process time constant, and $I(0)$ is the concentration of the component before the step change. More sophisticated models were considered [19, 45], but in this case the FOPDT model was found to adequately describe the data. The best fit global FOPDT model parameters are shown in Table 19.1. The process time constants are similar for all three components in the formulation. This is consistent with expectations, because the extruder acts as a CSTR-PFR model in the sense that all mass elements will experience the same environment as they pass through the extruder. The processing environment is fixed by the extruder process parameters (screw profile, barrel temperature profile, mass flow rate, screw speed) and the formulation rheology. The process dead times are also similar, but their variation is consistent with the location in which they are introduced into the process. Example model fits are shown in Figure 19.17 (panels a, b, and c are the responses for each formulation component). The high speed of data acquisition of the spectrometer (one prediction every

TABLE 19.1 First-Order Plus Dead-Time Model Parameters

Model Parameter	API	Surfactant	Polymer
τ (time constant second ⁻¹)	11.1	10.5	10.3
t_0 (dead time second)	59.4	66.2	62.2

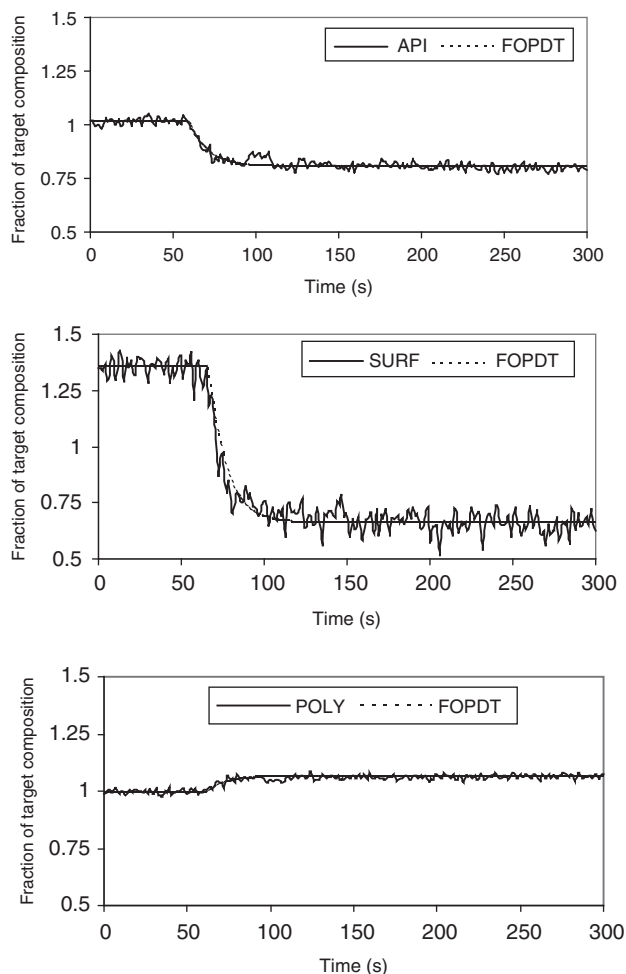


FIGURE 19.17 Example FOPDT Model fits of the extruder process response to simultaneous mass feed rate step changes in API (upper), Surfactant (middle), and Polymer (lower) as measured by in-line transmission NIR.

1.31 seconds) was well suited for model parameter identification, in particular the process dead time:

$$I(t) = K \cdot \Delta x \cdot u(t - t_0) \cdot (1 - \exp(-(t - t_0)/\tau)) + I(0) \quad (19.2)$$

19.5.3 Process Disturbance Rejection Capability

Disturbance rejection capability of the process can be assessed by calculating the periodogram of the time derivative of the identified process model. The periodogram [46]

is defined as the absolute value of the square of the finite Fourier transform (FFT) versus a frequency vector as described by Equation 19.3, where U is the FFT of the time series $u(t)$, N is the number of elements in the time series, ω is the frequency in inverse time, i is the square root of negative one, and t is time:

$$|U_N(\omega)|^2 = \left| \frac{1}{\sqrt{N}} \sum_{t=1}^N u(t) e^{-i\omega t} \right|^2 \quad (19.3)$$

The time derivative of the identified process model is an estimate of the finite impulse response function for the process. The periodogram that is a plot of signal power vs. frequency is the frequency response function of the process. Figure 19.18 shows the periodogram of the process model. This plot describes the extruder's ability to dampen mass feeder input disturbances and shows that the extruder acts as a low-pass filter. Specifically, input signals that contain energy content at frequencies above 0.05 cycles/s will be nearly entirely damped. Conversely, input signals that contain energy content at frequencies below 0.05 cycles/s will pass through the extruder virtually undamped and will affect product uniformity. Additionally, the amplitude of the instantaneous

variability in mass flow rate is essentially irrelevant as long as the variability frequency is above the critical frequency (ca. 0.05 cycles/s in this work) and that the mean flow rate is on target. This knowledge can be used to either set mass feeder performance specifications or could be used to redesign the extrusion process (screw profile, or screw speed, total mass flow rate) to be compatible with known mass feeder performance. The mass feeder performance is highly dependent on the feed material properties that could vary from lot-to-lot including particle size, Carr index, and compactibility. A possible control strategy for monitoring mass feeder performance would be numerically calculating the periodogram from the real-time mass flow rate data (from loss on weight) over a moving window (one minute for example) that would trigger an alarm when the time series has frequency contributions below the critical frequency or a defined threshold value.

The process model can also be used for real-time predictive process monitoring. This could be achieved by numerical convolution of the differential form of the process model with the time-discretized mass flow rate data coming from the mass feeders calculated over a moving window. This enables predicting the extruder outlet composition one mean residence time in the future and triggering quality control actions in a feed-forward manner.

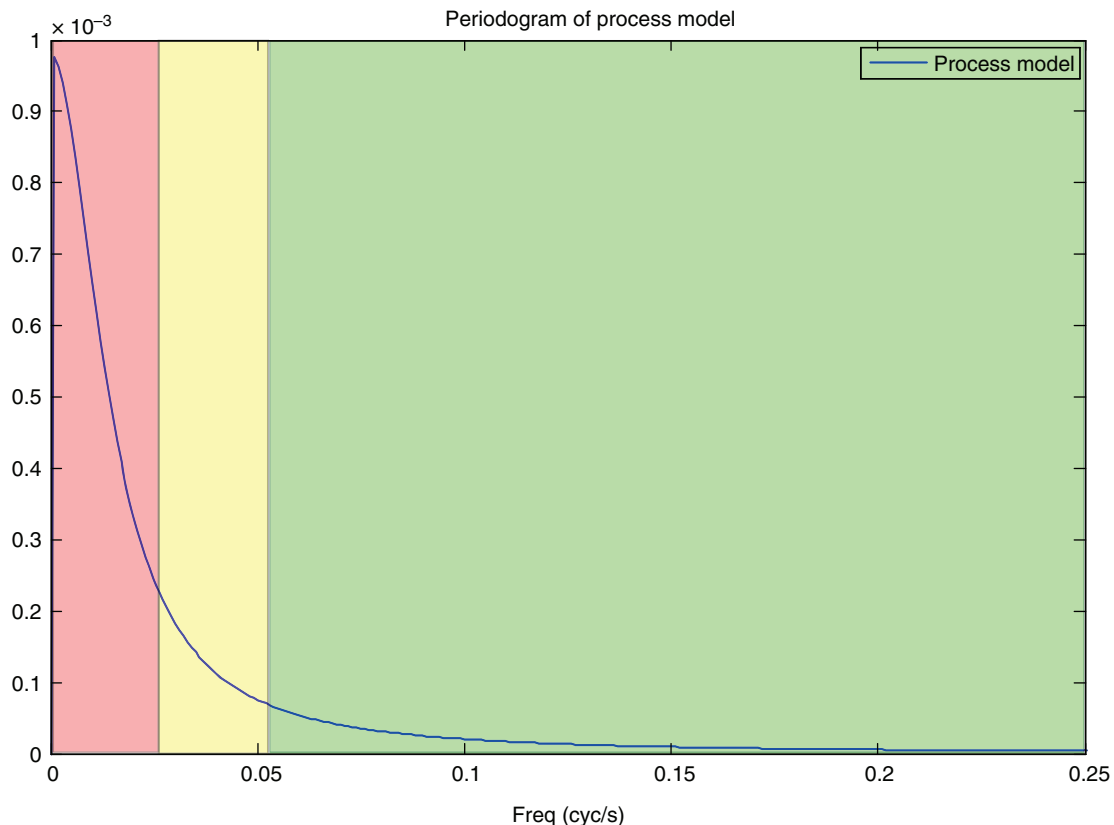


FIGURE 19.18 Periodogram of the process model showing that the process acts as a low-pass filter capable of the rejecting high frequency noise, but that input disturbances with a frequency of 0.05 cycles/s and lower will pass through the process potentially impacting product quality.

19.5.4 The Future of Quality Control and Process Understanding for HME

The future of quality control and process understanding is the integration of PAT tools and model-based process knowledge. Large progress toward process understanding can be made during development, but additional information required to increase robustness and optimize the process can only be addressed in the supply phase. For example, during the product and process development cycle, it is not practical to study every factor with DOEs. Additionally, the full range of raw material variability has not been experienced by the process, as development typically involves only a few lots of API. Lastly, unmeasured disturbances can pose a threat to product quality. PAT tools that include process sensors to measure physical and chemical information in real time together with the implementation of multivariate data analysis techniques can be used to enhance process understanding.

Multivariate statistical process control (MSPC) provides an efficient way of reducing all of the real-time process data streams into one convenient control chart that also captures all of the variable interactions as well. MSPC models the covariance pattern of the real-time process data and signals an alarm when the base covariance pattern is broken. When faults and deviations are detected, the variables causing the breakdown in the covariance pattern are identified. The implementation

of MSPC to a manufacturing line can improve process robustness in several ways. MSPC can detect both sensor faults and deviations from normal process behavior. Early detection of sensor faults and process abnormalities improves process robustness by enabling intervention and rapid root cause identification. The extrusion process contains a multitude of sensors that could potentially fail, and monitoring each sensor individually is impractical, making it a good fit with MSPC.

Holistic process analysis consists of aggregating all of the data sources related to the process and the product and analyzing the data with multivariate tools. This type of data aggregation and analysis will detect any existing correlations between process inputs such as raw materials or equipment used and process outputs like product quality or process performance. This type of analysis is also capable of detecting process drifts and differences on a batch-to-batch and a campaign-to-campaign basis. To facilitate holistic process analysis, it is necessary to have an IT infrastructure capable of aggregating raw material data, real-time process data from multiple unit operations and process trains, in-process testing, PAT data, equipment status and system suitability data, and quality testing and product release data. These IT systems need to be able to trace material and product genealogy as well. Figure 19.19 shows an overview of the holistic analysis process, and Figure 19.20 shows a process flow diagram for

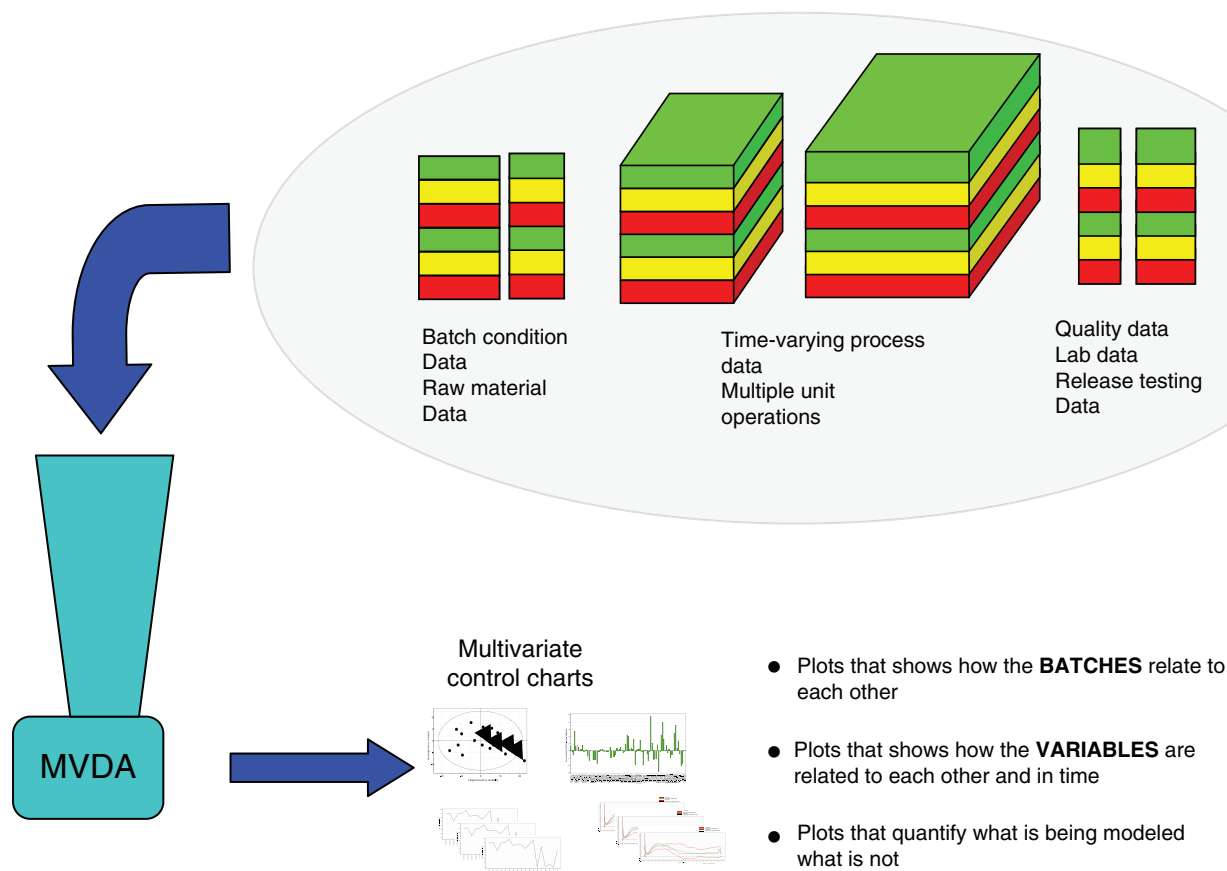


FIGURE 19.19 Overview of applying multivariate data analysis for holistic process analysis. MVDA techniques enable fast and efficient analysis of large and complex data sets.

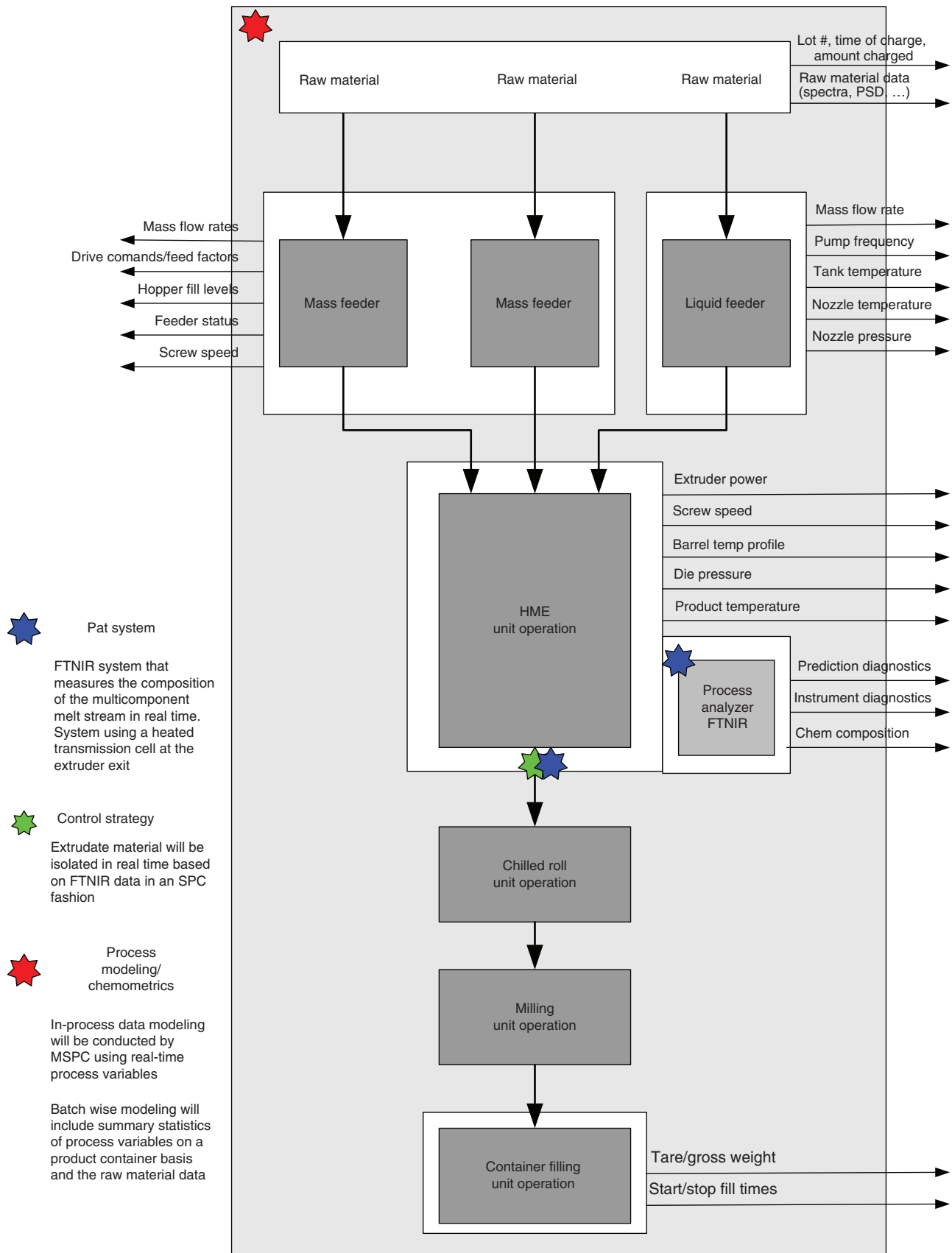


FIGURE 19.20 Material and information process flow diagram for an HME process train that depicts the use of a PAT analyzer to trigger an automated waste diverter and the implementation of multivariate statistical process control for fault detection and isolation as well as holistic process analysis that compares process performance on a batch-to-batch or campaign-to-campaign level.

an extrusion intermediate production process. The process depicted contains (i) a PAT system that measures the composition of the extrudate in real time and triggers a diverter system to isolate off-specification material, (ii) an MSPC system for sensor and process fault detection, and (iii) a PAT-IT system that aggregates raw material data, lot/batch genealogy, process data, and release data for post batch holistic multivariate analysis. The implementation of these tools in manufacturing facilitates expanding the process knowledge space by detecting process upsets and deviations in real time and providing that data and information to rapidly identify correlations between process inputs and process performance. A logical progression of process understanding is identification of correlations, establishment of causation, and model-based understanding. Ideally, model-based process understanding will be based on first principles, but empirical, hybrid, and statistically based models can often be sufficient for improving product quality and rejecting process disturbances. Identified process models can be used to develop feed-forward control systems that make processes robust to measured input variability.

This work describes the application of QbD principles and design space development to one unit operation in a pharmaceutical process train. A more complete process design space needs to be holistic, going from raw materials all the way to the final product image, with model-based process knowledge describing the relationships between process inputs, process set points, and final product attributes.

19.6 CONCLUSION

Extrusion is a pharmaceutical process technology that meets a growing need to enable oral delivery of insoluble candidates and is particularly well suited to the quality by design approach. The application of DFSS methodology helped manage the technical complexity of developing the hot melt extrusion process and deliver on QbD. The knowledge gained from this development exercise facilitated process optimization and clear definition of a robust design space. The integration of PAT enabled added process robustness, process understanding, and derivation of a multifactor quadratic model of the process parameters most responsible for influencing the CQAs. The future of pharmaceutical extrusion will include fully integrated process analysis and control beyond single system PAT approaches and the continuous delivery of final drug product via a small footprint seamless process.

ACKNOWLEDGMENTS

This work is the composite output of numerous functional areas and individuals within Merck. The authors would like

to thank our colleagues for their inspiration, managerial support, analytical efforts, technical input, hard work, and discussions on quality by design and DFSS methodology.

REFERENCES

1. Repka, M., Majumdar, S., Battu, S.K. et al. (2008). Application of hot-melt extrusion for drug delivery. *Expert Opin. Drug Deliv.* 5 (12): 1357–1376.
2. Lipinski, C. (2002). Poor aqueous solubility – an industry wide problem in drug discovery. *Am. Pharm. Rev.* 5: 82–85.
3. Benet, L.Z. and Wu, C.Y. (2006). *Using a Biopharmaceutics Drug Disposition Classification System to Predict Bioavailability and Elimination Characteristics of New Molecular Entities*. Somerset, NJ: NJDMDG.
4. Leuner, C. and Dressman, J. (2000). Improving drug solubility for oral delivery using solid dispersion. *Eur. J. Pharm. Biopharm. Sci.* 50 (1): 47–60.
5. Serajuddin, A.T.M. (1999). Solid dispersion of poorly water-soluble drugs: early promises, subsequent problems, and recent breakthroughs. *J. Pharm. Sci.* 88 (10): 1058–1066.
6. Chiou, W.L. and Riegelman, S. (1971). Pharmaceutical applications of solid dispersion systems. *J. Pharm. Sci.* 60 (9): 1281–1130.
7. Yu, L. (2001). Amorphous pharmaceutical solids: preparation, characterization and stabilization. *Adv. Drug Deliv. Rev.* 48 (1): 27–42.
8. Friesen, D.T., Shanker, R., Crew, M. et al. (2008). Hydroxypropyl methylcellulose acetate succinate-based spray-dried dispersions: an overview. *Mol. Pharm.* 5 (6): 1003–1019.
9. Curatolo, W., Nightingale, J.A., and Herbig, S.M. (2009). Utility of hydroxypropylmethylcellulose acetate succinate (HPMCAS) for initiation and maintenance of drug supersaturation in the GI milieu. *Pharm. Res.* 26 (6): 1419–1431.
10. Oh, D.M., Curl, R.L., and Amidon, G.L. (1993). Estimating the fraction of dose absorbed from suspensions of poorly soluble compounds in humans: a mathematical model. *Pharm. Res.* 10 (2): 264–270.
11. Moser, J.D., Broyles, J., Liu, L. et al. (2008). Enhancing bioavailability of poorly soluble drugs using spray dried solid dispersions part I. *Am. Pharm. Rev.* 11 (6): 68, 70–71, 73.
12. den Mooter, V., Weuts, I., De Ridder, T., et al. (2006). Evaluation of inutec SP1 as a new carrier in the formulation of solid dispersions for poorly soluble drugs. *Int. J. Pharm.* 316 (1–2): 1–6.
13. Janssens, S., de Armas, H.N., Roberts, C.J., den Mooter, V. (2008). Characterization of Ternary Solid Dispersions of Itraconazole, PEG 6000, and HPMC 2910. *J. Pharm. Sci.* 97 (6): 2110–2120.
14. Patterson, J.E. et al. (2007). Preparation of glass solutions of three poorly soluble drugs by spray drying, melt extrusion and ball milling. *J. Pharm. Sci.* 336 (1): 22–34.
15. Dong, Z. et al. (2008). Evaluation of solid state properties of solid dispersions prepared by hot-melt extrusion and solvent co-precipitation. *Int. J. Pharm.* 355 (1–2): 141–149.

16. Patterson, J.E. et al. (2008). Melt extrusion and spray drying of carbamazepine and dipyridamole with polyvinylpyrrolidone/vinyl acetate copolymers. *Drug Dev. Ind. Pharm.* 34 (1): 95–106.
17. Hauser, J.R. (1988). The house of quality. *Harv. Bus. Rev.* May–June: 3–13.
18. International Conference on Harmonization (2005). Q9: Quality Risk Management. https://www.ich.org/fileadmin/Public_Web_Site/ICH_Products/Guidelines/Quality/Q9/Step4/Q9_Guideline.pdf 26 Nov 2018.
19. Kim, E.K. and White, J.L. (2002). Transient compositional effects from feeders in a starved flow modular co-rotating twin-screw extruder. *Polym. Eng. Sci.* November: 2084–2093.
20. Todd, D. (1998). Practical aspects of processing in intermeshing twin screw extruders. *J. Reinf. Plast. Compos.* 17: 1607–1616.
21. Lee, S.H. (1997). Continuous mixing of low viscosity and high viscosity polymer melts in a modular co-rotating twin screw extruder. *Int. J. Polym. Proc.* 12: 316–322.
22. Ishibashi, J. and Kikutani, T. (2005). Experimental study of factors influencing throughput rate and process of polymer-mineral filler mixing in a twin screw extruder. *Int. Polym. Process.* 388–397.
23. Todd, D. (1998). *Plastics Compounding: Equipment & Processing*. Hanser.
24. Todd, D. (1993). Melting of plastics in kneading blocks. *Int. Polym. Process.* 8 (2): 113–118.
25. Tadmor, Z. and Klein, I. (1970). *Engineering Principles of Plasticating Extrusion*. Van Nostrand Reinhold.
26. Jung, J. and White, J.L. (2003). Investigation of melting phenomena in modular co-rotating twin screw extrusion. *Int. Polym. Process.*, XVIII 127–132.
27. Rauwendaal C., *Polymer Extrusion*, 4th Ed (2001) Hanser Gardner Publications, Inc.. 463–476.
28. Potluri, R., Todd, D., and Gogos, C. (2006). Mixing immiscible blends in an intermeshing counter-rotating twin screw extruder. *Adv. Polym. Tech.* 25 (2): 81–89.
29. Lim, S. and White, J.L. (1994). Flow mechanisms, material distributions and phase morphology development in a modular intermeshing counter-rotating twin screw extruder of Leistritz design. *Int. Polym. Process.* IX: 33–45.
30. Denn, M.M. (2009). Simulation of polymer melt processing. *AIChE J* 55: 1641–1647.
31. Brouwer, T., Todd, D., and Janssen, L.P.B. (2002). Flow characteristics of screws and special mixing enhancers in a co-rotating twin screw extruder. *Int. Polym. Process.* XVII: 26–32.
32. Tadmor Z., Gogos C., *Principles of Polymer Processing*, 2nd Ed. (2006) John Wiley & Sons, Inc.. 322–354.
33. Sekiguchi, K. and Obi, N. (1961). Studies on absorption of eutectic mixtures. *Chem. Pharm. Bull.* 9: 866–872.
34. Breitenbach, J. (2002). Melt extrusion: from process to drug delivery technology. *Eur. J. Pharm. Biopharm.* 54 (2): 107–117.
35. International Conference on Harmonization (2016). Q3C(R6): Impurities: Guidelines for Residual Solvents. https://www.ich.org/fileadmin/Public_Web_Site/ICH_Products/Guidelines/Quality/Q3C/Q3C__R6__Step_4.pdf
36. Young, C.R. et al. (2002). Production of spherical pellets by a hot-melt extrusion and spheronization process. *Int. J. Pharm.* 242: 87–92.
37. Rauwendaal, C., del, P., and Noriega, M. (2001). *Troubleshooting the Extrusion Process*, 67–70. Hanser Gardner Publications, Inc.
38. Chong, J. (1968). Calendering thermoplastic materials. *J. Appl. Polym. Sci.* 12 (1): 19.
39. Isayev, A. (1987). *Injection and Compression Molding Fundamentals*. CRC Press.
40. Nakamichi, K., Nakano, T., Yasuura, J. et al. (2002). The role of the kneading paddle and the effects of screw revolution speed and water content on the preparation of solid dispersions using a twin-screw extruder. *Int. J. Pharm.* 241: 203–211.
41. Verreck, G. et al. (2006). Hot stage extrusion of P-amino salicylic acid with EC using CO₂ as temporary plasticizer. *J. Supercrit. Fluids* 327: 45–50.
42. Munjal, M., Elsohly, M.A., and Repka, M.A. (2006). Polymeric systems for amorphous Δ^9 -tetrahydrocannabinol produced by hot-melt method. Part II: effect of oxidation mechanisms and chemical interactions on stability. *J. Pharm. Sci.* 95 (11): 2473–2485.
43. Villetti, M.A. et al. (2002). Thermal degradation of natural polymers. *J. Therm. Anal. Calorim.* 67: 295–303.
44. Chen, S.U., Nandakumar, K., and Wetzel, M.D. (2004). Investigation of the melting mechanism in an twin-screw extruder using a pulse method and online measurement. *Ind. Eng. Chem. Res.* 43: 6822–6683.
45. Puaux, J.P., Bozga, G., and Ainsler, A. (2000). Residence time distribution in an corotating twin-screw extruder. *Chem. Eng. Sci.* 55: 1641–1651.
46. Zhu, Y. (2001). *Multivariable System Identification for Process Control*. New York: Pergamon.

20

DRUG PRODUCT PROCESS MODELING

MARY T. AM ENDE,* WILLIAM KETTERHAGEN,** ANDREW PRPICH, AND PANKAJ DOSHI
Process Modeling & Engineering Technology Group, Pfizer Inc., Groton, CT, USA

SALVADOR GARCÍA-MUÑOZ
Small Molecule Design and Development, Eli Lilly and Company, Indianapolis, IN, USA

RAHUL BHARADWAJH
Engineering & Business Development, Engineering Simulation and Scientific Software (ESSS), Woburn, MA, USA

20.1 INTRODUCTION

The budget-restricted pharmaceutical environment is countered by the heightened expectations for drug products to be developed with more intensive use of science-based principles. The issues that arise during drug product development are often attributed to the impact of changing batch size and equipment type or scale on the formulation and process. In the commercial arena, one of the more prevalent causes for batch failures, or product not meeting specifications, is property shifts in the excipients or active pharmaceutical ingredient (API). During development and commercialization of a drug product, it is important to design a robust dosage form that is minimally sensitive to raw material variations and equipment scale.

The fundamental principles taught in the chemical engineering curriculum equip the chemical engineer with the skill base that allows them to address these complicated process operations and the understanding to connect the material properties to the processing equipment and design. Courses on the fundamental laws of heat transfer, mass transfer, momentum transfer, transport phenomena, and physical

chemistry allow the chemical engineer to mathematically describe the process. For example, an impeller used to mix API and excipients imparts energy and momentum on the material to achieve uniformity of the blend. In this chapter, three modeling examples of powder blending, lubrication, and powder discharging from a bin are monitored through computational methods to predict measurements of product quality. Utilizing the basics of the process calculation course enables students to break down the system into a control volume and solve mass and energy balances to determine the solution. An example of this type will be illustrated in this chapter during the derivation of the thermodynamic film coating model. Process control principles allow the chemical engineer to develop models to accommodate variations in the inlet stream properties and can adjust the process through feed-forward or feedback control to produce consistent quality product. In this chapter, this approach is used in reverse to set specifications on the raw material properties to ensure product quality using empirical models. The undergraduate curriculum, including many other courses not specifically highlighted here, provides a well-rounded understanding of how to approach solving process problems and how to break down a problem into its fundamental parts. In addition it provides the science-based hypothesis testing principles that are important to understanding the solution to the problem. Chemical engineers are skilled at writing in mathematical terms the driving forces affecting processes and are capable of modeling a process using first principles. They are able

*Current address: Lyndra Therapeutics, Watertown, MA, USA

**Current address: Drug Product Development, Research and Development, AbbVie Inc., North Chicago, Illinois 60064, USA
Sections 20.2, 20.3.2, and 20.3.3 are reproductions of the original book chapter [1].

to construct a control volume for engineering balance determination across inlet and outlet streams and use empirical methods, such as traditional regression polynomials, neural networks, or multivariate latent variable models (LVM), to understand complex processes. The two opposing external environmental factors of setting expectations to reduce costs from consumers and the heightened scientific expectations from the regulators have created a crucial opportunity for chemical engineering principles to be applied and implemented across the industry.

In this chapter, modeling techniques applied to drug product formulation and processing operations are discussed as support to the design, development, and scale-up for solid oral drug products. These process modeling techniques are discussed and exemplified with case studies ranging from raw material specifications to process parameter predictions. In general, the main unit operations utilized to produce tablets or capsules include (i) blending and other powder processing, (ii) dry or wet granulation, (iii) tablet compression (powder compaction) or encapsulation, (iv) lubrication final blending, and possibly (v) film coating. Specific consideration is given in this chapter to transfer and scale-up issues along with general process design-related challenges to pharmaceutical process R&D.

20.1.1 Benefits of Using Modeling Tools to Design, Develop, and Optimize Drug Products

Over the last decade pharmaceutical companies, in an effort to reduce costs, have embarked on bulk conserving methods for drug product design and development. These efforts have resulted in significant advances in process scale-up that utilize science of scale tools and predictive models. The major benefits of using modeling during development of pharmaceutical products have been previously highlighted by Wassgren and Curtis [2] who illustrated how employing reliable models can improve understanding of critical processes that may rapidly accelerate process improvements. An economic analysis of one specific engineering company utilizing computational fluid dynamics (CFD) modeling for a six-year period revealed a sixfold return on investment (ROI) [3].

A more comprehensive economic analysis was conducted by Louie et al. [4] for the modeling of API and materials science properties within the pharmaceutical development. Examples of the modeling capabilities considered in this analysis included API material properties (crystal morphology, surface area, powder X-ray diffraction patterns), solubility, polymorphism, breakage planes, and refractive index molecular and solvent interactions. The benefits considered in the Louie analysis [4] included (i) improved experimental effectiveness, (ii) broader/deeper understanding in the exploration of solution to a problem,

(iii) improved productivity by employing knowledge-based reasons for moving forward, (iv) reduced time to market for new products (IP/exclusivity), and (v) fewer unknowns with a reduced risk for failures.

The analysis indicated that the use of modeling and simulation tools in pharmaceutical development is producing an ROI of \$4–\$10 per dollar invested for an occasional user to a super-user, respectively. The greatest impact on ROI was found to originate from employing super-users (or subject matter experts) in materials sciences and API development. A similar analysis would be beneficial for the modeling of drug product processes during development and commercialization. The modeling capabilities available and applied to solid oral drug products consist of a balance between fundamental models, engineering-based models, and empirical models with intentional focus on applied use (Table 20.1). There appears to be a similar level of modeling and

TABLE 20.1 Comparison of Modeling and Simulation Capabilities for API Material Properties Versus Solid Oral Drug Product Properties and Processes

Materials Science Modeling and Simulation of API [4]	Solid Drug Product Modeling and Simulation
Structural properties (crystal morphologies, orientations, attachment energy, surface energy, PXRD, possible API crystal forms)	Raw material properties (particle size distribution, particle shape, density, material properties) [2, 5]
Physical properties (solubility, hydration, predicting preliminary physical data)	Blend properties (flow, velocity, segregation) [2, 6–11]
Molecular interactions (hydrogen bonding, solvent interactions)	Agglomeration properties (population balances of wet granulation) [12]
Purity (polymorphism, impurities, predicting stability of crystal forms)	Breakage properties (population balances of milling) [13]
Mechanical properties (shear strength, hardness)	Fluid dynamics (turbulent fluid flows in spray dryers, dry powder inhalers) [2] (agitated vessels, fluidized beds) [14–16]
Thermal properties	Solid mechanics (stress analysis and density distribution for tablet and tooling) [14, 17–20]
Optical properties (refractive index, spectral absorption, circular dichroism)	Mass and energy balances (film coating pan, fluid bed dryer) [21]
Electrical properties (conductivity, resistivity, dielectric behavior)	
Empirical models [22, 23]	

simulation capabilities for materials science of API and solid oral drug products; therefore the authors would anticipate the ROI for drug products to be positive also.

20.1.2 Summary of Modeling Approaches for Solid Oral Drug Products

Kremer and Hancock [14] were first to review modeling in the pharmaceutical industry and point out that process modeling can be considered as numerical simulations of the underlying physical processes. The process models in this category, which are based on first principles, can be expressed by the governing equations that are solved either analytically or numerically. The primary modeling examples presented in this chapter for the physics-based models include fluid dynamics, solids mechanics for tooling design, and particle-based models of powder discharging from hoppers.

Fluid dynamics is the study of flowing media such as gases, liquids, and certain types of solids such as dense, rapidly flowing powders. Pharmaceutical researchers have utilized commercially available CFD software packages to simulate a variety of applications including spray drying, inhalation, mixing in agitated vessels, and flow of granular material. The performance of several unit operations has been investigated and optimized using CFD.

Discrete element method (DEM) is a particle-scale modeling approach in which the motion and forces associated with each particle are tracked individually. Commercial and open-source DEM software packages are available. Each has a wide range of technical capabilities that are continually expanding, leading to higher fidelity simulations. Despite the technical advances, a specialist is often required to develop particle-based modeling using DEM. The primary disadvantage for DEM is the significant computational resources required to compute and track the wealth of particle-level information produced: particle velocities, forces, residence times, and stresses. The maximum number of particles, N , which can be modeled for reasonable simulation times, is typically on the order of $N \sim 10^6$. However, this guideline is routinely exceeded with the use of graphical processing units (GPUs) that can offer significant increases in computational speed relative to traditional central processing unit (CPU).

Engineering models can be considered a subclass of physics-based models because they are based on first principles; however they are applied to a defined control volume typically encompassing a unit operation. These types of models are often built upon mass, momentum, and/or energy balances across the control volume or derived from nondimensional analysis of the driving forces involved in a certain process. One example of the former (tablet film coating) and two examples of the latter approaches (wet granulation and fluid bed drying) will be provided in the subsequent sections of this chapter. The thermodynamic film coating model is

used to scale up the tablet film coating process based on matching exhaust air temperature and humidity as a representation of the tablets in the coating pan (environmental similarity). Nondimensional analysis (e.g. Froude or Reynolds number) can be employed to blending, milling, and wet granulation processes to examine the specific driving forces. In addition, momentum transfer of drying air to the wet granules can be analyzed for corresponding fluidization conditions based on the granule particle size as an applied engineering approach to predict acceptable drying air flow rates.

Tablet film coating models [21, 24–26] have been used for process scale-up based on environmental similarity in the coating pan and maintaining constant droplet size from the spray guns. The thermodynamic film coating model has been validated across lab to production scales and can be used to predict the temperature and relative humidity of outlet airstream, determine process set points based on desired exhaust air conditions, and minimize or eliminate the need for scale-up trials. The film coating atomization model is a validated theoretical model that describes the film coater atomizer performance and allows for the prediction of droplet diameter at the spray guns. The model requirements for the thermodynamic model include previous process data for the coating pan in question to generate the heat loss factor (HLF), and those for the atomization model include the nozzle specifications of the spray guns and rheological properties of coating solution (viscosity, surface tension, density).

Empirical modeling approaches are typically based on existing process data that can yield a set of parameters that can then be interpreted from a fundamental deterministic knowledge of the process. Multivariate latent variable modeling (LVM) is one specific empirical approach that has been proven to deliver a deep process understanding for unit operations that are particularly difficult to abstract in a first principles model and in situation where there is a wealth of data from the process under study [23].

The application of LVM can be better appreciated when it is mapped along the life cycle of a product/process. Although the methods can be widely applied, their application is limited by the availability of data. LVM can be applied at early stages of development, where materials (and ratios of) are being selected along with processing conditions [27]. These methods can also be applied during process design and scale-up [28] to minimize experimental work at the larger scale. For a commercial process, LVM can be used to troubleshoot/diagnose issues [29–31] and to optimize [32, 33] and to control [34–36] the operation. Other applications of LVM include the analysis of images [37, 38] and the establishment of multivariate specifications for incoming materials [39].

The first stage in the general strategy for these applications is to (i) fit a model to data that is considered relevant to the particular application (e.g. for a monitoring application, the data will correspond to normal operating conditions that will serve as a basis), and (ii) once the model is fitted and deemed

valid, the parameters of the model are interpreted (if possible) from a fundamental perspective, trying to associate each of the identified principal components (or latent variables) with a driving force acting upon a system.

A model that is considered valid and representative of a given system can be used in either passive or active mode. In passive mode, the model does not influence the process directly (e.g. troubleshooting or monitoring applications). In active mode, the model is actually influencing a decision either through a feedback control or by a design exercise.

Multivariate latent variable modeling is an alternative approach when fundamental modeling is not an option due to timing or investment constraints or lack of available fundamentals. The main difference between these types of latent variable models and other empirical approaches (such as neural networks or traditional ordinary least squares fitting) is the capabilities embedded in the method to handle massive amounts of incomplete and ill-conditioned data (which is common from an industrial process).

20.1.3 Overview of Modeling Approaches Mapped onto Unit Operations

In order to design dosage forms effectively, the engineer must be aware of the potential issues that might occur and the

driving forces associated with these issues. For solid oral dosage forms produced through a dry granulation process (Figure 20.1) [40], the potential issues that can occur during process scale-up can be considered from an engineering perspective. Blending operations are intended to evenly disperse API within the excipient powder by transferring mass and energy from the mixer or impeller to the powder. Between unit operations, there often exists a transfer step that can create an opportunity for the uniform blend to segregate resulting from mass transfer. Dry granulation and milling processing is intended to reduce this segregation potential by altering the particle size distributions (PSD) of active powder blend through a combination of mass and energy transfer. The goal of the tablet compression process is to produce tablets with the target product properties such as tablet weight, hardness, potency, and dissolution. This process can also be susceptible to tablet weight variations or powder segregation issues, all of which are related to the differential mass transfer, energy transfer, and momentum transfer of the materials. Film coating of tablets can exhibit issues of over-drying, over-wetting, or attrition that cause defects in the final product. These issues are related to their associated driving forces of energy transfer (high drying flow rates), mass transfer (accumulation of moisture and coating from high spray rates), and momentum transfer (impacting low-density

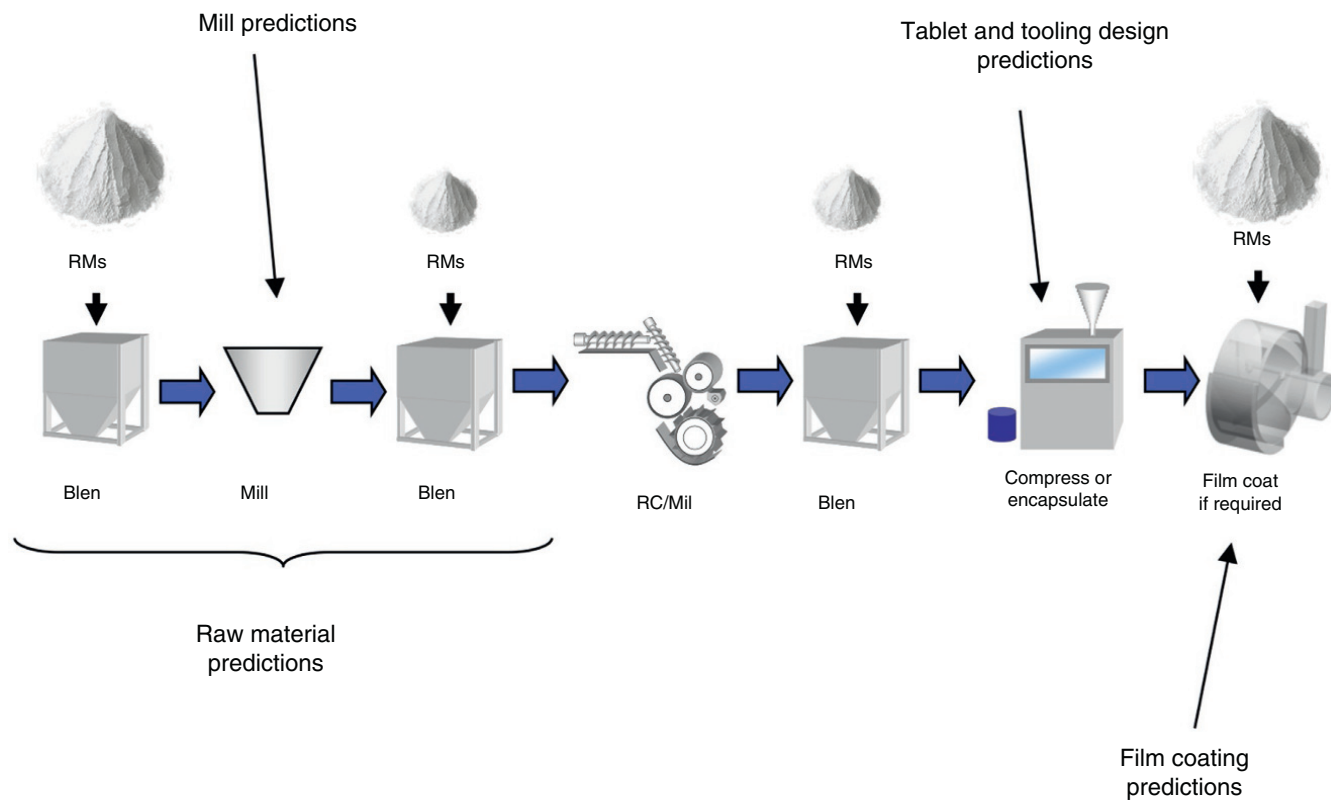


FIGURE 20.1 Modeling approaches mapped onto solid oral drug product processing. *Source:* Reprinted with permissions from am Ende et al. [40], Copyright (2009) John Wiley & Sons, Inc.

regions of tablet surfaces that attrite). The dosage form design criteria must ensure the stability, performance, and manufacturing capability of the final product. The source of the energy and/or momentum imparted onto the drug product formulation is through the processing equipment. Therefore, engineers can use energy/momentum transfer analysis as the design levers to adjust the product to the desired result in the quality attributes (e.g. blend and granulation content uniformity, tablet potency, dissolution, and stability profile).

20.2 FORMULATION MODELING

20.2.1 Empirical/Statistical Models for Raw Materials

Material properties are inherently variable, and therefore understanding the impact of these variations is an important factor in the formulation and process development of a drug product. A well-designed, robust product will be minimally sensitive to ingoing raw material changes. This section will address the relationship between the raw material variability and the drug product quality attributes, demonstrated both conceptually and through an example.

Quality should not be qualitative but quantitative and, even more, should be multidimensionally quantitative. As Duchesne and MacGregor write, "Quality is a multivariate property requiring the correct combination of all measured characteristics" [39]. So from this point and beyond, the concept of quality will not refer to a single numeric value representing a measured attribute of a product, but rather to a vector of multiple values (of multiple attributes of a product) that represent quality as a set.

Recent guidance documents from the Food and Drug Administration emphasize that quality should be built into, rather than tested on, the product [41–44]. This should be pursued from the design stages all the way to commercialization to varying degrees. Quality is defined on a case-specific basis and in principle should be guided by the ultimate effect on product safety and efficacy and performance [45]. This guidance is referred to as the "quality by design" paradigm and has established a science and risk-based approach to pharmaceutical development and commercialization.

A process-driven design exercise is executed downstream and in sequence. Such a scenario may start with the effect of the starting material onto the first unit operation and will continue downstream with emphasis remaining on the process. It is only at the end of the process (the design of the last few unit operations of the train) that the product quality will be considered.

A quality by design exercise implies upstream design, where the design starts with a focus on the product quality and how it is impacted by the process. It continues upstream, with emphasis on the product, and ends with the analysis of

the effect of the starting materials on the complete manufacturing train. This analysis will eventually end in the establishment of an acceptance criterion for each of the raw materials. These acceptance criteria will be mostly based on statistical diagnostics.

Typical statistical tools focus on testing one variable at a time in order to accept/reject a given lot of raw material. Recent trends in data analysis suggest performing this diagnosis on multiple variables simultaneously. First, a mathematical description of each of the statistical tools is developed, followed by an illustration of these concepts highlighting the benefits of each technique.

20.2.2 Traditional Statistical Tools

As a well-established field, statistical diagnostics are now well accepted and widely applied in practically all areas of engineering. This section is not intended to serve as a reference in statistics; the reader is referred to other texts for this purpose [32]. What is included is a high-level description of some of the concepts mentioned in statistics books so that the reader can better interpret the information presented in these.

20.2.3 A High-Level View of Hypothesis Testing and Significance Levels

A set of experiments should always be carried out with a purpose in mind. The purpose is usually to obtain information from a given system and learn from it. And this learning usually comes from proving (or disproving) a preconceived idea about the system. This is referred to as hypothesis testing. The type of test used will depend on the particular hypothesis being tested and the available data.

In engineering, hypothesis tests are usually of the quantitative nature. An inequality test, e.g. is the density greater than 1.5 g/ml, would require a one-sided hypothesis test, whereas a range (is the density greater than 1.5 and lower than 1.8) would require a two-sided test.

The reader should be aware that in statistics all statements involve a probability, usually referred to the level of significance and quantified as $100(1 - \alpha)$ for a given conclusion. Typical values chosen for $100(1 - \alpha)$ are 95, 99, or 99.73% (which is the probability associated with six standard deviations or six sigma in a normal distribution). The choice of the level of significance for a given test usually depends on the consequences and the implications (sometimes legal) of drawing the wrong conclusion (99% confidence on the conclusion that a plane engine will not fail means 1 in 100 times it might!). Obvious to mention is that a 95% significance level implies a value of 0.05 for α . If the test to be used is a single-sided test, α is taken as is. For a double-sided test, α is usually divided by two to allow the test to be centered on the 50% probability.

Often in the establishment of specifications for materials, the hypothesis to be tested is a double-sided one. In these cases, the value of property A for a new material is compared against a reference set of values for property A to verify that it lies within a given range. The reference values for property A are usually chosen from materials used in the past due to desirability of those materials in terms of quality of the product or cost of manufacturing.

20.3 THE ESTIMATE OF A MEAN VALUE AND ITS CONFIDENCE INTERVALS

A simple tool to establish a double-sided diagnostic to test property A for new materials is to estimate the mean value of property A and estimate the upper and lower confidence intervals for this mean value. This estimation will yield lower and upper bounds for the mean value of property A in the reference set. A new lot of raw material will be tested to determine its value of property A and then decided if the mean value of property A for this new lot is also a plausible (probable) mean value for property A in the reference materials.

This confidence interval is a function of the degrees of freedom (number of samples n used to estimate the mean of property A minus 1), the standard deviation of the reference values for property A, and the t value from statistical tables. This t value is a function of the desired significance level (α) and a number of degrees of freedom ($n - 1$). And although this t value is a strong function of α , the α value is often the less questioned parameter. For all practical purposes, α can easily be fixed (e.g. 99.73%) for the sake of testing if the mean value of a given property of a new material is the same as the one in the reference set.

Equation (20.1) describes the calculation of the confidence intervals for the mean value of a normally distributed population with unknown (only estimated) variance σ^2 . In practice, the factor that has the greatest impact on this calculation is the number of samples (n). The range of acceptance (upper minus lower bound) will be large for a small number of samples and will asymptotically narrow as the number of samples increases:

$$\bar{x} - t_{\alpha/2, n-1} \sigma / \sqrt{n} \leq \mu \leq \bar{x} + t_{\alpha/2, n-1} \sigma / \sqrt{n} \quad (20.1)$$

In this formula, σ is the calculated standard deviation, \bar{x} is the calculated average, n is the number of samples considered, and $t_{\alpha/2, n-1}$ refers to the value of t for a significance level of $100(1 - \alpha)$ and $n - 1$ degrees of freedom; this value is taken from statistical tables.

For example, consider three sample sets taken from the same population. All sets with an average of 4.25 and standard deviations of 0.0470, 0.0565, and 0.0895, respectively, which were calculated using 30, 10, and 3 samples. At a

TABLE 20.2 Confidence Intervals for the Mean Value of Three Sample Sets of the Same Population

Mean Value	Sample Sets of Same Population		
	$n = 30$	$n = 10$	$n = 3$
Mean lower bound	4.234	4.214	4.027
Mean upper bound	4.268	4.295	4.472
Range	0.03397	0.08088	0.44471
% Change from $n = 30$	0	138.12	1209.27

95% level of significance, the values of t for $n = 30$, 10, and 3 are 2.042, 2.26, and 4.3. With these values, the confidence intervals on the means for each population are given in Table 20.2. Notice the dramatic increase in the range, just due to the number of samples considered in the calculations. The data for this example is from a pharmaceutical-grade polymeric material and is a real illustration of how acceptance limits can vary in an application. The practitioner is encouraged to sample properly to avoid artificially large acceptance regions simply due to a limited number of samples available.

20.3.1 Emerging Multivariate Techniques

A natural implication of the evolution of analytical technology is the fact that a given material can be characterized by a large number of attributes. It is the duty of the engineer to determine which of these attributes are relevant to the product/process. And the answer to this question for a pharmaceutical product is rarely a single property (a scalar) and more often a set of properties (a vector) that will impact the product or process. The challenge now is how to establish a specification for multiple quantities.

The simple solution to this challenge is to establish multiple univariate diagnostics using traditional statistical tools described previously. This practice however implies that all the measured characteristics for the new material can be tested and assessed independently of each other. This assumption falls apart quite easily for complex materials where a large number of properties are related. For example, for a polymeric material, the molecular weight distributions (or compositional distributions for a copolymer) are not independent of viscosity or density, which can also be linked to cross-linking.

In such a situation, there is a need for a tool that will enable the establishment of acceptance criteria for multiple properties simultaneously, accounting for their correlated nature. It is no longer enough to know the desired level (mean) and tolerance for each property. Additional information is necessary to account for the dependencies across the multiple properties being tested.

Multivariate latent variable methods (LVM) have been proposed in literature to address this need. LVM will

empower the user to establish diagnostics and acceptance criteria based on a model of the data [39].

20.3.2 Using a Model to Establish a Test of Acceptance

Before describing the calculations behind a multivariate specification, it is important to discuss the overall strategy of using a model (any model!) to establish acceptance regions. To illustrate this point, consider the case of a material that is characterized by two properties, *A* and *B* (plotted in Figure 20.2 where each dot is plotting the numerical values of property *A* vs. *B* for a given lot of raw material). Assume also that there is enough data to conclude that the lots of material represented by gray markers are desirable and those represented by black dots correspond to undesirable material.

The challenge is to somehow delineate the region spanned by the gray markers. If univariate measures are taken (define a lower and upper bound for properties *A* and *B* separately), one would end with a region equivalent to the smallest square that fits in the “gray marker” zone (drawn with a dotted line square). Although feasible and simple, this may constrain the practitioner to a very small region of acceptance and will end in large amounts of rejected material.

The alternative is to use mathematics to delineate this region. Taking all the data that corresponds to the gray markers, a regression model can be fitted. The regression line given by $A_{\text{gray}} = mB_{\text{gray}} + \delta$ is illustrated with a black line in Figure 20.2 (where *m* and δ are fitted parameters from the data).

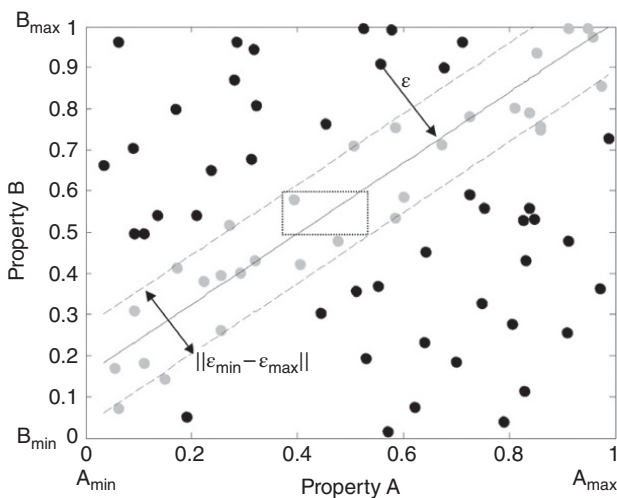


FIGURE 20.2 Conceptual illustration of a case where a multivariate specification is required. Dotted line square is a possible univariate dual specification, and solid line corresponds to a fitted line $A_{\text{gray}} = mB_{\text{gray}} + \delta$.

This simple model (a line) can then be used to quantify the perpendicular distance from any given point to the regression line (illustrated as ϵ in Figure 20.2).

The set of distances (ϵ) for all the gray points can be used to determine a maximum and minimum ϵ (ϵ_{min} and ϵ_{max}). These limits (ϵ_{min} and ϵ_{max}) on the perpendicular distance and the regression model and the upper–lower boundaries ($A_{\text{min}}, A_{\text{max}}, B_{\text{min}}, B_{\text{max}}$) can be used to establish a multivariate specification that will ensure properties *A* and *B* for a new lot are within the region spanned by the gray markers. There is still a delicate statistical exercise to determine the perpendicular distance to tolerate (given by $\|\epsilon_{\text{min}} - \epsilon_{\text{max}}\|$), and this will determine the width of the acceptance region (region bounded by the dashed lines in Figure 20.2) and the risk associated with the test.

For this conceptual case, the steps to accept a new lot of material (with properties *A* and *B* equal to a_{new} and b_{new} , respectively) using the overall bounds include a simple regression line as the model and then determining an acceptable perpendicular distance:

IF $A_{\text{min}} < a_{\text{new}} < A_{\text{max}}$ and $B_{\text{min}} < b_{\text{new}} < B_{\text{max}}$ continue;
 OTHERWISE reject a_{new} and b_{new} . Calculate $\epsilon = a_{\text{new}} - (mb_{\text{new}} + \delta)$.

IF $\epsilon_{\text{min}} \leq \epsilon \leq \epsilon_{\text{max}}$, then the new lot is not rejected (there is no statistical evidence to prove that this lot of material is any different from the population represented by the gray markers).

Step *i* will ensure that the values of a_{new} and b_{new} are at least within range, step *ii* uses the model to determine the perpendicular distance to the line that runs in the middle of the acceptance region, and step *iii* determines if this perpendicular distance is within tolerance. Notice that the model is not being used for predictive purposes.

Both properties (*A* and *B*) still need to be measured and none of them are being predicted from the other. The model in this case is just a mere geometrical tool to delineate a region that is one degree of complexity beyond a simple squared region (which is the result of two univariate specifications together). Also notice that there is still an exercise of probability and risk analysis associated with determining the upper and lower bounds for the key diagnostic(s) involved (in this case ϵ_{min} and ϵ_{max}). The strategy proposed here is multivariate in the sense that it handles more than one variable, but more importantly, it is multivariate simultaneous, which means it handles more than two properties at the same time.

For the conceptual case illustrated in Figure 20.2, it is easy to see how a line can be used as a model in the specification since the data is composed of two properties. As the dimensionality of the problem increases (the amount of variables to consider simultaneously), so does the need to have a model that considers all variables and their uncertainty levels and

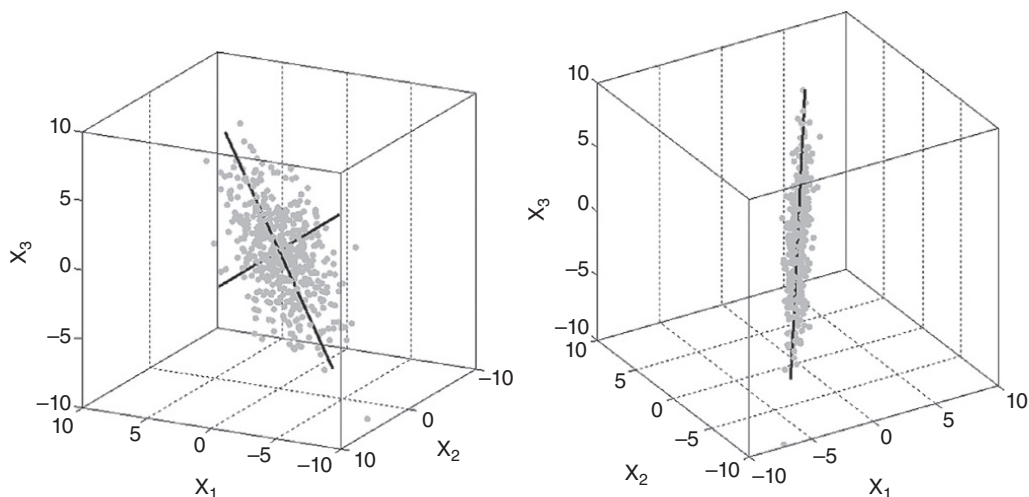


FIGURE 20.3 Dimension reduction of a 3-D data set to a 2-D plane by PCA.

correlation simultaneously. And for that, principal component analysis (PCA) is suggested.

20.3.3 Principal Component Analysis

PCA is a well-established technique to project or compress data to a lower number of new variables called principal components. Geometrically, the PCA works by identifying a new coordinate system within the data (see Figure 20.3) so that each point can be referred to by its coordinates with respect to this new system. Hopefully, the number of coordinates needed to span data well enough will be dramatically less than the original number of variables. The example in Figure 20.3 illustrates a data set with three variables (X_1 , X_2 , and X_3) that can significantly be represented with a two-dimensional coordinate system, assuming the deviation from this plane (see right plot Figure 20.3) is negligible.

Many software packages¹ are available in the market to fit a PCA model, and therefore such calculation is not discussed here. A PCA model is quite powerful as a tool to establish specifications due to the diagnostics provided by the model. This application is extensively discussed in literature [5, 39] and only summarized here.

Once a PCA model is fitted, each of the observations used in the model can be summarized by two overall diagnostics, the squared prediction error (SPE) and the Hotelling T^2 statistic.

20.3.4 The Squared Prediction Error (SPE)

This diagnostic is identical in interpretation to the perpendicular distance (ϵ) mentioned in the conceptual problem

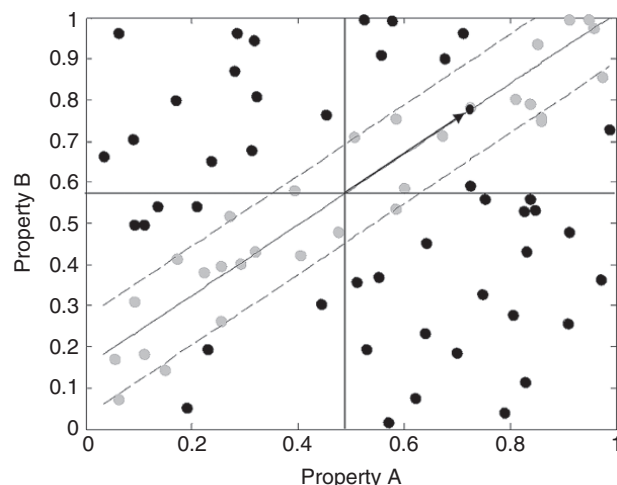


FIGURE 20.4 Conceptual problem illustrating the distance being diagnosed by the Hotelling T^2 in a two-dimensional problem.

presented previously in this section. This is a quantitative measure on how well the new sample adheres to the structure of the reference data. This is a positive number and is well accepted to follow a chi-squared distribution [46], which means that confidence intervals for a given level of significance can be computed.

20.3.5 The Hotelling T^2 Statistic

This diagnostic is a measure of the squared distance from the origin of the data (the mean values) to the expected conditional value for the properties of the new lot. In the conceptual problem, this would be equivalent to the square of the distance along the black line (illustrated with an arrow), from the center of the box to the intersection where b_{new} meets \hat{a} , which is given by the regression model (Figure 20.4). Since

¹www.umetrics.com; www.prosensus.ca; www.camo.com; www.eigen-vector.com

the model prediction is being used here, it is imperative to first assess that the SPE is within tolerance. Notice that imposing a bound on this diagnostic implicitly imposes a bound on the magnitude of properties A and B , and, hence, step i in the suggested sequence could be replaced by a simple one-sided test on the Hotelling T^2 diagnostic.

The use of these multivariate diagnostics is illustrated with an example taken from a real scenario in the pharmaceutical sector.

EXAMPLE 20.1

Setting Multivariate Raw Material Specifications Using PCA

Consider a polymeric ingredient that is characterized by three descriptors of its particle size: D10, D50, and D90 (each number represents the average size at the 10th, 50th, and 90th quantile from the distribution). Each quantity is reported in logarithmic scale in Table 20.3. This table also contains the lower and upper limits for the mean, as calculated from this table. The fundamental concept to illustrate is that these three descriptors for particle size are not independent of each other and a change in one of them will imply a change in the others (Figure 20.5).

Consider now six new lots of material (data provided in Table 20.4). If the three independent specifications are used to decide whether to accept or not these six lots of material, it is quite obvious that lots 1, 5, and 6 (marked by a ●, ▼, and ★ respectively) will be rejected. In contrast, a PCA model was fitted to the data using one significant component, the Hotelling T^2 , and the SPE diagnostic, and finally a multivariate specification was built (Figure 20.6 where the new lots are colored in red). Notice that this specification also rejects lots 1, 5, and 6; however lots 3 and 4 (marked by ▲ and ◇, respectively) are rejected, and only lot 2 (marked by ■) is accepted. The reason is simply to understand by looking at the three descriptors simultaneously, as shown in Figure 20.7.

In the top plot of Figure 20.7, lots 1, 5, and 6 were rejected because all three particle size descriptors are clearly different from the lots used as a reference. The bottom plot (which is a rotated version of the top) illustrates why lot 2 (marked by ■) was accepted, while lots 3 and 4 (marked by ▲ and ◇) were rejected. The particle size descriptors for lot 2 exhibit the same expected proportions between the $\log_{10}(\text{D10})$, $\log_{10}(\text{D50})$, and $\log_{10}(\text{D90})$ (referred to as covariance structure) as in the reference set, and hence it is safer to accept this material than other materials where the proportions between these properties are different. An added advantage of the multivariate specification (Figure 20.6) is that a single plot can be used to impose specifications on multiple properties simultaneously. For this case, it was possible to visualize the three particle size variables in a three-dimensional plot; a real case

TABLE 20.3 Example of Data from a Polymeric Pharmaceutical Excipient Used in Example 20.1

Sample	$\log_{10}(\text{D10})$	$\log_{10}(\text{D50})$	$\log_{10}(\text{D90})$
Lot 1	4.17	4.57	5.42
Lot 2	4.17	4.59	5.48
Lot 3	4.19	4.57	5.43
Lot 4	4.19	4.57	5.37
Lot 5	4.20	4.53	5.40
Lot 6	4.20	4.54	5.37
Lot 7	4.22	4.58	5.40
Lot 8	4.22	4.58	5.40
Lot 9	4.22	4.59	5.39
Lot 10	4.23	4.60	5.44
Lot 11	4.23	4.60	5.40
Lot 12	4.23	4.59	5.44
Lot 13	4.23	4.57	5.44
Lot 14	4.23	4.59	5.40
Lot 15	4.24	4.60	5.41
Lot 16	4.25	4.59	5.38
Lot 17	4.25	4.61	5.40
Lot 18	4.25	4.62	5.46
Lot 19	4.25	4.61	5.44
Lot 20	4.26	4.58	5.46
Lot 21	4.26	4.59	5.45
Lot 22	4.27	4.58	5.43
Lot 23	4.27	4.62	5.44
Lot 24	4.28	4.62	5.41
Lot 25	4.29	4.61	5.47
Lot 26	4.31	4.67	5.51
Lot 27	4.31	4.64	5.46
Lot 28	4.31	4.65	5.47
Lot 29	4.32	4.63	5.47
Lot 30	4.32	4.63	5.49
Lot 31	4.32	4.63	5.52
Lot 32	4.35	4.67	5.50
Mean	4.25	4.60	5.44
Standard deviation	0.047	0.032	0.041
n	32.00	32.00	32.00
$\tau(\alpha/2), \nu [\alpha = 0.05]$	2.042	2.042	2.042
Mean low limit	4.23	4.59	5.42
Mean upper limit	4.27	4.61	5.45

scenario may consist of several hundreds of variables, and then the power of a multivariate approach is the ability to still monitor the multivariate proportion of all properties, simultaneously in a couple of plots.

Ultimately, the impact of the raw material physical and chemical properties on the final drug product depends on the manufacturing process (e.g. wet granulation will be affected by properties that do not affect dry granulation). This section exemplified a general method to establish specifications on the properties of a given material. The next section focuses on unit operation-specific details where the relevance of certain physical/chemical properties of the material is discussed in the context of the processing route.

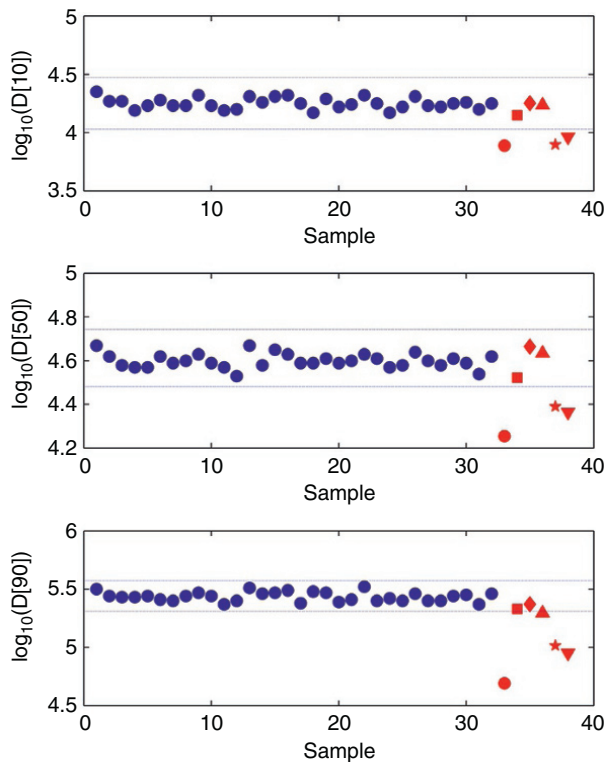


FIGURE 20.5 Univariate specs and acceptance criteria for particle size descriptors. ● markers denote reference lots, and ■, ◇, ▲, ★, ▼, ● markers denote new lots to be tested for acceptance.

TABLE 20.4 DATA FOR SIX NEW LOTS OF POLYMER FOR EXAMPLE 20.1

New Lot	Log ₁₀ (D10)	Log ₁₀ (D50)	Log ₁₀ (D90)
A	3.9	4.2	4.7
B	4.1	4.5	5.3
C	4.2	4.7	5.4
D	4.2	4.6	5.3
E	3.9	4.4	5.0
F	4.0	4.4	4.9

20.4 PROCESS MODELING FOR SOLID ORAL DRUG PRODUCT PROCESSES

20.4.1 Powder Flow Models

20.4.1.1 Model Development for Powder Processes Using Discrete Element Method The DEM is one possible approach to model powder flow in processing operations. DEM models are a computational approach whereby the state of each particle in the system is tracked over the time frame of interest. These models produce a wealth of fundamental particle-level data including particle positions and velocities as

well as the forces acting on each particle. These fundamental quantities can then be used to calculate many other useful quantities, or derived quantities, such as solid fractions, stress tensors, local mass concentration fractions, extents of lubrication, tablet coating uniformity, etc. This wealth of data is a key advantage of the DEM approach; many of these quantities are expensive and difficult, if not impossible, to measure experimentally. However, acquiring these data via DEM does have significant costs, primarily in long computational times. Depending on the number of particles modeled, the complexity of the simulation domain, and the length of time modeled, simulation times may range from a few hours to well over a month of computing time using traditional CPU processing. However, the increasing adoption of GPU processing has dramatically reduced simulation times and expanded the number of particles that can be simulated.

The algorithm of a typical DEM model is shown in Figure 20.8. The simulation is initiated by defining the computational domain and creating particles within it. Each particle is given a size, mass, and density and assigned a position and velocity. The simulation is started, and all contacting pairs (both particle–particle and particle–wall) are identified. This contact detection step is among the most time-consuming aspect of DEM programs. A brute force contact detection algorithm that searches between all possible pairs scales with N^2 , where N is the number of particles in the simulation. However, using techniques such as a neighbor searching algorithm can reduce the time to the order of $N \ln(N)$. Once each contacting pair is identified, force–displacement models are used to determine the contact forces acting on each of the particles. While several such models can be used, most models specify the normal and tangential forces as a function of the overlap distance between particles (the overlap approximates particle deformation during contact and is typically constrained to a small value [$<1\%$ of particle diameter]). For example, one such model for the normal force is based on the theoretical work by Hertz [47] in 1882. This model describes the elastic contact of a sphere and gives the normal force, \mathbf{F}_N , as

$$\mathbf{F}_N = k_N \delta^{3/2} \hat{\mathbf{n}} \quad (20.2)$$

where

k_N is a stiffness related to the radii and material properties of the contacting spheres.

δ is the overlap distance between the spheres.

$\hat{\mathbf{n}}$ is the unit normal vector.

Other more complex models build upon this and other theories to include dissipative effects for modeling inelastic contacts. In addition to the normal and tangential contact forces,

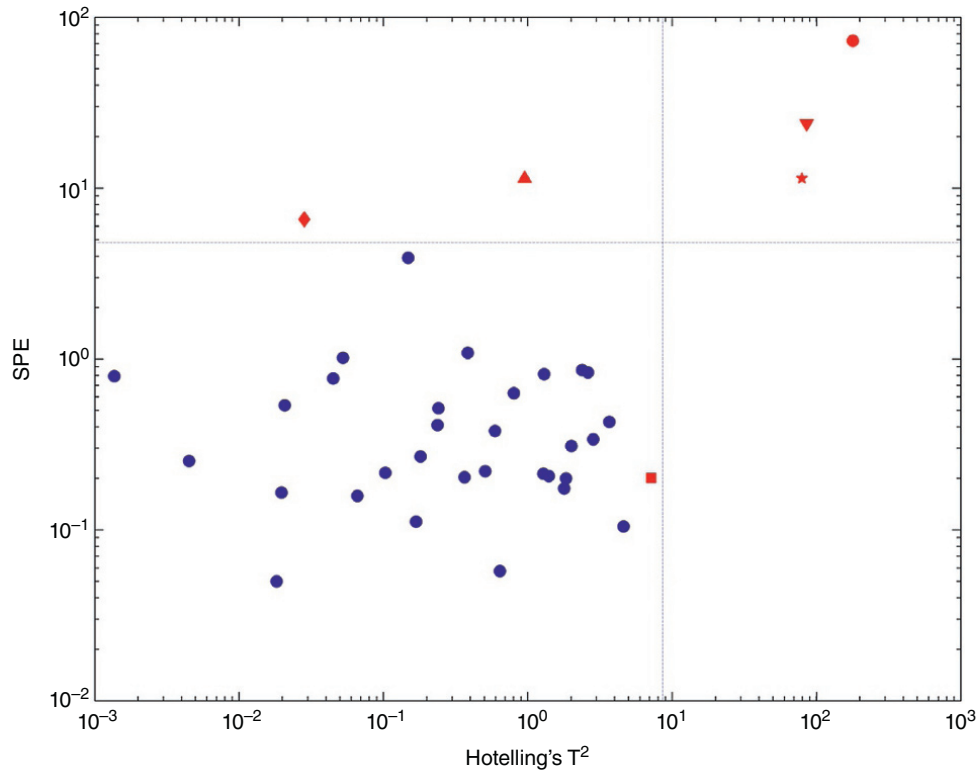


FIGURE 20.6 Multivariate specification for particle size data in Example 20.1 with acceptance limits. Blue markers denote reference lots, and red markers denote new lots to be tested for acceptance.

body forces – such as the acceleration due to gravity, electrostatics, or magnetic fields – can also be included. The resultant contact and body forces acting on each particle are summed. Newton’s second law is then used to calculate the translational and rotational accelerations, respectively:

$$\mathbf{F}_{N, \text{total}} = m \frac{\partial^2 \mathbf{x}}{\partial t^2} \tag{20.3}$$

$$\mathbf{M}_{\text{Total}} = I \frac{\partial^2 \boldsymbol{\theta}}{\partial t^2} \tag{20.4}$$

where

$\mathbf{F}_{N, \text{total}}$ is the total resultant normal force acting on a given particle.

m is the particle mass.

\mathbf{x} is the particle position vector.

t is time.

$\mathbf{M}_{\text{Total}}$ is the total resultant moment due to the tangential forces acting on a given particle.

I is the particle moment of inertia.

$\boldsymbol{\theta}$ is the particle orientation vector.

Subsequently, these accelerations are integrated in time to determine updated particle velocities and positions. At this point, virtually any quantity of interest may be measured, and the simulation then proceeds to the next iteration by incrementing the time step and repeating the necessary contact detection calculations. This procedure is repeated until the simulation has reached the desired end point, such as when sufficient data has been collected or a process has completed.

Oftentimes, certain assumptions are made in DEM models to simplify the computational demands. For example, assumptions of spherical particles, cohesionless particles, and negligible interstitial fluid effects have frequently been made. Each of these assumptions helps to not only make the simulations faster but also make the modeled system less representative of the real system of interest. It is common in contemporary DEM simulations to include some representation of non-spherical particles and/or cohesion. DEM simulations coupled to other physics-based simulations such as CFD or multibody dynamics (MBD) are becoming more common. These advances, especially when coupled with the increased computational power of GPU computing, have led to an explosion in simulation fidelity in recent years. Interested readers are referred to review papers [7, 48–50] and the references therein for more detailed information.

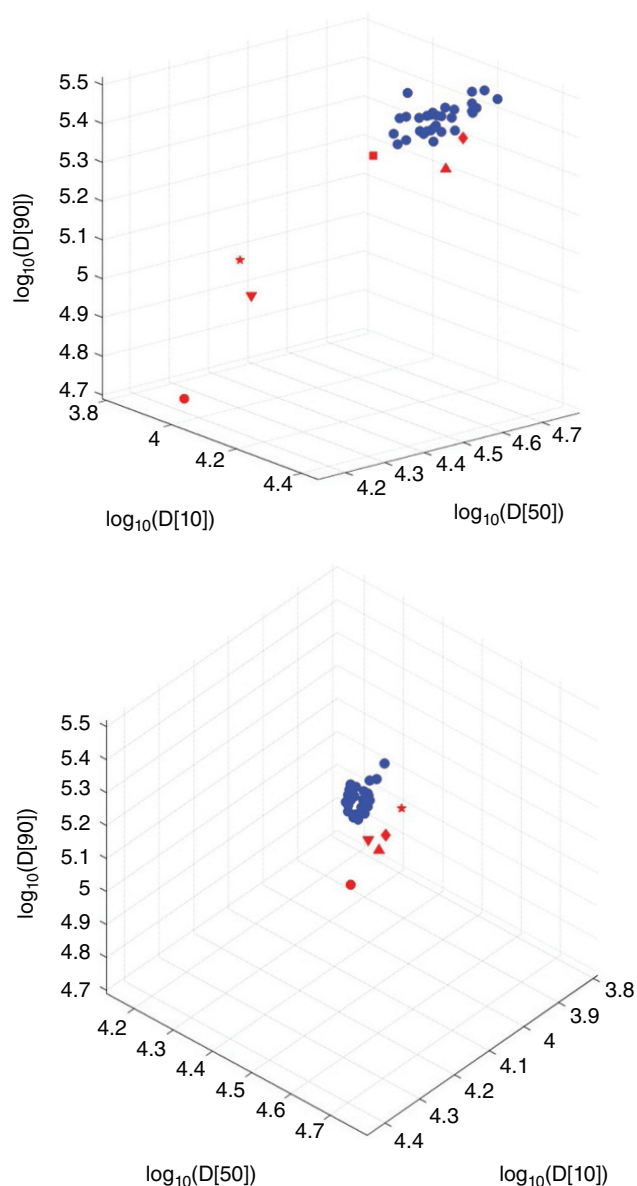


FIGURE 20.7 Three particle size descriptors plotted simultaneously. Blue markers denote reference lots, and red markers denote new lots to be tested for acceptance. Bottom plot is a rotation of the top plot.

20.4.1.2 Modeling Powder Blending, Fluidization Segregation, and Lubrication Using DEM The drug product manufacturing process typically involves several powder handling operations that are used to create the final product – a dosage form such as a tablet or capsule – from several raw materials, typically powders with varying physical and chemical properties. In all cases, content uniformity is a critical quality attribute of the final dosage form. The drug content in each dosage form is important, because if this were to vary considerably, patients could receive doses that might be ineffective or possibly result in undesired side effects. Detailed guidelines

issued by the US Pharmacopeia (USP) state the acceptance limits of variability in drug loading and prescribe the testing procedure used to determine the variability of a particular batch [51]. The most recent recommendation by the ICH Steering Committee recommends that the official pharmacopoeia from Europe, Japan, and the United States on Uniformity of Dosage Units and USP General Chapter <905> be used interchangeably in the ICH regions [52–58]. A parametric two one-sided tolerance interval (PTOSTI) test [59–61] analysis can be performed on the stratified core data to provide confidence in passing USP <905> acceptance criteria for content uniformity.

There are many powder processing unit operations used in the manufacture of tablets and capsules. These operations include blending, hopper filling and discharge, and flow through various feeding or dispensing devices, to name but a few. In this section, an example of a batch blending process will be presented along with a different example of a powder transfer process subject to fluidization segregation and, consequently, a detrimental impact on the content uniformity of the final dosage form. A third example will describe an approach to troubleshoot the impact of extensive shear on a lubricated blend and assess potential equipment modifications *in silico*.

The objective of the blending process is, as the name suggests, to combine the raw material powders into a well-mixed blend containing a uniform distribution of all materials, where each raw material consists of particles with a range of sizes and morphologies. Differences in these properties will cause particles to segregate or de-mix. Assuming that a uniform mixture can be obtained in a blender, segregation in subsequent operations can lead to content uniformity that is out of specification.

In this example, a pharmaceutical granulation was to be mixed with extra-granular excipients in a pilot-scale rotary bin blender shown in Figure 20.9. An unusual characteristic of this bin is that it tilts upward such that the axis of rotation is inclined at a 25° angle with respect to horizontal. Upon executing several batches, subpotent (API concentration < 100% label claim) thief samples were observed in the lower left corner of the bin for a subset of the batches. The root cause was unclear: there could be issues with the sampling procedure or the assay analysis, or there may have been poor mixing in this particular process.

The DEM approach was used to investigate the quality of mixing in this process. The material properties and the bin geometry and dynamics were all determined and used as inputs to the DEM simulations. The material properties were used to calibrate the DEM contact model parameters (i.e. the coefficient of friction) [62]. The PSD is scaled up to maintain a reasonable number of particles and hence computational time. The simulation setup (i.e. bin fill level, composition, etc.) was targeted to replicate the batch conditions as closely as possible. Simulation snapshots shown in Figure 20.10

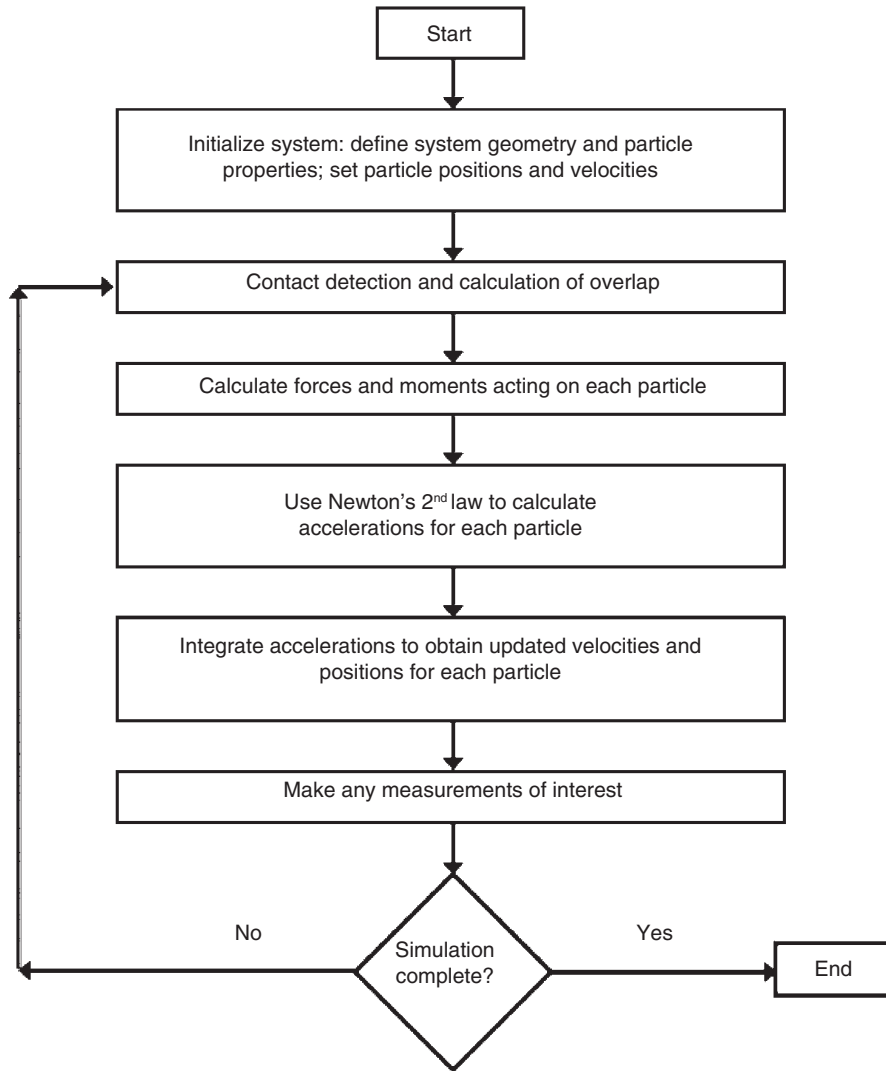


FIGURE 20.8 A flowchart showing the algorithm typically used in discrete element method simulations of powder flow. *Source:* Reprinted with permissions from Ketterhagen et al. [7], Copyright (2009) John Wiley & Sons, Inc.

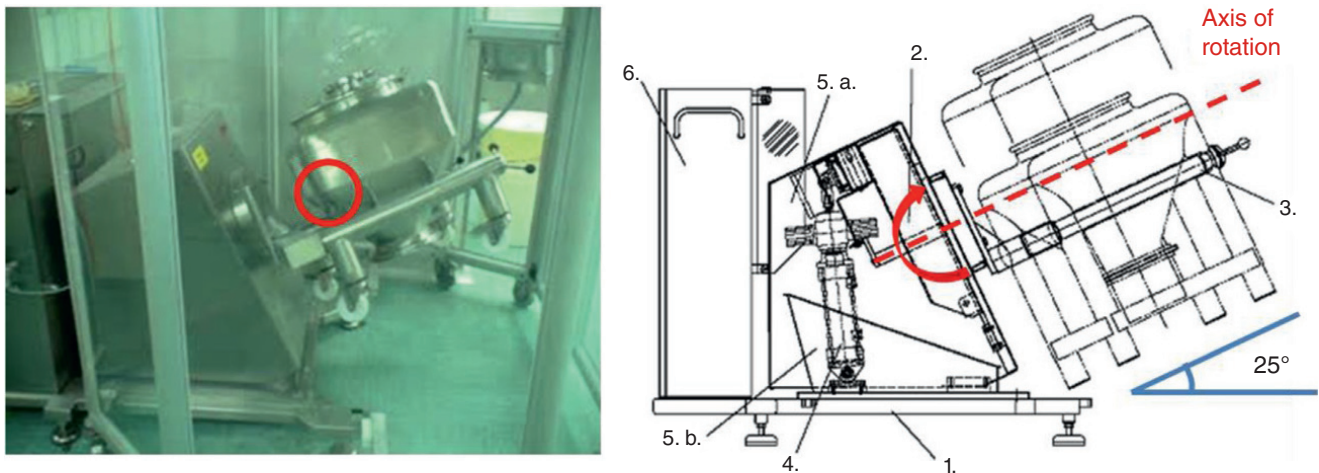


FIGURE 20.9 Photograph and schematic of bin blender.

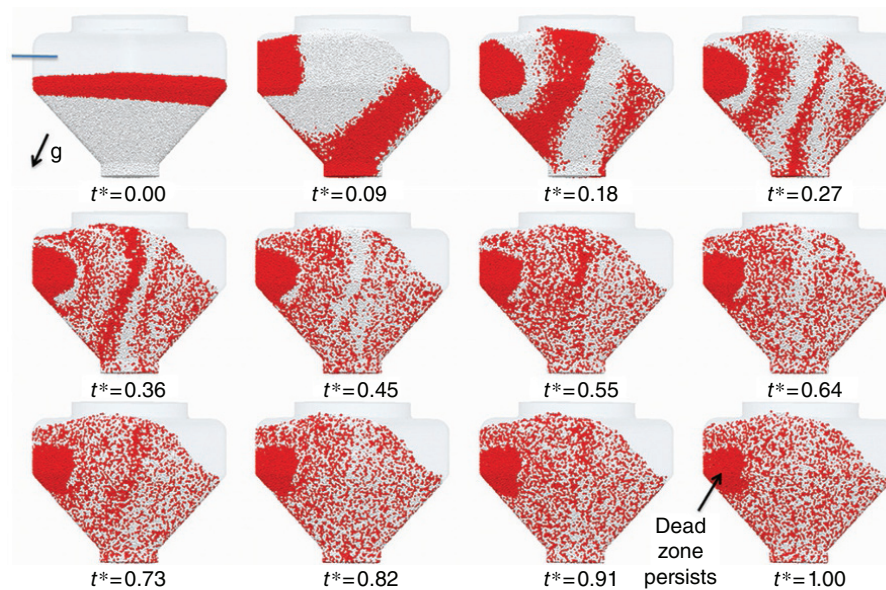


FIGURE 20.10 DEM simulation images showing the progression of mixing for the case of a bin blender with axis of rotation tilted to an incline of 25°. An unmixed “dead zone” is present due to the incline angle that causes the local fill level on the left side of the bin to be too large to allow for sufficient shearing for powder mixing.

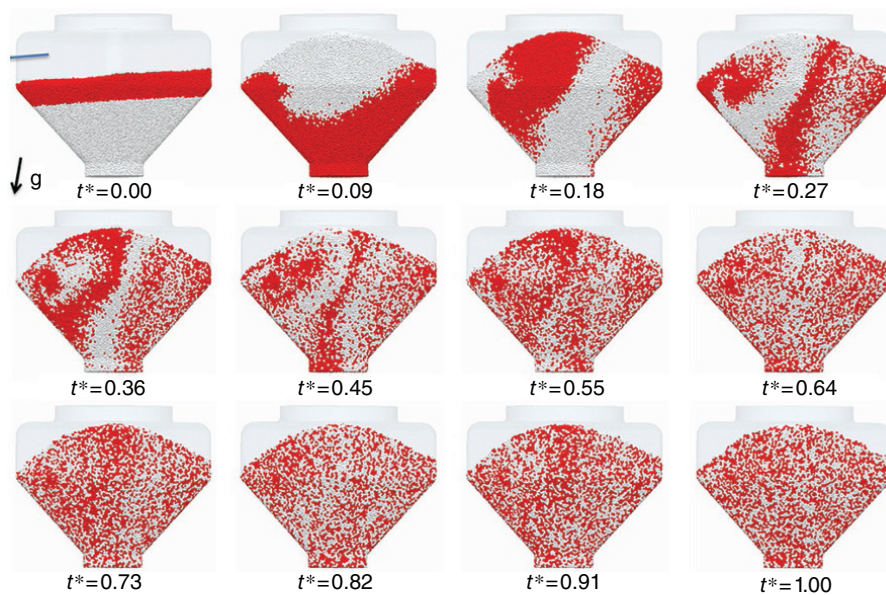


FIGURE 20.11 DEM simulation images showing the progression of mixing for the case of a bin blender with axis of rotation tilted to an incline of 15°. Improved mixing is observed throughout the bin.

depict the progression of the mixing. White particles model the active granulation, while black particles model the extra-granular excipients, and the gravitational vector indicates the tilt angle of the bin. As the simulation progresses, good mixing is predicted throughout most of the bin. However, the local fill level on the left side of the bin is too large due to the incline of the bin, and good mixing is not observed in this area. There is an unmixed “dead zone” or core of material surrounding the axis of rotation. This material is not

exposed to shear on the free surface and therefore does not mix. If a thief sample is to be taken from this area, the simulation would predict a subpotent sample, consistent with observations from the batches.

With the insight from the DEM simulations of the bin inclined at 25°, additional simulations were used to study modifications to the equipment design and operation in order to improve the mixing and reduce the impact on content uniformity. In the example shown in Figure 20.11, a reduced

incline angle (15°) is simulated. In this case, the local fill level throughout the bin is sufficient to permit good mixing in all areas, and no unmixed “dead zones” are present.

Through the efforts in this example, the DEM simulations first provided insight into the root cause of the observed issues in the experimental blending operation. Second, additional DEM simulations were used to understand the effects of modifications on the blending performance *in silico*. These simulation results were then used as a basis to guide decision making to overcome this issue and improve process robustness.

Once a powder is uniformly blended, it is desired to maintain this uniformity through the downstream manufacturing processes until the dosage form API content is fixed during the tableting or encapsulation processes. Often these processes will require handling of the powders via gravity discharge from the bin blender described previously and gravity flow through a chute to the tablet press or encapsulator. In this powder transfer process, a filled bin containing a well-mixed blend is mounted above a transfer chute depicted in Figure 20.12 [63]. Initially, the chute is filled with air, but after the outlet valve on the bin is opened, powder will

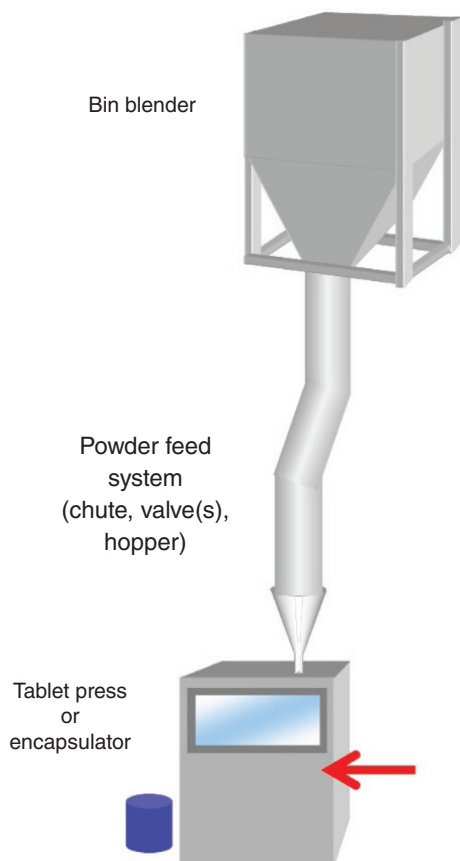


FIGURE 20.12 A schematic of a powder transfer system including a bin blender, transfer chute, and hopper of a tablet press or encapsulator [63].

discharge from the bin to the chute. The air inside the chute is then displaced by the falling powder with the flow affected by the presence of vents in the chute. It is during this initial powder flow phase where the powder becomes fluidized en route to the bottom of the chute and is subject to segregation due to a fluidization mechanism. Typically, in such a case, particles with larger aerodynamic diameters will tend to settle to the bottom of the chute most quickly, while smaller particles will tend to become entrained in the airflows. The smaller particles will tend to remain in the upper regions of the chute for this particular case. In this second example, a computational modeling approach is used to predict the extent of segregation for this powder transfer process and assess *in silico* the impact of changes to the PSD on the segregation.

Two simulation approaches are used in a coupled manner to simulate both the powder flow and the airflow in the system. In this case, the DEM approach used to model the powder is coupled with a CFD approach [64]. In this approach, a so-called two-way coupling is implemented; this refers to modeling both the effect of airflow on the particle flow and vice versa. In contrast to the previous example of the bin blending process where airflow could be neglected, the air must be simulated in order to capture its effects on the powder flow and the potential segregation.

In this example, the coupled CFD-DEM simulations are set up with an initial state consisting of a uniform blend of API and excipient particles in the bin blender and only air in the powder transfer chute. The DEM contact model parameters are calibrated by matching DEM results from simulations of lab-scale test equipment (e.g. shear cell, powder rheometer) to experimental measurements for the formulation of interest. At the start of the simulation, the bin valve is opened, and the flow of powder from the bin through the chute is simulated until it comes to rest at the bottom of the chute.

Once the powder has come to rest at the bottom of the chute, the simulation is stopped, and the potency, or fraction of API, is measured in horizontal slices through the chute (Figure 20.13a). These data are plotted in Figure 20.13b–d for varying API PSD. In the first case (Figure 20.13b), the API PSD is markedly larger than that of the excipient blend. As the large particles tend to travel down the chute faster than the smaller particles, the system segregates, and a superpotent region of powder is predicted to exist at the bottom of the chute. The potency in the rest of the chute is predicted to fluctuate around 100% drug loading. If the API PSD is reduced (Figure 20.13c) such that it more closely matches that of the excipient blend, a smaller extent of segregation is predicted with only a slightly subpotent region predicted at the bottom of the chute. Finally, if the API PSD is milled to a much smaller size, the segregation trend can reverse. The largest particles still tend to fall fastest through the chute, but in this case, the largest particles are from the excipient blend. Therefore, the blend at the bottom of the chute is highly subpotent.

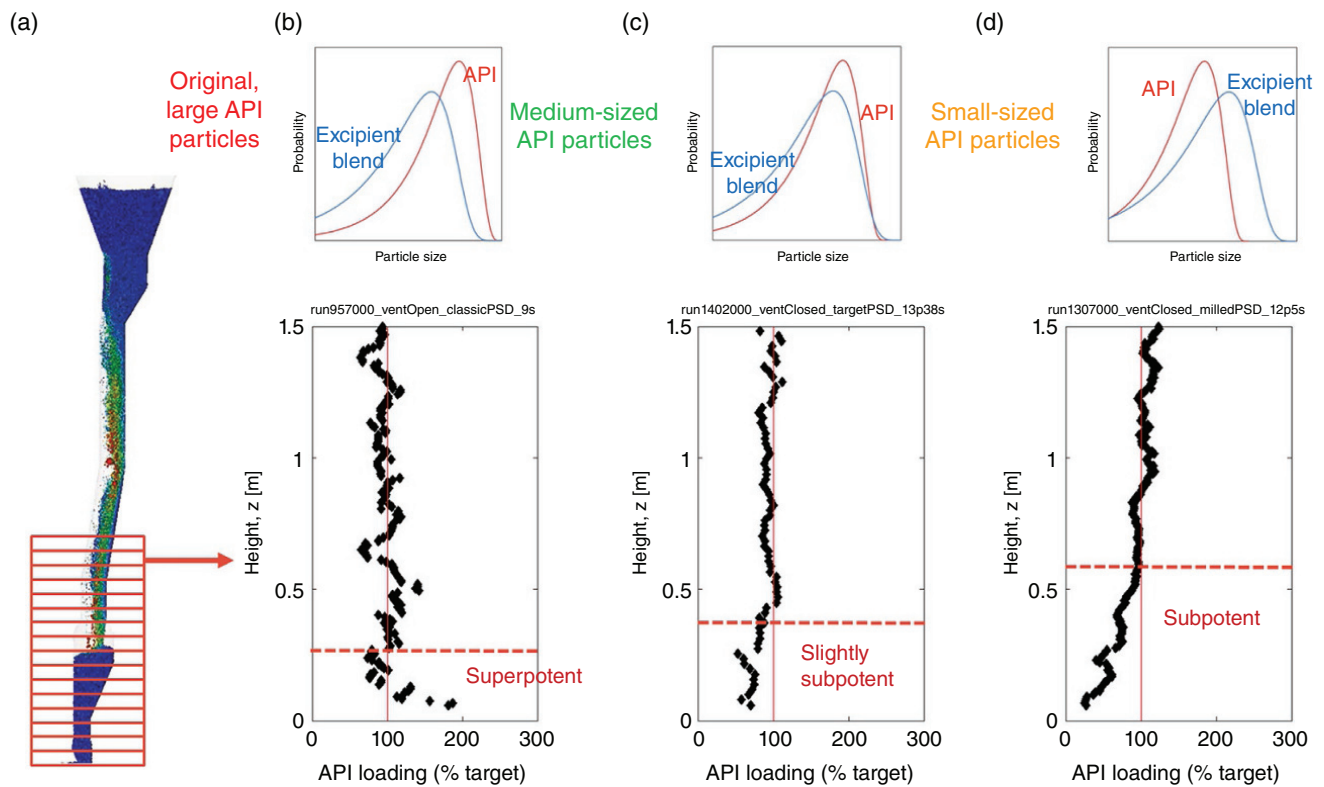


FIGURE 20.13 Simulation data showing (a) horizontal slicing for API loading analysis and (b–d) API loading for the three given API/excipient particle size distributions.

Images and videos from the simulations can greatly help to visualize the process and the underlying physics that can cause an issue like fluidization segregation. In Figure 20.14, simulation images from each of the three cases described in Figure 20.13 are shown. Here, the particles are colored by the local API mass fraction normalized by the target API mass fraction. Thus, particles colored gray are from a well-mixed region of the blend, while particles colored yellow and red (green and blue) are from superpotent (subpotent) regions of the blend. In the initial well-mixed state (not shown), the particles are nearly all colored gray. As the bin valve is opened, flow is initiated, and particles begin to segregate and change color. In Figure 20.14b, the highly segregated blend is apparent with red-colored particles at the bottom of the chute. This is problematic because the first tablets manufactured will consist of this powder and be superpotent. There is also evidence of trajectory-based segregation in the hopper (left-to-right decrease in potency) due to the inclined chute above the hopper. For the cases with the reduced API PSD (Figure 20.14c), nearly all the material is colored gray with only small amounts of green color showing up at the bottom of the chute. This is indicative of a well-mixed blend and ideal material from which to manufacture tablets. Finally, for the case of the very small API PSD (Figure 20.14d), the green color and extent of segregation

become much more significant. Simulation results in the form of quantitative data (Figure 20.13) and visualizations (Figure 20.14) can give helpful insight to better understand the flow and segregation of powder in processing equipment and, in this case, understand how changes to the API PSD can potentially impact the quality of the final product.

In addition to powder blending and segregation, DEM approaches can be used to make predictions of other powder blend properties. One such example is the state of lubrication within the blend. Dry powder lubricants, such as magnesium stearate, are typically added to most pharmaceutical formulations to improve manufacturability. The addition of a lubricant helps to reduce adhesion of powder to the metal surfaces of the processing equipment and can help to improve flowability of powders by reducing interparticle friction and increasing the powder bulk density. However, if a lubricated powder blend is exposed to excessive amounts of shear during processing prior to creation of the final dosage form, the properties or performance of that dosage form can be adversely impacted. A process-dependent K value has been proposed to quantify the extent of shear imparted to the powder by various drug product equipment such as batch bin blender [65], a tablet press feed frame [66, 67], and the encapsulator feed system described here [68–70].

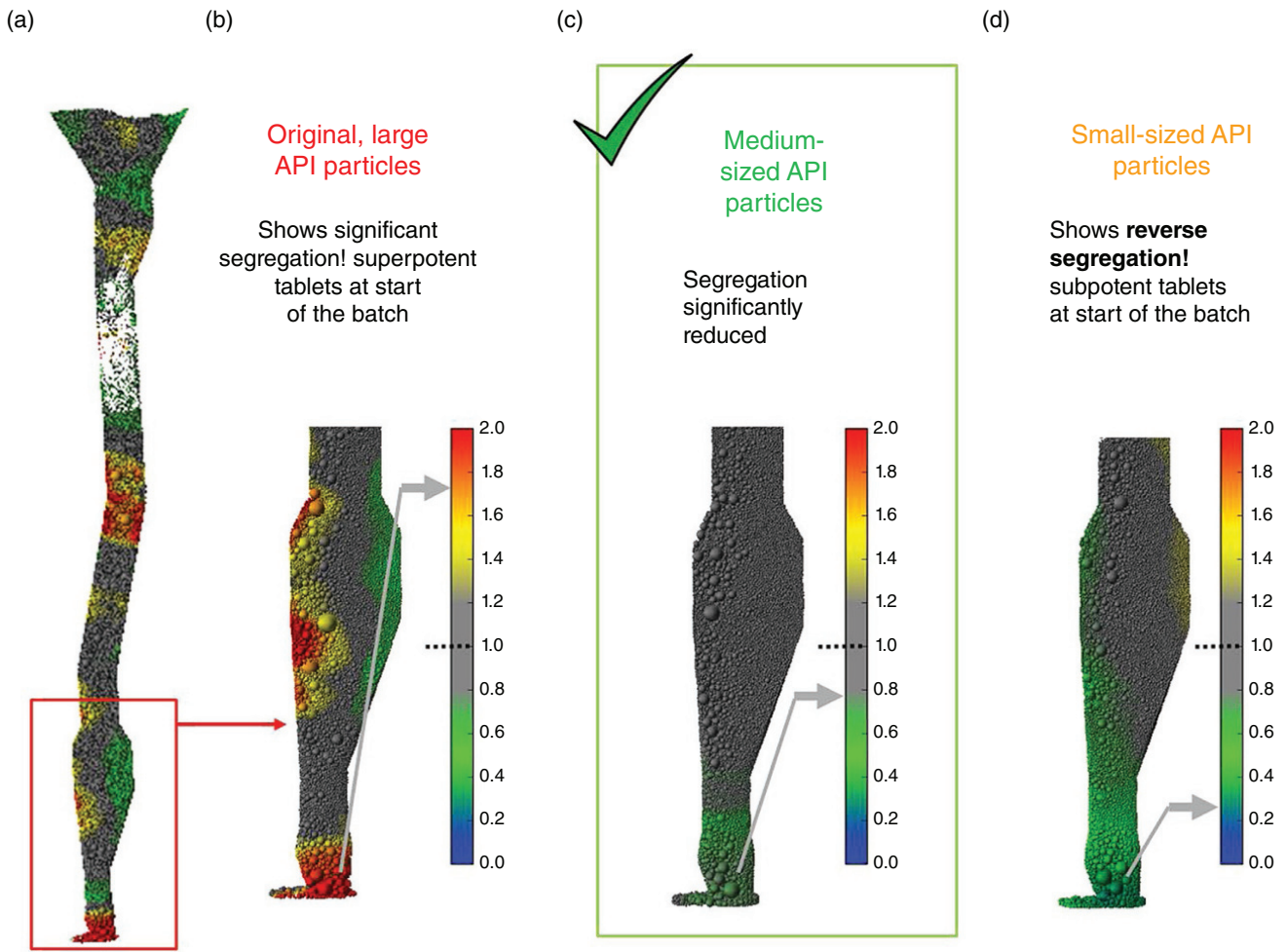


FIGURE 20.14 Simulation snapshots of (a) the complete powder transfer system and (b–d) close-ups of the tablet press feed hopper for the given API particle sizes. In each case, the particles are colored by the local API mass fraction normalized by the intended mass fraction [63].

When powders are exposed to shear within the process, the lubricant particles tend to be increasingly dispersed over the available surfaces of the particles within the blend. This lubrication effect often manifests itself in the final product as a decrease in the tensile strength of a compact or a reduction in the disintegration and/or dissolution rates. Figure 20.15 shows a dissolution profile for dosages manufactured from powder sheared to varying extents, given by the K value [68]. This lubrication sensitivity data shows that as the extent of shear applied increases (increasing K value), the dissolution becomes increasingly suppressed. The effect can be very significant. For example, the percent dissolved at $t = 30$ min can decrease from approximately 90 to approximately 30%. Therefore, to ensure consistent product quality, it is important to control or predict the extent of lubrication during processing.

In the following example, the shearing effects within the powder feed system of an encapsulator are examined. This particular system, shown in Figure 20.16, consists of a feed

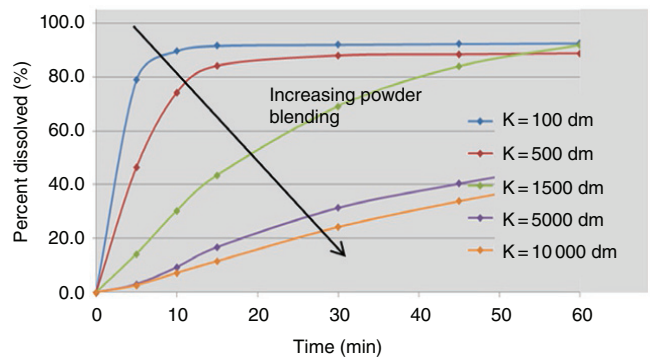


FIGURE 20.15 Dissolution profiles for capsules manufactured from blends having various extents of shear as given by the K value [68].

hopper with a rotating auger/impeller combination [68]. As the auger/impeller rotates, powder is dispensed into the encapsulator bowl, from which consolidated plugs of powder

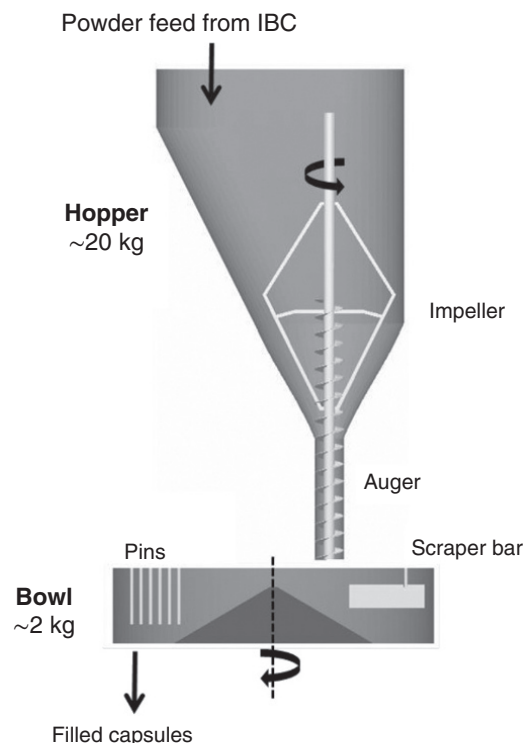


FIGURE 20.16 A schematic of an encapsulator powder feed system consisting of a feed hopper with a rotating auger/impeller that feeds powder to the bowl where it is formed into plugs and subsequently filled into capsules [68].

are formed and filled into capsules. However, the rotation of the impeller also imparts considerable shear into the powder. This effect was observed as a slowdown in the dissolution of filled capsules, particularly at the end of the batch. A DEM approach was used to better understand the powder flow kinematics in relation to the auger/impeller rotation.

The DEM approach is unable to directly simulate the lubricant particles due to their small size ($D[4,3] \approx 5\text{--}10\ \mu\text{m}$), which results in a very large number of particles, as well as the complex contact mechanics involving the deagglomeration of lubricant particles, their dispersion throughout the blend, and subsequent bonding to the surfaces of other particles. Instead, the total distance each particle has traveled is tracked as a cumulative measure of the shear strain in parallel to the experimental approaches described previously [66, 69]. In this way, the extent of shear the powder is exposed to can be characterized.

The DEM approach is used to simulate the original hopper design with a full impeller and four alternative designs all shown in Figure 20.17 (top row) [68]. Simulation snapshots with the particles colored by the distance they have traveled are shown in Figure 20.17 (bottom row) [68]. For the original design, the total distance traveled is extremely large (particles colored red), indicating a large extent of lubrication and a significant anticipated effect on the dosage form dissolution

following the lubrication sensitivity curves shown in Figure 20.15. For each of the design alternatives, the DEM simulations predict a smaller extent of lubrication (particles colored green and/or blue).

Based on this understanding of the powder lubrication in the feed hopper gained through the DEM simulations, modifications to the equipment design were carried out to alleviate suppression in dissolution that was observed at the end of a batch. The modifications to the equipment are shown schematically in Figure 20.18a where the original feed system is shown on the left and the modified feed system on the right [68]. Three modifications were made to reduce the impact of lubrication and shear on the drug product dissolution. First, the level of powder was reduced. This reduces the amount of powder exposed to shear at any given time. Second, the size of impeller was reduced to decrease the portion of powder exposed to shear during rotation. Importantly, the impeller was not eliminated completely due to concerns with powder potentially ratholing around the feed auger. Third, the design of the feed auger was adjusted from a constant volume per pitch design to an increasing volume per pitch design. This significantly improved the powder flow pattern within the hopper and helps to avoid a funnel flow pattern that can result in powder remaining resident in the hopper for long periods of time.

Each of the two geometries shown in Figure 20.18a was simulated using DEM. The mean distances traveled for each case, as measured at the hopper outlet, are shown in Figure 20.18b [68]. For the original equipment design, the mean distance traveled, or extent of shear, significantly increases at the end of the batch due to particles remaining resident in the hopper and exposed to shear for long durations prior to being discharged. In the case of the modified design, the three improvements described above significantly improve the powder flow pattern and reduce the exposure to shear. The extent of lubrication is significantly reduced at the end of the batch. While quantitative dissolution predictions are beyond the scope of this chapter, a sense of the impact on dissolution can be estimated through the use of the lubrication sensitivity shown in Figure 20.15. At the end of the batch, the K values for the original and modified equipment designs are 1100 and 350 dm, respectively. Referring to Figure 20.15, these correspond to approximately 75% dissolved at $t = 30$ min for the original design and 90% dissolved at $t = 30$ min for the modified design. Further details on quantitative predictions of lubrication effects are available elsewhere [70].

These examples have illustrated how the DEM approach can be used to understand issues in the blending and powder transfer processes, as well as visualize alternatives to either processing conditions or material properties to improve quality and process robustness. Oftentimes, the computational models do not need to precisely predict the phenomena of interest in quantitative way. Rather, value can be achieved

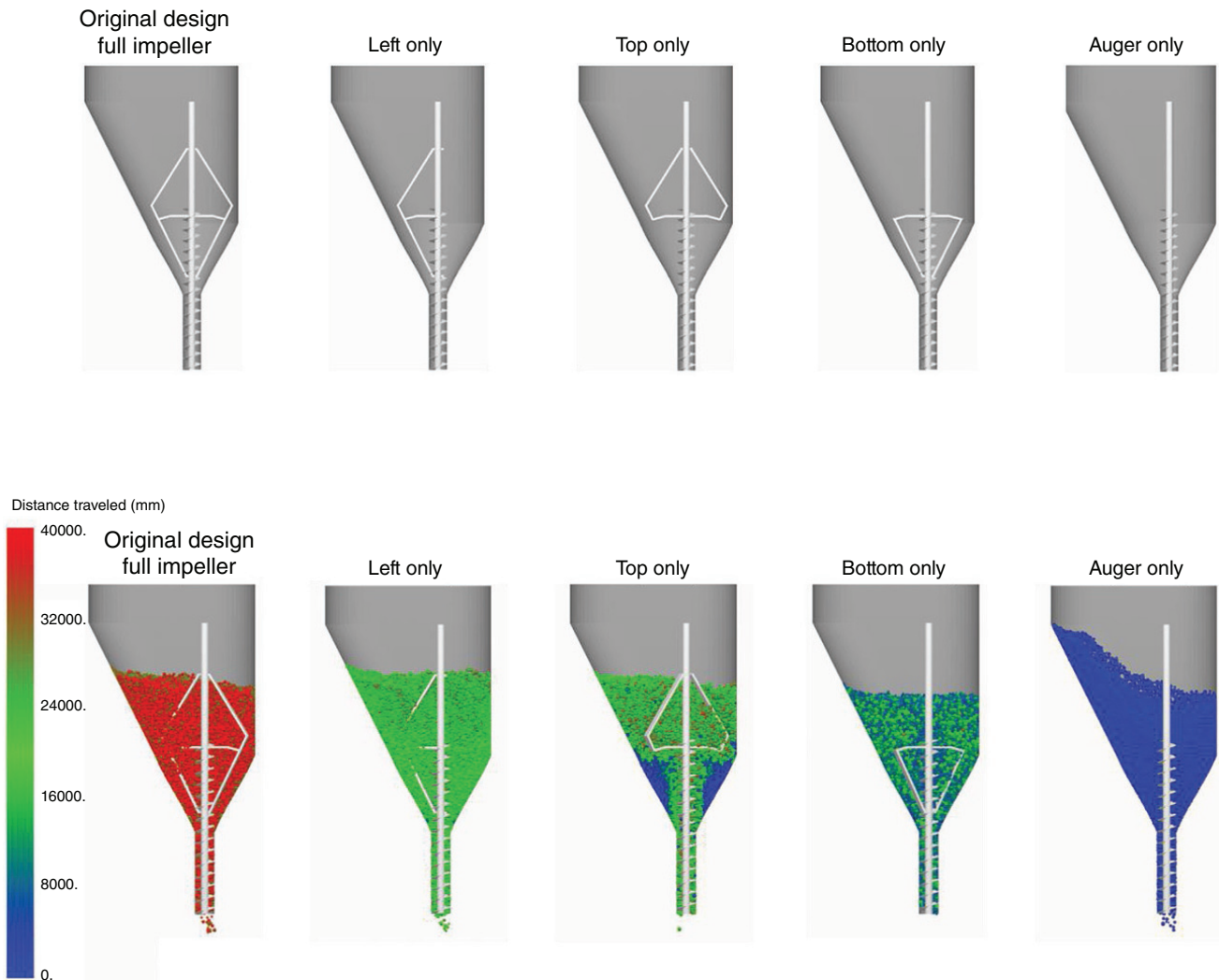


FIGURE 20.17 Schematics of the original agitated feed hopper design with the full impeller and four alternative designs. Each system (top row) was simulated in DEM with the particles colored by the respective distance traveled (bottom row) providing indication of the total extent of shear the particles have been exposed to [68].

by using the models, even with qualitative predictions, to support decision making during drug product process development and scale-up activities. The impact from these efforts includes accelerated drug product development timelines and improved consistency in product quality.

In this section, we have reviewed the importance of establishing and maintaining content uniformity and highlighted one of the methods by which segregation of materials can be modeled. In the next section, we discuss modeling of wet granulation, a process that helps to bind particles of different materials together, thereby reducing the potential for segregation in subsequent processing and handling operations and improving the likelihood of good content uniformity in the final dosage form.

20.4.2 Wet Granulation Process Models

20.4.2.1 Model Development for Wet Granulation Using Engineering Principles

Wet granulation is a particle size enlargement process that is commonly used in the manufacture of drug product dosage forms. There are several reasons to wet granulate pharmaceutical blends. Increasing the particle size will tend to improve flow. Fine powders usually have significant cohesive forces between the constituent particles, which act to retard flow. By enlarging the particle size, these cohesive forces become less significant compared with the particle mass, thereby improving the flowability of the bulk powder. Another reason to granulate includes reducing the potential for segregation of the API. The wet granulation

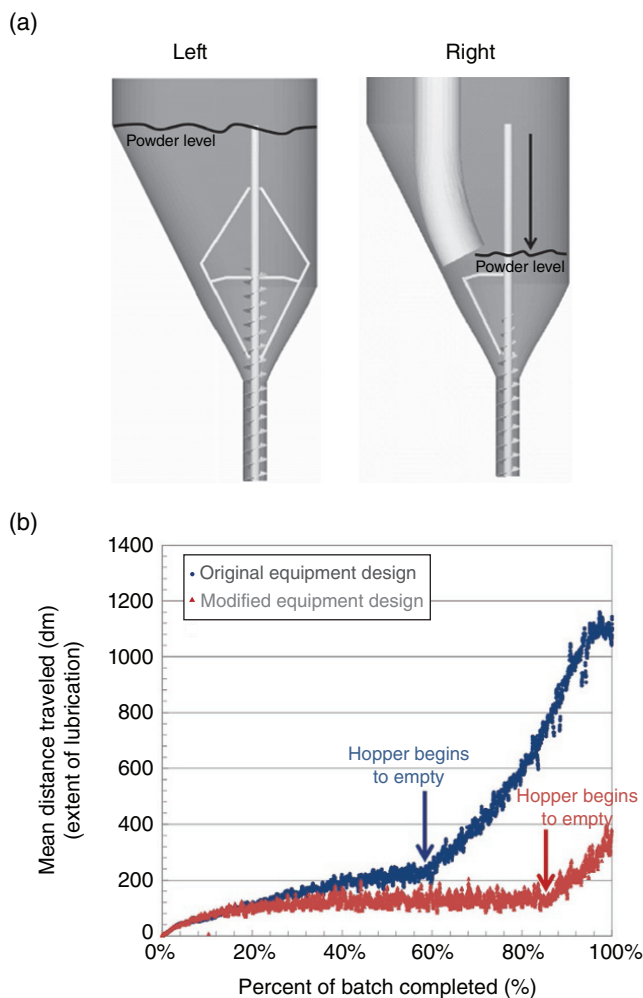


FIGURE 20.18 A comparison between the original equipment design on the left of (a) and the final modified equipment design on the right of (a). The effect of the throughput of the batch on the mean distance traveled for particles (extent of lubrication) for the original versus modified equipment designs is depicted in (b) [68].

process physically binds the blended particles together, thereby reducing the likelihood that one species will segregate and cause potential content uniformity problems.

Wet granulation processes are often carried out as batch processes using high-shear mixers as shown in Figure 20.19. While wet granulation can also be conducted in other equipment such as planetary mixers, fluidized beds, or extruders, high-shear mixers are the most common in the pharmaceutical industry and will be our focus in this section. In the high-shear mixer, a centrally located impeller (in this case a top-driven impeller is shown, although some high-shear mixers utilize a bottom-driven impeller) is used to mix and consolidate the granulation. A chopper (located on the left side of this schematic) spins at a high speed and helps to break up very large granules. Finally, a spray nozzle

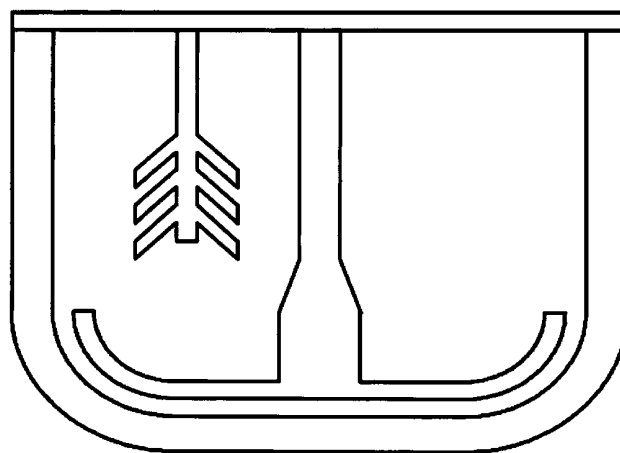


FIGURE 20.19 A schematic of a high-shear wet granulator. Source: Reprinted with permissions from Faure et al. [71], Copyright (1999) Elsevier.

(not shown in Figure 20.19) is used to add a liquid granulating agent. This liquid may contain a liquid binder or may consist only of water when a dry binder has already been added to the formulation.

Granulation in a high-shear mixer begins with the addition of the dry powder blend and dry mixing with the impeller for a short period of time. With both the impeller and chopper rotating, the liquid addition phase begins. After the desired amount of liquid has been added, the “wet massing” phase begins where the impeller and chopper continue to mix the granulation while the liquid addition is stopped. In general, the point at which to stop the granulation process (the process end point) is difficult to determine scientifically and is still a matter of ongoing research. In the past, skilled operators would deem a granulation complete if it passed the so-called squeeze test, where a small amount of material is squeezed in one’s hand and subjectively observed. Many different researchers have proposed various ways to monitor the high-shear wet granulation process in a more objective manner. Some of these approaches include impeller power or torque, offline measurement in a torque rheometer, and some more recent analytical techniques such as using near infrared (NIR) and focused beam reflectance measurement (FBRM). Several of these techniques are discussed further in a recent review [72].

A second area of difficulty with high-shear wet granulation processes revolves around process scale-up. Many of the process parameters such as impeller speed, water addition amount, water addition rate, and wet massing time are determined through experimental design of experiments (DOEs). However, the process dynamics change significantly as larger-scale granulators are used. Hence, similar DOEs are conducted at each of the scales during the process scale-up effort. These experiments consume significant labor

resources and also incur large raw material costs if granulations of proprietary APIs in limited quantities are being conducted. This is especially true at the largest of scales where batch sizes may be on the order of 1000 l. Thus the use of models to (i) predict process parameters for scaling up and (ii) determine when to stop the wet granulation process (end point) can be extremely useful.

20.4.2.2 Wet Granulation Scale-Up Process scale-up in pharmaceutical industry is driven by two important factors: (i) cost of API, which usually runs into several thousand dollars per kilogram of material, and (ii) tight product specifications as desired by various stages of clinical trials. The practical considerations demand that in a pharmaceutical industrial setting, a wet granulation process can be developed, which is cost efficient and robust and delivers products with high quality. The process development and scale-up from lab scale to pilot or commercial scale broadly takes place in two steps, which are described below [73]:

Formulation development and optimization: The process and formulations are developed and optimized in a small granulator, typically at 1 l lab scale using few hundred grams of API, through detailed experimentation. The main goal of this experimentation is to explore the design space of the process parameters such as the total amount of binder and its rate of addition, impeller speed, total processing time, dry powder fill height, end point, etc. This exploration should result in a process that yields granules with desired size and porosity distribution that can be compressed into tablets with the desired quality attributes including hardness-compression behavior and dissolution profiles.

Process development and scale-up: After the design space is explored and the formulation is optimized at lab scale, it is scaled up to pilot or production scale. Simply put, a process template is developed at lab scale, and it has to be replicated at a much larger scale, which results in product with similar quality attributes. Several routes to process scale-up have been demonstrated by various groups, mostly applicable to a specific set of granulators and limited formulations [72]. Almost all the procedures of scale-up are based on the similarity between the two granulators of interest. Formally, the similarity principle for a process is established using dimensional analysis. The application of the similarity principle begins with the recognition that any physical process can be represented by a dimensional relationship between n process variables and constants as shown below:

$$F(X_0, X_1, X_2, \dots, X_n) = 0 \tag{20.5}$$

The above relation can be reduced by applying Buckingham Π theorem, which simply states:

between $m = n - r$ mutually independent dimensionless groups

where

r = number of dimensional units, i.e. fundamental units (rank of the dimensional matrix).

Equation (20.5) can be reduced to the following relationship:

$$F(\Pi_0, \Pi_1, \dots, \Pi_n) = 0 \tag{20.6}$$

The above relationship can be rewritten by expressing first dimensionless group in terms of the rest of them as shown below:

$$\Pi_0 = f(\Pi_1, \Pi_2, \dots, \Pi_n) \tag{20.7}$$

It must be noted that the similarity analysis should be applied to processes where a clear understanding of process is established.

20.4.2.2.1 Scale-Up Approach 1 The earliest application of similarity analysis for the scale-up of wet granulation process was demonstrated by Leunberger and coworkers at the University of Basel and Sandoz AG [74–77]. The physical relationship used to describe the granulation process can be written as

$$P = f(\rho, D, \Omega, q, t_p, V_b, H, g) \tag{20.8}$$

The description of various physical quantities is shown in Table 20.5.

The above relationship is nondimensionalized using d , $1/\omega$, and ρd^3 as length, time, and mass scales, respectively. The dimensionless quantities are shown in Table 20.6.

These investigators performed wet granulation experiments using a placebo formulation (86% w/w lactose, 10% w/w corn starch, 4% w/w polyvinylpyrrolidone as binder) in mixers of planetary type (e.g. Dominici, Glen, and Molteni). The batch size ranged from 3.75 up to 60 kg. The impeller speed was scaled using a constant Froude

TABLE 20.5 List of Important Process Variables and Parameters That Define a Wet Granulation Process in Scale-Up Approach 1

No.	Quantity	Symbol	Units	Dimension
1	Power consumption	P	watt	ML^2T^{-3}
2	Specific density	P	kg/m ³	ML^{-3}
3	Impeller diameter	D	m	L
4	Revolution speed	Ω	rev/s	T^{-1}
5	Binder flow rate	Q	kg/s	MT^{-1}
6	Bowl volume	V_b	m ³	L^3
7	Gravitational constant	G	m/s ²	LT^{-2}
8	Bowl height	H	m	L
9	Process time	t_p	s	T

number ($d_1\omega_1^2 = d_2\omega_2^2$). The volume fraction and geometric ratio were also kept constant. It was seen that the power profile measured during the granulation can be divided into five different phases (S1–S5) as shown in Figure 20.20. It was found that the amount of binder liquid added during the process varies linearly with the batch as shown in Figure 20.21.

TABLE 20.6 Dimensionless Numbers and Groups Used in Scale-Up Approach 1

No.	Symbol	Dimensionless Group	Description
1	Π_0	$P/(d^5 \omega^3 \rho)$	Power number
2	Π_1	$qt_p/V_b\rho$	Specific amount of liquid binder
3	Π_2	V/d^3	Volume fraction of dry powder
4	Π_3	$(d \omega^2)/g$	Froude number (centrifugal force/gravitational force)
5	Π_4	H/d	Geometric ratio

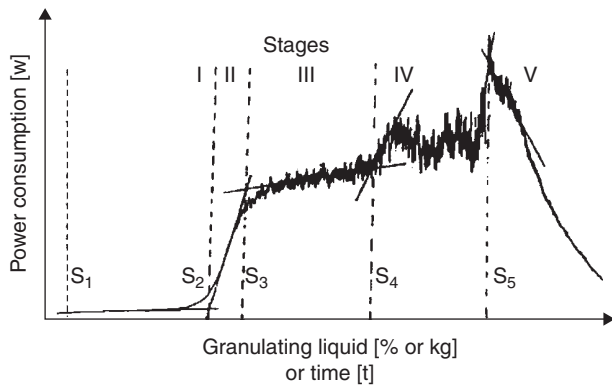


FIGURE 20.20 Division of a power consumption curve. *Source:* Reprinted from Leuenberger [76], Copyright (2001) Elsevier.

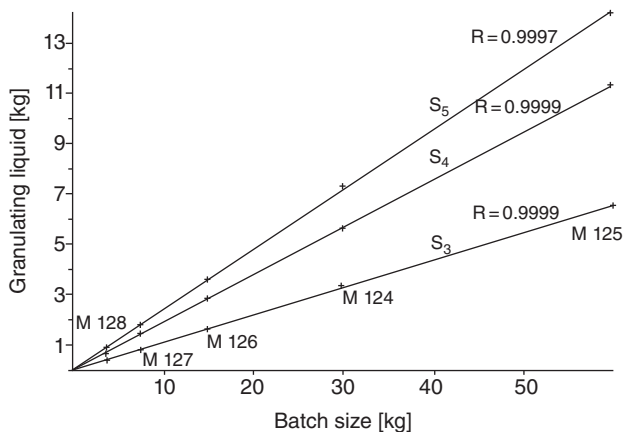


FIGURE 20.21 Scale-up precision measurements with identical charges [76]. *Source:* Reprinted from Leuenberger, H., *European Journal of Pharmaceutics and Biopharmaceutics* 52 (3), 279–288. Copyright (2001) Elsevier.

Therefore, the functional relationship between dimensionless groups is as follows:

$$\Pi_0 = f(\Pi_1) \tag{20.9}$$

From these findings, one can conclude that the correct amount of granulating liquid per amount of particles to be granulated is a scale-up variable. It is necessary, however, to mention that during this scale-up exercise, only a low-viscous granulating liquid was used. The exact behavior of a granulation process using high-viscous binders and different batch sizes is unknown. It is shown that the first derivative of the power consumption curve is a scale-up invariant, and it is proposed that it can be used as an in-process control or a fine-tuning of the correct amount of granulating liquid.

20.4.2.2.2 Scale-Up Approach 2 One of the key assumptions made in this study that the viscosity of the wet mass is unimportant may not hold true for many formulations and viscous binders. Rowe and coworkers [71, 77] developed a different approach to scale up wet granulation process. These authors defined the process by the following relationship:

$$\Delta P = f(\rho, R, \Omega, \mu, R_b, g, m) \tag{20.10}$$

The description of the various physical quantities is shown in Table 20.7. The above relationship can be nondimensionalized using d , $1/\omega$, and ρd^3 as length, time, and mass scales, respectively. The dimensionless quantities are shown in Table 20.8.

The power number is expressed as a function of the other dimensionless quantities as follows:

$$\text{Log}_{10}(N_P) = a \cdot \text{Log}_{10}(\psi \text{Re} \cdot \text{Fr} \cdot \text{fill ratio}) + b \tag{20.11}$$

TABLE 20.7 List of Important Process Variables and Parameters That Define a Wet Granulation Process in Scale-Up Approach 2

No.	Quantity	Symbol	Units	Dimension
1	Net power consumption	ΔP	watt	ML^2T^{-3}
2	Wet mass density	P	kg/m^3	ML^{-3}
3	Impeller radius	R	m	L
4	Revolution speed	Ω	rev/s	T^{-1}
5	Wet mass consistency/viscosity	μ	Nm	ML^2T^{-2}
6	Bowl radius	R_b	m	L
7	Gravitational constant	G	m/s^2	LT^{-2}
8	Amount of wet mass	M	kg	M

In the above equation, a and b are regression constants. Faure et al. [71] carried out wet granulation experiments of lactose and maize starch-based placebo formulations in a series of Collette Gral granulators with sizes of 8, 25, 75, and 600 L. They fitted the experimental data with Eq. (20.11) and found that the regression coefficient was $r^2 > 0.88$ using the data from the 8, 25, and 75 L bowls with PTFE lining and the 600 L bowl that did not require the lining (see Figure 20.22). The slope was found to be $a = -0.926$, and the intercept $b = 3.758$. This work shows that a nonlinear scale-up relationship exists between wet granulation carried out in geometrically similar granulators of different sizes. Moreover, this relationship can be effectively used when scaling up the process from one size to another size. In the next section, a case study is presented, which demonstrates the application of abovementioned scale-up approaches.

TABLE 20.8 Dimensionless Numbers and Groups Used in Scale-Up Approach 2

No.	Symbol	Dimensionless Group	Description
1	N_p	$\Delta P / (d^5 \omega^3 \rho)$	Power number
2	ψRe	$\rho R^2 \omega / \mu$	Pseudo-Reynolds number (inertial force/viscous force)
3	Fr	$(R \omega^2) / g$	Froude number (centrifugal force/gravitational force)
4	Fill ratio	$\rho R_b^3 / m$	Fill ratio (granulator volume/wet mass)

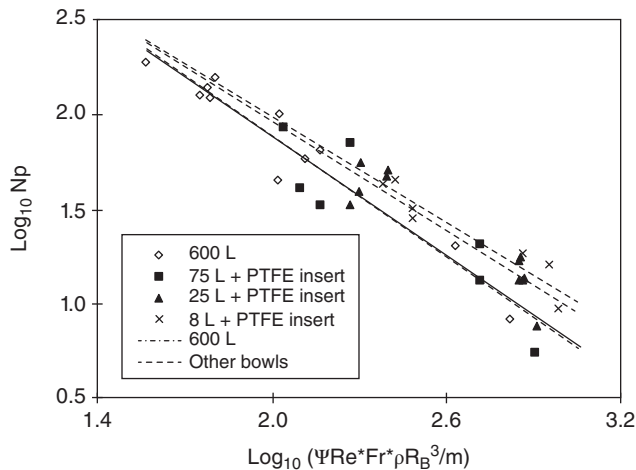


FIGURE 20.22 Dimensionless power relationship of the 600 L Collette Gral mixer–granulator. Source: Reprinted with permissions from Faure et al. [71], Copyright (1999) Elsevier.

EXAMPLE 20.2

Scale-Up of Wet Granulation Using Engineering Models

A wet granulation process is carried out for a placebo formulation with water as a binding liquid in an 8 L Collette Gral granulator. The process parameters are optimized for desirable end point through a DOE study. These parameters are listed in Table 20.9. This process is to be scaled up to a granulator with 75 L capacity such that the final product is of the same quality. In this case the product quality is identified by the specific density and wet mass consistency of granules at the end point. Hence, desirable end point is one that gives product with essentially the same specific density and wet mass consistency or viscosity obtained at 8 L scale. There are three main process parameters that are to be determined at 75 L scale: (i) impeller rotation speed, (ii) amount of water used, and (iii) power consumption near the end point. All these quantities will be calculated using above discussed scale-up approaches.

Solution

- a. The impeller rpm is scaled using constant impeller tip speed:

$$\Omega_2 = \Omega_1 \left(\frac{r_1}{r_2} \right) \tag{20.12}$$

$$\Omega_2 = 350 \left(\frac{0.12}{0.25} \right) = 164 \text{ rpm}$$

TABLE 20.9 Process Variables and Parameters for Two Different Collette Gral Bowl Sizes Used in Example 20.2

Process Parameters and Variables	Units	8 1 Bowl	75 1 Bowl
Impeller radius	m	0.119	0.254
Bowl radius	m	0.123	0.262
Revolution speed	1/s	5.83	2.73
Revolution speed	1/min	350.0	164.0
Dry powder weight	kg	1.50	14.00
Total water added	kg	0.50	4.67
Gravitational constant	N/m ²	9.81	9.81
Bulk density	kg/m ³	400.0	400.0
Viscosity	N/m	0.30	0.30
Power consumption	watt	127.7	3503
Froude number		0.41	0.19
Pseudo-Reynolds number		110.1	235.1
Fill ratio		0.37	0.38
Power number		67.41	405.8
Log(N_p)		1.83	2.61
Log($Re * Fr * \text{Fill Ratio}$)		1.23	1.24
Scale-up constant a		-0.93	-0.93
Scale-up constant b		3.76	3.76

Bolded values are calculated by scale-up rules.

- b. The amount of water used at 75 l scale is determined by assuming water to dry powder weight ratio is invariant across the scales:

$$\begin{aligned} \text{Water used at 75 l} &= \frac{(\text{amount of water used at 8 l})(\text{dry powder mass at 75 l})}{(\text{dry powder weight at 8 l})} \\ \text{Water used at 75 l} &= \frac{(0.5 \text{ kg})(14 \text{ kg})}{(1.5 \text{ kg})} = 4.67 \text{ kg} \end{aligned} \quad (20.13)$$

- c. Finally, the power consumption at the end point is determined using scaling relationship given by Eq. (20.11) as follows:

$$\begin{aligned} Fr &= 0.19 \\ \psi Re &= 235.1 \\ \text{Fill ratio} &= 0.38 \\ \log_{10}(N_p) &= -0.926 \times \log_{10}(0.19 \times 235.1 \times 0.38) + 3.758 \\ N_p &= 405.8 \\ \Delta P &= N_p \times (\rho \Omega^3 R^5) = 3503 \text{ watts} \end{aligned} \quad (20.14)$$

The process parameters for both scales are listed in Table 20.9. The power calculated using the above scale-up approach is used to guide determination of the wet granulation end point. However, it must be kept in mind that this scale-up approach, like any other approach based on dimensional analysis, is semiempirical in nature and needs some experimental work to achieve optimum scale-up and process design. This approach should be contrasted from a process model based on fundamental principles that requires material properties and process parameters to achieve optimum design.

20.4.2.3 Fluidization Regime During Granule Drying In the area of fluid bed drying of granules produced by wet granulation processing, there is an engineering approach to assess fluidization regimes a priori if the PSD and equipment airflows are known. A semiempirical approach, in that empirical heat transfer data is required for the simulations to compare favorably with experiment, has been used to predict process parameters for fluidization properties in a fluid bed dryer. Based on the mean diameter of the granule distribution, the process map shows that the proposed equipment is adequate to fluidize the granules in the desired bubbling regime (Figure 20.23). Granule characteristics, such as moisture content and granule size distribution, are known for a particular product entering the fluid bed dryer. The range of volumetric flow rates of the drying gas and the dimensions of the air inlet were obtained from the equipment manufacturer.

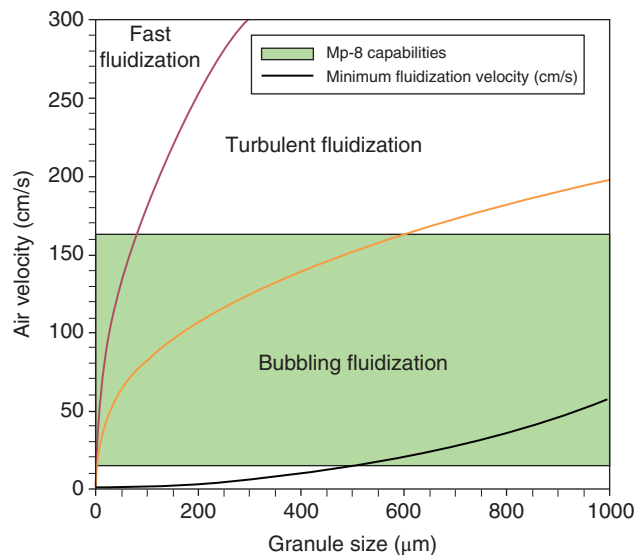


FIGURE 20.23 Engineering model of granule fluidization using empirical relationships found in literature based on granule distribution and equipment airflow [14, 15]. *Source:* Courtesy of D.M. Kremer.

Finally, a review of the literature indicated that researchers had constructed models to predict both the minimum fluidization velocity and the transitions to turbulent and fast fluidization.

The granules produced through wet granulation and dried through fluid bed drying are finally compressed into tablets (Figure 20.1). The focus of this section was on the use of modeling to achieve consistent granulation properties on scale-up to assure consistent input to the tableting process. Tableting is an important unit operation for solid dosage manufacturing since it defines the dosage strength and performance. In the next section some of these models employed to understand compaction of blends into tablet will be discussed in detail.

20.4.3 Tablet Compression Models

20.4.3.1 Principles of Finite Element Analysis (FEA) A finite element analysis or method (FEA/FEM) is a numerical approach to solve a partial differential equation. It is widely used in engineering and science as many physical phenomena can be described in terms of a partial differential equation. The technique consists of the following steps:

1. Subdividing the problem domain (or geometry) into finite elements connected together by nodes. These finite elements are commonly termed as a *mesh*.
2. Developing equations (such as force and mass balance) for each element and then assembling them for the entire domain or system of elements.

3. Solving the resulting system of equations.
4. Analyzing quantities of interest such as stresses and strains and obtaining visualizations of the response of the system to the applied loadings.

Powder Compaction: Tablet compression is an important unit operation in the pharmaceutical industry as it significantly affects the mechanical strength and relative density of the drug product. FEA are a common method that is employed to study the tableting process where the formulation powder is assumed to be a continuum material [78–81]. The approach is based on the following components:

- Continuity equations (e.g. conservation of mass).
- Equilibrium equations (e.g. force balance on the material).
- Initial and boundary conditions of the problem.
- Dynamics of the loading and the geometry of the problem.
- Constitutive behavior of the powder (stress–strain relationships).

Due to the availability of powerful, inexpensive computers and commercial finite element software, the continuity and equilibrium equations can be solved accurately and quickly after the appropriate boundary conditions are defined. Moreover, it is possible to define a complex sequence of loading and unloading steps such as the compression, decompression, and ejection during model setup. However, it is a challenge to obtain model inputs to the FEA solver [77–81] such as the constitutive relationships

of the formulation powder (e.g. stress–strain relationships) and friction between powder and die wall to get accurate estimation of the powder stress levels during tableting. The following section highlights some of the commonly used constitutive relationships for the powder continuum.

20.4.3.2 Tablet Finite Element Analysis Powder Material Models: The stress–strain relationships for powders were originally developed for classical soil mechanics applications and were used to simulate compaction of ceramic powders. They were assumed to be elastic–plastic materials, and appropriate relationships were developed to describe the yield surface of the material. There are a number of phenomenological models to describe the yield surface of the powder materials such as the Gurson model [17], Cam-Clay model [18], and Drucker–Prager cap plasticity (DPC) model [19, 20]. The DPC model has been widely used [78, 82–85, 91] in comparison with the other models for two main reasons: firstly, it can efficiently capture the shear failure during the decompression and ejection phase of the tableting process, and secondly, it is because experiments on real powders can be designed to efficiently characterize its parameters. The stress–strain relationship and the yield loci used in the DPC model are shown in Figure 20.24. The yield surface consists of three segments: a shear failure surface, a “cap” surface that represents plastic compaction or inelastic hardening, and a transition surface between them. The transition surface is introduced for smooth numerical implementation. For a detailed description of the equations of the different yield surfaces of the DPC model and the experimental procedure to obtain its parameters, refer to Han et al. [79] or Cunningham et al. [78].

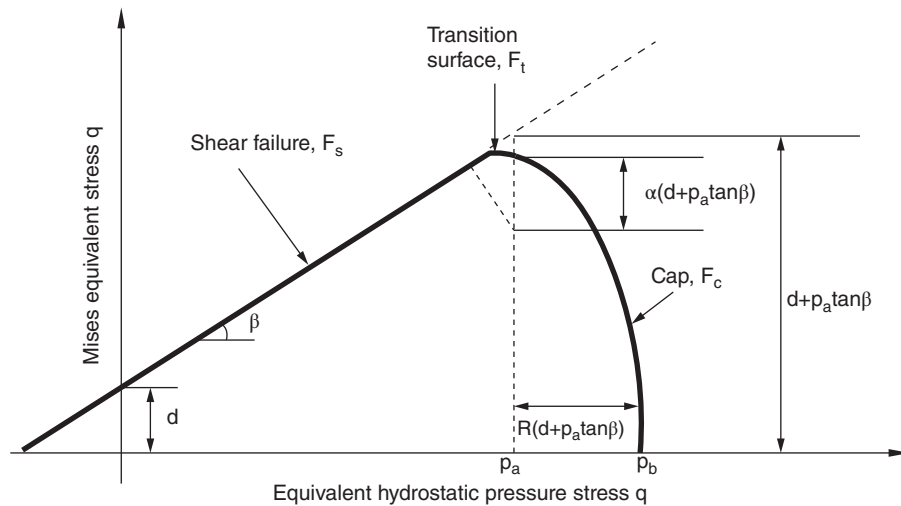


FIGURE 20.24 Drucker–Prager cap model: yield surface in the p – q plane. *Source:* Reprinted with permissions from Han et al. [79], Copyright (2008) Elsevier.

Applications: FEA performed using the DPC material model has been used to study the elastic recovery or “spring-back” of material during the compression and ejection phases of tableting that leads to capping incidence in tablets [82, 85]. The relative density distribution after ejection [81, 85] and temperature distribution in the compact and tooling during compaction [86] have also been investigated using FEA methods. Moreover, the stresses during the different tableting stages (see Figure 20.25) have been analyzed and correlated to possible tablet failure mechanics [79, 85].

This section summarized the concept of applying a continuum-based finite element model to predict the stresses on a tablet during tableting operations. Further investigation is necessary to establish if these predicted stresses and density distributions by FEA have an implication in understanding possible failures that might occur in other unit operations, such as during tablet film coating. The next

section explains the tablet coating operation and available engineering models used to predict the coating uniformity.

20.4.4 Tablet Film Coating Models

20.4.4.1 Model Development for Film Coating Using First Law of Thermodynamics Tablet film coating is a widely used unit operation within the pharmaceutical industry for applying both aesthetic and functional coatings on tablets. Color coating is often used in combination with tablet shape to enable manufacturers, pharmacists, and patients to distinguish between different products as well as different dosage strengths within the same product. From a physical standpoint, a thin film coat layer can improve the mechanical integrity of the tablets and also make them smoother, which improves tablet flowability, enhances packaging efficiency, and increases palatability for patients. Film coating can

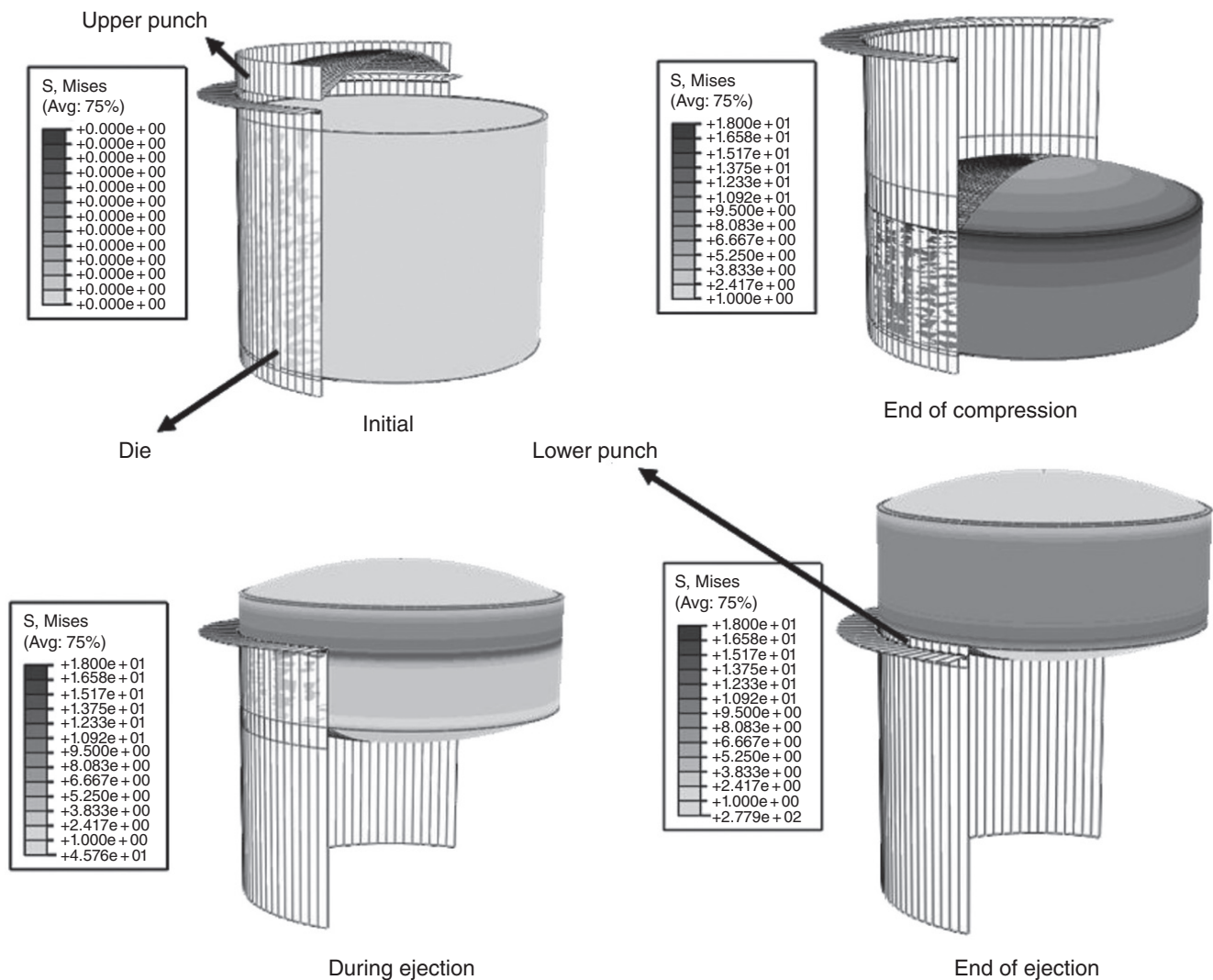


FIGURE 20.25 Stress distributions on the tablet during compaction using a standard round concave (SRC) punch.

improve functionality by providing a barrier against environmental exposure to moisture, light, or air. This can enhance product stability and reduce the requirement for more expensive packaging materials. In some cases the color of a tablet can be an important part of brand recognition and can even be used to trademark.

Typically coating formulations can either be aqueous or solvent-based systems. While aqueous coatings are rapidly becoming the preferred method for many applications, solvent-based coatings are still used to apply many functional coatings onto tablet cores for controlled drug delivery (e.g. semipermeable or delayed/sustained release membranes) [87]. Although solvent-based systems hold a number of advantages in terms of application and flexibility, the move to aqueous coatings has largely been driven by factors such as cost, safety concerns, more stringent regulations on effluent discharge, and the broad variety of film formulations that have been developed recently for aqueous application [88].

The film coating process can essentially be considered an adiabatic evaporative cooling process. The driving energy for fluid evaporation is a combination of the airflow volume, temperature, and moisture content of the air. This can be considered as the bulk gas phase in total. The underlying law controlling the thermodynamic environment within the process is the first law of thermodynamics (conservation of energy). It is important to understand that the three principles of driving energy are linked by the operating parameters within the process. For example, an increase in the inlet air temperature will lower the relative humidity of the drying air into the pan (although the absolute water content remains the same). Increasing the spray rate of the coating solution will increase the moisture content of the air in the pan to which the tablets are exposed. Once stabilized, however, the process will remain in equilibrium unless disturbed by the alteration of a process parameter or external condition. Since the quality of the overall coating is greatly influenced by the thermodynamic conditions inside the pan, it is of great importance to understand these relationships and how to control them.

Thermodynamic models utilizing material and energy balances have been used in the past with some success to model the aqueous film coating process [24–26, 88, 89]. These models are used to predict the key process parameters that impact the quality of the film coat, mainly exhaust air temperature and exhaust air relative humidity. Film coating models are particularly important in pharmaceutical development where process conditions vary greatly for the purpose of design of experiment, scale-up, and coating formulation changes. Most of the previous models have been restricted to aqueous film coatings, and one coating pan type or scale, making them limited in their scope and applicability. More recently, am Ende and Berchielli developed a universal thermodynamic model that is applicable to both aqueous and organic film coating systems to aid in process optimization and scale-up [21]. This

model will be the basis for the discussion and calculations laid out in the following section. Whether the system in question is aqueous or organic based, the film coating process is a delicate balance that requires a high level of control over the process conditions to produce films that provide the required aesthetics or functionality.

20.4.4.2 Model System, Assumptions, and Limitations A typical tablet film coater schematic is shown in Figure 20.26 for a perforated coating pan. The drying air is heated by an external heating source to a target temperature, $T_{\text{air,in}}$, which is typically controlled by the operator to target a desired exhaust air temperature, $T_{\text{air,out}}$. The airflows through the tablet bed where it serves to dry the damp tablets as the coating solution droplets impact and spread on the tablet face. The air exits the pan through the exhaust air duct at a lower temperature due to the evaporative cooling effect from the volatile components (aqueous or organic) in the coating solution. The coating solution is supplied to the spray nozzle(s) by a pump where it is atomized by a compressed airstream into a pattern of tiny droplets that are propelled from the nozzle toward the tablet bed.

The rotating tablet bed defines the control volume for the material and energy balances. The model applies to steady-state conditions where the heat, temperature, and mass do not change with time:

$$\frac{dq}{dt} = \frac{dT}{dt} = \frac{dm}{dt} = 0 \quad (20.15)$$

The inlet streams to the system include the drying air and the film coating solution. The compressed airstream is neglected in the overall airflow through the pan since it is only a minor component compared with the drying air. The outlet stream consists of the drying air exhaust, and it is assumed that the volatile components in the coating solution exit the coater through this stream as vapor (i.e. the tablets do not retain any moisture). This is a reasonable assumption since coating pans are typically overdesigned with regard to drying capacity. The model was developed for a closed but not isolated system since energy exchange occurs as heat loss from the pan during operation and sampling. The model also neglects the humidity from the compressed air as well as the sensible heat term for the solid components of the coating solution.

This film coating model is a macroscopic analysis of the process, and, therefore, the physical operating conditions such as the spray gun-to-bed distance, pan speed, pan load, and spray-zone coverage are not considered. It is well known that these parameters also influence key tablet attributes such as coating uniformity and elegance, and, therefore, they should be monitored carefully during process optimization.

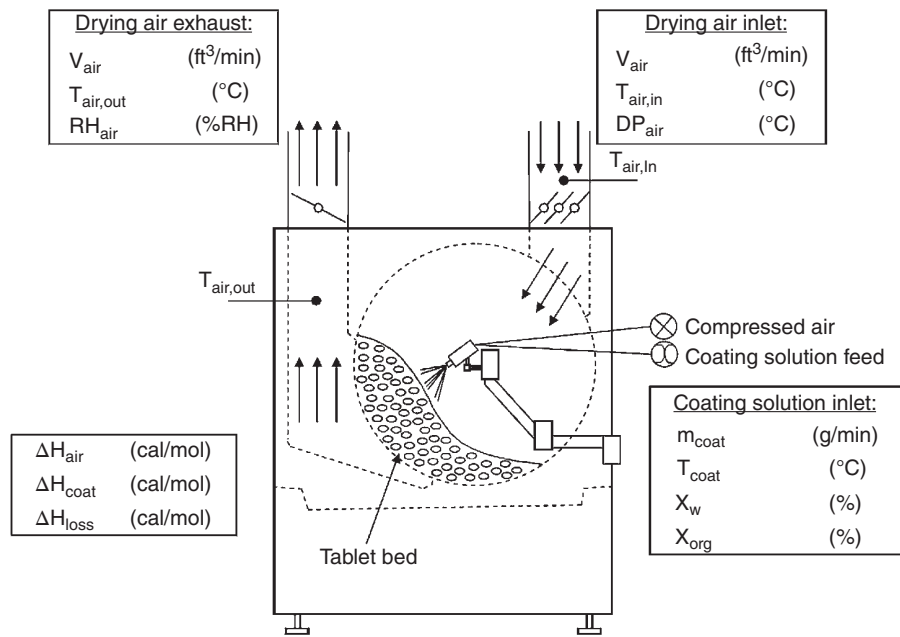


FIGURE 20.26 Schematic of a typical perforated tablet film coating pan.

20.4.4.3 Material Balance The material balance for the tablet bed control volume can be expressed in terms of each of the three components involved (e.g. water, organic solvent, and air). Assuming no reaction or accumulation, the total mass entering the system should be equal to the total mass exiting the system:

$$\sum m_{\text{in}} + \sum m_{\text{coat}} = \sum m_{\text{out}} \quad (20.16)$$

The material balance for water takes into account the humidity from the inlet drying airstream as well as the water in the coating solution:

$$m_{w,\text{in}} + m_{w,\text{coat}} = m_{w,\text{out}} \quad (20.17)$$

where

$m_{w,\text{in}}$ is the mass flow rate of water in the inlet airstream.

$m_{w,\text{coat}}$ is the mass flow rate of water in the coating solution.

$m_{w,\text{out}}$ is the mass flow rate of water in the outlet airstream.

For coating formulations with organic components, the material balance can be expressed as

$$m_{\text{org,coat}} = m_{\text{org,out}} \quad (20.18)$$

where

$m_{\text{org,coat}}$ is the mass flow rate of organic in the coating solution.

$m_{\text{org,out}}$ is the mass flow rate of organic in the outlet airstream.

Neglecting the contribution from the compressed air line to the spray nozzle, the material balance for air can be expressed as follows:

$$m_{\text{air,in}} = m_{\text{air,out}} = V_{\text{air,in}} \text{ ft}^3/\text{min} \times \left(\frac{28.31/\text{ft}^3 \times 29 \text{ g/mol}}{22.41/\text{mol}} \right) \times \left(\frac{273 \text{ K}}{273 \text{ K} + T_{\text{air,in}}} \right) \quad (20.19)$$

where

$m_{\text{air,in}}$ is the mass flow rate of air in the inlet airstream.

$m_{\text{air,out}}$ is the mass flow rate of air in the outlet airstream.

$V_{\text{air,in}}$ is the volumetric flow rate of air in the inlet airstream.

This equation converts the volumetric flow rate of the inlet airstream to a mass flow rate and incorporates a correction to account for change in molar volume of air due to the elevated temperature, $T_{\text{air,in}}$, at the inlet flow meter.

20.4.4.4 Energy Balance The overall energy balance for the system can be expressed based on the first law of thermodynamics as follows:

$$\Delta H = \Delta H_{\text{air}} + \Delta H_{\text{coat}} + \Delta H_{\text{loss}} = 0 \quad (20.20)$$

where

ΔH is the overall enthalpy change across the control volume.

ΔH_{air} , ΔH_{coat} , and ΔH_{loss} represent the enthalpy change across the control volume due to the drying airflow, the coating solution, and heat loss, respectively.

The individual enthalpy terms can be expressed in terms of a sensible heat term and, where applicable, a latent heat of vaporization:

$$\Delta H = mC_p\Delta T + m\Delta\hat{H}_{\text{vap}} \quad (20.21)$$

The enthalpy change for the airflow then becomes

$$\Delta H_{\text{air}} = m_{\text{air,in}}C_{p,\text{air}}(T_{\text{air,out}} - T_{\text{air,in}}) \quad (20.22)$$

where

$C_{p,\text{air}}$ is the heat capacity of air.

$T_{\text{air,out}}$ is the temperature of the exhaust air.

$T_{\text{air,in}}$ is the temperature of the inlet air.

Since the tablet bed temperature, T_{tablet} , typically is not measured during normal operation, it is assumed to be the same or similar as the exhaust air temperature. Therefore, the sensible heat term in the coating solution enthalpy change is defined in terms of the temperature difference between the exhaust air and the coating solution temperature, T_{coat} . Thus, the enthalpy change for the coating solution can be written as

$$\begin{aligned} \Delta H_{\text{coat}} = & m_{w,\text{coat}}C_{p,w}(T_{\text{air,out}} - T_{\text{coat}}) + m_{w,\text{coat}}\Delta\hat{H}_{\text{vap,w}} \\ & + m_{\text{org,coat}}C_{p,\text{org}}(T_{\text{air,out}} - T_{\text{coat}}) + m_{\text{org,coat}}\Delta\hat{H}_{\text{vap,org}} \end{aligned} \quad (20.23)$$

$$m_{w,\text{coat}} = \chi_w m_{\text{coat}} \quad (20.24)$$

$$m_{\text{org,coat}} = \chi_{\text{org}} m_{\text{coat}} \quad (20.25)$$

where

$$T_{\text{air,out}} = \frac{m_{\text{air,in}}C_{p,\text{air}}T_{\text{air,in}} + \chi_w m_{\text{coat}}C_{p,w}T_{\text{coat}} - \chi_w m_{\text{coat}}\Delta\hat{H}_{\text{vap,w}} + \chi_{\text{org}} m_{\text{coat}}C_{p,\text{org}}T_{\text{coat}} - \chi_{\text{org}} m_{\text{coat}}\Delta\hat{H}_{\text{vap,org}} + \text{HLF} \times T_{\text{RT}}}{m_{\text{air,in}}C_{p,\text{air}} + \chi_w m_{\text{coat}}C_{p,w} + \chi_{\text{org}} m_{\text{coat}}C_{p,\text{org}} + \text{HLF}} \quad (20.29)$$

The thermodynamic film coating model detailed above provides a direct relationship between the inlet air temperature, drying air flow rate, coating solution spray rate, and composition to the temperature of the exhaust airstream. Once the HLF is determined for a specific coating pan, the model can be used to predict the exhaust air temperature

$C_{p,w}$ is the heat capacity of water.

$C_{p,\text{org}}$ is the heat capacity of the organic component.

$\Delta\hat{H}_{\text{vap,w}}$ is the latent heat of vaporization for water.

$\Delta\hat{H}_{\text{vap,org}}$ is the latent heat of vaporization for the organic component.

T_{coat} is the temperature of the coating solution (assumed to be room temperature).

x_w is the mass fraction of water in the coating solution.

x_{org} is the mass fraction of organic in the coating solution.

m_{coat} is the mass flow rate of the coating solution.

The enthalpy change for the heat loss to the surroundings can be expressed as

$$\Delta H_{\text{loss}} = h_{\text{loss}}A(T_{\text{air,out}} - T_{\text{RT}}) \quad (20.26)$$

where

h_{loss} is the heat transfer coefficient.

A is the surface area for heat loss.

T_{RT} is the room temperature.

Since the heat transfer coefficient and surface area for heat loss can differ greatly between coating pans, these two terms are lumped together into an empirically determined HLF:

$$\Delta H_{\text{loss}} = \text{HLF}(T_{\text{air,out}} - T_{\text{RT}}) \quad (20.27)$$

The overall energy balance can be obtained by substituting the individual enthalpy terms into Eq. (20.20) as follows:

$$\begin{aligned} \Delta H = & m_{\text{air,in}}C_{p,\text{air}}(T_{\text{air,out}} - T_{\text{air,in}}) + \chi_w m_{\text{coat}}C_{p,w}(T_{\text{air,out}} - T_{\text{coat}}) \\ & + \chi_w m_{\text{coat}}\Delta\hat{H}_{\text{vap,w}} + \chi_{\text{org}} m_{\text{coat}}C_{p,\text{org}}(T_{\text{air,out}} - T_{\text{coat}}) \\ & + \chi_{\text{org}} m_{\text{coat}}\Delta\hat{H}_{\text{vap,org}} + \text{HLF}(T_{\text{air,out}} - T_{\text{RT}}) = 0 \end{aligned} \quad (20.28)$$

The energy balance equation can then be rearranged to solve for the unknown exhaust air temperature:

based on the operating conditions of the coater. The percent relative humidity of the exhaust airstream ($\%RH_{\text{out}}$) can be calculated based on the material balance for water around the control volume and then taking the ratio of the partial pressure of water vapor in the exhaust air ($P_{w,\text{out}}$) to the

vapor pressure of water at the exhaust air temperature ($P_{w,T_{air,out}}^*$):

$$\%RH_{out} = \frac{P_{w,out}}{P_w^*(at\ T_{air,out})} \times 100\% \quad (20.30)$$

20.4.4.5 Prediction of Target Film Coating Parameters Using Thermodynamic Model Before the predictive capabilities of the thermodynamic film coating model can be utilized, a value for the HLF must be determined for the specific coating pan in question. The model outlined in Eq. (20.29) has two unknowns, the HLF and the exhaust air temperature, $T_{air,out}$. The HLF can be determined empirically by comparing Eq. (20.29) to a set of experimental data where the exhaust air temperature has been measured. The HLF is used as a variable fitting parameter to minimize residual sum of squared error between the experimental data and the predicted exhaust air temperature from the model. This can be done with any simple optimization function such as Solver in Microsoft Excel[®]. A sample data set with an optimized HLF calculation for a Vector HCT-30 model film coater is shown in Table 20.10. The data includes both aqueous and organic coating formulations for a wide range of operating conditions (inlet temperature, airflow, and spray rate). By minimizing the sum of squared error between the actual and predicted exhaust air temperature, the HLF was determined to be 150 cal/min °C.

Once the HLF is determined, the model can be used for process optimization and scale-up predictions for that specific pan. For example, an operator can see how a change in spray rate or drying airflow will affect the exhaust air temperature of the process. The operator can also determine what inlet air temperature set point will be required to target a specific exhaust air temperature in the coater.

EXAMPLE 20.3

Thermodynamic Film Coating

A small-scale Vector LDCS-5 film coater has a heat loss factor (HLF) of 282 cal/min °C. (i) Determine the required inlet air temperature set point to achieve a target an exhaust air temperature of 50 °C at an airflow of 40 ft³/min and a spray rate of 4 g/min. The coating solution is an aqueous formulation with 20 wt % solids, and the room temperature is 22 °C. The heat capacity of air and water and the latent heat of vaporization of water can be readily found in any physical chemistry textbook. (ii) What is the relative humidity of the exhaust air at these conditions if the dew point of the inlet air is 10 °C?

Solution

- a. Equation (20.29) cannot easily be rearranged to solve for the inlet air temperature, $T_{air,in}$, since this term is embedded in the denominator of Eq. (20.19) to calculate the mass flow rate of air in the coater. Therefore, the equation must be solved iteratively by guessing a value of $T_{air,in}$ and solving for $T_{air,out}$. Since we know $T_{air,in}$ must be higher than $T_{air,out}$, a good initial guess might be 20 °C higher than the target exhaust temperature (i.e. $T_{air,in} = 70$ °C). Thus, from Eq. (20.19),

$$m_{air,in} = 40\text{ ft}^3/\text{min} \times \left(\frac{28.31\text{ l/ft}^3 \times 29\text{ g/mol}}{22.41\text{ l/mol}} \right) \times \left(\frac{273\text{ K}}{(273 + 70)\text{ °C K}} \right) = 1166.45\text{ g/min}$$

and from Eq. (20.29),

TABLE 20.10 Sample Data Set with Heat Loss Factor Determination for an HCT-30 Film Coater

Trial	Acetone (%)	Water (%)	Room Temp (°C)	Inlet Temp (°C)	Drying Airflow (ft ³ /min)	Spray Rate (g/min)	Actual $T_{air,out}$ (°C)	Predicted $T_{air,out}$ (°C)	Difference Predicted–Actual
1	85	5	22	41	32	20	27.6	26.3	–1.3
2	85	5	18	61	24	22	30.3	31.1	0.8
3	85	5	18	50	37	21	31.4	31.0	–0.4
4	85	5	18	49	37	21	30.3	30.9	0.6
5	0	94	18	74	38	7	45.2	44.2	–1.0
6	0	94	18	77	38	8	44.9	45.5	0.6
7	0	94	18	75	38	8	44.6	43.7	–0.9
8	0	94	18	71	38	7	41.6	42.6	1.0
9	0	94	21	78	35	8	44.1	44.6	0.5
10	0	94	18	79	35	8	45.0	44.4	–0.6
Group average									–0.06
Standard deviation									0.85
Sum of squares									6.58
HLF (cal/min °C)									150

$$T_{\text{air,out}} = \frac{(1166.45 \text{ g/min})(0.238 \text{ cal/g}^\circ\text{C})(70^\circ\text{C}) + 0.8(4 \text{ g/min})(1.0 \text{ cal/g}^\circ\text{C})(22^\circ\text{C})}{(1166.45 \text{ g/min})(0.238 \text{ cal/g}^\circ\text{C}) + 0.8(4 \text{ g/min})(1.0 \text{ cal/g}^\circ\text{C}) + 282 \text{ cal/min}^\circ\text{C}}$$

$$+ \frac{-0.8(4 \text{ g/min})(540 \text{ cal/g}) + (282 \text{ cal/min}^\circ\text{C})(22^\circ\text{C})}{(1166.45 \text{ g/min})(0.238 \text{ cal/g}^\circ\text{C}) + 0.8(4 \text{ g/min})(1.0 \text{ cal/g}^\circ\text{C}) + 282 \text{ cal/min}^\circ\text{C}}$$

$$T_{\text{air,out}} = 42.6^\circ\text{C}$$

Based on this result the inlet air temperature must be higher than 70 °C. Increasing the estimate to $T_{\text{air,in}} = 85^\circ\text{C}$ gives

$$m_{\text{air,in}} = 1117.57 \text{ g/min}$$

and

$$T_{\text{air,out}} = 49.3^\circ\text{C}$$

This could be iterated further to get an exhaust air temperature closer to 50 °C, but this estimate is well within the accuracy of the model and the variability of the actual equipment. Therefore, the inlet air temperature set point required to achieve a target exhaust air temperature of 50 °C is approximately 85 °C.

- b. To determine the relative humidity of the exhaust air, we need to determine the molar flow rate of water vapor exiting the coater. Based on the material balance for water around the control volume, the molar flow rate of water vapor in the exhaust stream should be equal to the amount of water vapor entering in the drying air plus the amount of water added through the coating solution. First calculate the water vapor partial pressure in the inlet air, which is, by definition, the vapor pressure at the dew point temperature. This can be determined using the Arden Buck equation:

$$P_{\text{w,in}} = P_{\text{w}}^*(\text{at } T_{\text{dew}}) = 0.012 \text{ atm}$$

Assuming an ideal gas mixture, the mole fraction of water vapor in the inlet stream can be calculated by the ratio of the partial pressure to the total pressure (which is assumed to be atmospheric):

$$y_{\text{w,in}} = \frac{P_{\text{w,in}}}{P_{\text{total}}} = \frac{0.012 \text{ atm}}{1 \text{ atm}} = 0.012$$

and, thus, the mole fraction of dry air is

$$y_{\text{air,in}} = 1 - y_{\text{w,in}} = 0.988$$

The overall molar flow rate of the inlet airstream is the mass flow rate divided by the average molecular weight of the inlet air:

$$\dot{n}_{\text{inlet}} = \frac{m_{\text{air}}}{\text{MW}_{\text{inlet}}} = \frac{m_{\text{air}}}{y_{\text{w,in}}\text{MW}_{\text{w}} + y_{\text{air,in}}\text{MW}_{\text{air}}}$$

$$= \frac{1117.57 \text{ g/min}}{0.012 \times 18 \text{ g/mol} + 0.988 \times 29 \text{ g/mol}} = 38.713 \text{ mol/min}$$

and the molar flow rate of water vapor in the inlet air is

$$\dot{n}_{\text{w,inlet}} = y_{\text{w,in}} \times \dot{n}_{\text{inlet}} = 0.012 \times 38.713 = 0.4646 \text{ mol/min}$$

The contribution of water vapor from the coating solution is

$$\dot{n}_{\text{w,spray}} = m_{\text{coat}} \left(\frac{\chi_{\text{w}}}{\text{MW}_{\text{w}}} \right) = 4 \text{ g solution/min}$$

$$\times \left(\frac{0.8 \text{ g water/g solution}}{18 \text{ g water/mol}} \right) = 0.1778 \text{ mol/min}$$

By conservation of mass the molar flow rate of water vapor in the exhaust airstream is

$$\dot{n}_{\text{w,outlet}} = \dot{n}_{\text{w,inlet}} + \dot{n}_{\text{w,spray}} = 0.4646 + 0.1778 = 0.6424 \text{ mol/min}$$

and the overall molar flow rate of the exhaust airstream is

$$\dot{n}_{\text{outlet}} = \dot{n}_{\text{inlet}} + \dot{n}_{\text{w,spray}} = 38.713 + 0.1778 = 38.891 \text{ mol/min}$$

Finally, the partial pressure and vapor pressure of water in the exhaust airstream can be calculated as follows:

$$P_{\text{w,out}} = y_{\text{w,out}} P_{\text{total}} = \frac{\dot{n}_{\text{w,outlet}}}{\dot{n}_{\text{outlet}}} P_{\text{total}} = \frac{0.6424 \text{ mol/min}}{38.891 \text{ mol/min}} \times 1 \text{ atm}$$

$$= 0.0165 \text{ atm}$$

$$P_{\text{w}}^*(\text{at } T_{\text{air,out}}) = 0.1177 \text{ atm}$$

and thus the relative humidity of the exhaust airstream is

$$\%RH_{\text{out}} = \frac{P_{\text{w,out}}}{P_{\text{w}}^*(\text{at } T_{\text{air,out}})} \times 100\% = \frac{0.0165 \text{ atm}}{0.1177 \text{ atm}} \times 100\% = 14\%$$

20.4.4.6 Process Scale-Up Based on Model Predictions

An important feature of the thermodynamic film coating model is in its predictive capabilities, which can be used for process simulation and to assist with scale-up from one film coater to another. Typically, when a film coating process is scaled to a larger unit, it is desirable to maintain the same thermodynamic conditions in the pan that the tablets are exposed to during coating. This can be accomplished by matching the temperature and relative humidity of the exhaust airstream across the two coaters when the HLF for each coater is known. The thermodynamic model can be used

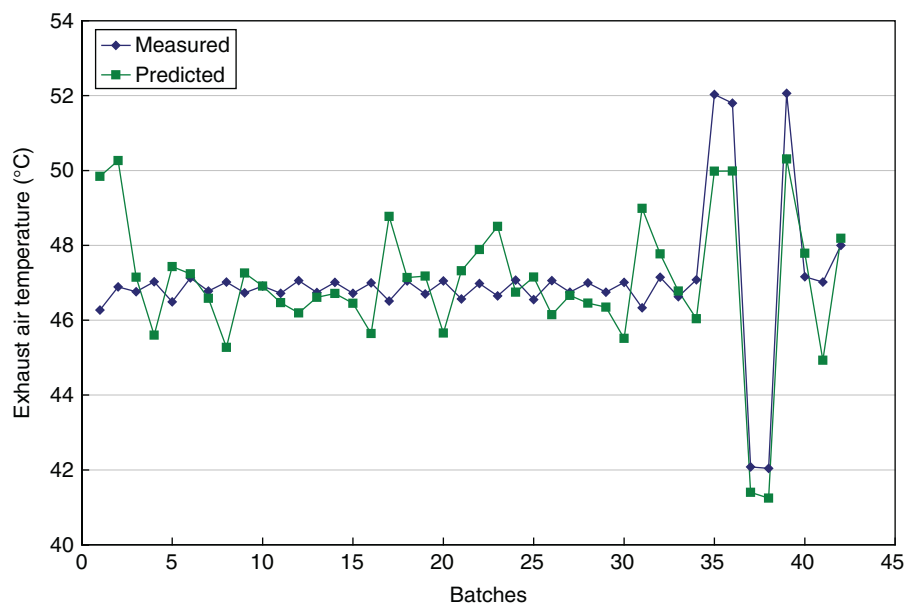


FIGURE 20.27 Comparison of predicted exhaust air temperature from the thermodynamic model to process data for drug A from the manufacturing site. *Source:* Reprinted with permissions from Ketterhagen et al. [90], Copyright (2017) Elsevier.

to determine the process parameters on the new coater that will result in similar exhaust air conditions to the proven operation of the original coater. Ideally, this will reduce the number of trials required during scale-up and minimize any failed batches during validation. This concept is best illustrated through a simple case study.

20.4.4.7 Case Study: Immediate Release Film Coating of Drug A Tablets

The initial clinical tablet film coating for drug A was executed on a Glatt GC 1000 pan coater at a batch size of 60 kg. For commercial launch the film coating process was transferred to a separate area of the manufacturing site and would utilize two coating pans that had not previously been used for this product. Based on commercial demand, one dosage strength would be manufactured on a Glatt GC 1000 coating pan (60 kg batch size), and the other dosage strength would be manufactured on a smaller Glatt GC 750 (30 kg batch size). The thermodynamic model described above was used to correlate the thermodynamic operating space to the process parameter ranges for the new coating pans to assist in defining the design space.

First, the model was used to predict the exhaust air temperature and relative humidity based on existing process data from the clinical GC 1000 coater as shown in Figure 20.27 [90]. This data was also used to generate the HLF estimate for that coater. The HLF estimates for the two commercial coaters (GC 750 and GC 1000) were generated based on historical coating process data from other products manufactured in the same facility and equipment using a similar aqueous film coating system.

Based on the existing process data from drug A, the model was utilized to map the coating process design space to the two

commercial coaters as shown in Figure 20.28 [90]. The design space is expressed as a thermodynamic operating space, which is independent of any specific equipment. The four points of the operating spaces for the GC 750 and GC 1000 are based on the thermodynamic modeling data, and the central region summarizes the ranges typically used during actual manufacture. Table 20.11 summarizes the acceptable ranges of coating process parameters based on thermodynamic consideration for maintaining constant exhaust air temperature between scales, although a multitude of solutions for the inlet variables are possible to meet the same exhaust conditions.

By mapping the thermodynamic conditions from the clinical coating pan, we have defined the potential design space for drug A on the commercial coaters without running any trials at scale. The operating space can be verified with as many or as few confirmation batches as the operators and formulators deem necessary. We can also map out a design space plot that allows one to visualize the relationship between changes in spray rate, inlet air temperature, and exhaust air temperature for a given drying air flow rate. This type of plot for the GC 1000 film coater is shown in Figure 20.29. The contour plot allows one to map out any combination of operating conditions (inlet air temperature, exhaust air temperature, and spray rate) within the design space for the given drying air flow rate. It also allows operators and formulators to visualize how the exhaust air temperature changes as a function of both spray rate and inlet air temperature across the entire design space.

This exercise has illustrated how the thermodynamic film coating model can be used to predict specific operating conditions, assist with scale-up from one coater to another, and also simulate different operation scenarios on a given film coater. The model is an extremely versatile predictive tool

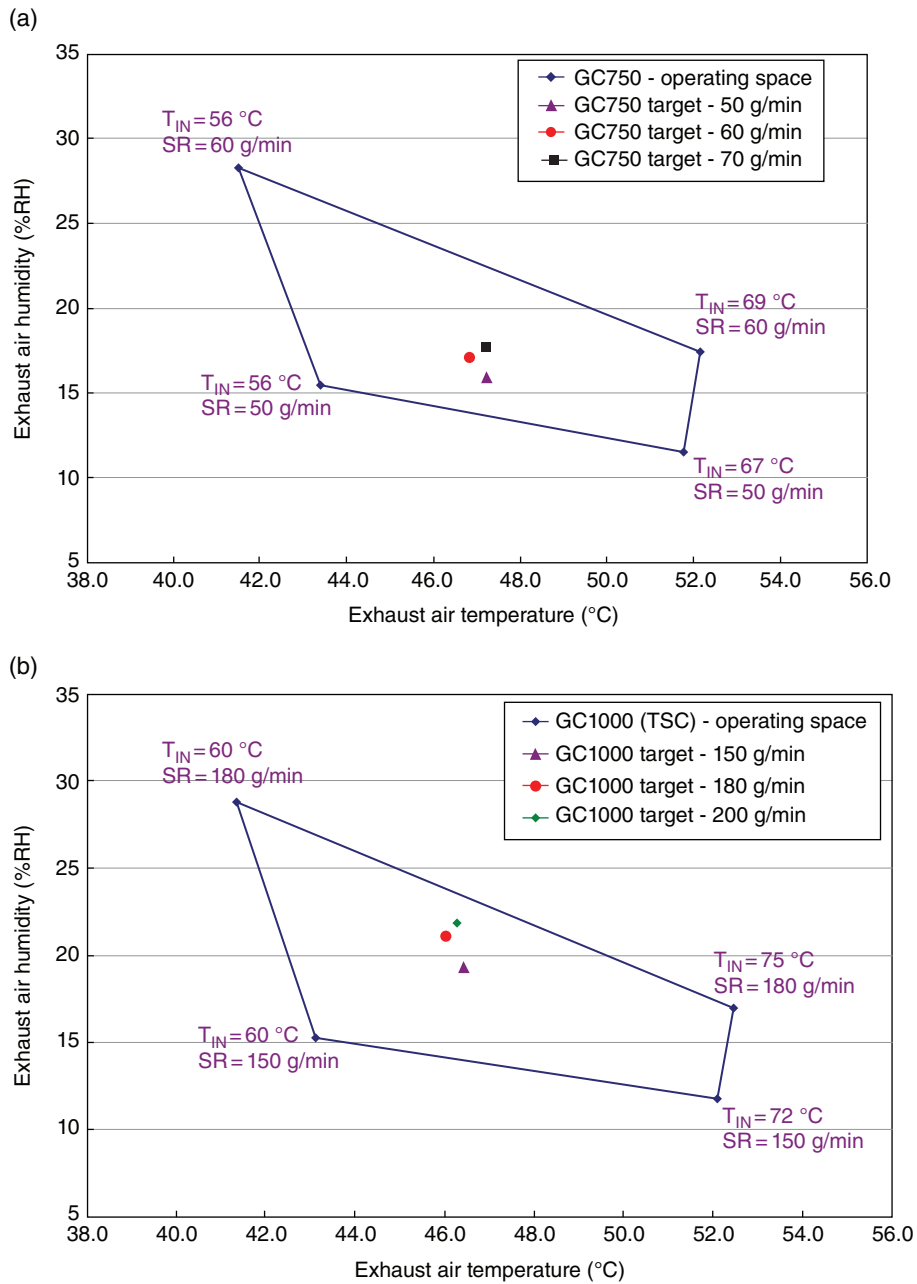


FIGURE 20.28 Thermodynamic operating space for the drug A film coating process in (a) a Glatt GC 750 coating pan and (b) a Glatt GC 1000 coating pan. *Source:* Reprinted with permissions from Ketterhagen et al. [90], Copyright (2017) Elsevier.

TABLE 20.11 Acceptable Ranges of Coating Process Parameters Based on the Coating Thermodynamic Model [90]

Coating Process Parameter	Glatt GC 750	Glatt GC 1000
Exhaust temperature (°C)	42–52	42–52
Exhaust humidity (%RH)	10–30	10–30
Inlet air temperature (°C)	55–70	60–75
Drying air flow rate (CFM)	470	860–900
Spray rate (ml/min)	50–70	140–200

Source: Reproduced with permission of Elsevier.

that is applicable to a wide range of coating formulations on units ranging from lab-scale coaters to commercial-sized equipment.

20.5 SUMMARY

This chapter highlighted several modeling techniques applied to the design, development, and scale-up for solid oral drug

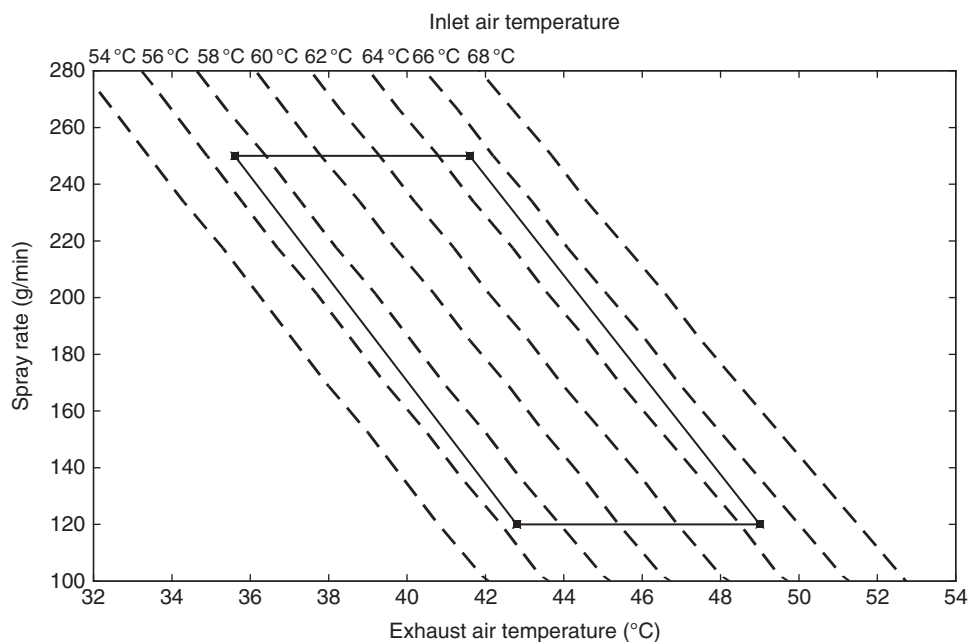


FIGURE 20.29 Design space contour plot for the Glatt GC 1000 film coater for a drying air flow rate of 900 ft³/min.

products. These process modeling techniques were discussed and exemplified with case studies ranging from raw material specifications to process parameter predictions in wet granulation and film coating. Specific consideration was given to transfer and scale-up issues along with general process design-related challenges to pharmaceutical process R&D. There are many other modeling approaches available to formulation and process scientists that were not covered in this chapter. However, the purpose was to demonstrate that the fundamental principles taught in the chemical engineering curriculum ensure the chemical engineer is well poised to apply and implement modeling techniques to solve challenging issues in the pharmaceutical industry.

REFERENCES

1. am Ende, M.T., Bharadwaj, R., Doshi, P. et al. (2010). Process modeling techniques and applications for solid oral drug products. In: *Chemical Engineering in the Pharmaceutical Industry: R&D to Manufacturing* (ed. D.J. am Ende). New York: Wiley.
2. Wassgren, C. and Curtis, J.S. (2006). The application of computational modeling to pharmaceutical materials science. *MRS Bulletin* 31 (11): 900–904.
3. Davidson, D.L. (2001). In the enterprise-wide application of computational fluid dynamics in the chemicals industry. 6th World Congress of Chemical Engineering, Melbourne, Australia.
4. Louie, A.S., Brown, M.S., and Kim, A. (2007). Measuring the return on modeling and simulation tools in pharmaceutical development. *White Paper HI204892*. Health Industry Insights.
5. Garcia-Munoz, S. (2009). Establishing multivariate specifications for incoming materials using data from multiple scales. *Chemometrics and Intelligent Laboratory Systems* 98 (1): 51–57.
6. Ketterhagen, W.R. (2006). Modeling granular segregation during hopper discharge. AICHE Spring Meeting & Global Congress on Process Safety, Proceedings – 5th World Conference of Particle Technology (WCPT5), Orlando, Florida, April 23–27.
7. Ketterhagen, W.R., am Ende, M.T., and Hancock, B.C. (2009). Process modeling in the pharmaceutical industry using the discrete element method. *Journal of Pharmaceutical Sciences* 98 (2): 442–470.
8. Ketterhagen, W.R., Curtis, J.S., and Wassgren, C.R. (2005). Stress results from two-dimensional granular shear flow simulations using various collision models. *Physical Review E: Statistical, Nonlinear, and Soft Matter Physics* 71 (6-1): 061307/1–061307/11.
9. Ketterhagen, W.R., Curtis, J.S., Wassgren, C.R., and Hancock, B.C. (2008). Modeling granular segregation in flow from quasi-three-dimensional, wedge-shaped hoppers. *Powder Technology* 179 (3): 126–143.
10. Ketterhagen, W.R., Curtis, J.S., Wassgren, C.R., and Hancock, B.C. (2009). Predicting the flow mode from hoppers using the discrete element method. *Powder Technology* 195 (1): 1–10.
11. Ketterhagen, W.R., Curtis, J.S., Wassgren, C.R. et al. (2007). Granular segregation in discharging cylindrical hoppers: a discrete element and experimental study. *Chemical Engineering Science* 62 (22): 6423–6439.

12. Iveson, S.M., Litster, J.D., Hapgood, K., and Ennis, B.J. (2001). Nucleation, growth and breakage phenomena in agitated wet granulation processes: a review. *Powder Technology* 117 (1–2): 3–39.
13. Bilgili, E. and Scarlett, B. (2005). Population balance modeling of non-linear effects in milling processes. *Powder Technology* 153 (1): 59–71.
14. Kremer, D.M. and Hancock, B.C. (2006). Process simulation in the pharmaceutical industry: a review of some basic physical models. *Journal of Pharmaceutical Sciences* 95 (3): 517–529.
15. Kunii, D. and Levenspiel, O. (1991). *Fluidization Engineering*. Newton, MA: Butterworth-Heinemann.
16. Bi, H.T., Ellis, N., Abba, I.A., and Grace, J.R. (2000). A state-of-the-art review of gas-solid turbulent fluidization. *Chemical Engineering Science* 55 (21): 4789–4825.
17. Gurson, A.L. Continuum theory of ductile rupture by void nucleation and growth: part I-Yield criteria and flow rules for porous ductile media. *Journal of Engineering Material Technology (Transaction of ASME)* 1977: 2–15.
18. Schofield, A.N. and Wroth, C.P. (1968). *Critical State Soil Mechanics*. London: McGraw-Hill.
19. DiMaggio, F.L. and Sandler, I.S. (1971). Material model for granular soils. *Journal of the Engineering Mechanics Division* 97: 935–950.
20. Drucker, D.C., Gibson, R.E., and Henkel, D.J. (1957). Soil mechanics and work hardening theories of plasticity. *Transactions American Society of Civil Engineering* 122: 338–346.
21. am Ende, M.T. and Berchielli, A. (2005). A thermodynamic model for organic and aqueous tablet film coating. *Pharmaceutical Development and Technology* 10 (1): 47–58.
22. Box, G.E.P., Hunter, W.G., and Hunter, J.S. (1978). *Statistics for Experimenters: An Introduction to Design, Data Analysis, and Model Building*. Wiley.
23. MacGregor, J.F., Yu, H., Garcia Munoz, S., and Flores-Cerrillo, J. (2005). Data-based latent variable methods for process analysis, monitoring and control. *Computers & Chemical Engineering* 29 (6): 1217–1223.
24. Page, S., Baumann, K.H., and Kleinebudde, P. (2006). Mathematical modeling of an aqueous film coating process in a Bohle Lab-Coater, part 1: development of the model. *AAPS Pharm. Sci. Tech.* 7 (2): 42.
25. Page, S. and Baumann, K.-H. (2006). Mathematical modeling of an aqueous film coating process in a Bohle Lab-Coater: part 2: application of the model. *AAPS Pharm. Sci. Tech.* 7 (2): 43.
26. Stetsko, G., Banker, G.S., and Peck, G.E. (1983). Mathematical modeling of an aqueous film coating process. *Pharmaceutical Technology* 7 (11): 50–62.
27. Muteki, K. and MacGregor, J.F. (2007). Multi-block PLS modeling for L-shape data structures with applications to mixture modeling. *Chemometrics and Intelligent Laboratory Systems* 85: 186.
28. Garcia Munoz, S., MacGregor, J.F., and Kourti, T. (2005). Product transfer between sites using Joint-Y PLS. *Chemometrics and Intelligent Laboratory Systems* 79 (1–2): 101–114.
29. Garcia-Munoz, S., Kourti, T., MacGregor, J.F. et al. (2003). Troubleshooting of an industrial batch process using multivariate methods. *Industrial & Engineering Chemistry Research* 42 (15): 3592–3601.
30. Kourti, T., Lee, J., and MacGregor, J.F. (1996). Experiences with industrial applications of projection methods for multivariate statistical process control. *Computers and Chemical Engineering* 20S: S745.
31. Neogi, D. and Schlags, C.E. (1998). Multivariate statistical analysis of an emulsion batch process. *Industrial & Engineering Chemistry Research* 37 (10): 3971–3979.
32. Garcia-Munoz, S., Kourti, T., MacGregor, J.F. et al. (2006). Optimization of batch operating policies. Part I. Handling multiple solutions. *Industrial & Engineering Chemistry Research* 45 (23): 7856–7866.
33. Garcia-Munoz, S., MacGregor, J.F., Neogi, D. et al. (2008). Optimization of batch operating policies. Part II. Incorporating process constraints and industrial applications. *Industrial & Engineering Chemistry Research* 47 (12): 4202–4208.
34. Doyle, F.J., Harrison, C.A., and Crowley, T.J. (2003). Hybrid model-based approach to batch-to-batch control of particle size distribution in emulsion polymerization. *Computers & Chemical Engineering* 27: 1153–1163.
35. Flores-Cerrillo, J. and MacGregor, J.F. (2003). Within-batch and batch-to-batch inferential-adaptive control of semi-batch reactors: a partial least squares approach. *Industrial & Engineering Chemistry Research* 42 (14): 3334–3345.
36. Flores-Cerrillo, J. and MacGregor, J.F. (2005). Latent variable MPC for trajectory tracking in batch processes. *Journal of Process Control* 15 (6): 651–663.
37. Bharati, M.H., Liu, J.J., and MacGregor, J.F. (2004). Image texture analysis: methods and comparisons. *Chemometrics and Intelligent Laboratory Systems* 72 (1): 57–71.
38. Yu, H. and MacGregor, J.F. (2003). Multivariate image analysis and regression for prediction of coating content and distribution in the production of snack foods. *Chemometrics and Intelligent Laboratory Systems* 67 (2): 125–144.
39. Duchesne, C. and MacGregor, J.F. (2004). Establishing multivariate specification regions for incoming materials. *J. Quality Technology* 36 (1): 78–94.
40. am Ende, M.T., Blackwood, D.O., Gierer, D.S., and Neu, C.P. (2009). Challenges in development and scale-up of low dose dry granulation products: a case study. In: *Analytical and Formulation Development For Low-Dose Oral Drug Products* (ed. J. Zheng), 117–157. New York: Wiley.
41. FDA (2006). Draft Guidance for Industry: Quality Systems Approach to Pharmaceutical CGMP Regulations.
42. FDA (2009). International conference on harmonisation: guidance on Q8(R1) pharmaceutical development. *Federal Register* 74 (109): 27325–27326.
43. FDA (2006). International conference on harmonisation: guidance on Q9 quality risk management. *Federal Register* 71 (106): 32105–32106.
44. FDA (2009). International conference on harmonisation: guidance on Q10 pharmaceutical quality system. *Federal Register* 74 (66): 15990–15991.
45. Nosal, R. and Schultz, T. (2008). PQLI definition of criticality. *J. Pharm. Innov.* 3: 79–87.

46. Nomikos, P. and MacGregor, J.F. (1995). Multivariate SPC charts for monitoring batch processes. *Technometrics* 37 (1): 41–58.
47. Hertz, H. (1882). Über die Berührung fester elastischer Körper. *J. für die reine u. angew. Math* 92: 136.
48. Zhu, H.P., Zhou, Z.Y., Yang, R.Y., and Yu, A.B. (2007). Discrete particle simulation of particulate systems: theoretical developments. *Chemical Engineering Science* 62 (13): 3378–3396.
49. Zhu, H.P., Zhou, Z.Y., Yang, R.Y., and Yu, A.B. (2008). Discrete particle simulation of particulate systems: a review of major applications and findings. *Chemical Engineering Science* 63 (23): 5728–5770.
50. Lu, G., Third, J.R., and Müller, C.R. (2015). Discrete element models for non-spherical particle systems: from theoretical developments to applications. *Chemical Engineering Science* 127: 425–465.
51. Uniformity of Dosage Units, *Pharmacopeia U.S.*, Forum, P., Ed. 2006, pp 1653–1659.
52. European Pharmacopoeia (Ph. Eur.): Supplement 6.1, Uniformity of Dosage Units (reference 04/2008:20940), April 2008.
53. European Pharmacopoeia (Ph. Eur.): Supplement 7.4, Uniformity of Dosage Units, April 2012.
54. Japanese Pharmacopoeia (JP): 6.02 Uniformity of Dosage Units, Fifteenth Edition, The Ministry of Health, Labour and Welfare Ministerial Notification No. 285, 31 March 2006.
55. Japanese Pharmacopoeia (JP): 6.02 Uniformity of Dosage Units, Fifteenth Edition, The Ministry of Health, Labour and Welfare Ministerial Notification No. 285, 5 November 2008.
56. Japanese Pharmacopoeia (JP): 6.02 Uniformity of Dosage Units, Fifteenth Edition, The Ministry of Health, Labour and Welfare Ministerial Notification No. 190, 31 May 2013.
57. United States Pharmacopoeia (USP): <905> Uniformity of Dosage Units, Pharmacopoeial Forum, Volume 35, Number 3, October 2010.
58. United States Pharmacopoeia (USP): <905> Uniformity of Dosage Units, Pharmacopoeial Forum, Volume 35, Number 3, December 2011.
59. Novick, S., Christopher, D., Dey, M. et al. (2009). A two one-sided parametric tolerance interval test for control of delivered dose uniformity – part 1 – characterization of FDA proposed test. *AAPS PharmSciTech* 10 (3): 820–828.
60. Novick, S., Christopher, D., Dey, M. et al. (2009). A two one-sided parametric tolerance interval test for control of delivered dose uniformity – part 2 – effect of changing parameters. *AAPS PharmSciTech*. 10 (3): 841–849.
61. Novick, S., Christopher, D., Dey, M. et al. (2009). A two one-sided parametric tolerance interval test for control of delivered dose uniformity – part 3 – investigation of robustness to deviations from normality. *AAPS PharmSciTech*. 10 (3): 829–840.
62. Ketterhagen, W.R., Curtis, J.S., Wassgren, C.R., and Hancock, B.C. (2009). Predicting the flow mode from hoppers using the discrete element method. *Powder Technology* 195: 1–10.
63. Ketterhagen, W.R., Sarkar, A., Kerbl, J., et al. (2016). CFD-DEM model for fluidization segregation during transfer of pharmaceutical blends. AICHe Annual Meeting, San Francisco, CA (15 November 2016).
64. Zhu, H.P., Zhou, Z.Y., Yang, R.Y., and Yu, A.B. (2007). Discrete particle simulation of particulate systems: theoretical developments. *Chemical Engineering Science* 62 (13): 3378–3396.
65. Kushner, J. IV and Moore, F. (2010). Scale-up model describing the impact of lubrication on tablet tensile strength. *Int. J. Pharm.* 399: 19–30.
66. Blackwood, D.O., Hancock, B.C., Polizzi, M., et al. (2011). A powder mixing model for a small-scale continuous mixing/direct compression manufacturing process. AICHe Annual Meeting, Minneapolis, MN (17 October 2011).
67. Ketterhagen, W.R. (2015). Simulation of powder flow in a lab-scale tablet press feed frame: effects of design and operating parameters on measures of tablet quality. *Powder Technology* 275: 361–374.
68. Ketterhagen, W.R. and Blackwood, D.O. (2014). Reducing powder lubrication in an encapsulation process for improved dissolution performance. AICHe Annual Meeting, Atlanta, GA (17 November 2014).
69. Blackwood, D.O., Ketterhagen, W.R., Kresevic, J. et al. (2018). Quantifying and reducing powder shear sensitivity when manufacturing capsules with lubricants. *Drug Development & Industrial Pharmacy* 44: 1350–1356.
70. Ketterhagen, W.R., Mullarney, M.P., Kresevic, J., and Blackwood, D.O. (2018). Computational approaches to predict the effect of shear during processing of lubricated pharmaceutical blends. *Powder Technology* 335: 427–439.
71. Faure, A., Grimsey, I.M., Rowe, R.C. et al. (1999). Applicability of a scale-up methodology for wet granulation processes in Collette Gral high shear mixer-granulators. *European Journal of Pharmaceutical Sciences* 8 (2): 85–93.
72. Levin, M. (2006). Wet granulation: end-point determination and scale-up. In: *Encyclopedia of Pharmaceutical Technology* (ed. J. Swarbrick), 4078–4098. Boca Raton, FL: Taylor & Francis.
73. He, X., Lunday, K.A., Li, L.-C., and Sacchetti, M.J. (2008). Formulation development and process scale up of a high shear wet granulation formulation containing a poorly wettable drug. *Journal of Pharmaceutical Sciences* 97 (12): 5274–5289.
74. Bier, H.P., Leuenberger, H., and Sucker, H. (1979). Determination of the uncritical quantity of granulating liquid by power measurements on planetary mixers. *Pharmazeutische Industrie* 41 (4): 375–380.
75. Leuenberger, H. (1982). Granulation, new techniques. *Pharmaceutica Acta Helveticae* 57 (3): 72–82.
76. Leuenberger, H. (2001). New trends in the production of pharmaceutical granules: the classical batch concept and the problem of scale-up. *European Journal of Pharmaceutics and Biopharmaceutics* 52 (3): 279–288.
77. Landin, M., York, P., Cliff, M.J. et al. (1996). Scale-up of a pharmaceutical granulation in fixed bowl mixer-granulators. *International Journal of Pharmaceutics* 133 (1,2): 127–131.

78. Cunningham, J.C., Sinka, I.C., and Zavaliangos, A. (2004). Analysis of tablet compaction. I. Characterization of mechanical behavior of powder and powder/tooling friction. *Journal of Pharmaceutical Sciences* 93: 2022–2039.
79. Han, L., Elliott, J., Bentham, A. et al. (2008). A modified Drucker-Prager Cap model for die compaction simulation of pharmaceutical powders. *Intl. Journal of Solids and Structures* 45 (10): 3088–3106.
80. Sinka, I.C. (2007). Modelling powder compaction. *KONA* 25: 4–22.
81. Sinka, I.C., Cunningham, J.C., and Zavaliangos, A. (2004). Analysis of tablet compaction. II. Finite element analysis of density distributions in convex tablets. *Journal of Pharmaceutical Sciences* 93 (8): 2040–2053.
82. Wu, C.Y., Ruddy, O.M., Bentham, A.C. et al. (2005). Modeling the mechanical behaviour of pharmaceutical powders during compaction. *Powder Technology* 152 (1–3): 107–117.
83. Frenning, G. (2007). Analysis of pharmaceutical powder compaction using multiplicative hyperelasto-plastic theory. *Powder Technology* 172 (2): 103–112.
84. Michrafy, A., Ringenbacher, D., and Tchoreloff, P. (2002). Modeling the compaction behavior of powders: application to pharmaceutical powders. *Powder Technology* 127 (3): 257–266.
85. Wu, C.Y., Hancock, B.C., Mills, A. et al. (2008). Numerical and experimental investigation of capping mechanisms during pharmaceutical tablet compaction. *Powder Technology* 181 (2): 121–129.
86. Zavaliangos, A., Galen, S., Cunningham, J., and Winstead, D. (2008). Temperature evolution during compaction of pharmaceutical powders. *Journal of Pharmaceutical Sciences* 97 (8): 3291–3304.
87. am Ende, M.T., Herbig, S.M., Korsmeyer, R.W., and Chidlaw, M.B. (2000). Osmotic drug delivery from asymmetric membrane film-coated dosage forms. In: *Handbook of Pharmaceutical Controlled Release Technology* (ed. D.L. Wise), 751–785. New York: Marcel Dekker, Inc.
88. Ebey, G.C. (1987). A thermodynamic model for aqueous film-coating. *Pharmaceutical Technology* 11 (4): 40–50.
89. Rodriguez, L., Grecchi, R., Cini, M. et al. (1996). Variation of operational parameters and process optimization in aqueous film coating. *Pharmaceutical Technology* 10: 76–86.
90. Ketterhagen, W., Aliseda, A., am Ende, M.T. et al. (2017). Modeling tablet film coating processes. In: *Predictive Modeling of Pharmaceutical Unit Operations* (ed. P. Pandey and R. Bhargadwaj), 273–316. New York: Elsevier.
91. Sinka, I.C., Cunningham, J.C., and Zavaliangos, A. (2003). The effect of wall friction in the compaction of pharmaceutical tablets with curved faces: a validation study of the Drucker-Prager Cap model. *Powder Technology* 133 (1–3): 33–43.

PART III

CONTINUOUS MANUFACTURING

21

CONTINUOUS MANUFACTURING IN SECONDARY PRODUCTION

MARTIN WARMAN

Martin Warman Consultancy Ltd, Chestfield, UK

21.1 INTRODUCTION

For the entire history of pharmaceutical manufacturing, secondary production processes have been carried out in batches. There are deep-rooted reasons for this [1], some historical (the ancestry of many pharmaceutical processes come from food or confectionary) and some founded in the need to track the “history” of the dosage form (in terms of materials and processes, i.e. the “batch record”). However, even though pharmaceutical secondary production shares many unit operations with other industries, the reality is, many of those industries have already realized the commercial and operational benefits of continuous manufacturing and have started running those unit operations continuously, and have abandoned batch production.

It would be easy to simply justify the continued use of batch approaches for these reasons; however, the pharmaceutical industry is coming under huge quality, efficiency, financial and business pressures (which has even reached the public’s attention, for example, in this article in Wall Street Journal [2]) but also the regulatory landscape has changed. The initial PAT Framework [3], the 21st Century Initiative Final Report [4], the ICH Quality Trio (ICH Q8, Q9, and Q10) as well as the current guidance on Process Validation and the FDA white paper Modernizing Pharmaceutical Manufacturing: From Batch to Continuous Production [5] have all put a focus on the application of new technologies with a science-based approach. They also introduce a new term to the Pharmaceutical industry; that of quality by design (QbD).

The use of continuous manufacturing during development and commercial manufacturing facilitates both the application of QbD approaches but also (as will be shown later in this text) the ability to implement these approaches in a highly efficient way.

21.2 DEFINITIONS

To be clear from the start, when describing batch production, we are describing an overall production system whereby the entire mass/volume goes through the unit operation at the same time, normal in one “container”; as an example, unit operations such as bin blending, wet granulation (WG), fluid-bed operation (drying, granulation, or coating), and coating, where we start with individual components being added to a single processing unit and during the production process the entire batch changes to reach a single end point – the process is only complete when the entire mass of the blend is uniform. Whereas continuous manufacturing describes where the input materials continually enter into the unit operation and output materials continually exit, under a “first in, first out principle,” taking the blending example, a continuous blend operation is where the input materials are continually being fed into a continuous mixing process. The outcome of the process is not only continuous but also in a state of control, resulting in a uniform output where each unit dose mass is not only the same/having the correct concentration of each component (inter-dose uniformity) but also that those components are optimally dispersed

within the unit dose, thus ensuring correct delivery performance. In the case of continuous blending the properties of the individual dosage are “generated” early on, and in some ways the process consists of a stream of individual unit doses, such that verification of performance of continuous systems has to be considered against this production paradigm. It is also very important that continuous manufacturing processes are not to be confused with the situation where standard batch operations are simply linked together. For example, it is very common for consecutive unit operations in secondary production to be connected together, for example, a bin blend discharging into a high-shear granulator, into a fluidized bed dryer, etc. – the bulk material flow maybe linked but the entire batch goes through a transformation at the same time within each unit operation.

21.3 REVIEW OF TYPICAL UNIT OPERATIONS

As previously indicated, many pharmaceutical unit operations are shared with other industries; however, we also have to acknowledge that many of the individual unit operations used are themselves continuous operations – as an industry we simply chose to collect the output and form/maintain the batch. So let us first look at traditional solid dosage operations and consider if they are “batch”, “continuous,” or could be made continuous.

21.3.1 Typical Batch Process Operations

For more details on the types of secondary production, please refer to the specific sections of this publication; they are listed here purely for background before discussing the individual unit operations.

The vast majority of current pharmaceutical products are currently “solid oral doses,” commonly known as tablets or capsules. Typically, production of solid dosage forms is carried out in three types of process streams. The simplest is well described and commonly called direct compression (Figure 21.1). In a simple direct compression process, the

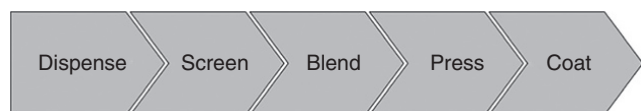


FIGURE 21.1 Schematic of the direct compression process.



FIGURE 21.2 Schematic of the dry granulation process.

active pharmaceutical ingredient (API)/excipients are dispensed (via a screen to de-lump) into a v-shell or bin blender. Post blend operation, the material is transferred to the feed hopper of the tablet press, post compression the batch is coated “en-mass” in a pan coater, before packaging.

Small variations in this basic process workflow may occur, for example, replacing the compression/coating steps with an encapsulation step during capsule production, but the general workflow stays the same.

However, if the particle size and/or physical characteristics of the individual API/excipient powders are likely to cause segregation due to equipment design or transfer or compression issues during tableting, it is common to introduce some form of granulation step post blending. This is often then followed by an additional (second) blend step when a lubricant excipient is needed to improve flowability within the process and to prevent sticking/chipping during compression.

If the granulation step is “dry” (also known as roller compaction or RC), the premixed materials are forced through two counter-rotating rollers which exert mechanical pressure on the powders during a high-pressure agglomeration or “compact.” The compact can be of several forms; however, in each case the true granulate is formed by milling. Typically, the high-pressure compact formation and subsequent milling are together described as a single-unit operation (Figure 21.2).

In comparison, the alternate is WG, in which shear and compression forces are used along with the wet/massing forces during addition of a binder to firstly generate agglomerated particles using the three phases shown in Figure 21.3.

Leuenberger [6] identified the optimal granulation point as being the initiation of the capillary state. However, granulates in this phase then need to be dried. For this reason WG is typically described as a two-part process with the initial WG being followed by drying process (even though the two steps can and sometimes are carried out in the same

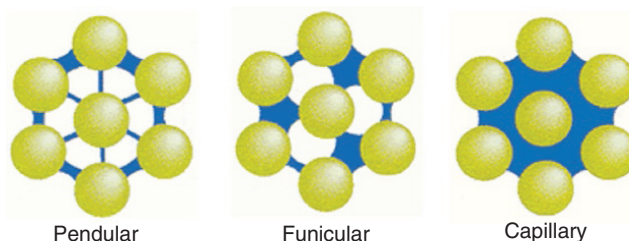


FIGURE 21.3 Three phases of granulate wetting.



FIGURE 21.4 Schematic of the wet granulation process.

vessel – commonly described as a “single-pot” granulator (see Figure 21.4).

21.4 SOLID DOSAGE UNIT OPERATION

From the process workflows given above it is apparent that several unit operations are common, even repeated. But let us consider each of these “common steps” in turn and consider how they are run now, and what are the opportunities for continuous processing.

21.4.1 Dispensing

Under a traditional batch paradigm the production material is typically weighed in the pharmacy, verified, and released, having already been individually bagged ready to be loaded into the production process. However, it is not uncommon for the occupational safety requirements, e.g. toxicity, sensitization, explosivity, and therefore containment, of some materials to cause manual handling issues. These are often overcome by automated dispensing systems. In some cases these dispensing systems are connected directly to the production process. Although these systems are typically used to initiate a batch process, they are (in themselves) continuous systems – the first powder into the feeder is typically the first powder out. In reality, pharmaceutical production may use the current range of volumetric and loss-in weight feeders to deliver in an automated way to ensure containment, they were actually designed for, and able to run, as part of continuous systems. Several of the feeder suppliers have even extended their product range forward, such that their control system becomes the basis for controlling material flow through the entire continuous system.

Although the basic principle of using existing automated feeders as part of a continuous manufacturing system is sound, what we are trying to achieve is very different. A feeder on a batch-dispensing line simply has to deliver, reproducibly, the correct mass of powder. The mass is simply what is needed for an entire batch. So, for example, an automatic feeder dispensing 20% API to a 1000 kg batch process simply has to reproducibly deliver 200 kg of API. The automatic feeder is able to accelerate during bulk dispensing, slow as it approaches the end point, and operate what can only be described as fine-step control to end up at the predefined mass. Taking the same 20% API example, on a typical continuous process producing 40 000 500 mg image tablets



FIGURE 21.5 K-Tron MT12™ twin-screw microfeeder, capable of both batch and continuous operation.

per hour, the feeder is required to deliver not only exactly 4 kg/h but also needs to deliver approximately 1.1 g of API every second, reproducibly, across the entire production run (see Figure 21.5). From a mechanical engineering perspective, the two are very different challenges (and even worse if the API or excipient concentration is lower (for example, a typical lubricant addition rate of 1% equates to 55 mg/sec addition rate or rather 5 mg/90 msec). For these reasons it is possible that a mini-batch approach could be included in a continuous manufacturing line for low-concentration components and where the line rate is low (<10 kg/h). A mini-batch approach utilizes in-line a small batch blender and the repeated production of small sub-batches, of, for example, 1–3 kg. At a line rate of 5 kg/h, this could mean a mini-batch is proposed every 12–20 minutes, effectively allowing the feeders to operate in a consecutive batch mode and achieve much higher precision than when running in continuous mode.

In general, the only additional consideration for continuous use is the maintenance of an acceptable level of materials in the feeder-charge hopper, but recognizing that the period

while the feeder is being refilled causes a disturbance in the gravimetric mode/feeder feed factor. This means there is typically a need to define an operating space for the feeder refill and to verify the process stays in a state of control during these refills.

21.4.2 Screening

Even when running under a continuous paradigm, it is anticipated that many raw materials will be delivered in as drums/lots/batches. It is a straightforward logistic operation to track the use of the material to final dosage form and in effect

allowing traceability of lots to whatever is defined as final batch integrity for compliance purposes. However, in many cases the performance of the feeder is impacted directly by fluctuations in the hopper level, so simply, manually charging the hoppers is not an option. Material-handling solutions have been developed for other industries such as food and food ingredients which are even capable of receiving raw materials on rail or by road, transferring the local storage facilities before charging feed hoppers at local unit operations. These systems can also include flow aids, filtered venting/exhaust systems as well as redundancy/parallel storage to ensure supply (see Figure 21.6).

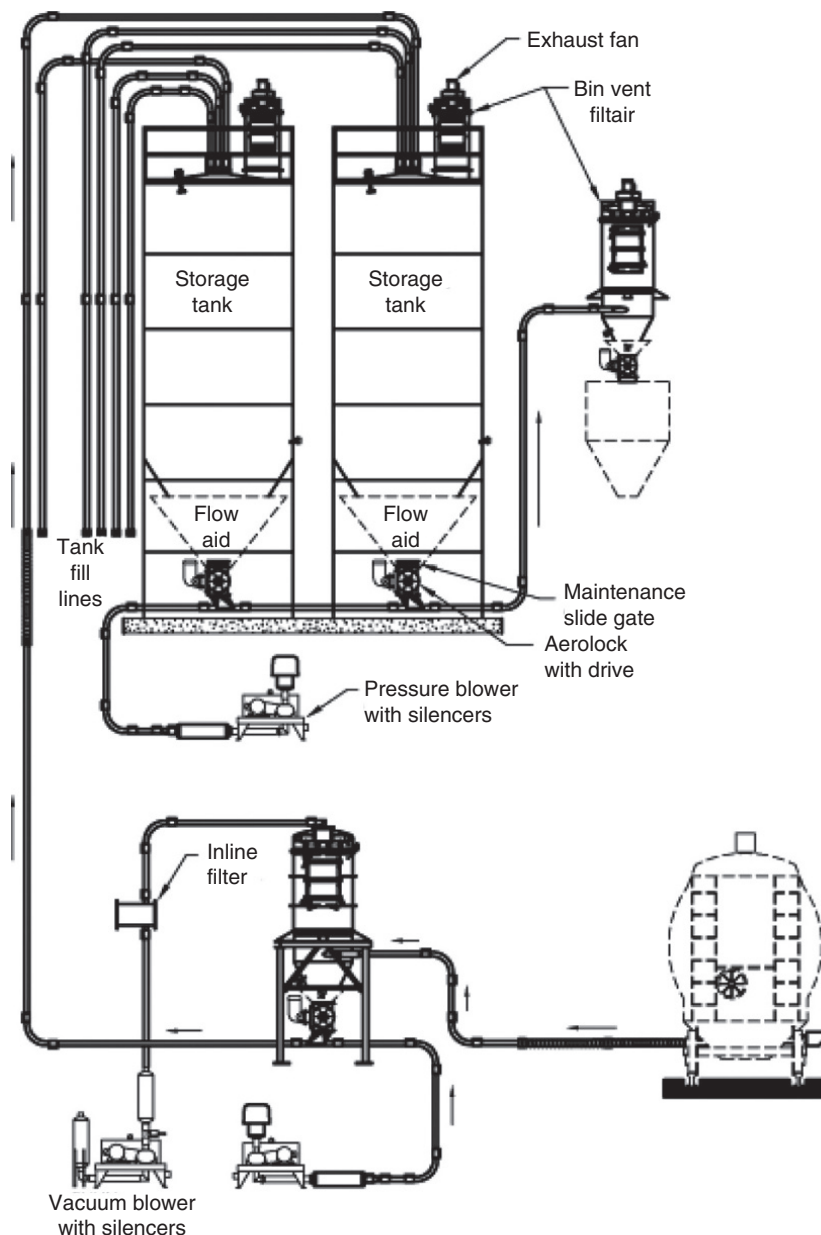


FIGURE 21.6 Example pneumatic conveying system courtesy of K-Tron Premier™.

The next step in all typical batch operations is to screen the input materials commonly described as de-lumping. This is often carried out using a screen mill but is in itself a continuous unit operation (it operates on a first in/first out principle) so although used with a batch paradigm could very easily be used as part of a continuous process. The main considerations when doing so are does the performance of the screen mill changes over time (for example, does the screen become blocked or does the mechanical action actually wear or cause the screen to break). The rate of screening (including removal of “overs” retained on screen and screen cleaning if the system is not designed for automatic removal of “overs”) must be appropriate for the overall system rate. These considerations are often not important within a batch paradigm simply because the effects are not seen but critical to a continuous manufacturing system.

However, a more elegant approach is to combine the screen into the dispensing operation. Many feeders can have mixers, micro-centrifugal feeders, or even screens incorporated within their design. Combining the two unit operations in this way simplifies the production process but also has similarities with a dry granulation unit operation where milling is regarded as integral to the granulation process.

21.4.3 Blending

There can be little doubt the blending unit operation is the most common unit operation in pharmaceutical secondary production, it is also the operation with the fewest comparisons between batch and continuous (unless running in a mini-batch mode) (because of how the operation is carried out). However, continuous blending is common place in other industries and in the last few years, continuous blenders which claim compliance to Code of Federal Regulations (CFR) and current Good Manufacturing Practice (cGMP) have also become available and have been used by all the approved continuous manufacturing products to date and in the foreseeable future.

Because of the widespread use of this unit operation but also because of the lack of experience in applying continuous blending to pharmaceutical processes, this processing step has been subject to the greatest intensity in academic and theoretical research in recent years. Laurent and Bridgwater [7–9] were one of the first to investigate the flow patterns within a continuous blender, using techniques such as tracking radioactive tracer; this allowed them to generate the axial and radial displacements as well as velocity fields with respect to time. This was followed by Marikh et al. [10], where the focus is on the characterization and quantification of the stirring action, relating it empirically to the flow rate and the rotational speed of the continuous blender. In doing so it systematically investigates the effects of operating conditions (such as rotational speed and processing angle) and

design parameters (such as blade design) on the mixing efficiency.

However, the key to the successful use of continuous blending is recognition that the blender actually has to fulfill more than one purpose. Put simply, it must dampen time-domain fluctuations in feed rate from the feeders and simultaneously achieve a homogenous blend to ensure content uniformity of dosage units is achieved. A further refinement by Pfizer is to also be able to adjust distance traveled and hence the extent of lubrication. Its primary role is to take the variation in the disparate individual feeds (API and excipients) and generate a single uniform blend, such that each and every individual unit dose is of appropriate quality. However, in order to achieve the blender’s primary role, the continuous blender has to remove any time-domain variability remaining from the dosing operation. As such, a continuous blender has previously been described as “variability reduction ratio” (VRR) device. Williams and Rahman [11] proposed a mathematical approach to predict the VVR, utilizing data generated from a residence time distribution (RTD) test for both an “ideal” and “nonideal” blender. The metric of “ideality” is defined by a mixing efficiency proposed by Beaudry [12]. In another publication, Williams and Rahman [13] investigated this mathematical methodology by using a salt/sand formulation of different compositional ratios. They verified the predicted VRR with experiments and suggested that the results were comparable. They also illustrated that (over at least typical conditions) the mixing speed and VRR were directly correlated. Harwood et al. [14] studied the performance of seven continuous mixers as well as the outflow sample size effect of sand and sugar mixtures. All of these activities were reviewed and then additionally verified by P.M. Portillo et al. [15], including experimental investigation of operation and design parameters such as processing angle, impeller rotation rate, and blade design are examined.

In summary, all these investigations show that the powder’s residence time and number of blade passes it experiences were affected not only by rotation rate but also by the processing angle, and that an upward processing angle and low impeller rotation rate are the optimal processing settings, when combined with optimal blade design. These generate a slight backflow between blade rotation and a turbulent flow within the linear flow of the process.

In this last paper [15] a type of continuous/in-line blender (manufactured by GEA) called the Continuous Dry Blender is used – this is the first dedicated, purpose-designed for the pharmaceutical industry, continuous blender (see Figure 21.7).

Fundamental research is still ongoing into continuous blending with the primary focus being the addition order of individual components. This can be achieved by adding different ingredients at different points along the path of blending. For instance, adding an ingredient near the end of blender results in less mixing. Thus, for the first time

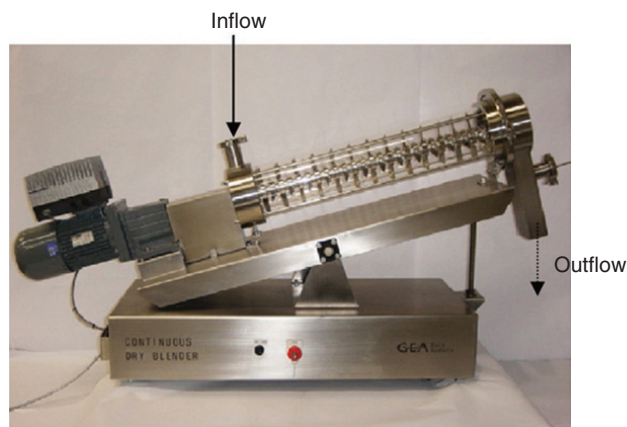


FIGURE 21.7 Early version of the GEA continuous dry blender [14]. *Source:* Reprinted with permissions from Portillo et al. [15]. Copyright (2008) Elsevier.

the dry blending process can be engineered to allow optimized mixing/interaction of components to effectively “build” the formulation in a structured way. One area where this is critical is around the addition of the lubricant component; or more accurately what type of effect is trying to be achieved by the addition of the lubricant, i.e. do we want the lubricant to be distributed within the blend but remain as a discrete powder or do we want the lubricant to be smeared over particles of the other components? In the past we could change blend time and rotational speed but little else. In addition, the recent emergence of continuous mixing technology (CMT) mixers has also allowed control of the relationship of hold-up mass and impellor speed on lubrication effects.

Not only are continuous dry blenders now commercially available so are specialized PAT measurement systems to monitor the blending process in the same way that online near infrared (NIR) systems are commonly used to monitor the batch-blending process [16], ultrafast scanning (diode array) NIR linked to optimized sample presentation systems are available. Using these systems it is relatively easy to get comparative data on the trajectory and end point of both batch and continuous systems and therefore even compare the output from both. If we first look at a simple development-scale batch process, an NIR prediction model can be generated to trend how each component changes over the blend process over time (Figure 21.8).

In the example above (carried out at in a Paterson Kelly 4qt V-shell blender) we can see the batch process reaches uniformity after around 20 rotations, with one component taking longer to reach uniformity than the other two. The drawback of batch mixing is the extreme difficulty in scaling-up from small blenders such as this approximately 4 l blender to production size units of 300–1800 l. This typically requires blender trials at intermediate and full scale which consume

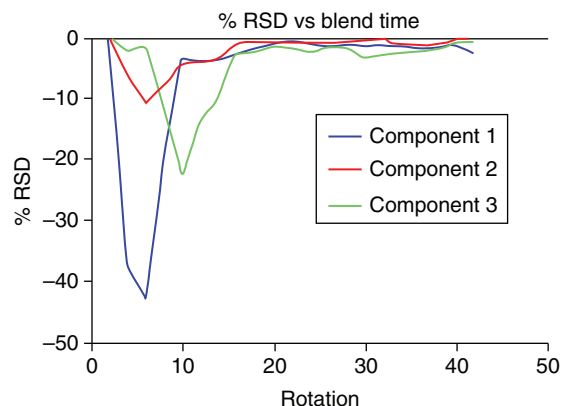


FIGURE 21.8 Typical batch blend plot.

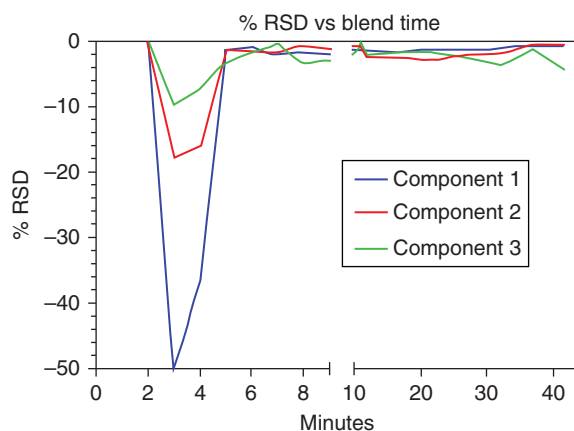


FIGURE 21.9 Typical continuous blend plot.

very large quantities of API. If we look at exactly the same composition running in a continuous mixer from start-up, the plot is slightly different (Figure 21.9).

This data is from a “dry” start-up, i.e. from when the feeders themselves are started and with the blender empty. In this case the continuous blender has a volume of 500 ml and at the powder flow rate used (20 kg/h), the blender has a mean residence time (MRT) of around 90 seconds – meaning it therefore takes 90 seconds before the powder starts to exit the blender. From this time point it then takes approximately 3 minutes to reach a % RSD equivalent to the batch process.

The significant detail in this case is the start-up process used approximately 2 kg of powder to reach this steady state (approximately the same weight as used in the 4qt development batch). While this may not sound like a compelling reason to switch from batch to continuous, this is an enormous benefit to continuous blending. There is no scale-up involved moving to commercial scale; the powder flow rate used (20 kg/h) is equivalent to 140 000 kg/year at expected equipment utilization rates (80–85%). We are able to go from

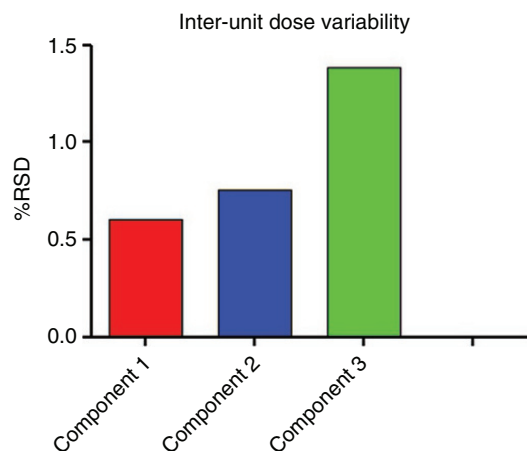


FIGURE 21.10 Typical inter-unit dose variability plot.

development straight to commercial scale because unlike batch processes with scale in “volume,” continuous manufacturing scale in “time”.

However, we also have to consider the scale of scrutiny when describing measurement systems. In a batch process we can statistically sample (spatially) the output of the blend process; in continuous blending we have a different and more smaller scrutiny scale, effectively the uniformity of the individual/consecutive unit doses is generated here (and in the feed frame in the case of tablets) and simply doing a unit-dose-scale measurement at a fixed time interval across the batch can miss unit dose to unit-dose variability (in much the same way that inadequate or poorly specified sampling will miss variability in a batch process). If we look at the unit dose (in this case the product is 500 mg image) to unit-dose variance and calculate % RSD, we get a very good demonstration of the high-frequency variability in the system (Figure 21.10).

In general, the application of PAT to the blending step of a CM process, for both in-process control (IPC) and to provide data for product release, can be described as gravimetric (using the weight data from the feeders and possibly having the blender itself on a load cell/weighing the material leaving the blender) or based on the measurement of the powder itself using NIR, sometimes even both in the form of a redundant measurement. Sample presentation is critical to the success or failure of any powder or multiphase system. In addition, one needs to determine whether the objective is to measure the output of the mixer or the input to the dosage-formation operation. In the latter case, blending in the feed frame may drive sampling in the feed frame rather than in the powder chute. When describing feed rate data as a form of PAT, it is critical to understand that, for example, feeder data can be used in multiple ways. Instantaneous data can be used to determine feeder status (warning/alarm), in case of out-of-limit value or an unacceptable signal-to-noise measurement

(perturbation). Yet, the same data can be averaged over time to either reflect the mass contribution to an attribute, for example, to give the mass composition relating to a unit dose, in the case of plug flow, a fixed mass of material. This is a good case study of an overall control strategy, where the performance of the feeder cannot be looked at in isolation. Rather than the feed rate from the feeders, the effect of the VRR of the blender and also the weight of the final tablets giving rise to the attributes of assay and content uniformity (CU) must all be looked at holistically.

When considering NIR (or other spectroscopic) solutions to measure the concentration of components in the powder blend, it is very important to understand that the NIR signal is simply a response generated from the mass of material it sees. Whichever approach is used, it is essential that the signal is related to a mass of powder, and that means carrying out the measurement on powder where the flow is controlled and consistent. There are currently two widely used approaches. Firstly, the measurement can be carried out in the inlet chute of the next unit operation; here we will consider a situation where a blender is directly feeding a press. If the press is set to run at a constant turret speed, the linear velocity of the powder in the inlet chute is consistent if the chute is operating in mass flow rather than funnel flow. This constancy can be used to move the powder across a NIR probe mounted across the inlet chute. This transverse configuration allows representative data to be collected across the whole process flow. However, the mass of the powder being measured could still change somewhat (even if the flow rate is constant) if the bulk density is being impacted. For this reason, a measurement in the inlet chute is typically accompanied by the generation of a (powder) flooded section above the feed frame, with powder level controlled. Using this setup, a simple calculation can be made to estimate the measured mass (optical area per acquisition \times number of acquisitions \times depth of penetration = sample area, sample area/BD = mass contribution), but more importantly changing the number of acquisitions gives flexibility to change the mass contribution to match the attribute being measured, e.g. equivalent to unit dose. The selection of the number of acquisitions should be related to the unit dose mass.

The second approach tries to include the additional intermixing that happens in the press feed frame itself. A probe with end-facing optics can be placed looking directly onto the powder in the feed frame. Here, the powder flow is constrained by the movement of the feed-frame paddle wheels, and this movement creates the consistent powder flow. Here, the optical area per acquisition is much smaller ($\sim 1/7$ th of the transverse probe) but the same calculation of mass contribution can be made. When using this location it is important to understand that there will be some compression effect as the paddle wheel passes the probe (and thus a change in powder density) but also that the design of the paddle wheel has an impact on the movement of powder (for example, a star

paddle wheel consistently moves the powder but causes compression at the dies and so is often replaced with a bar or “spaghetti” or “wire” paddle for cohesive blends; these paddles are designed to move through the powder, not push the powder and so the mass contribution will be very different as the powder in front of the probe maybe much more static). All of these factors must be taken into consideration when building the model.

Regardless of where a probe might be located, consideration must be given to how that information will be used in the overall control strategy; whether it is used at all and if so for monitoring purposes only, for powder or tablet rejection, or for feedback control. Consideration should also be given to the sampling frequency in relation to the rate of change and duration of disturbances. Too long an averaging time or too low a frequency risks missing disturbances. However, it is possible these disturbances may be detected by other means at other points in the process.

Although direct compression is the simplest form of solid dosage production, it has significant restrictions in use in batch operation. Typically, solid dosage formulations do not use API and excipients with comparable physical properties such as particle size, which means there is a tendency to segregate post blending but before the final dosage form is made. This means in batch production it is very common to follow the first blend step with some form of granulation. However, CM direct compression minimizes the opportunity to segregate (because the blender directly discharges on a continuous or mini-batch basis into the feed chute of the press/encapsulator). Current industry estimates indicate that in the future up to 70% of products will be continuous direct compression, and granulation only being applied where there is a need to modify solubility or permeability.

21.4.4 Dry Granulation

As described in Section 21.4.3, in dry granulation the blending step is typically followed by another unit operation originally developed for another industry: RC. In typical pharmaceutical processes the RC step is carried out as part of a batch production line; however, the activity follows a first in/first out principle and is inherently “continuous.” Powder is fed from a feed or charge hopper into the RC unit and between two counter-rotating rolls. The compression force (and utilizing the elastic strength of the individual particles) causes the free-flow blend to form solid compacts (sometimes ribbons, sometimes briquettes); see Figure 21.11.

Even though the actual RC activity is “continuous,” variations in powder flow into the feed hopper will impact the uniformity of the compacts produced, both in terms of physical (i.e. tensile strength) and chemical (segregation). For this reason most commercial-scale and many development-scale RC units include a mechanical system (similar in design to an automated feeder) to deliver a constant feed rate at the rollers.

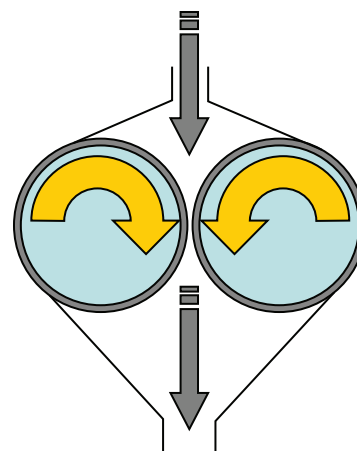


FIGURE 21.11 Schematic of roller compaction.

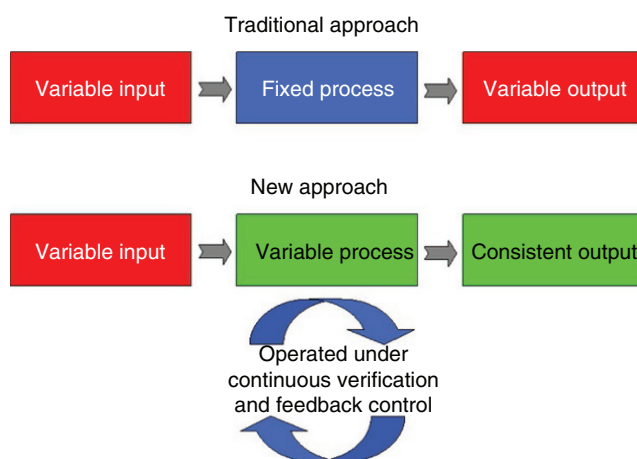


FIGURE 21.12 Continuous quality verification.

Most RC units used in the pharmaceutical industry were themselves developed/optimized with sophisticated feedback controls (for speed, press, and even torque) to function with little variability.

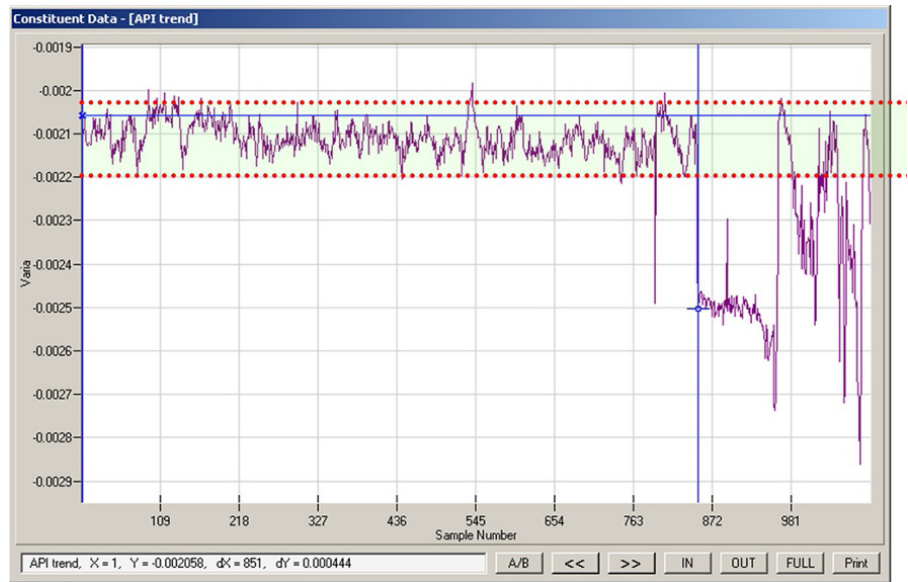
Because the mechanism used when reducing variability in the output of the blending step may actually cause variability in the RC step, when using RC as part of a continuous process specific consideration has to be given to changing this paradigm. It is possible that the RC process itself will need to be adjusted in order to cope with varying input and thus ensure a constant output. This fundamental change to the way pharmaceutical processes are run is consistent with the current initiatives and is often represented graphically, and is known as Continued Process Verification and is particularly important to highly constrained traditional processes such as RC (Figure 21.12).

There are PAT measurement systems now available that have the capability to monitor both physical and chemical changes (e.g. density, which in turn impacts tensile strength)

TABLE 21.1 Online Measurement of Variability

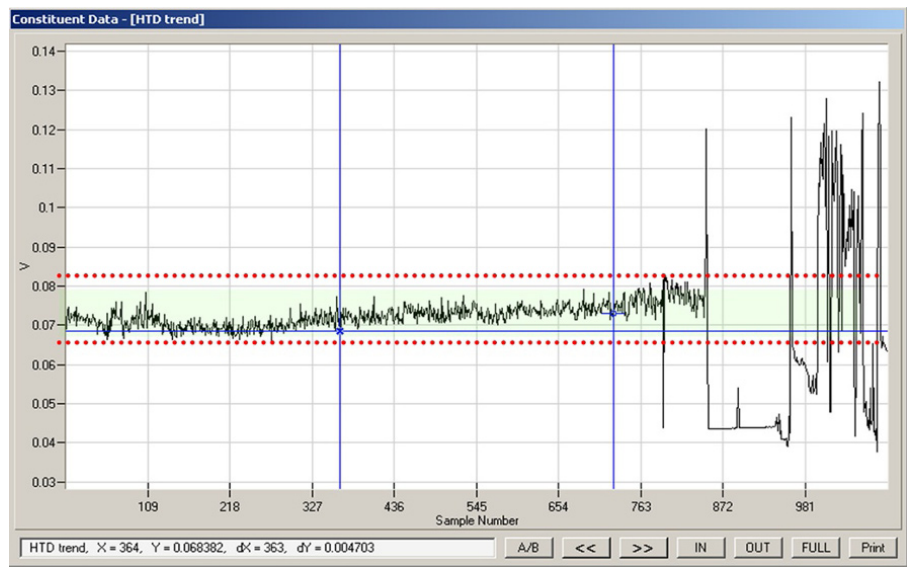
API Trend

Showing constituent alarm states and the effect seen when a process parameter changes to take the process variance outside expected norms



Physical Variance

Showing constituent alarm states and the effect seen when a process parameter changes to take the process variance outside expected norms



of RC ribbons. These allow continuous online measurement and real-time prediction (Table 21.1). Specific optical measuring heads have been developed which allows the use of NIR directly onto the compacts as they come off the RC rollers (see Figure 21.13, Table 21.1, and inset figures). Research has also been done on using process parameters to predict ribbon density.

In this case study, intact ribbons are generated as a result of compaction. This is not always the case and it is essential that the capabilities of the measurement systems are matched to the application in terms of speed of analysis, rate of analysis, reproducibility, and sample size.

The second component of dry granulation is the milling or granulation step. Here, the compacts are typically put through

a screen mill, again on a first in/first out basis so inherently continuous. When running this sort of mill under batch conditions (much like the example given for the roll-compaction process) the main considerations are around the performance of the screen mill, and does it change over time. Particle characterization post milling can be carried out using focused beam reflectance measurement (FBRM). FBRM utilizes a spinning laser (of known rotational speed) to measure the chord length across any particle, by simply back-calculating the duration of reflection of the laser off the particle (see Figure 21.14).

The laser light is delivered via fiber-optic probe and so is relatively easy to install in the output stream of the RC granulator, often fitted with optional gas purge or mechanical “wiper” to keep the tip clear (see Figures 21.15 and 21.16).

21.4.5 Wet Granulation

As described in Section 21.3.1, the need to granulate pharmaceutical powders is common; however, it is not always possible to dry granulate, because the powders may not have sufficient elasticity (they are too brittle) or because the differences in particle size distribution or physical properties are too great. In these cases it is very common to wet granulate.

Unlike dry granulation, traditional pharmaceutical WG is not continuous. Continuous granulation has been widely practiced in the consumer goods, agriculture, and specialty chemical industries since at least the mid 1980s. In pharmaceuticals there are examples dating back to the mid 1980s,

suggesting alternate approaches to traditional WG which could be run continuously or semi continuously, for example Koblitz and Ehrhardt [18] published on WG and using continuous variable frequency fluid-bed drying.

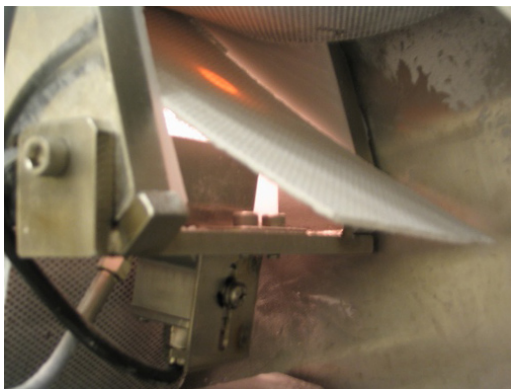


FIGURE 21.13 A NIR measurement head installed in Gerteis Macropactor.



FIGURE 21.15 FBRM with purge tip installed on the Gerteis Macropactor.

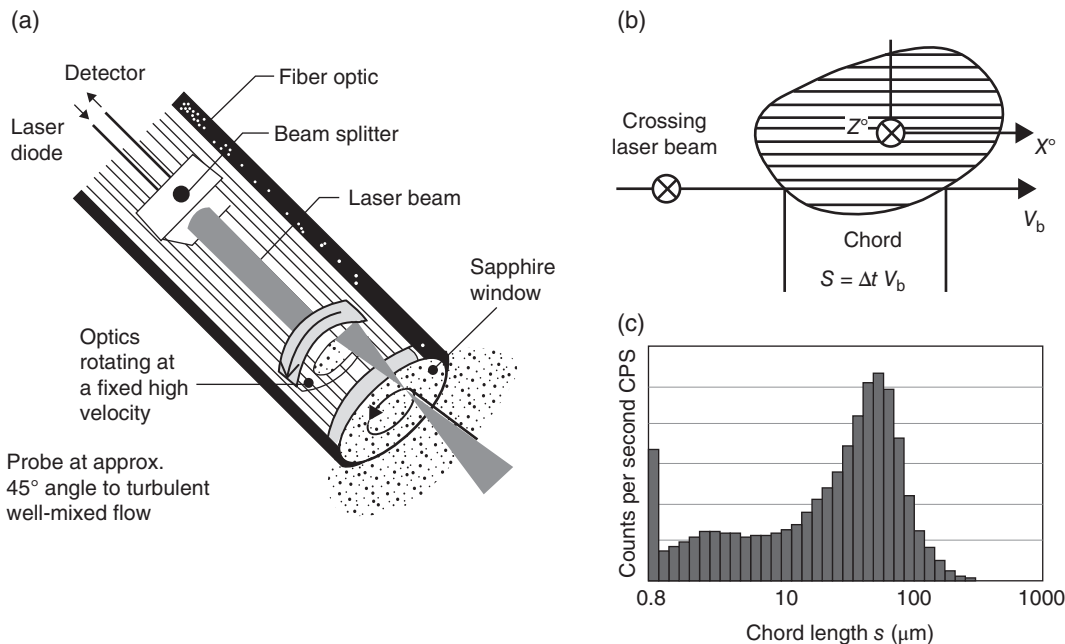


FIGURE 21.14 Theory of FBRM measurement including: (a) FBRM probe, (b) chord measurement (the laser beam direction is perpendicular to the paper) with coordinate system corresponding to the measurement, and (c) histogram of the chord length counts. *Source:* Reprinted with permissions from Ruf et al. [17]. Copyright (2000) John Wiley & Sons, Inc.

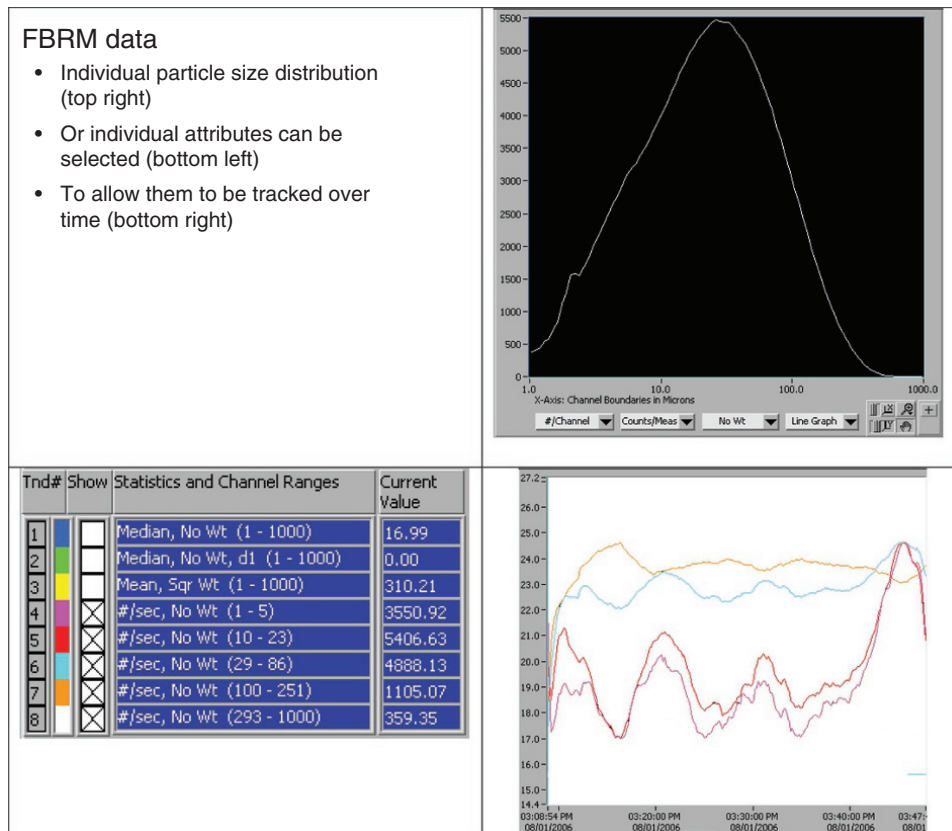


FIGURE 21.16 Monitoring the RC milling process using FBRM.

A breakthrough approach came from Glatt with the launch of their Glatt Multicell (GMC) in the late 1990s. The technology has not seen widespread adoption but is well documented including several publications by Leuenberger [19].

The GMC 30 is comprised of a 271 High-Speed Plough-Shear granulator which equates to a 5–9 kg sub-batch (Figure 21.17). The granulator “self-discharges” via a de-lumping system into a multistage fluidized bed dryer. After granulation/wet massing, the material is conveyed sequentially through three stages of drying. In this way, four small batches (one in granulator and three in drying) are processed simultaneously and the cycle repeats for semi-continuous operation (Figures 21.18 and 21.19).

Although this system is best described as a mini-batch system, the technical significance of this system should be recognized, especially as it was one of the earliest examples of continuous verification and feedback control; each of the three fluidized bed towers can be fitted with a noninvasive NIR measurement system, which simply views the drying process through the preexisting inspection windows. A control strategy is then put in place balancing the sub-batch throughput (Figure 21.20).

Once dried, the sub-batches could be discharged directly into a second continuous blender for lubricant addition

although all known implementations currently collect the sub-batches to form a single batch which moves forward.

In their White Paper for PharmaManufacturing.com, Mol-lan and Lodaya [20] identified that a continuous fluidized-bed granulation system would have five or more functional zones. These are product in-feed zone, product mixing and preheating zone, spraying zone, drying and cooling zone, and discharge zone, with a more detailed explanation being given by Paul et al. [21]. Continuous versions of some of these individual functions are available and have been published. Lindberg [22] used an inversion mixer (where powders and liquid are metered into a narrow space at the periphery of the grooved disc, which rotates at high speed) to study WG of placebo. Applegren et al. [23] used a similar system to study continuous melt granulation, and a system is commercially available that uses a planetary extruder to granulate and a microwave tube to dry continuously. As an alternate to microwave energy there is also a commercial system based on using radiofrequency energy (using wavelengths specific for aqueous drying). Until recently, radiofrequency heating has been used mainly in other industries such as food, paper, and ceramic. Jones and Rowley [24] have reviewed several applications for drying where dielectric heating is used by itself or in combination with other methods.

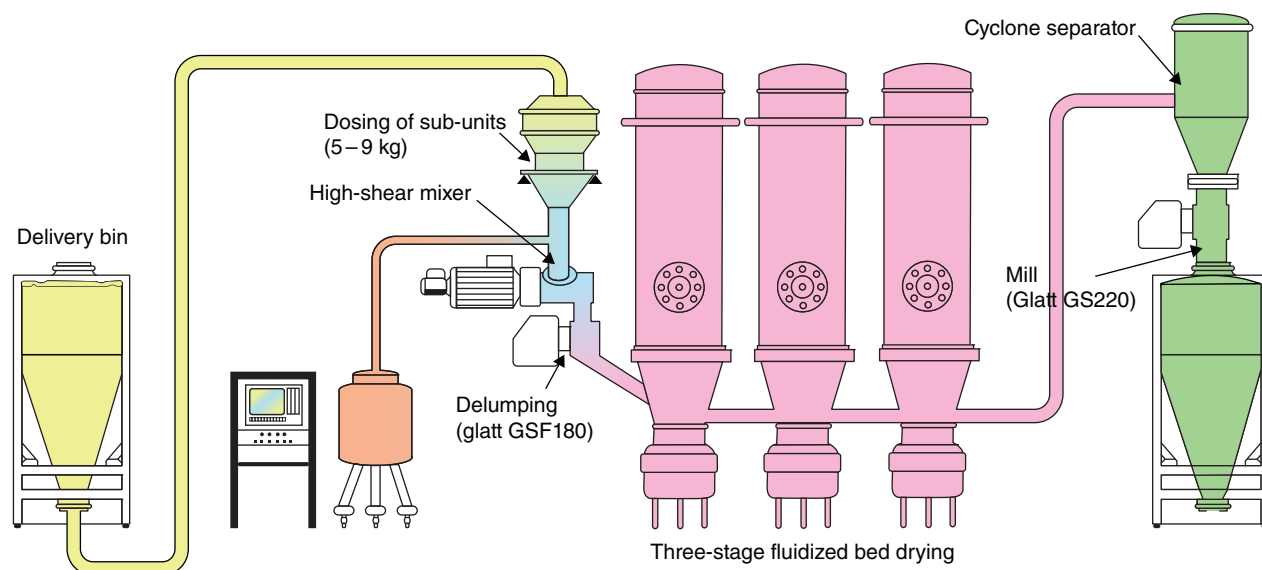


FIGURE 21.17 Glatt Multicell GMC 30.

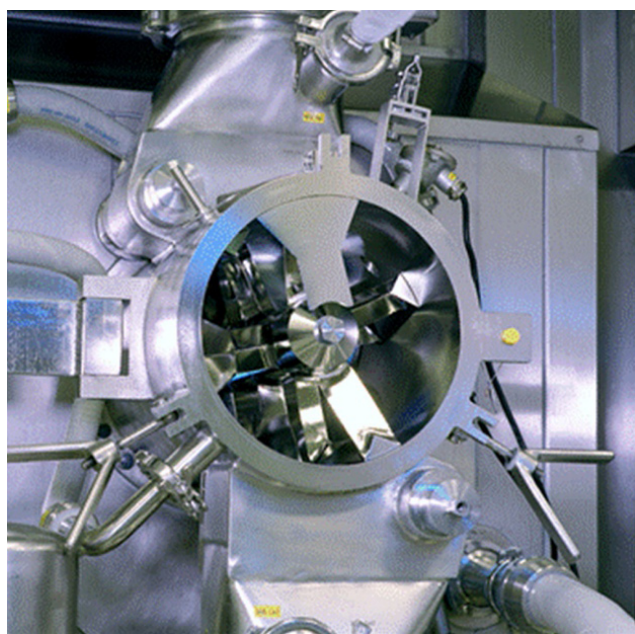


FIGURE 21.18 HSP-S mixer.



FIGURE 21.19 GMC 30.

However, the type of WG which has the greatest potential for continuous operation is using a twin-screw granulator/mixer. This is a modified twin-screw extruder with the extruder plate removed and relies on twin intermeshing screws that convey, mix, wet granulate, and wet mass the powder blend. These systems offer several advantages over traditional WG processes, and the interchangeability of screw elements ensures flexibility. Twin-screw extruders themselves have been utilized for WG since the 1980s [25, 26], and some

aspects of the application are also covered by patent [27]. In addition, Ghebre-Sellassie et al. [28] have published on a continuous WG and drying system that combines twin-screw mixer (for WG and wet massing) with radiofrequency energy (for drying).

More recently, commercial systems have become available from GEA called the ConsiGma™, Glatt called the MODCOS®, LB Bohle called the QbCon®, as well as others.

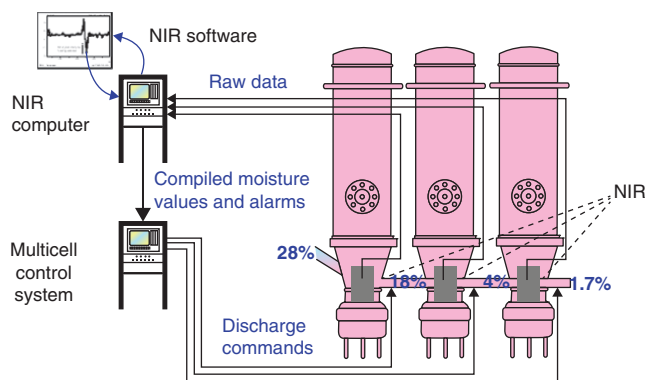


FIGURE 21.20 GMC 30 NIR control strategy.



FIGURE 21.21 GEA Pharma Systems CONSIGMA™.

The GonsiGma™ system is at the center of an integrated tablet line (continuous tableting line (CTL)) and starts with separate liquid (binder) and powder-feed systems (the powder feed either being metered directly or coming from the outlet of one for their continuous dry blender systems). These feeds come together in a twin-screw granulator, which continuously outputs into either a linear or segmented dryer (Figure 21.21).

Interestingly, the ConsiGma™ does not utilize a continuous dry blender for addition of the lubricate post drying.

Instead, after drying, the product is held in a dedicated discharge hopper/mill and a mini-batch approach is taken using a small-batch “lubricant” blender/feeder (called the Press Preparation Unit). All of this is under an integrated control system.

The focus on twin-screw WG has not been exclusive, and other manufacturers, for example, Bosch Packaging Technology, have launched mini-batch fluid-bed granulation, which includes direct transfer from single-cell development to multiple cells run sequentially to deliver a continuous fluid-bed granulation in their Xelum® processing line.

These different approaches to blending and granulation also lead to different approaches around control strategy. In general, processes that maintain plug flow (like mini-batch operations) rely on MRT for material tracking and segregation decisions; however, unit operations which are fully continuous rely on RTD, and therefore the intermixing between materials produced at different times, for control decisions. In order to justify the use of RTD as a basis for control strategy, a detailed knowledge of powder/process behavior is needed. This can be achieved by either spike or step-change studies to introduce a transient and track its propagation through the system. An example of a spike study would be the addition of a fixed mass of material before an in-line blender, then collecting data (either using PAT or from physical sampling) to show how the mean retention time of the material changes but also the distribution around the mean. As an illustration, the RTD produced by an in-line blender can be shown to have a fairly symmetrical distribution around the mean, this is because a linear blender tends to generate equal amounts of forward and back mixing. However, the RTD of a CMT blender is different and the spike is rapidly distributed, and the RTD is mainly described by the dilution effect as subsequent material added to the blender washes the spike out, this generates a RTD with a very sharp leading edge and a tail. These RTD can be compared with that of a mini-batch blending step where there is uniformity across the mini-batch and little/no intermixing between mini-batches. Thus, the RTD of a mini-batch blend is a block where the width of the block is the mass of the mini-batch as a function of line rate, for example, a 1 kg mass mini-batch running at 20 kg/h will have a block width of 1/20th of an hour or 3 minutes. Because the RTD of mini-batch blends is a block function, it is desirable to keep the mini-batch mass to a minimum, and generate multiple, discrete, mini-batches. Because each mini-batch represents a segment of production, and that production could be a small number of mini-batches (for example, in development or for clinical supply) or a large number of mini-batches (for commercial supply), this has to be reflected in the control strategy and demonstrated during validation.

The complications come when plug-flow processes also include fully continuous steps (for example, encapsulation or compression) and then a mixed approach based on plug flow (RT) overlaid with RTD needs to be executed. Also

often missed in these residence time studies is the impact of feeders when used as part of the main process flow. As an example, consider a segmented dryer discharging via a mill to a blend step; if the whole output of the dryer cell is used to charge the blender, the integrity of that mass of material is maintained and there is no intermixing with adjacent material. If, however, the material charges into a feeder which then discharges into the blender, the agitator with the feeder hopper will mix that material with material already in the hopper and any additional material added before the original mass is discharged. This leads to a situation that for weight control it is better to accurately dispense the material (i.e. use a feeder) but for control strategy/material traceability it is best to simply weigh the material that arrives and adjust the extragranulated components accordingly.

21.4.6 Compression/Encapsulation

No matter what the route taken to reach the compression or encapsulation unit operation, and no matter what vendor system is adopted in the step, the compression step has generally been a continuous process. As an industry we traditionally have chosen to charge the feed hopper as a batch and collect the output as a single entity. If this step would run truly continuously, some consideration needs to be given to a reduced line rate (often pilot-scale equipment is better suited to CM processes), dust buildup (even if preceded by a de-dusting step), product buildup on press tooling (which will cause issues with subsequent defects on tablet cores), and (in the case of encapsulation) the replenishment of raw materials for encapsulation (the empty capsule shells).

The addition of automated tablet testers and capsule checkers for feedback control also allows the compression/encapsulation process to be adjusted in real-time (for tablets, any one or two of weight, thickness, and hardness). The addition of spectroscopic to these automated tablet testers allows confirmation of attributes such as assay, content uniformity, and identity for IPC and real-time release testing (RTRT). The frequency and periodicity of these tests need to be defined within the overall control strategy.

21.4.7 Coating

When describing the coating process, we have to remember there are two input streams. The first is the uncoated cores but we also have to consider the spray solution itself, so even if a continuous coater was fed directly from the press, the spray solution preparation also needs to be addressed. That said, continuous coating is performed in food, flavor, and nutraceutical processing and the first commercially available continuous coater with claimed compliance to current CFR for use in pharmaceutical production has been launched by both O'Hara Technologies (fully continuous) and LB Bohle called KoCo[®] (mini-batch). Other companies are in late-stage development.

Their designs resemble an extension (in depth) of a standard batch-rotating drum coater and spray manifold, but with the addition of inlet and outlet chutes to create a flow-through process. There is still the same level of development needed and adjustments to tablet feed rate, pan RPM, and residence amount and/or time all affect the coating uniformity. The unit is quite new to the market but details from the vendor indicate that tablet cooling, elevating, and/or waxing can be added at the discharge end of the coater. The unit also includes some improvements to their spray manifold design, improving solution distribution and an update of their air caps, providing better anti-bearding properties (which are required for continuous operation). The coater is also designed to be run in batch mode during start-up and development.

However, a different approach has been taken by GEA with their Omega[™] coater. This is a mini-batch approach that utilizes centrifugal force to first pin the uncoated tablet cores to the perforated drum, but then break that force to allow the cores to free fall through the coating spray. This approach generates a faster cycle time (coating times as low as 3 minutes/3 kg have been demonstrated) which allow them the Omega to keep pace with the typical line rates of a OSD CM line, and even if a single Omega is not shown to be able to keep pace, a second can be added in parallel, allowing duty-cycle operation, thanks to its mini-batch approach. This has the advantage of using less material during development than the larger continuous mixers.

21.4.8 Packaging (Including Printing of Final Dosage Form)

All current pharmaceutical packaging processes are run as batch processes but are inherently continuous. However, much like encapsulation, consideration has to be given to the multiple input streams (product and packaging materials) and the biggest issue is maintaining a continuous, traceable, supply of package materials.

21.5 CREAMS, LIQUIDS, AND SUSPENSIONS

Continuous processing has been widely used in the soft drink and consumer product industries for decades. In pharmaceuticals these concepts have also been implemented in the area of sterilization and solution manufacturing. In addition, it is normal for the containers/bottles/ampoules/pouches used to be manufactured alongside the actual product, even under the same sterile conditions. It is also common for the product strength or even alternate products to be run concurrently (with the appropriate change-over procedures being run automatically – including appropriate PAT measurement systems to provide verification of the change). This class of products comes as close as any to realizing true continuous processing in pharmaceutical production; although they typically start

and finish with a batch solution/suspension preparation, all other unit operations (including things like “blow fill” container formations and dosing) run continuously until the product is collected into batches at the end.

21.6 LYOPHILIZATION

Currently, lyophilization is carried out in very large batch sizes (based on the number of individual samples) and there are no indications/research ongoing to suggest this will change in the near future. However, Rey [28] proposed some very interesting concepts on continuous or semi-continuous lyophilization technology based on practices from the food industry where continuous freeze-drying is deployed, and there are currently several academic and academic/consortium activities focused on developing continuous lyophilization operations.

21.7 NOVEL UNIT OPERATIONS

21.7.1 Spray Dried Dispersion (SDD)

These are common in food and other industries and becoming more so in pharmaceutical, primarily as they provide a way to alter/control the bioavailability of certain API. They also provide a mechanism for holding the API in a specific state/form. During the SDD process the API and a waxy polymer are dissolved in solvent before the solution is sprayed under controlled conditions to generate a modified API, with defined particle characteristics (which actually make secondary formulation more straightforward – often direct compression). The actual spraying process is continuous (first in/first out) and even current manufacturing approaches could easily be adapted to flow production but also modified to be truly continuous. Adaption of API to give better material-processing characteristics will also open more opportunities for incorporation into OSD CM lines and it is possible to visualize a flow chemistry line feeding a continuous SDD or extrusion step for API physical property modification directly into a continuous direct compression process running at 1–5 kg/h; these component parts are available now.

21.7.2 Melt Congeal Extrusion/Spinning Disk Extrusion

Much like SDD production, the melt congeal/spinning disk extrusion process is deployed to modify the availability of the API; however, in this case it is normally to modify the rate of release and provides the basis of slow/sustained release formulations. The actual extrusion process is continuous and the batch nature of production comes not even from the initial feeder hopper but from the collection into batch post

processing. Melt extrusion processes also can be set at line rates of the same order used in flow chemistry and so a melt extrusion/flow/simple OSD process is possible, and also has been demonstrated by Novartis-MIT in their collaborations.

21.7.3 Webs/Oral Care Strips

Web-based products such as oral care strips bear more resemblance to screen printing than pharmaceutical manufacturing. Their production is continuous but the two input streams are both batch (the support/paper backing and gel-like product suspension). Even though the suspension preparation could be made continuous, this process is more easily adapted to flow production than continuous.

21.7.4 Transdermal Patches

Much like web-based products, transdermal patches have more in common with printing than pharmaceuticals. Typically, they are produced by deposition (sometimes spraying, more often roller deposition) into a permeable support medium, over which a protective coat is then applied – forming a sandwich. The support medium and protective coating come on long rolls (much like the paper used in a cash register and used to provide a till receipt). The solution preparation (typically, purely a dilution of the API in a carrier) is batch and normally highly toxic (for example, nicotine solution used in nicotine patches is classified as an occupational exposure band (OEB) level 4/5 because in solution form it is not only toxic but also readily absorbed. It is not likely that production of these types of products will become truly continuous in the near future.

21.8 WHY CONSIDER CONTINUOUS PROCESS FOR DRUG PRODUCT OPERATIONS?

21.8.1 Benefits of Continuous

The two biggest advantages in developing continuous manufacturing rather than batch are the elimination of time and material consuming scale-up and an overall reduction in the amount of API consumed during development. Additionally, developing a direct compression process tends to be faster than a RC or WG process. So to the extent of CM enables DC, this is a third benefit of continuous in development.

Processes are developed at the same process flow rate as commercial manufacturing; it is purely that the process runs for a longer period of time in commercial production. This is key; the batch process performance changes with scale, plus often development activities are not even carried out on the same design-process equipment (for example, a V-shell blender being used in development but a bin blender used in commercial manufacture). These types of dramatic changes in equipment

geometry and scale result in unavoidable differences of scale-up criteria such as surface area (proportional to L^2) and volume (proportional to L^3), which lead to significant differences in the way the process to make the product performs.

Typically, this goes hand in hand with a larger equipment footprint at commercial scales, e.g. a development-scale blender is around 3 ft tall, while a production-size V-blender can be one to two stories high, and this is just the blend step – a complete direct compression equipment train with gravity flow between production steps typically requires a building three stories high. The same annual output can be achieved from a self-contained, typically wall-mounted, process suite occupying only one room.

Continuous processes also provide the ability to vary batch sizes based on product and demand – we simply run longer. This enables lower inventory, fresher product, less expired product when used in a just in time (JIT) or make to order (MTO) manufacturing paradigm which is routine outside of pharmaceuticals. Having a smaller footprint in a cGMP space is a huge cost savings because the cost of construction and operation are typically proportional to the area or volume of the manufacturing space. Furthermore, if the equipment could be “skid mounted” and pulled out of storage only when needed for use, then this introduces the idea of the equipment being housed in a cGMP bubble which could (in theory) be dropped into any cGMP facility (for example, a contract manufacturing organization (CMO)) and run under that facility-compliant processes.

Smaller equipment also typically means cheaper equipment; certainly comparing the cost of the large V-shell to a typical continuous blender has the V-shell costing around ten times more.

Because these systems are designed to run continuously (with 80–85% availability) they have much higher equipment utilization rates (a typical batch blender has 25–30% utilization). They are also (typically) highly automated, resulting in lower labor costs and higher operating efficiencies. Another advantage of continuous processing is a reduction in work in progress (WIP) time and therefore inventory that needs to be held, leading to JIT manufacturing.

In commercial manufacturing there is also a significant advantage in running processes continually in a state of control (rather than those which progress towards an end point); there is a reduction in variance but also it is simple to introduce PAT measurement systems to increase quality and reduce waste through continuous improvement (CI). Especially, when we also consider that these systems are typically contained, from start to finish and therefore more applicable to high-potency products but also often include automated clean in place (CIP) systems which allow automated changeover between products, which is particularly important when you consider the benefit of efficient start-up and shutdown.

This last statement regarding CIP systems is key if we consider the benefit of continuous purely from a development

view point. Part of the 21st Century Quality Initiatives is the principle of establishing Process Understanding using tools such as design of experiments (DoE). To run a DoE even at development scale will take multiple small-scale batches which consume a lot of material and also time. Whereas running the DoE (automated) on a continuous system simply means “driving” the continuous process around process space while tracking/isolating the product produced (so that the impact on the product performance can be determined). This could be carried out in two ways; the most basic is where the process simply drives to the next set of DoE conditions, waits to achieve a state of control, collects product, and then moves again; the more complicated and more rich information is there the process trajectory is investigated between the points on the DoE, this allows for a more detailed surface response curve to be generated and the uncertainty within the process space to be lowered. Typically, product development for CM has become a three-stage activity; stage one is working on the individual unit operations to establish their impact on overall product performance (i.e. identify the critical process steps). Stage two is to show the interdependency between critical unit operations; this is often done without this operation being connected, for example, the impact of granulation and tablet compression on the dissolution properties of a tablet core. Stage three is where a process design space (DS) has been identified and the overall performance of that DS can be verified and the proposed control strategy demonstrated. Typically, the number of data points carried out at each stage drops significantly with only critical unit operations needing to be considered in stage two, and a predefined series of set points demonstrated in stage three. With direct compression, lab-scale evaluation of formulation and tablet design is possible with material usage measured in grams or 10 seconds of grams. Progress is being made on being able to predict feeder performance from lab-scale data but in many cases full-scale testing is still required. The advantage of single-feeder testing is it is nondestructive for most materials so that the same material can subsequently be used for other purposes.

21.8.2 Cost Analysis

It is possible to quantify possible cost savings by comparing continuous to batch activities based on yield increases (a 2% yield improvement is common simply from start-up and shutdown savings). As an example, a typical batch direct compression solid dosage formulation requiring 80 000 kg/year could be achieved by running 100 × 800 kg batches (about the maximum number of batches possible through a single commercial blender). Start-up and shutdown of the 100 batches will account for approximately 2% or 1600 kg of waste, whereas the same volume could be delivered by running four separate 52-day production cycles (208 days in total) of a continuous system running 20 kg/h. The continuous process would waste only 64 kg. This could be further

improved if production was carried out in a single production run; however, this would be product being held on inventory (impacting shelf life) for up to five months. In addition to the yield improvements, the improved equipment efficiencies leave the equipment available for other use equivalent to an additional 50 000 kg of production.

Another approach to economics is to look at development time where a single day sooner in approval translates to millions of dollars in profit over the product lifecycle. In addition, innovations in clinical-study design to slash clinical timelines, often the critical path in a project timeline, such as adaptive designs, require a very nimble supply chain in development. Scaling up from very small to small-to-medium-scale batches is slow and expensive. With continuous, it is a matter of running longer or shorter periods of time to be responsive to the needs of a rapidly changing clinic.

21.9 IMPLEMENTATION OF CONTINUOUS MANUFACTURING

21.9.1 Regulatory Implications

In the past, one of the main reservations when considering developing and implementing continuous processes is regulatory burden. The first element often to be considered is the traceability provided by running a “batch.” However, if we refer to the CFR and look at the definition of a batch:

A specific quantity of a drug or other material that is intended to have uniform character and quality, within specified limits, and is produced according to a single manufacturing order during the same cycle of manufacture [29].

It is clear that the regulatory definitions are already in place to support the concept of a batch being a period of time, whether that time period is very short (possibly even an individual dosage form), per day, or even a longer period if the process output can be adequately controlled using continuous process verification (CPV).

“Continued Process Verification (CPV) is described as an approach to process validation where manufacturing process (or supporting utility system) performance is continuously monitored, evaluated and adjusted as necessary”.

More specifically, “it is a science-based approach to verify that a process is capable and will consistently produce product meeting its pre-determined critical quality attributes. With real time quality assurance (that CPV will provide), the desired quality attributes are ensured through continuous assessment during manufacture. Data from production batches can serve to validate the process and reflect the total system design concept, essentially supporting validation with each manufacturing batch” [30, 31].

Under this paradigm the idea of a “batch” becomes no more than form of tracking and quality assurance, i.e. CM process but batch release. Also of note, currently there is an ASTM standard published under E55 called E2668 Standard Guide for the Application of Continuous Manufacturing to the Manufacture of Pharmaceutical Products and gives clarity and guidance around process development, development of control strategy, and regulatory implications.

21.9.2 Validation

One of the latest documents being drafted as part of the FDA’s 21st Century Quality Initiatives is an updated Process Validation Guidance. This guidance divides Process Validation into three component parts: (i) “Process Design” – establishing process understanding during development, (ii) “Process Qualification” – including process performance qualification (PPQ) of the process (this step replaces the old three batch validation activities), and (iii) CPV. Developing a CM process under a QbD paradigm actually leads to a process with significantly higher level of process understanding (because we would have been able to investigate the impact on the product of many more process conditions, as we drive between points on the DoE). There is even a possibility that our confidence in how the process will run “in commercial manufacturing” (because there is no scale-up) will be so high that we could only carry out the process performance qualification (PPQ) immediately before launch, reducing the financial burden of holding registration/validation material on inventor.

There is also an expectation that CM processes will be adaptive and under continuous quality verification/feedback control which is aligned with the principle of Continued Verification. It also supports the principle of RTRT where the process is under feedback control ensuring the output quality.

However, we have another opportunity around validation; if we recognize that a CM process has various phases, start-up, shutdown, pause/restart, and operation. A validation activity has to demonstrate each phase, but if we recognize that the operation phase can be broken down into repeating cycles/operations, we can describe “an operation module” (so for a WG process running on a segmented dryer, an operation module includes running the dryer through a whole cycle, top-ups of any feeders, product transfers, etc.) which gets repeated multiple times to give the operation phase. This is significant with regard to flexible batch size as the PPQ step could include a batch consisting of just start-up, a single (or limited number) or iterations of the operation module, and shutdown (representing a development of clinical supply batch), a repeat of this including a pause/restart to validate a hold-time, and a final batch could be the same as the development/clinical batch size just with “*n*” iterations of the operation module (maybe several hundreds) to give an appropriate maximum batch size.

REFERENCES

- Kossik, J. (2002). Think small: pharmaceutical facility could boost capacity and slash costs by trading in certain batch operations for continuous versions. Pharmamag.com, article ID/DDAS-SEX 52B.
- Abboud, L. and Hensley, S. (2003). Factory shift: new prescription for drug makers: update the plants; After years of neglect, industry focuses on manufacturing; FDA acts as a catalyst; The three-story blender. *Wall Street Journal (Eastern edition)* A.1.
- US Food and Drug Administration, Center for Drug Evaluation and Research (2003). Guidance for industry PAT: a framework for innovative pharmaceutical manufacturing and quality assurance. <https://www.fda.gov/downloads/drugs/guidances/ucm070305.pdf> (accessed September 2004).
- US Food and Drug Administration, Center for Drug Evaluation and Research (2004). IPharmaceutical cGMPS for the 21st Century: a risk-based approach. Second progress report and implementation plan. <https://www.fda.gov/downloads/drugs/developmentapprovalprocess/manufacturing/questionsandanswersoncurrentgoodmanufacturingpracticescgmppfor drugs/ucm176374.pdf> (accessed September 2004).
- Lee, S.L., O'Connor, T.F., Yang, X. et al. (2015). Modernizing pharmaceutical manufacturing: from batch to continuous production. *Journal of Pharmaceutical Innovation* 10 (3).
- Leuenberger, H. (1994). Moist agglomeration of pharmaceutical powders (size enlargement of particulate material): the production of granules by moist agglomeration of powders in mixers/kneaders. In: *Powder Technology and Pharmaceutical Processes, Handbook of Powder Technology*, vol. 9 (ed. D. Chulia, M. Deleuil and Y. Pourcelot), 377–389. Amsterdam: Elsevier.
- Laurent, B.F.C. and Bridgwater, J. (2002). Convection and segregation in a horizontal mixer. *Powder Technology* 123: 9–18.
- Laurent, B.F.C. and Bridgwater, J. (2002). Performance of single and size-bladed powder mixers. *Chemical Engineering Science* 57: 1695–1709.
- Laurent, B.F.C. and Bridgwater, J. (2002). Influence of agitator design on powder flow. *Chemical Engineering Science* 57: 3781–3793.
- Marikh, K., Berthiaux, H., Mizonov, V., and Barantseva, E. (2005). Experimental study of the stirring conditions taking place in a pilot plant continuous mixer of particulate solids. *Powder Technology* 157: 138–143.
- Williams, J. and Rahman, M. (1971). Prediction of the performance of continuous mixers for particulate solids using residence time distributions. Part I: Theoretical. *Powder Technology* 5: 87–92.
- Beaudry, J.P. (1948). Blender efficiency. *Chemical Engineer* 55: 112–113.
- Williams, J. and Rahman, M. (1971). Prediction of the performance of continuous mixers for particulate solids using residence time distributions. Part II: Experimental. *Powder Technology* 5: 307–316.
- Harwood, C., Walanski, K., Luebecke, E., and Swanstrom, C. (1975). The performance of continuous mixers for dry powders. *Powder Technology* 11: 289–296.
- Portillo, P.M., Ierapetritou, M.G., and Muzzio, F.J. (2008). Characterization of continuous convective powder mixing processes. *Powder Technology* 182: 368–378.
- Warman, M. (2004). Using near infrared spectroscopy to unlock the pharmaceutical blending process. *American Pharmaceutical Review* 7 (2): 54–57.
- Ruf, A., Worlitschek, J., and Mazzotti, M. (2000). Modeling and experimental analysis of PSD measurements through FBRM. *Particle and Particle Systems Characterization* 17: 167–179.
- Koblitz, T. and Ehrhardt, L. (1985). Continuous variable-frequency fluid bed drying of pharmaceutical granulations. *Pharmaceutical Technology*.
- Leunberger, H. (2001). New trends in the production of pharmaceutical granules: batch vs. continuous processing. *European Journal of Pharmaceutics and Biopharmaceutics* 52: 289–296.
- Mollan, M.J. Jr. and Lodaya, M. Continuous processing in pharmaceutical manufacturing. *American Pharmaceutical Review* 7: 70–75.
- Paul, S., Knoch, A., and Lee, G. (1997). Continuous granulation: review. *PZ Prisma* 4: 112–124.
- Lindberg, N.O. (1988). Some experiences of continuous wet granulation. *Acta Pharmaceutica Suecica* 25: 239–246.
- Appelgren, C., Eskilson, C., and Medical, L. (1990). A novel method for the granulation and coating of pharmacologically active substances. *Drug Development and Industrial Pharmacy* 16 (15): 2345–2351.
- Jones, P.L. and Rawley, A.T. (1996). Dielectric drying. *Drying Technology* 14 (5): 1063–1098.
- Lindberg, N.O., Turfvesson, C., and Olbjer, L. (1987). Extrusion of an effervescent granulation with a twin screw extruder. *Drug Development and Industrial Pharmacy* 13: 1891–1913.
- Gamlen, M. and Eardly, C. (1986). Continuous granulation using a baker Perkins MP50 (multipurpose) extruder. *Drug Development and Industrial Pharmacy* 12: 1710–1713.
- Ghebre-Sellassie, I., Mollan, M.J., Jr., Pathak, N., et al. (2002). Continuous production of pharmaceutical granulation. United States Patent US 6,499,984 B1, 31 December 2002.
- Lodaya, M., Mollan, M., and Ghebre-Sellassie, I. (2003). Twin screw wet granulation. In: *Pharmaceutical Extrusion Technology, Drugs and the Pharmaceutical Sciences Series*, 1e, vol. 133 (ed. I. Ghebre-Sellassie and C. Martin). Informa Healthcare.
- Rey, L. (2003). Some leading edge prospects in lyophilization. *American Pharmaceutical Review* 6 (2): 32–44.
- US Food and Drug Administration, Center for Drug Evaluation and Research. 21 Code of Federal Regulations, Parts 210 and 211 current good manufacturing practice for manufacturing, processing, packing, or holding of drugs, Volume 4. https://www.ecfr.gov/cgi-bin/text-idx?SID=a723e65af3f2ff6251cb488305768760&mc=true&tpl=/ecfrbrowse/Title21/21cfrv4_02.tpl#0 (accessed 1 April 2018).
- ASTM E2537-16 (2016). Standard Guide for Application of Continuous Process Verification to Pharmaceutical and Biopharmaceutical Manufacturing. <https://www.astm.org/Standards/E2656.htm> accessed on 30 Jan 2019.

CONTINUOUS DIRECT COMPRESSION USING PORTABLE CONTINUOUS MINIATURE MODULAR & MANUFACTURING (PCM&M)

DANIEL O. BLACKWOOD AND ALEXANDRE BONNASSIEUX
Drug Product Design, Pfizer, Groton, CT, USA

GIUSEPPE COGONI
Analytical R&D, Pfizer, Groton, CT, USA

22.1 INTRODUCTION

Over the past decade, continuous processing of solid oral dosage forms has emerged as a leading technology used in the development and manufacture of solid oral dosage forms. Regulatory approvals in major world markets have been achieved for drug products developed with both new chemical entities and established products (previously manufactured using batch processes) [1, 2]. These drug products have been successfully manufactured using continuous direct compression, continuous dry granulation, and continuous wet granulation processes. For each of these continuous manufacturing processes, gravimetric feeding and continuous in-line blending of the formulation components serve as critical steps in achieving a uniform blend at target formulation concentrations, prior to subsequent downstream processing steps.

Early research in this area centered on evaluating the performance of gravimetric feeders in combination with horizontal in-line mixers from various equipment suppliers, such as Glatt, Gericke, and others. Recently, a vertically designed in-line powder mixer has been designed and developed as part of the Portable Continuous Miniature Modular (PCMM) program. Following a brief discussion and overview of the key features of the PCMM system, the features of the horizontal and vertical in-line mixers will be described and compared. Additionally, empirical studies on the impact of material properties and process parameters (such as impeller rotational speed [rpm], impeller geometric design, mass throughput, and

mass of powder resident within the mixer) on the residence time distribution (RTD) of powders within each of these systems will be described.

22.2 PORTABLE CONTINUOUS MINIATURE MODULAR (PCMM) TECHNOLOGY

PCMM technology has recently been created for the development and manufacture of solid oral dosage forms using continuous direct compression or continuous wet granulation technologies [3]. Figure 22.1 summarizes key features of the continuous direct compression system, which consist of:

- Six gravimetric feeders using loss-in-weight technology,
- A vertical in-line powder mixer (continuous mixing technology [CMT]).
- A rotary tablet press (GEA Courtoy MODUL™ P).
- An in-line near-infrared (NIR) system that is able to continuously monitor the blend inside of the feed frame immediately before tablet compression.
- Tablet metal detector and deduster.

22.2.1 Gravimetric Feeders

Figure 22.2 shows the key components of a gravimetric loss-in-weight feeder and volumetric refill system. Initially,

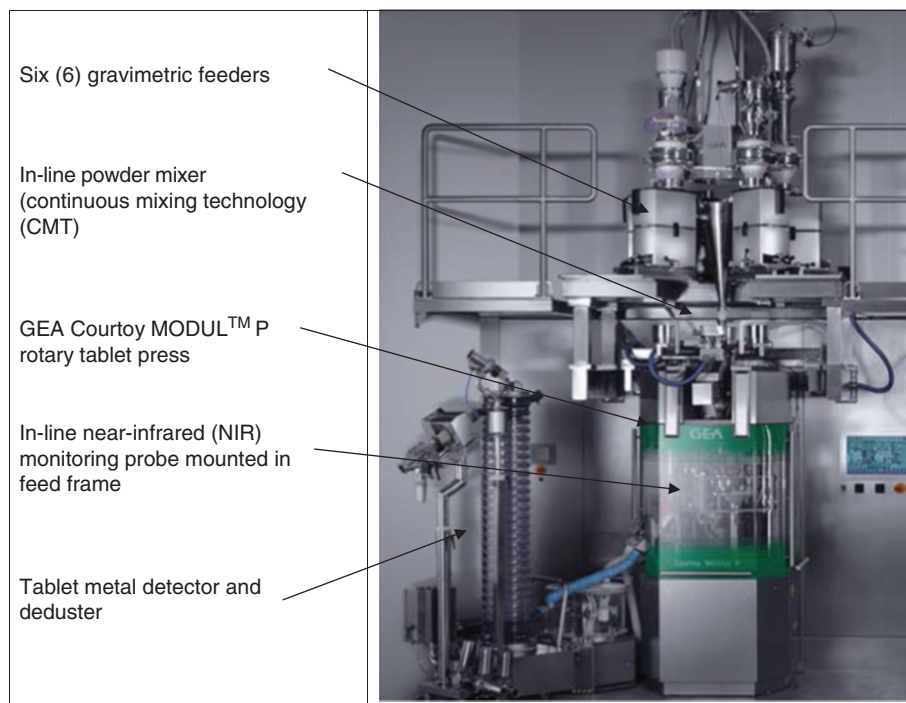


FIGURE 22.1 Overview of PCMM continuous manufacturing system.

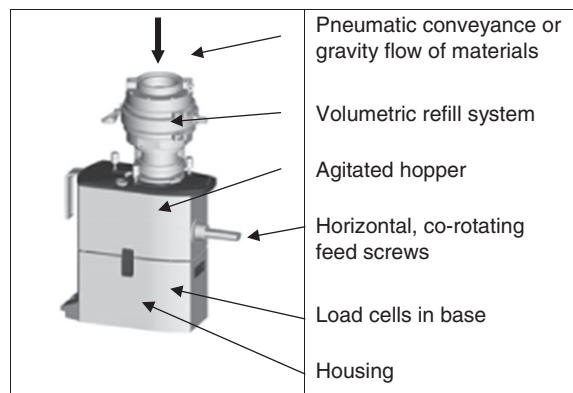


FIGURE 22.2 Illustration of a gravimetric loss-in-weight feeder and volumetric refill system. *Source:* Courtesy of GEA Inc.

formulation components are introduced into the volumetric refill system either by gravity or pneumatic conveyance. The volumetric refill system is inverted to deliver a fixed volume of powder into the agitated hopper of the gravimetric feeder. A drive system and two horizontally mounted co-rotating screws reside beneath this hopper. The weight of these components (agitated hopper, horizontal screws, and drive system) and the formulation component contained within the hopper rests on the feeder's load cell.

The gravimetric feeder set point (SP) (in kg/h), for each formulation component, is established by the product of the target overall line mass throughput and the target

formulation composition of a given component. As the horizontal screws rotate, material is removed from the agitated hopper and conveyed along the length of the screws. At the end of the screws, material is metered, on a continuous basis, into the CMT inlet hopper. This loss in weight from the system is measured on a continuous basis by the feeder load cells. Using a control algorithm, the screw rotational speed is adjusted (increased/decreased) to achieve a process value (PV) or target mass feed rate (in kg/h), for a given formulation component's SP, by evaluating the load cell's change in mass versus change in time (Figure 22.3).

As the level of material in the hopper decreases, a volumetric refill automatically occurs to refill the hopper with additional material. During this period, the feeder operates at fixed screw rotational speed in volumetric mode until the transient disturbance on the load cells diminishes and monitoring of loss-in-weight capabilities is restored (Figure 22.4)

22.2.2 Continuous Mixing Technology (CMT)

A hopper, located above the CMT, directs the gravimetric feed streams of the formulation components into the inlet of the CMT. The key components of the vertically designed CMT in-line powder mixers are illustrated in Figures 22.5 and 22.6.

The design features of the CMT mixer ensure robust control of the continuous mixing process:

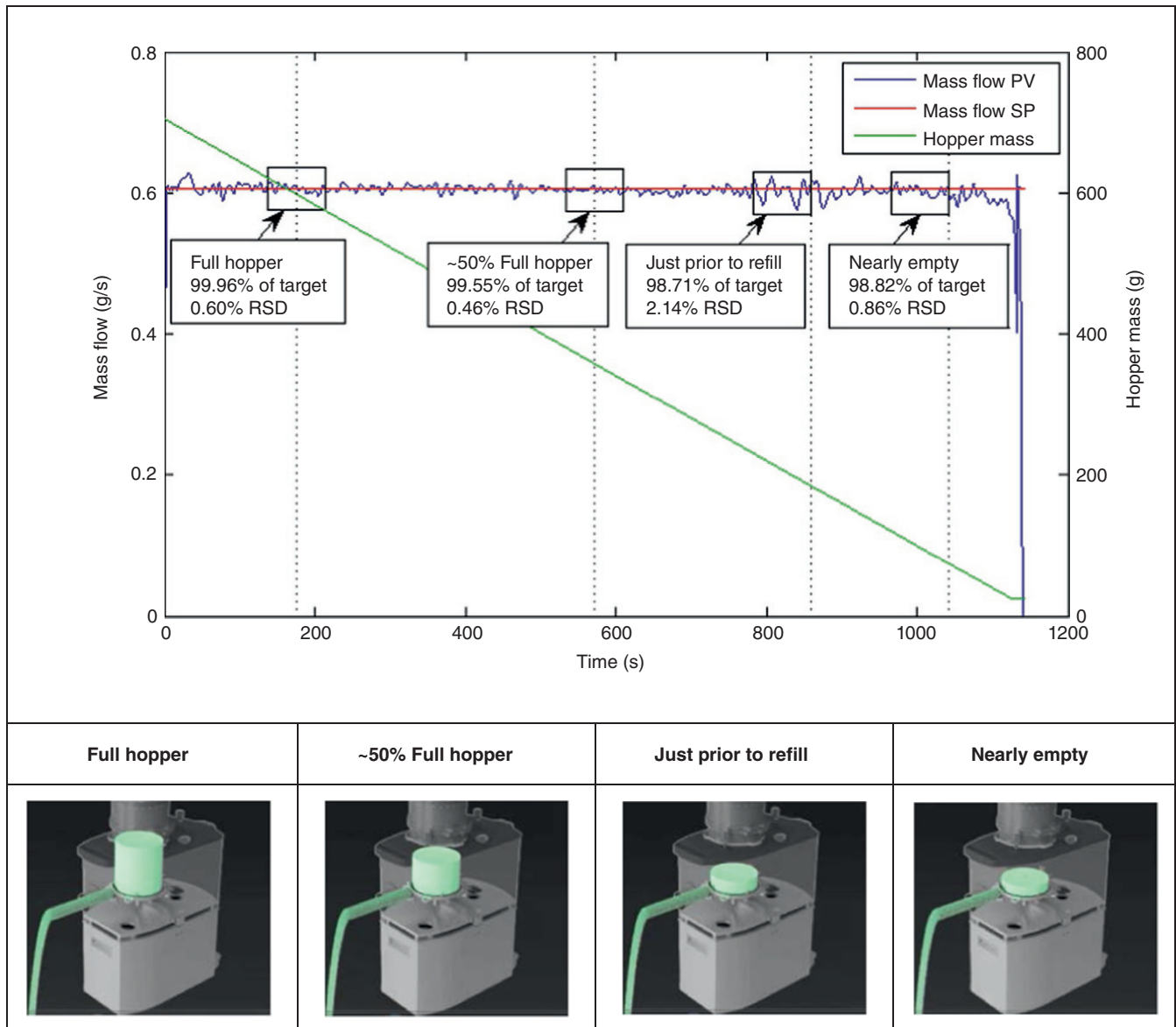


FIGURE 22.3 Example gravimetric feeder performance as a function of hopper fill level at a target feed rate of 0.606 g/s. PV, process value; SP, set point.

- The inlet hopper to the CMT mixer is designed to function as a vertical chute with steep walls (78° from horizontal) to allow for efficient and effective mass flow of the powder streams from the gravimetric feeders into the inlet of the CMT mixer.
- A delumping screen is present in the upper powder delumping region of the CMT mixer and is used to delump powder agglomerates in the feed stream. The delumping screen is provided as a change part with various screen hole size openings. The smallest screen contains 0.6 mm diameter holes, and the largest screen contains 2.1 mm diameter holes.
- Change parts for the upper and lower impeller with various cross-sectional areas (e.g. rectangular or triangular).
- Independent recipe SP control of the rotational speed and direction of the upper and lower impellers.
- Recipe SP control to define the amount of powder mass (in grams) is referred to as the *holdup mass (HUM)*, retained within the CMT, subject to shear and mixing energy from the impellers.
- A control system continuously adjusts the motor-driven exit valve, located at the bottom of the lower mixing chamber, which opens to create a flexible sized

“rectangular slit” through which the blend flows out of the CMT. If the measured HUM is above/below SP, the distance open (in mm) of the exit valve is gradually increased/decreased to ensure that the HUM is maintained at the target HUM SP (Figure 22.7). Additionally, the exit valve opening distance (in mm) is used

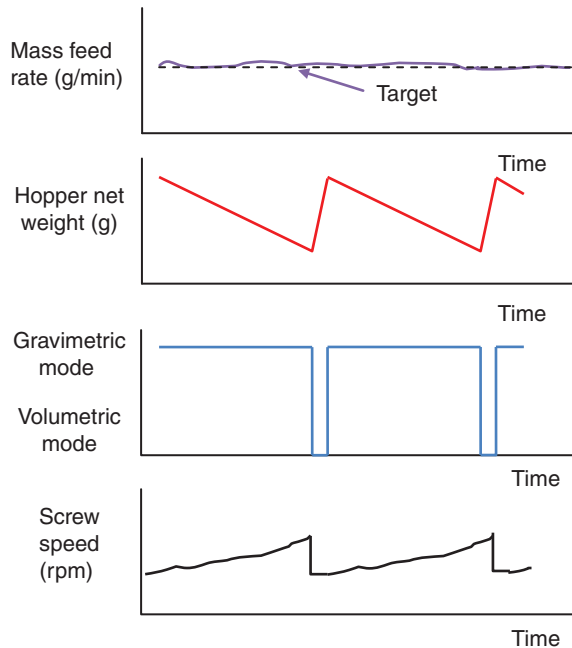


FIGURE 22.4 Schematic of a gravimetric feeder mass feed rate (g/min) and screw speed (RPM) as a function of hopper level.

to ensure that mass balance is preserved for the overall line mass throughput SP (e.g. mass flow rate in to the CMT equals the mass flow rate out of the CMT and therefore no accumulation).

- During a process pause, the exit valve shuts fully closed to keep the HUM retained within the CMT; the gravimetric feeders and CMT impellers are turned off. Upon restart, the gravimetric feeders and CMT impellers are turned back on to operate at SP, and the exit valve begins to open to control to target HUM.

22.2.3 Rotary Tablet Press and Integrated Feed Frame NIR Probe

The CMT is integrated above a conventional rotary tablet press (GEA Courtoy MODUL P, $n = 26$ stations). Homogeneous blend at target concentration flows from the discharge of the CMT through the feed chute and into the feed frame. Powder level sensors, including low and high level sensors, are used to slightly increase/decrease the turret rotational speed (RPM) in order to keep powder level within the feed chute at a consistent level. In the feed frame below, an in-line NIR probe (shown in Figures 22.8 and 22.9) continuously monitors the blend composition inside of the feed frame immediately before tablet compression. This NIR in-process control system allows for monitoring and potential diversion of nonconforming materials. Following compression, tablet cores are dedusted and passed through a metal detector. Additionally, tablet core physical properties, including weight, thickness, and hardness, are monitored during the tablet press operation.

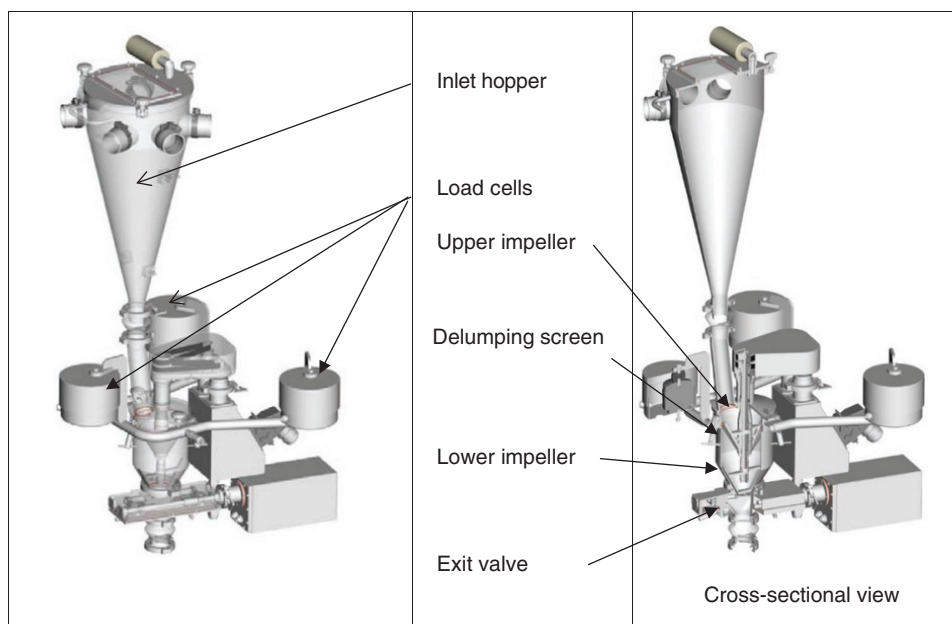


FIGURE 22.5 Illustration of the continuous mixing technology (CMT). *Source:* Courtesy of GEA Inc.

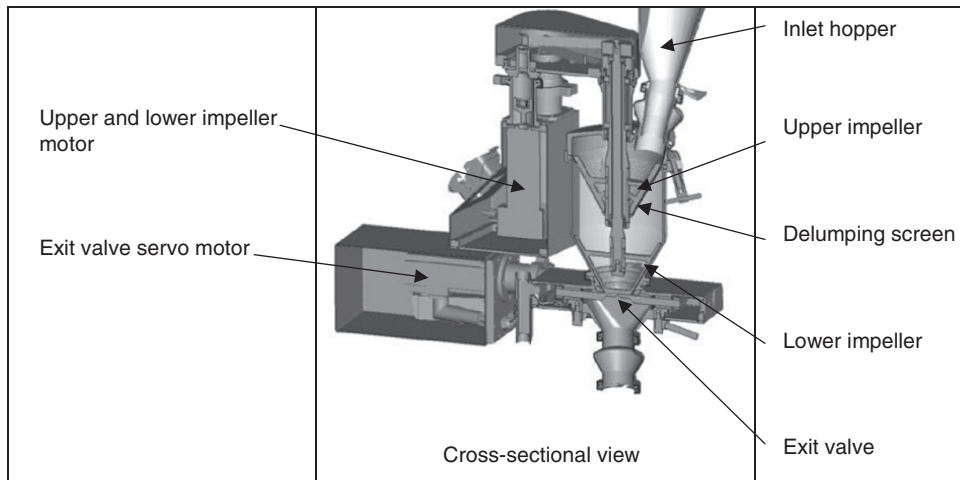


FIGURE 22.6 Cross-sectional view of the continuous mixing technology (CMT). *Source:* Courtesy of GEA Inc.

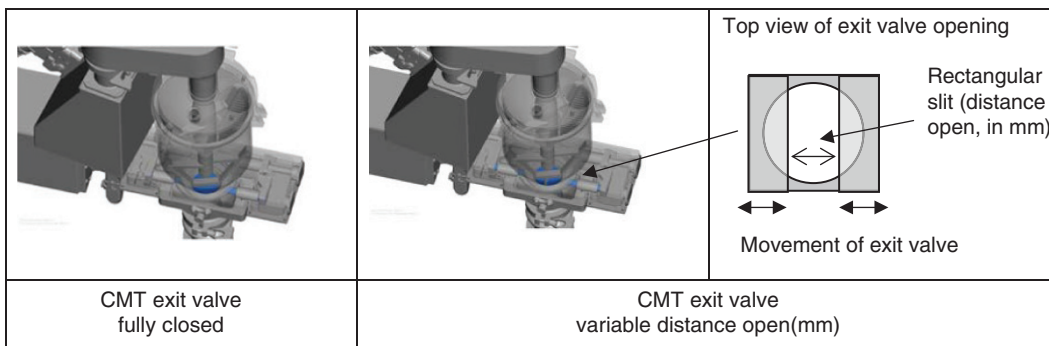


FIGURE 22.7 Overview of the key features of the CMT exit valve. *Source:* Courtesy of GEA Inc.

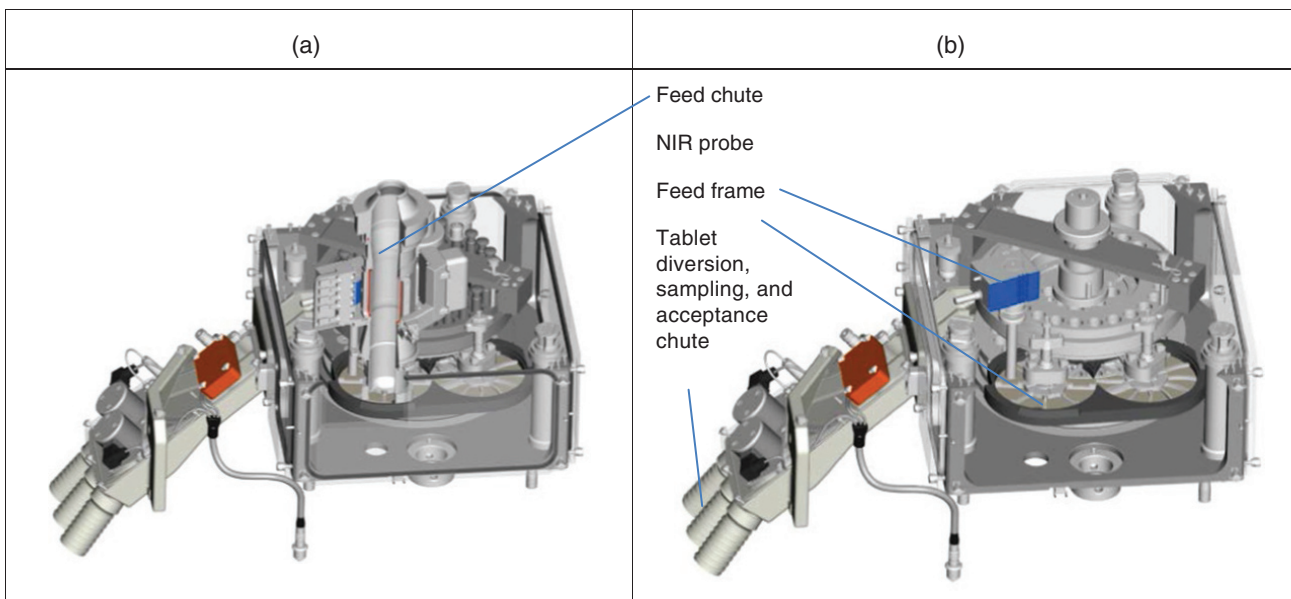


FIGURE 22.8 Design illustration of the GEA Courtoy MODUL P tablet press. (a) Shown with cross-sectional view of feed chute and (b) shown with NIR probe inserted into feed frame. *Source:* Courtesy of GEA Inc.

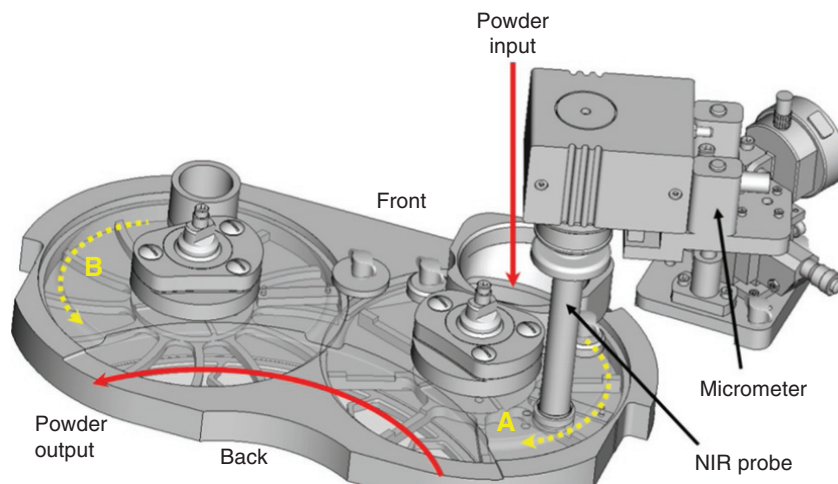


FIGURE 22.9 The GEA Courtoy MODUL P tablet press feed frame with integrated NIR system. *Source:* Courtesy of GEA Inc.

22.2.4 System Start-Up, Pauses, and Shutdown

The PCMM continuous manufacturing process has been designed for normal process start-up and shutdown as well as for system pauses that are either planned or unplanned.

22.2.4.1 System Start-Up At the start of a batch, the tared CMT is empty of powder and the exit valve is closed. The upper delumping impeller is turned on to delump powder through the upper screen into the lower mixing chamber. Initially, the lower impeller is turned off. Sequentially, one by one, each feeder dispenses a specified quantity of material into the CMT. At the end of this sequence, the formulation components are layered within the lower mixing chamber of the CMT at the target formulation composition. The lower impeller is then turned on, and the components are blended together for a recipe-driven amount of time and impeller rotational speed to create a homogeneous blend. Subsequently, the gravimetric feeders and the CMT upper and lower impellers are simultaneously turned on. The individual powder streams from the gravimetric feeder enter the inlet of the CMT and begin mixing with the blend of formulation components present within the CMT. Simultaneously, the exit valve begins to open and is controlled as the HUM reaches and maintains the target SP as well as the target mass throughput (mass in equal mass out). The homogeneous blend, discharging from the CMT, falls through the feed chute into the tablet press feed frame.

When a sufficient amount of blend is in the feed chute, the feed frame paddle wheels are turned on to prime the feed frame with blend and tablet compression commences to dial in the target tablet properties (e.g. weight and hardness). Simultaneously, NIR monitoring and trending of the blend composition circulating within the feed frame begins.

The process is paused to confirm that the target tablet core properties have been reached, prior to transitioning into

continuous manufacturing. This start-up procedure is designed to create tablet cores with acceptable quality attributes after dialing in the desired tablet physical properties.

22.2.4.2 System Pauses (Planned or Unplanned) During system start-up, or during routine operation, a planned or unplanned system pause may occur. The design of the system enables the process to be paused without impacting batch quality. For example, an issue with the tablet press will cause the turret and feed frame paddle wheels to stop. The feeders and CMT mixer will continue to operate until the powder level in the tablet press feed chute reaches the high level sensor. Simultaneously, each feeder will stop and the CMT exit valve closes and impellers turn off. Following resolution of the issue with the tablet press, the process is resumed with all feeders simultaneously turning back on, the CMT impellers turning on to the respective SP, and exit valve opening to achieve the target HUM. Alternatively, an issue with either a feeder or CMT may cause these systems to pause. The tablet press will continue to operate until the powder level in the feed chute reaches the low level sensor, at which point the turret and feed frame paddle wheel are stopped. Similarly, upon resolution of the issue, the feeders and CMT are simultaneously turned back on, and the tablet press operation is resumed.

22.2.4.3 System Pauses Requiring Diversion of Material Alternatively, the system may be paused due to a process-related alarm requiring diversion of nonconforming material (e.g. CMT mass balance alarm). In this case, the gravimetric feeders and CMT impellers are turned off, and the CMT exit valve is closed, and an investigation is opened. The material contained in the feed chute and tablet press feed frame is manually diverted to waste via compression of powder into the reject container. The system is manually restarted

(gravimetric feeders and CMT impellers turned on, and exit valve opened) to confirm acceptable gravimetric feeder and CMT mixer performance. The blend flows into the feed frame and feed chute, and tablet cores produced are manually diverted to waste, until acceptable quality attributes are confirmed. Subsequently, the tablet core physical properties are confirmed, and collection of tablet cores is resumed.

22.2.4.4 System Shutdown System shutdown is triggered when the drug substance low limit in the feeder hopper is reached, meaning the drug substance feeder is out of material. During shutdown, all feeders turn off simultaneously, and the CMT exit valve closes. The CMT impellers will continue to operate for a recipe-driven period of time and impeller rotational speed to create a homogeneous blend of the formulation components remaining within the CMT. After the set period of time, the exit valve begins to open, and the blend exits the CMT at a target mass throughput. The end of the batch is determined by either loss of tablet weight control or based on the NIR monitoring of the composition of the blend in the feed frame.

22.3 CHARACTERIZATION OF IN-LINE POWDER MIXERS PERFORMANCE

The characterization of in-line powder mixers has been described previously in literature with three key statistics highlighted below.

22.3.1 Mean Residence Time

The mean residence time [4] is a ratio of the HUM of powder contained within the mixer and overall line mass throughput passing through the mixer (Eq. 22.1). The mean residence time describes the amount of time an average particle will spend within the mixer, prior to exiting the system. In general, larger mean residence time offers increased dampening capabilities of fluctuations from the gravimetric feeders.

Mean residence time (in minutes)

$$= \frac{\text{Holdup mass (in kg)}}{\text{Overall line mass throughput (in kg/h)}} \times \left[\frac{60 \text{ min}}{\text{h}} \right] \quad (22.1)$$

22.3.2 Total Blade Passes

Total blade passes [5] is the product of the mean residence time and the impeller rotational speed (Eq. 22.2) and is an expression of impeller revolutions. Total blade passes describes the amount of shear applied to the powder over the average duration of time that powder spends within the mixer.

$$\text{Total blade passes (revs)} = \text{mean residence time (min)} \times \text{impeller speed (rpm)} \quad (22.2)$$

22.3.3 Residence Time Distribution

The RTD has been broadly used to characterize continuous powder mixing in pharmaceutical processes [6, 7]. In these studies, gravimetric feeders typically deliver the formulation components into the inlet of the in-line mixer at the target overall mass throughput. The in-line mixer impeller(s) is turned on, and the “weir” or exit valve is adjusted/controlled to achieve a target HUM in the mixer. The blend discharging from the in-line mixer is presented to a PAT sensor (typically NIR) via a sample interface device (Figure 22.10).

At an arbitrary time point, a tracer element is inserted into the inlet of the in-line mixer. The tracer element is blended into the formulation components residing within the mixer and soon thereafter exits from the mixer. The concentration, $C(t)$, of the tracer in the output stream is trended as a function of time until the concentration signal returns back to baseline and is used to determine the RTD $E(t)$ function (Eq. 22.3):

$$E(t) = \frac{C(t)}{\int_0^{\infty} C(t) dt} \quad (22.3)$$

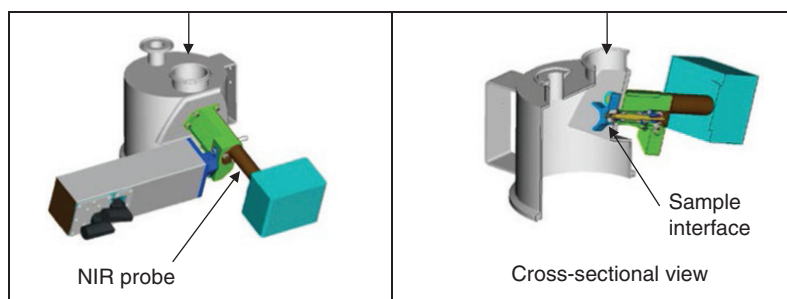


FIGURE 22.10 Example NIR probe and sample interface for in-line mixer RTD characterization. *Source:* Courtesy of GEA Inc.

Figure 22.11 introduces example RTD curves, assuming an ideal continuous stirred tank reactor (CSTR) with mean residence time (τ) of 60, 120, and 240 seconds using Eqs. (22.4) and (22.5) [8]. As shown in this figure, the variability in the input mass feed rate signal is significantly dampened via the mixer's RTD, with increasing smoothing of the output signal observed with increased residence time:

$$E(t) = \frac{1}{\tau} \exp\left(\frac{-t}{\tau}\right) \quad (22.4)$$

$$\tau = \frac{m}{\dot{m}} \quad (22.5)$$

where

m is the system HUM (in units of mass).

\dot{m} is the system mass throughput (in units of mass per time).

22.4 HORIZONTAL IN-LINE POWDER MIXER RTD CHARACTERIZATION

Figure 22.12 introduces a schematic representing key features of commonly available horizontal in-line powder mixing system from major pharmaceutical equipment suppliers (e.g. Gericke, Glatt, Ajax). These systems are designed with some basic features, including:

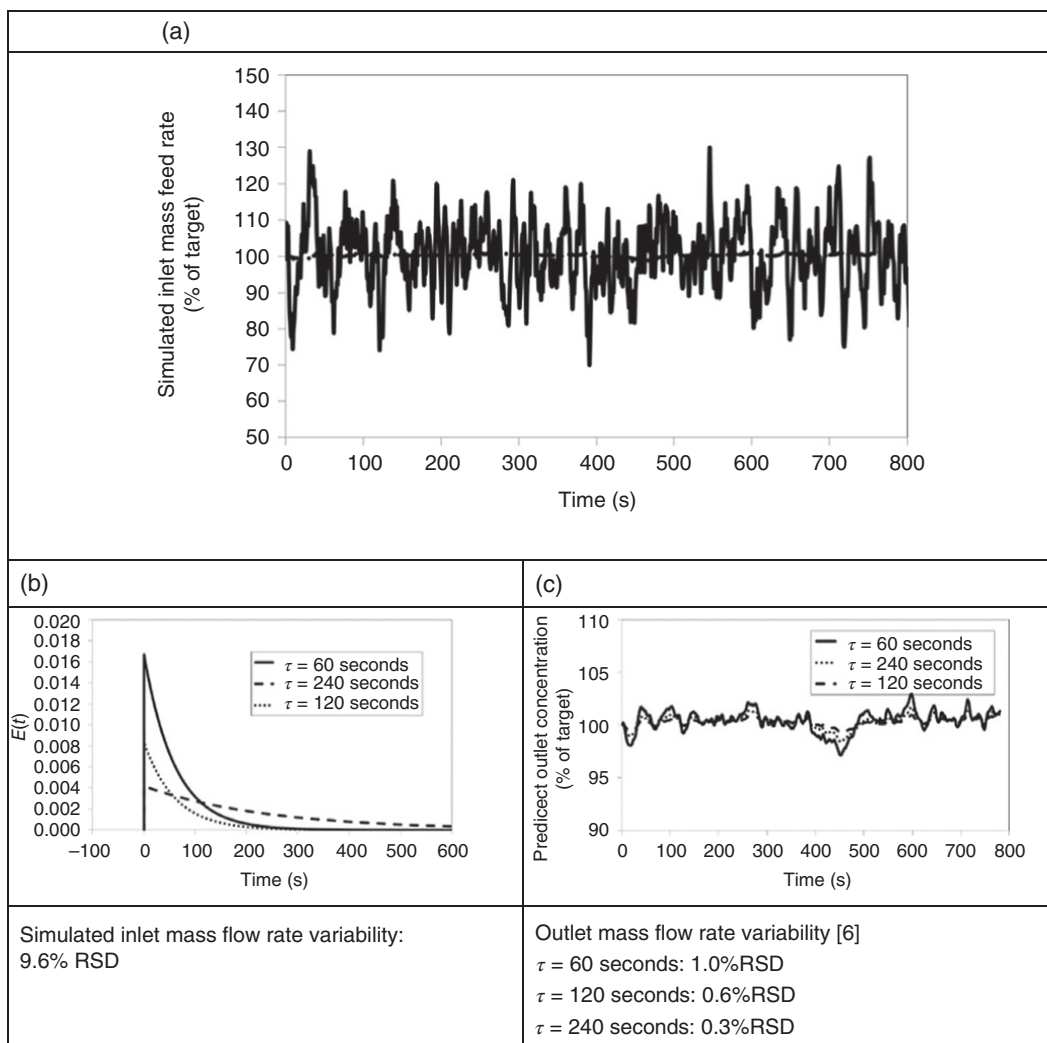


FIGURE 22.11 Modeling the dampening capability of the characteristic residence time distribution of the CMT mixer using offline gravimetric feeder trials. (a) Simulated gravimetric feeder data, (b) residence time distribution, $E(t)$, and (c) prediction of outlet mass flow rate using convolution of input inlet mass flow rate and residence time distribution.

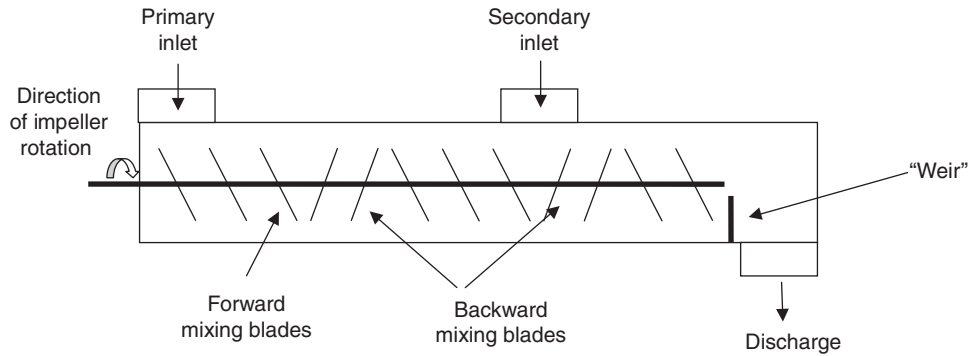


FIGURE 22.12 Schematic of a representative in-line horizontal powder mixing systems.

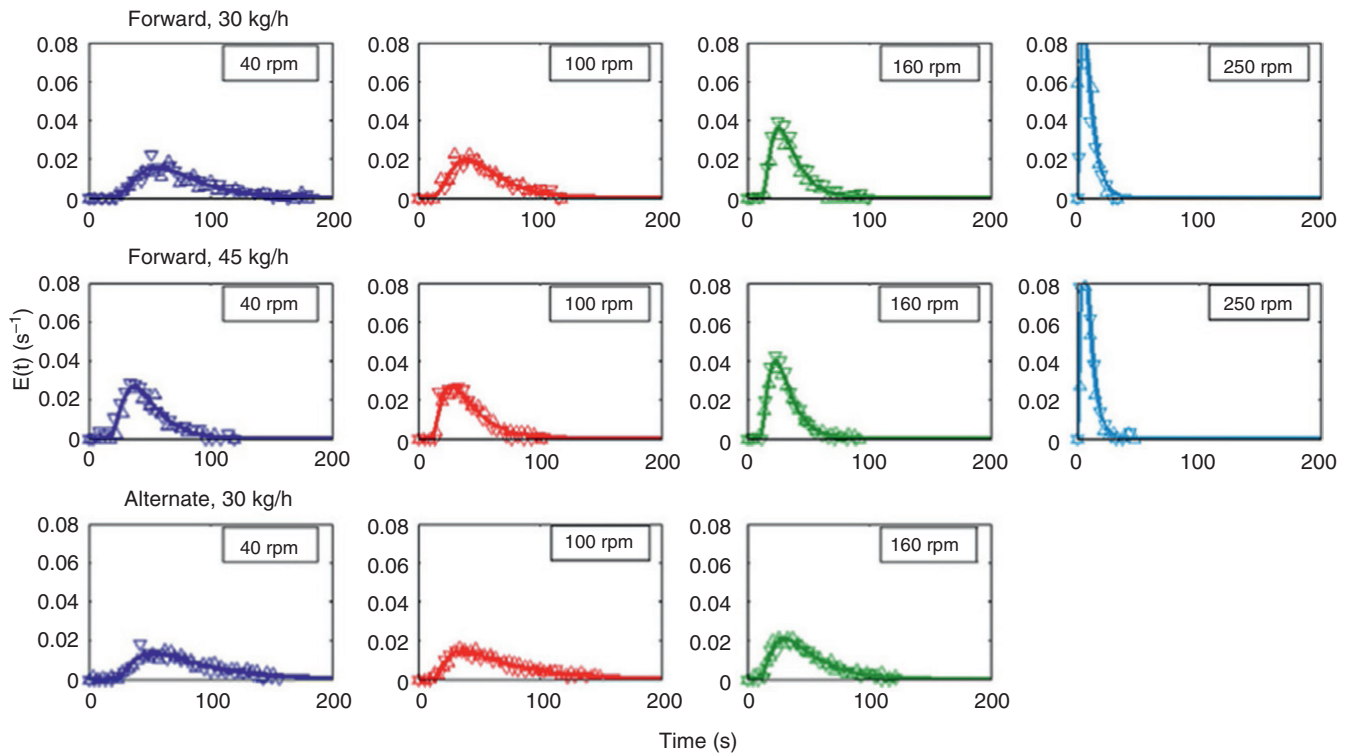


FIGURE 22.13 Example residence time distributions (RTDs) for a horizontal in-line powder mixer as a function of various process parameters (Gericke GCM 250). *Source:* Reprinted from Gao et al. [9], with permission from Elsevier.

- One or more powder inlet locations to combine streams of two or more components (or blends of components) together.
- An agitated horizontal mixing chamber, with a high length-to-diameter ratio. This chamber provides a fixed volume for retaining powder (HUM) to come into contact with rotating impeller(s) for blending components together.
- A single (Glatt, Gericke) or dual shaft (Ajax) horizontal rotating impellers with adjustable or fixed “weirs” design configurations. For some horizontal in-line blenders, the angle of inclination can be increased to induce additional back mixing.
- An adjustable impeller blade configurations and designs that can be installed in different configurations (e.g. forward, backward, or alternating).
- A powder exit port to discharge blended components into downstream processing equipment.

The RTDs of various horizontal in-line powder mixers have been described in previous research. Based on these studies, example RTDs for the Gericke GCM 250 horizontal powder mixing system are reproduced in Figure 22.13. As shown in this figure, the RTD of horizontal mixers varies significantly as a function of process parameters, such as impeller blade configuration (e.g. forward, backward, or

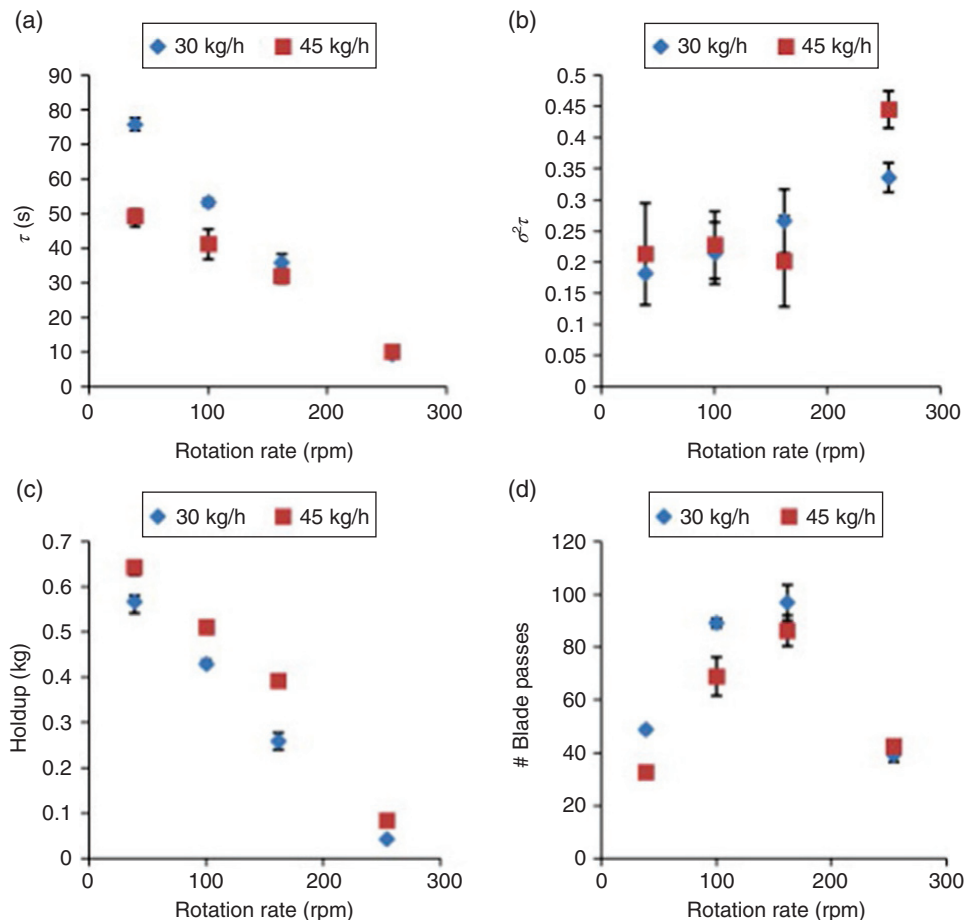


FIGURE 22.14 Impact of impeller rotational speed (RPM) and mass throughput (kg/h) on (a) mean residence time, (b) mean centered variance, (c) holdup mass, and (d) number of blade passes. *Source:* Reprinted from Vanarase and Muzzio [10], with permission from Elsevier.

alternating), overall mass throughput (kg/h), and impeller rotational speed (RPM). At high rotational speeds, powder is conveyed swiftly down the length of the mixer, thus reducing the time that powder is residing within the mixer. During this shorter period of time, powder is exposed to greater shear rates from the rotating mixing impeller. At slower rotational speeds, increased residence times are observed at the expense of lower shear rates being applied to the powder. Similarly, Figure 22.14 graphically reproduces this concept describing the impact of impeller rotational speed on the mean residence time, mean centered variance, HUM, and number of blade passes for the Gericke GCM 250 mixer. As shown in the work, a maximum of total blade passes is achieved at a moderate impeller rotational speed.

22.5 VERTICAL IN-LINE POWDER MIXER RTD CHARACTERIZATION

The design features of the CMT allow for selection via recipe SP and control of the mixer processing parameters: HUM,

impeller rotational speed, and overall line mass throughput. Together, these process parameters impact the mixer's mean residence time and the RTD. The optimization of these parameters (Figure 22.15) allows for the selection of a target HUM, which influences the system mean residence time, to dampen minor fluctuations in the input feed rates of each formulation component. Independently, the selection of a target impeller rotational speed ensures that mixing energy/shear is applied to the powder over a sufficient period of time to reach the desired blend uniformity and degree of powder lubrication.

22.5.1 CMT Residence Time Distribution

A model based on a plug-flow reactor (PFR) in series with two CSTR has been developed to describe the RTD of the CMT (PFR \rightarrow CSTR₁ \rightarrow CSTR₂). As shown in Figure 22.16, the model captures the residence time of powder flowing through the three regions of the CMT: inlet hopper (PFR), upper delumping region (CSTR₁), and lower mixing region (CSTR₂).

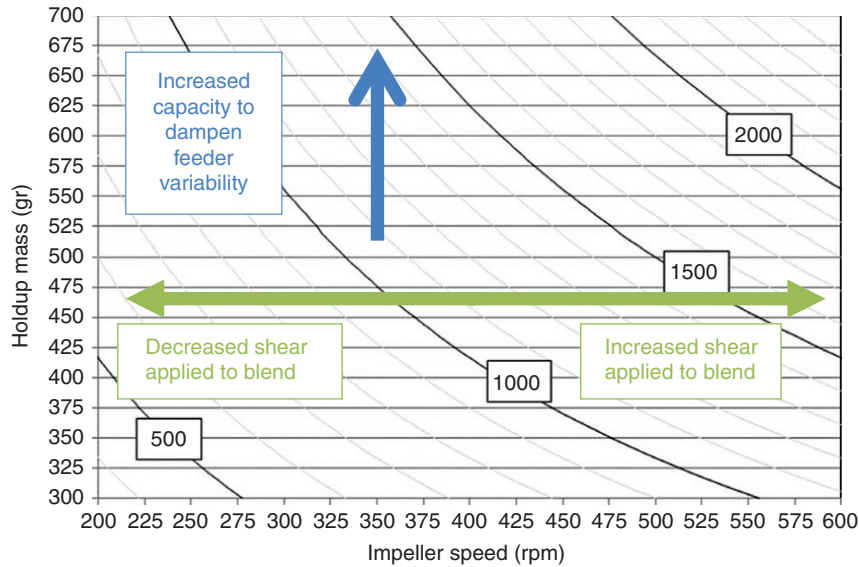


FIGURE 22.15 Total blade passes as a function of CMT mixer impeller speed and holdup mass at an overall mass throughput of 10 kg/h.

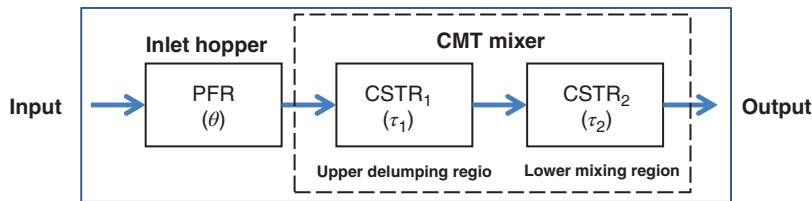


FIGURE 22.16 Schematic of residence time distribution model for the CMT mass balance. Source: Courtesy of GEA Inc.

TABLE 22.1 Summary of Model Parameters for the CMT Residence Time Distribution

Model Parameter	Value	Description
θ^*	0.001 seconds	Short time delay to account for powder flow within the inlet hopper (gravimetric feeder to the CMT inlet)
τ^*	$\frac{m}{\dot{m}}$	From Eq. (22.5): mean residence time: ratio of holdup mass set point to the overall line mass throughput rate
$r = 0.013$	$\frac{\tau_1}{\tau_2}$	Tau ratio based on relative amount of mass held in the upper delumping versus the lower mixing region
τ_1	$\tau \frac{r}{r+1}$	Portion of the mean residence time belonging to the first part of the model that physically corresponds to the upper delumping region of the CMT
τ_2	$\frac{\tau}{r+1}$	Portion of the mean residence time belonging to the second part of the model that physically corresponds to the lower mixing region of the CMT

*Source: Courtesy of GEA Inc.

The model parameter for the PFR region has been minimized (fixed to $\theta = 0.001$ seconds) to ensure that the calculated concentration signal reacts quickly to composition changes in the system, while the ratio of τ_1 to τ_2 ($r = 0.013$) has been determined and optimized through a series of experimental studies, highlighted in the next section. Using least squares, the model parameter (r), for CSTR₁ and CSTR₂, was selected by comparing the RTD model to single-

wavelength (NIR) trending of the tracer concentration in the output stream of the CMT as a function of time (Table 22.1).

Figure 22.17 highlights RTD examples using an ibuprofen (D50: 38 μm) tracer and placebo blend being fed through the CMT mixer at 10 kg/h. Comparable RTDs, following the model for the CMT, are demonstrated across the various combinations of HUM and impeller rotational speed studied. Similarly, Figures 22.18 and 22.19

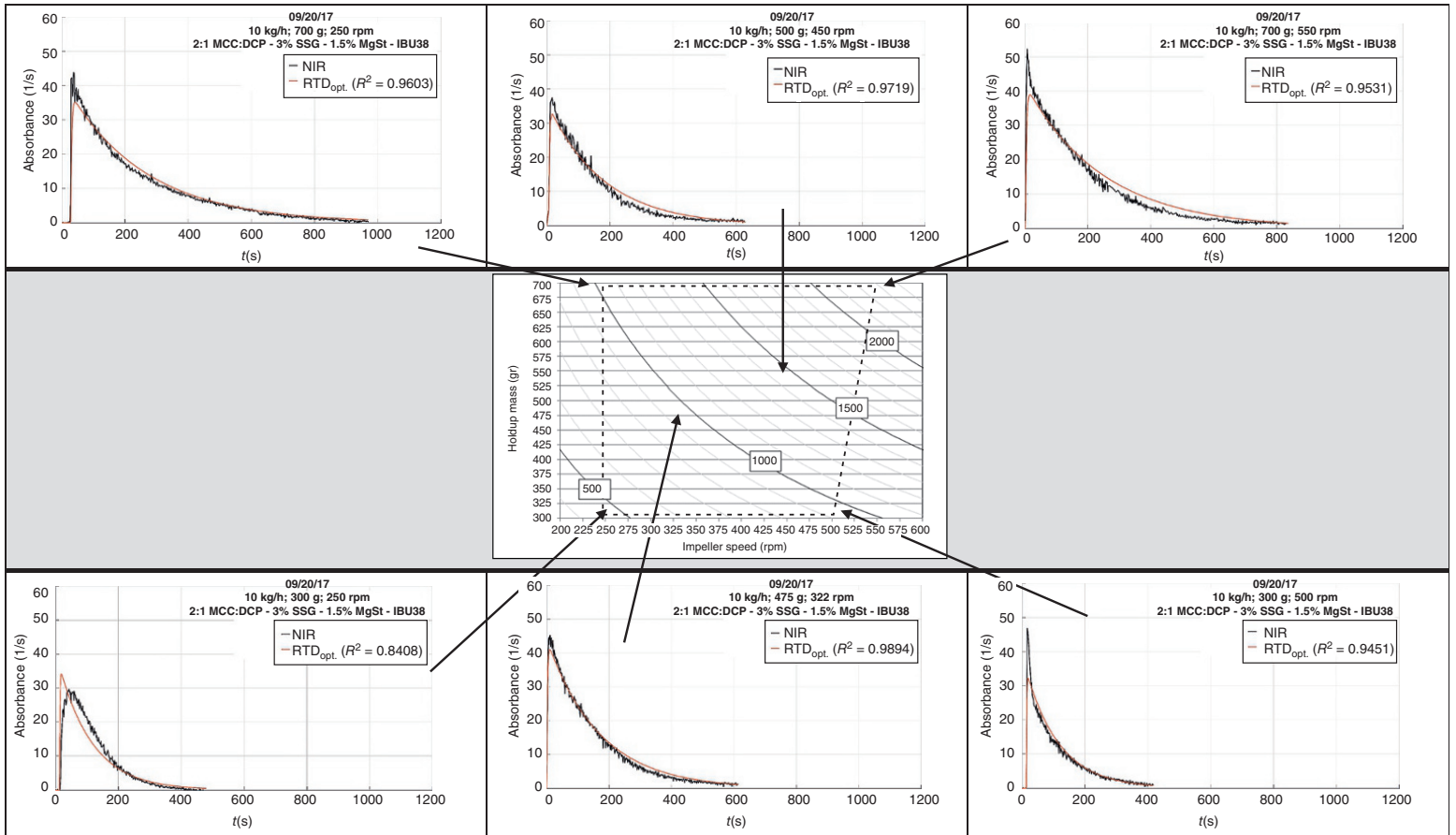


FIGURE 22.17 Residence time distribution – spike study with 38 μm D50 ibuprofen and 2 : 1 MCC/DCP, 3% SSG, 1.5% MgSt as a function of HUM (g) and impeller RPM at 10 kg/h.

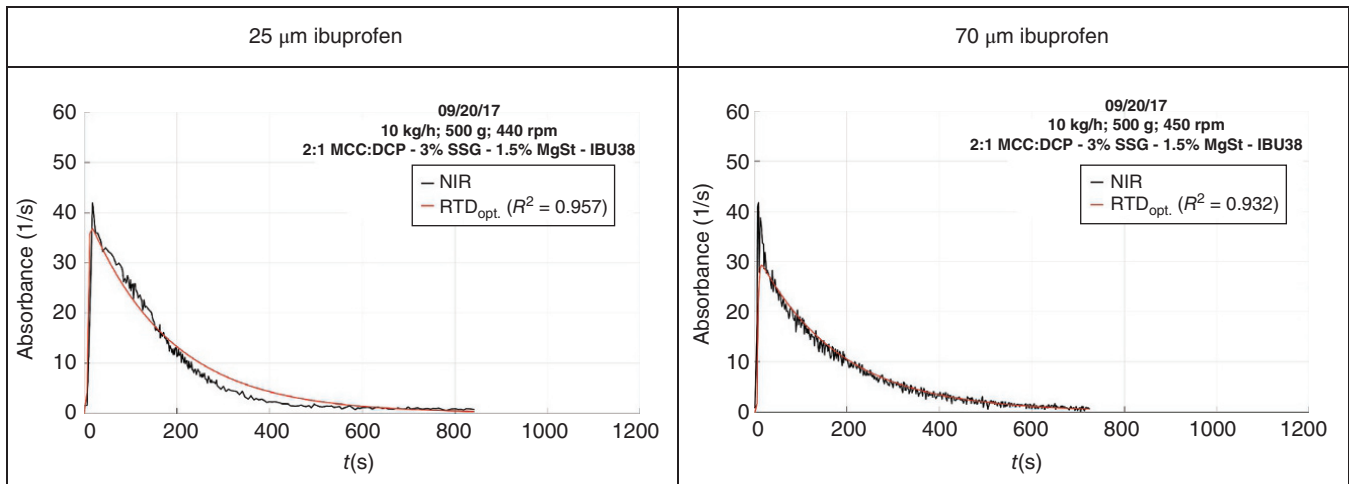


FIGURE 22.18 Residence time distribution – spike study with 25 and 70 μm D50 ibuprofen and 2 : 1 MCC/DCP, 3% SSG, 1.5% MgSt as a function of HUM (g) and impeller RPM at 10 kg/h.

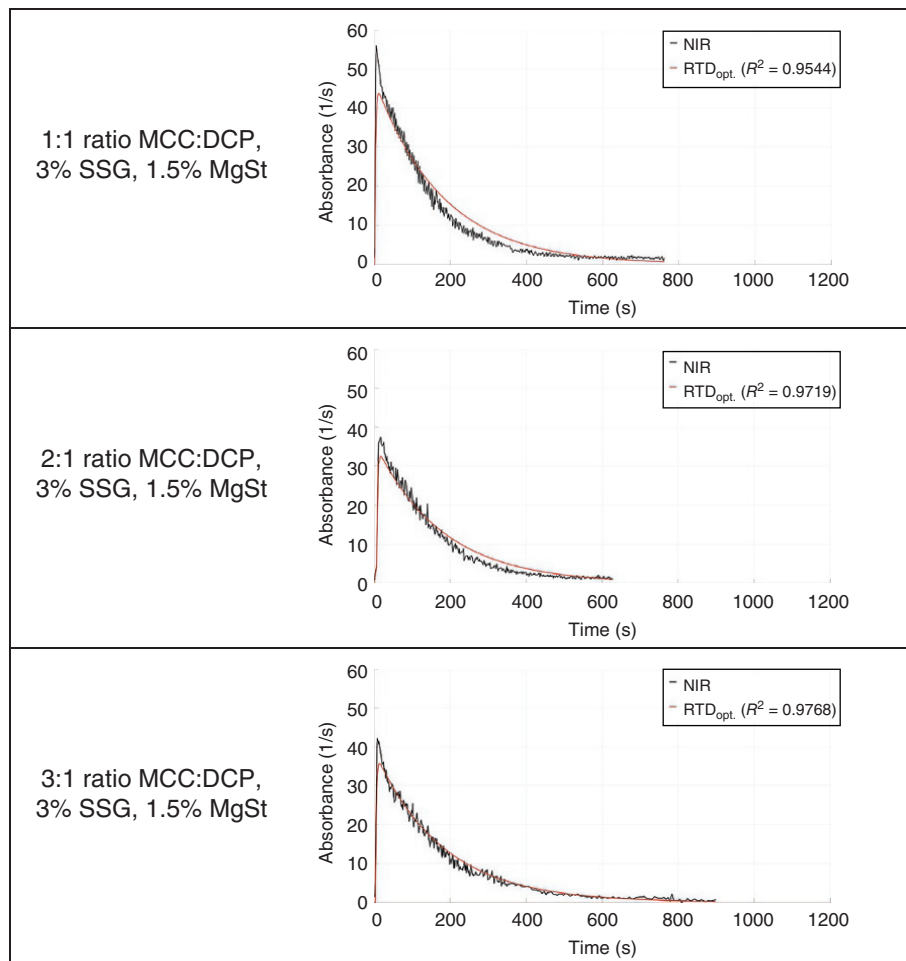


FIGURE 22.19 Residence time distribution – spike study with ibuprofen (D50 38 μm) and 1 : 1, 2 : 1 and 3 : 1 MCC/DCP, 3.0% SSG, 1.5% Mg St at 500 g HUM, 450 impeller RPM, and 10 kg/h.

demonstrate comparable RTDs for two different ibuprofen tracer particles sizes (D50: 25 and 70 μm) and in different diluent blend ratios, respectively.

22.6 SUMMARY

The recent introduction of vertical in-line powder mixers provides new capabilities for continuously blending powder streams as part of a continuous drug product manufacturing systems. These mixers offer new features to ensure robust control of the continuous mixing process, including integrated delumping screen to deagglomerate powders prior to mixing, recipe SP control of impeller rotational speed, recipe SP control of HUM with control system on the exit valve, and change parts for delumping and mixing impellers with various cross-sectional designs (e.g. rectangular or triangular shapes).

These in-line vertical mixers offer an additional means of blending powders streams together as part of a continuous drug product manufacturing process and complement the existing array of horizontal in-line mixers currently in use within the industry. The impact of mixer physical design and change parts, process parameters such as impeller speed, HUM, and overall line throughput, and material attributes on the RTDs from both these systems remains an active area of scholarship. In conjunction with gravimetric feeders, the in-line powder mixers serve as a critical unit operation in achieving a uniform blend prior to downstream granulation steps or conversion into a dosage form.

REFERENCES

1. Markarian, J. (2 April 2018). Modernizing Pharma Manufacturing. www.pharmtech.com (accessed 2 July 2018).
2. Pharmaceutical Technology Editors. (26 June 2017). EMA Approves Janssen Drug Made Via Continuous Manufacturing. www.pharmtech.com (accessed 2 July 2018).
3. (2016). *2016 Facility of the Year Award Category Winner for Equipment Innovation*. Pfizer Inc. <https://ispe.org> (accessed 26 July 2018)
4. *Elements of Chemical Reaction Engineering* by H. Scott Fogler., Chapter 16, 5th edition. Prentice Hall: Kendallville, IN, 767
5. Vanarase, A.U., Osorio, J.G., and Muzzio, F.J. (September 2013). Effects of powder flow properties and shear environment on the performance of continuous mixing of pharmaceutical powders. *Powder Technology* 246: 63–72. <https://doi.org/10.1016/j.powtec.2013.05.002>.
6. Engisch, W. and Muzzio, F. (2016). Using residence time distributions (RTDs) to address the traceability of raw materials in continuous pharmaceutical manufacturing. *Journal of Pharmaceutical Innovation* 11: 64–81. <https://doi.org/10.1007/s12247-015-9238-1>.
7. Gao, Y., Muzzio, F.J., and Ierapetritou, M.G. (2012). A review of the residence time distribution (RTD) applications in solid unit operations. *Powder Technology* 228: 416–423.
8. *Perry's Chemical Engineers' Handbook*, 6th Edition, Robert H. Perry, Don Green 1984. New York: McGraw-Hill
9. Gao, Y., Vanarase, A., Muzzio, F., and Ierapetritou, M. (2011). Characterizing continuous powder mixing using residence time distribution. *Chemical Engineering Science* 66: 422.
10. Vanarase, A.U. and Muzzio, F.J. (2011). Effect of operating conditions and design parameters in a continuous powder mixer. *Powder Technology* 208 (1): 30.

PROCESS CONTROL LEVELS FOR CONTINUOUS PHARMACEUTICAL TABLET MANUFACTURING

NIELS NICOLAÏ AND INGMAR NOPENS
BIOMATH, Ghent University, Ghent, Belgium

MAXIM VERSTRAETEN AND THOMAS DE BEER
Laboratory of Pharmaceutical Process Analytical Technology, Ghent University, Ghent, Belgium

23.1 CONTINUOUS MANUFACTURING: OPPORTUNITIES AND CHALLENGES

Over the last decades the pharmaceutical industry has relied on batch processing for the manufacturing of its oral solid dosage products. This conventional manufacturing approach consists of sequential processes with lengthy storage times and tedious offline quality testing of intermediate products, hence introducing a lot of inefficiencies and delays into manufacturing route. Yet, the pharmaceutical industry has been rather conservative toward the introduction of new technologies to speed up its manufacturing and release processes. Such reasoning was made possible because of large profit margins on blockbuster solid dosage products in the past, leaving pharmaceutical companies rather unmotivated to innovate their manufacturing technologies.

Recent times have changed for this industry. The ever-increasing competition between pharmaceutical companies of the twenty-first century demands for more cost-effective manufacturing. In fact, the current competition is largely driven by the socioeconomic pressures to control drug prices, tighter quality assurance of the final drug products, the presence of generic competitors, loss of exclusivity for blockbuster medicines, and the increasing development costs for new drug products. In addition, poor drug product quality and inflexible supply chains as a result of traditional batch manufacturing have caused major drug recalls and shortages in the past with direct impacts on the public healthcare system and individual patients [1–7].

As a result, pharmaceutical companies have recognized the need to modernize drug development and manufacturing with the goal of averting costs and risks associated with the inefficiencies of the conventional batch approach. Continuous manufacturing is identified as a valuable alternative [8]. It is characterized by a continuous inflow of input material and outflow of final products without the need of isolating intermediate processed material, hence drastically reducing the holdup times associated with the manufacturing steps. Within such a process, all unit operations are fully integrated into one single manufacturing equipment, therefore minimizing the risk for human intervention and contamination, while maximizing the utilization of the available workspace to reduce the manufacturing footprint [9]. Ultimately continuous manufacturing should allow for varying throughputs and process durations. This may result in a more demand-driven manufacturing strategy, with more flexible supply chains that have a faster response to the market demands and prevent drug shortages [10, 11].

As the intermediate products are not isolated anymore for quality control in these fully integrated continuous manufacturing processes, alternative strategies need to be explored to guarantee the quality specifications of the final drug product. Hence, a significant amount of resources is allocated to investigate process analytical technology (PAT) opportunities well suited to the continuous manufacturing field. PAT is defined here as a system for the design, analysis, and control of manufacturing processes based on timely measurements of critical quality attributes (CQAs) of raw and in-process

materials to ensure final product quality. The implementation of PAT methodologies was strongly encouraged by regulators such as the Food and Drug Administration (FDA), which launched its industry guidance “PAT – A Framework for Innovative Pharmaceutical Manufacture and Quality Assurance” in 2004 [12]. The framework outlined in this guidance intends to facilitate innovation (i.e. alleviating the regulatory deadlock) in pharmaceutical manufacturing by outlining scientific principles, tools, and strategies to support the implementation of PAT.

More specifically, PAT is described as an assembly of methods and instrumentation, which can be implemented during the different phases of a drug product cycle [13]. For example, in the early stages of process development, i.e. **the design phase**, the process should be characterized and optimized. Here, well-chosen experiments, typically based on statistical principles of orthogonality, are conducted to identify and evaluate material properties and process settings that may be critical to product quality and performance (Figure 23.1). This is also in line with the current paradigm shift from “testing for quality” toward “quality by design.” Clearly, building quality into drug products can only be achieved by revealing the interactions between critical material attributes (CMAs), critical process parameters (CPPs), and CQAs of the intermediate and final drug product [15] (Figure 23.1).

In a next phase, i.e. **the analysis phase**, suitable process sensors, real-time analyzers, and process models are selected in order to monitor the desired CMAs, CPPs, and quality attributes or CQAs during production. Such tools can either provide univariate process measurements (e.g. pH, temperature, pressure, and model predictions) or multivariate information related to biological, physical, and chemical attributes of the material being processed (e.g. NIR and Raman spectrometers).

Ultimately, the PAT approach gives revolutionary opportunities to enhance fundamental process understanding and to unravel several process phenomena and dynamics by providing in-process measurements. In the final stage, the

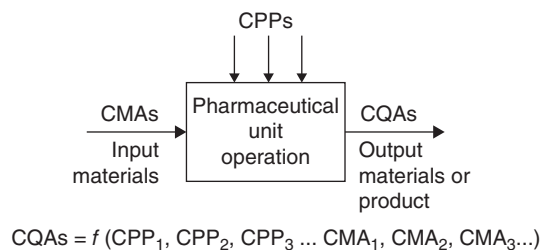


FIGURE 23.1 Design phase: establishing causal multifactorial relationships between input critical material attributes (CMAs), critical process parameters (CPPs), and output critical quality attributes (CQAs) for a unit operation. *Source:* Reprinted with permissions from Yu et al. [14]. Copyright (2014) Elsevier.

knowledge and measurement tools defined earlier may allow for the design of an effective quality control strategy, i.e. **the control phase**. Such a control strategy should yield lower final product variability while mitigating risks by rejecting out-of-spec (OOS) product.

At the moment, the PAT cycle has been employed for real-time monitoring, and thus understanding, of intermediate and final product CQAs in the pharmaceutical industry (often during post-approval stages). While such applications have gradually increased in the last decade, the majority focuses on improved understanding through more detailed analysis of the manufacturing processes. The opportunities offered by automatic process control have therefore not yet been exploited to their fullest potential at commercial scale. However, the current awareness of process control’s importance in the pharmaceutical industry, especially for continuous manufacturing, has resulted in an increasingly demanding knowledge of the control terminology [5, 16, 17]. Indeed, most industrial pharmacy departments are not yet fully up to speed with the potential advantages process control offers. The aim of this chapter is therefore to clarify the concept of process control by providing a structured overview combined with essential literature references that relate to the field of continuous pharmaceutical tablet manufacturing. Besides, collaborating with process control engineers is considered crucial to shape the continuous drug manufacturing industry.

23.2 DEFINING PROCESS CONTROL

Process control refers to a broad range of methods that are used to regulate different continuous variables, such as process operational settings, product characteristics, energy consumption, and economic performance of industrial and societal processes. In practically all cases, the common thread is the comparison of the actual process behavior with a desired process behavior. The controller will subsequently generate an action that intends to minimize or even eliminate the deviation between the desired and the actual behavior [18]. While industrial process control is often used in an automation context, it does not necessarily exclude manual operations by process operators and engineers. Hence, a process controller includes hardware, software, and human components. Automation and control for the process industry are thus fundamentally different concepts. Yet, faster and cheaper computation by means of networked digital computers, for increasingly complex and agile processes, combined with more demanding industry regulations has led to an irreversible entanglement between automation and process control. This way, manual operations induced by human process operators are avoided. Clearly, this will also be the case when pursuing continuous manufacturing and PAT systems in the pharmaceutical manufacturing industry.

23.3 NEED FOR PROCESS CONTROL

23.3.1 Ensuring Functional Process Operation

When operating a continuous pharmaceutical drug product manufacturing line, it is important to maintain a well-defined environment to which the product is exposed. Indeed, when CPPs such as temperature, pressure, speed, level and mass flow rate, etc. are not continuously monitored and controlled, the manufacturing line is doomed to result in failures and production losses. In process control terminology, this requirement is translated into maintaining the desired set points of the low level process components. To this end, different univariate sensors are typically in place to ensure the functioning of each operation in the process. For example, measurements coming from the sensing elements can be coupled to an automatic control loop that ensures a consistent set point for a particular part of the process.

23.3.2 Handling Process Disturbances

Just like most industrial processes, secondary drug product manufacturing is subject to external disturbances or failure modes in the CMAs and CPPs that could potentially cause undesired variability in the CQAs of the final drug product (see Section 23.4.1). In the case of oral solid dosage forms, the CQAs are typically those aspects affecting the potency, weight, uniformity, strength, and stability of the product as well as the release of the active pharmaceutical ingredient (API) in terms of disintegration and dissolution ([19], CFR 21 Part 211 Subpart F). Note that although the terms *disturbance* and *failure mode* can be interchanged in the case of continuous manufacturing, the first is typically used in a control engineering context, whereas the term *failure mode* originates from the field of risk management.

23.3.3 Economic Optimization of the Process

Nowadays, cost reduction is a major focus in the process industries. The same applies for the pharmaceutical drug development and manufacturing industries that are confronted with increasing product and process development costs, increased risk of failure, and inefficient quality by testing. At the same time revenues from existing drug products on the market stagnate or drastically reduce due to competition with generic pharmaceutical companies and the push to hold down healthcare costs. From an industry perspective it is therefore becoming more and more important to optimize the economics of pharmaceutical manufacturing processes. For continuous pharmaceutical processes in particular, market demand, energy consumption, raw material prices, utility prices, and product price have to be taken into account. Again, process control can offer solutions by steering a

manufacturing process toward an economic sweet spot, which guarantees maximal profitability.

23.4 DISTURBANCES AND RISKS

23.4.1 Sources of Product Variability

Process disturbances causing variability in the end product are often difficult to identify and may not be directly measurable. To facilitate identification, distinct categories of variability sources can be distinguished depending on their common origin in the manufacturing process. Within the scope of continuous oral solid dosage manufacturing, the following six categories are to be investigated for potential sources of CQA variability. Note the resemblance with the *Six Sigma 6M* method that is based on the initial work of Ishikawa [20]:

- **Raw materials:** The influence of both physical and chemical characteristics of the raw materials has been thoroughly studied for different unit operations and formulations. An overview for roller compaction is provided by Hlinak et al. [21], and examples for twin-screw wet granulation are given by El Hagrasy et al. [22]. Although strictly regulated and controlled within predefined tolerance limits, excipient and API properties of the same brand and grade might show batch-to-batch variability as a result of disturbances in their own individual manufacturing routes, especially for those derived from natural sources. An example study provided by Fonteyne et al. [23] assesses the impact of unspecified variability in microcrystalline cellulose (MCC) samples of the same grade on granule formation by means of twin-screw wet granulation. Other examples relevant to drug product manufacturing are provided in the work of García Muñoz et al. [24].
- **Process operation:** Deterministic disturbances that are often reflected as non-periodic or sudden step changes during processing are of major concern. A common example is related to the refill strategy of continuous low throughput loss-in-weight feeding systems. During normal operation accurate feeding of such systems is guaranteed by means of a gravimetric controller. However, upon refill a sudden disturbance in material weight, compression, density, and thus feed factor is introduced. As shown by Engisch and Muzzio [25], the refill strategy, which directly relates to how the process is operated, determines the level of disturbance in the material flow rate. Another example that demonstrates the importance of the process operational strategy on CQA variability is provided by Peeters et al. [26]. This study concerns a high-speed tablet press and illustrates the effect of tablet weight variability at

different operational settings, i.e. tableting speed and paddle speed.

- **Process design:** In an early stage of the process life cycle, most of the industrial manufacturing equipment is designed based on steady-state conservation laws, heuristic rules, and economic calculations with off-the-shelf material and utility properties. It is only in a later stage that a suitable operational and control strategy for the manufacturing of a particular drug product with this equipment is synthesized. Hence, there might exist a mismatch between the process design and process control development phase, which could result in troublesome uncontrolled sources of CQA variability. An example is given in the work of Vercruysse et al. [27] on continuous twin-screw wet granulation. As it turns out, the variability in liquid distribution for twin-screw wet granules is highly dependent on the method of liquid addition, i.e. in-phase/out-of-phase pump head configuration. For another example one can think of nonuniform heat transfer in a segmented fluid bed drying operation, resulting in under- and over-dried granules that may have an impact on the final tablet CQAs. From a process control point of view, such sources of variability are difficult to control because they are inherent to the design. Indeed, such design flaws should already be dealt with during the initial process life cycle stages in order to prevent cumbersome process control strategies aimed at minimizing the impact of those deviations. A more extensive elaboration that focuses on applications within the chemical process industry is given by Seferlis and Georgiadis [28].
- **Equipment status:** Wear of moving parts as a result of mechanical friction is not uncommon in the process industry. On the long term, mechanical changes of high-shear equipment (e.g. impeller, screws, and mill screens) might induce a drift in the intermediate and final CQAs of the product. Another important aspect in solids handling processes is the gradual fouling of process equipment and measurement instrumentation. An example of drifts because of process equipment fouling is provided by Vercruysse et al. [29]. More specifically, this study visualized the impact of gradual drying on air filter fouling for a fluid bed unit operation. The work presented by Saleh et al. [30] visualizes the formation of a powder cake during twin-screw wet granulation. Also, in the case of measurement probe fouling, a bias may be observed in the measurements. Therefore, correction mechanisms will miss their intended target and introduce unintended variability. A solution to probe window fouling for a fluid bed coating system is provided by Scheibelhofer et al. [31].
- **Human operator:** Manufacturing lines that require a great deal of manual interventions are prone to operator-induced variability despite the existence of training

programs and documented standard operating procedures (SOPs). Potential reasons for incorrect execution of existing protocols are operator fatigue and distraction. Matthews et al. [32] provided an extensive work on the nature of operator fatigue across industries, as well as the resulting failures and countermeasures.

- **Environment and utilities:** Changes in the production environment such as ambient temperature, humidity, and pressure as well as man- and machine-induced vibrations can have an impact on a production process. Indeed, damping machine vibration is especially important when working with load cells for low throughput dosing systems and to prevent particle segregation. The impact of vibrations on production segregations in a rotary tablet press is illustrated by Staniforth [33]. In addition, uncontrolled characteristics of utilities such as makeup water, compressed air, drying air, and steam might change over time and therefore result in quality variations of a final drug product.

23.4.2 Risk Assessment

Clearly, an important step in pharmaceutical process development is to identify these potential sources of CQA variability as well as their associated probability and severity, i.e. the process quality risk. To assess such risks to the final drug product, different risk management methodologies and tools can be found in regulatory guidances as well as the scientific literature [34]. An extensively used tool coming from the field of reliability engineering is failure mode and effects analysis (FMEA) applied to manufacturing processes [35]. One of the aims of this method is to determine for each failure mode a risk priority number (RPN). The RPN reflects a combination of the severity of the consequences (SEV), the anticipated frequency of occurring (OCC), and the ability/effectiveness to detect (DET) each failure mode. The three scores are multiplied to get the RPN. The higher the final score, the higher the need to take this risk into account in an overall control strategy of the process. Clearly, the scaling of the different coefficients differs across organizations. This is understandable since determining the risks associated with a specific process depends on in-house expertise, internal guidelines, and prior knowledge obtained from previous operations and/or trials. Other risk assessment methodologies such as fault tree analysis (FTA), hazard and operability analysis (HAZOP), and risk ranking and filtering can also be used to assess the risks to a drug product manufacturing process [34].

Table 23.1 demonstrates a non-exhaustive FMEA matrix for the continuous feeding and twin-screw granulation unit. For this particular process the FMEA study indicates that the most important risks for the feeding and granulation unit operations are related to pre-blend uniformity changes, irreversible material buildup on the granulation screws, feeder

TABLE 23.1 FMEA Matrix of the Feeding and Granulation Unit of a ConsiGma™-25 System (GEA Pharma Systems, Wommelgem, Belgium) Located at Ghent University, Ghent, Belgium

No.	Unit Operation	Potential Failure Modes (Process Defects)	Potential Effect(s) of Failure	SEV	Class	Potential Cause(s) of Failure	OCC	Existing Controls in place	Capabilities Current Process Controls	DET	RPN
1	Feeding	Feeder refill strategy	Mass flow rate disturbance	10	Process operation	Inherent to the operation of the process	10	Feeder MFR controller	100	3	300
2	Feeding	Electrostatic sticking of the powder in the bridge breaking part	Mass flow rate disturbance	10	Process design	Design limitations	3	None	30	5	150
3	Feeding	Irreversible material buildup on the feeding screws	Mass flow rate disturbance	4	Equipment status	Fouling	10	None	40	1	40
4	Feeding	External vibrations	Mass flow rate disturbance	10	Environment and utilities	Operator induced, external machinery	3	None	30	5	150
5	Feeding and twin-screw wet granulation	Blend uniformity primary powder blend	Nonuniform content distribution in granules	10	Raw materials	Blending, segregation	7	Offline confirmation at start	70	9	630
6	Feeding and twin-screw wet granulation	Density primary powder blend	Mass flow rate disturbance	10	Raw materials	Manufacturing, transport, storage	3	None	30	5	150
7	Feeding and twin-screw wet granulation	Temperature production environment	Aggregation rate changes because of changes in solubility and viscosity in the granulation liquid	5	Environment and utilities	Seasonal fluctuations	7	Building HVAC	35	1	35
8	Feeding and twin-screw wet granulation	Relative humidity production environment	Aggregation rate changes because of changes in water binding capacity	5	Environment and utilities	Seasonal fluctuations	7	Building HVAC	35	1	35
9	Twin-screw wet granulation	Viscosity granulation liquid	Aggregation rate changes	3	Raw materials	Manufacturing, transport, storage	6	None	18	5	90
10	Twin-screw wet granulation	Moisture content primary powder blend	Aggregation rate changes because of changes in water binding capacity	3	Raw materials	Manufacturing, transport, storage	6	None	18	5	90
11	Twin-screw wet granulation	Particle size distribution primary powder blend	Change in wet granule size	4	Raw materials	Manufacturing	3	None	12	5	60

(continued)

TABLE 23.1 (Continued)

No.	Unit Operation	Potential Failure Modes (Process Defects)	Potential Effect(s) of Failure	SEV	Class	Potential Cause(s) of Failure	OCC	Existing Controls in place	Capabilities Current Process Controls	DET	RPN
12	Twin-screw wet granulation	Particle morphology primary powder blend	Aggregation rate changes	3	Raw materials	Manufacturing	3	None	9	9	81
13	Twin-screw wet granulation	Degree of crystallinity primary powder blend	Change in wet granule size	4	Raw materials	Manufacturing	5	None	20	9	180
14	Twin-screw wet granulation	Unstable granulation screw speed	Aggregation rate changes	3	Process design	Time varying torque inside granulation barrel	7	Granulator speed controller	21	1	21
15	Twin-screw wet granulation	Barrel temperature fluctuations	Aggregation rate changes because of changes in solubility and viscosity in the granulation liquid	5	Process design	Time varying torque inside granulation barrel	8	Granulator temperature controller	40	2	80
16	Twin-screw wet granulation	Noninteracting control loops for liquid and powder feeding	Change in wet granule size and moisture content	9	Process design	Poor design	10	None	90	3	270
17	Twin-screw wet granulation	Breakage of the granulation liquid addition nozzle O-ring	Change in wet granule size and moisture content	9	Process design	Manufacturing, placement	4	Visual inspection at start run	36	5	180
18	Twin-screw wet granulation	Blockage of the granulation liquid injection nozzle	Change in wet granule size and moisture content	9	Process design	Manufacturing, failed cleaning procedure	1	Visual inspection at start run	9	9	81
19	Twin-screw wet granulation	Pulsation during granulation liquid addition	High frequent change in wet granule size and moisture content	4	Process design	Poor design	4	Out of phase positioning of the pump heads	16	9	144
20	Twin-screw wet granulation	Wear on the granulation screw elements	Aggregation rate changes	3	Equipment status	Wear	4	Visual inspection at start run	12	4	48
21	Twin-screw wet granulation	Increased spacing between holes of the granulator outlet cartridge	Aggregation rate changes	3	Equipment status	Wear, failed operator procedure	4	Visual inspection at start run	12	4	48
22	Twin-screw wet granulation	Irreversible material build up on the granulation screw elements	Change in wet granule size	4	Equipment status	Fouling	10	None	40	9	360

refill strategy, the decoupling of powder and liquid flow rate, breakage of an O-ring on the liquid injection nozzle, and the crystallinity degree when working with polymers such as MCC. It is up to the process engineers to come up with a suitable process control strategy that mitigates these process-specific risks.

23.5 CONTROL SYSTEMS HIERARCHY

As explained in the previous sections, a process control strategy encompasses different tasks. Having a clear overview on the different levels of complexity is proven to be crucial when evaluating existing or implementing new control loops within a continuous drug product manufacturing line. Figure 23.2 shows the different technological layers that belong to the enterprise-wide control system [36]. As the figure suggests, the lower levels are the fundamentals for the higher levels of the pyramid. Hence, measurement data is typically exchanged from the bottom up, whereas reference values are moving from the top toward the bottom. In addition, the time scale also varies from layer to layer. The bottom layers are typically operated at high frequency (seconds, minutes), while the higher layers have a relatively low update frequency (days, weeks). Although different ways exist to divide an enterprise in its functional categories, the control hierarchy presented in Figure 23.2 is especially valid for the process industry dominated by continuous operations. Discrete parts manufacturing lines (e.g. automotive assembly lines) and batch production processes (e.g. specialty chemicals) have slightly different control architectures as a result of their discontinuous operation. A more generic functional model, including a detailed enterprise management level, can be found in more elaborate references, e.g. ANSI/ISA-

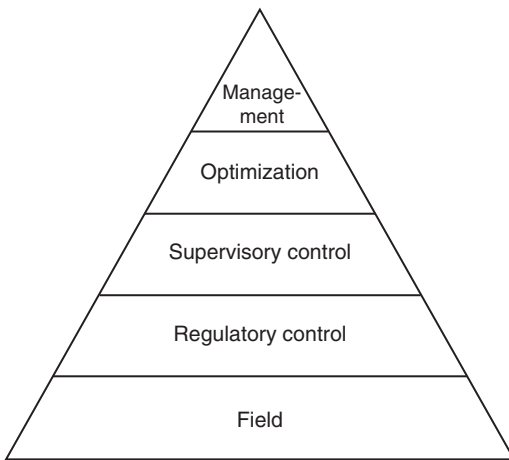


FIGURE 23.2 The different levels of complexity in the control hierarchy.

95.00.03-2005 – Enterprise-Control System Integration, Part 3: Models of Manufacturing Operations Management [37].

23.5.1 Field Level

At the lowest level or **field level**, the process sensors or primary sensing elements such as temperature probes, flow sensors, level sensors, and load cells are located together with their respective signal transmitters to communicate the data in a standardized format. Typically, the field-level sensors are of a univariate nature, i.e. they only measure a single variable of interest. Discrete sensors such as proximity switches that provide information on the state of a certain operation (e.g. motor is on/off, equipment is present and in correct position, etc.) are also part of this lower layer. Hence, these sensors provide the necessary information required for safe operation of the process and therefore ensure the limits and/or constraints to which the process can be operated and controlled (e.g. level low/high sensor). In addition, the field level also includes actuating devices. Process actuators are hardware components used to manipulate the operation of the process, e.g. motors, pumps, valves, and heating elements. In control engineering terms, such actuators are typically referred to as the final control elements as they are in direct control of the manipulated variable of a process. Common manipulated variables are flow rates, temperatures, and pressures. While the field level does not provide a control strategy itself, it does provide the necessary instruments to measure the process and/or manipulate it.

23.5.2 Regulatory Level

Located just one step higher in the control systems hierarchy is the **regulatory level**, sometimes referred to as the basic control level. The main aim of the regulatory control layer is (i) safe sequential operation and (ii) maintaining a constant environment to which the product is exposed. While the former task is part of a discrete control system, the latter is guided by a process control strategy at the regulatory level. Indeed, the technologies found at the field and regulatory level are making up the physical shop floor of a process plant and are characterized by a high degree of automation provided by the equipment manufacturer.

To a less specialized audience, process control systems are typically regarded as those hardware and software components that relate to the primary task of the regulatory layer, i.e. automatically ensuring a predefined sequence of operations. Although this is an important task in discrete manufacturing, it is of less importance for the control of continuous industrial processes. In fact, a difference should be made between discrete “factory” automation and continuous “process” control [37, 38]. The focus of this chapter is primarily on continuous control since it is of major importance for continuous drug product manufacturing. Note that at the present

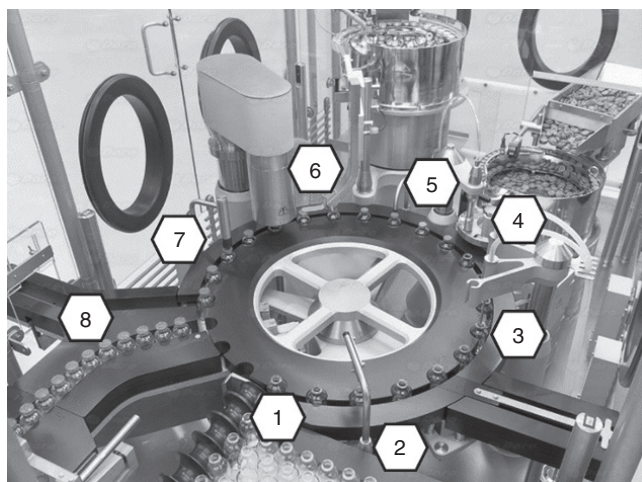


FIGURE 23.3 A fully automated vial filling line using a discrete control system. *Source:* Courtesy of Maquinaria Industrial Dara, SL.

time, most industrial processes are automatically controlled by combining both discrete and continuous control operations to reduce the number of manual process operations and manual data interpretation.

An example of a unit operation controlled only by **discrete controls** is an automatic vial filling line (Figure 23.3). Here, the aim is to continuously guarantee a sequence of operations: (i) bring an empty vial into the workbench, (ii) ensure that a vial is in place, (iii) fill the vial to the desired level, (iv) place a vial stopper, (v) ensure that the stopper is in place, (vi) place a metal seal, (vii) ensure that the metal seal is in place, and (viii) release the vial from the workbench. To orchestrate the synchronization between all these steps, a discrete control system is in place to handle this sequence of operations. Moreover, the control system will also react when one of the steps is not performed as specified due to failure. This is handled by having intermediate checks in place that verify the outcome of each step. Such checks are performed with the help of a few discrete sensors such as limit switches. Whenever a vial, stopper, or cap would be missing, the unit is diverted from its normal route to avoid reaching its intended destination, i.e. the patient. In the case of a vial filling line, the discrete control system is completely automated and adaptable. Indeed, no operators are involved at this stage to check each individual vial nor do they perform the corrective actions. To this end, a programmable logic controller (PLC) or another kind of microcontroller is in place. The PLC is programmed to ensure a certain flow or sequential logic of the units using a combination of basic if-else structures that respond to discrete events and exact timing functionalities. Other discrete technologies at the regulatory level include motion control, the control of pneumatic and hydraulic cylinders, unit inspection through machine vision, serialization, and identification of the product. While it may seem that discrete control is only useful when

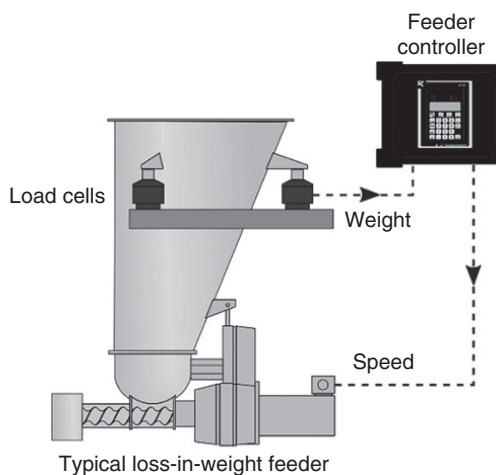


FIGURE 23.4 A loss-in-weight powder feeder that uses a continuous process control system. *Source:* Courtesy of Coperion GmbH.

dealing with processes that handle products composed of discrete units or parts, it is also used for batch processes and processes operating on continuous product streams such as the typical unit operations found in drug product manufacturing facilities. For those, discrete controls are used to control the start-up and shutdown of the continuous process equipment, handling of process, and safety interlocks, providing process alarms upon faults or violations of operating constraints and diagnose process instrumentation.

In addition to discrete control functionalities, the regulatory layer may also carry out **continuous control**. In process control terminology, this is referred to as regulatory process control. The aim here is to maintain a reference or set point value for the different continuous low level or regulatory processes despite external disturbances. This way, the product handled in that particular part of the process is exposed to an exactly determined environment, e.g. temperature, shear, and pressure.

An illustration of a regulatory process control loop in continuous drug product manufacturing is, for example, the control of material throughput with a loss-in-weight powder feeder (Figure 23.4). When operated in its gravimetric mode, the weight decrease in the hopper due to material feeding is continuously monitored over time using one or more load cells. This information is subsequently used by a controller (programmed on a microcontroller, PLC, or DSC system) to calculate the rate at which powdered material is fed into the next unit operation. Through automatic manipulation of the feeding screw speed, the controller is able to keep a specified set point over time for the material flow rate regardless of the level of powder inside the hopper.

Currently, regulatory process control of most pharmaceutical processes is fully automated to increase production rates and prevent inconsistent manipulations of the CPPs of a unit

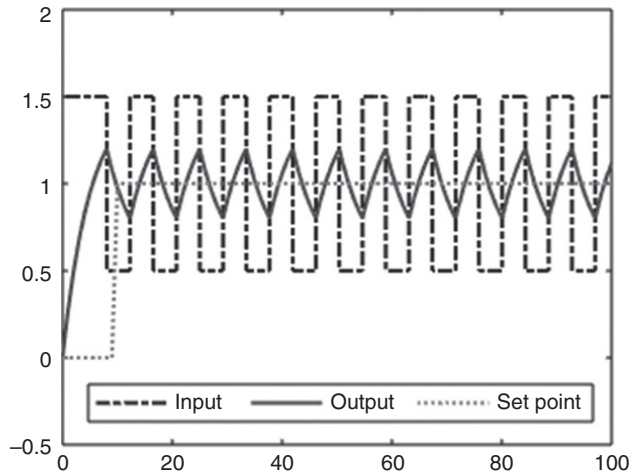


FIGURE 23.5 ON-OFF control law acting on a continuous process.

operation. Depending on the problem at hand, different control laws can be used inside the controller to obtain satisfactory results. One of the simplest control laws is the **ON-OFF control law** (Figure 23.5). Here, the control action $u(t)$ applied to the manipulated variable can assume just a minimum u_{\min} or a maximum u_{\max} value depending on how the value of the controlled variables $y(t)$ relates to the set point $y_{\text{sp}}(t)$, i.e. the control error $e(t) = y_{\text{sp}}(t) - y(t)$. If u_{\min} is chosen to be 0 or OFF, the result is similar to that of a relay switch. For an ON-OFF controller, the control action is thus calculated in the following way:

$$u(t) = \begin{cases} u_{\min}, & \text{if } e(t) < 0 \\ u_{\max}, & \text{if } e(t) > 0 \end{cases} \quad (23.1)$$

In this case, the only question that arises is how to define u_{\min} and u_{\max} or more general $\Delta u(t) = u_{\max} - u_{\min}$. On the one hand, if $\Delta u(t)$ is substantial, the response of the controlled variable will be fast, and therefore it might significantly overshoot the desired set point value. On the other hand, having a very small $\Delta u(t)$ will result in very fast oscillations around the target, which might not be desired. Although the ON-OFF control law is widely adopted, e.g. in heating, ventilation, and air conditioning applications, its main disadvantage is the persistent oscillation of the controlled variable [39].

Another widely adopted control law at the regulatory level is the **PID control law** (Figure 23.6). It usually consists of applying the sum of three types of control actions in a feedback loop: a proportional action (P), an integral action (I), and a derivative action (D). Again, these actions relate to the error $e(t)$ observed between the set point $y_{\text{sp}}(t)$ and the measured process value $y(t)$, i.e. the control error $e(t) = y_{\text{sp}}(t) - y(t)$. The control action for the noninteracting PID control law is theoretically calculated as follows:

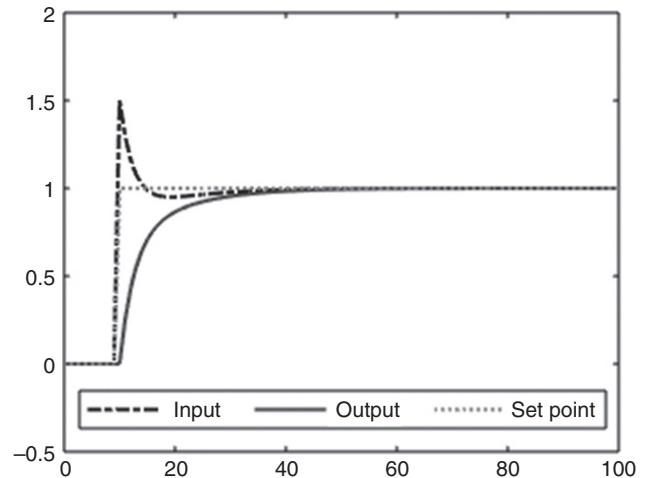


FIGURE 23.6 PID control law acting on a continuous process.

$$u(t) = P + I + D = K_p \left(e(t) + \frac{1}{T_i} \int_0^t e(\tau) d\tau + T_d \frac{de(t)}{dt} \right) \quad (23.2)$$

In fact, the three terms in this equation can be physically interpreted. The P action denotes a direct proportional response of the control whenever the measured value $y(t)$ deviates from the set point $y_{\text{sp}}(t)$ as defined by $e(t)$. The I action takes into account the past behavior of the process by accumulating the errors over the progress of time $\int_0^t e(\tau) d\tau$. But the D action accounts for future errors by calculating the rate at which the error is changing $\frac{de(t)}{dt}$. For satisfactory PID control, good values for the control parameters K_p , T_i , and T_d have to be assigned. In control terminology this is referred to as the act of tuning a process controller. Different methods exist for controller tuning ranging from rules of thumb to mathematically substantiated theories. Hence, in the past the performance of a PID control loop depended on the expertise of the user. However, today's software tools, including process models and auto-tuning hardware, allow for fast and effective tuning of different regulatory control loops without requiring specialist knowledge [40]. Note that the ON-OFF and PID controller do not necessarily need to know which process loop they are actually controlling. This makes these controllers empirical in nature, which comes with drawbacks when handling more complex interacting, nonlinear, and time-varying processes.

It is known that the standard feedback PID controller might not perform as desired for more complex processes such as those characterized by a high degree of noise, constraints on the manipulated variables, interactions with other processes, multiple control variables, and substantial time delays. In such cases the PID control law needs to be modified such that only some terms are selectively used (e.g. P, PI, and PD controller) or a modified strategy is required (e.g. anti-reset windup, cascade, and feed-forward strategies) to

overcome specific problems. Indeed such modifications to the standard control strategy are likely to be found at the regulatory level. In general, this is typically referred to as **advanced regulatory control** within process control engineering practice [18]. Note that although new and effective control laws are continuously being developed, feedback PID controllers are still by far the most widely adopted controllers in industry – some even estimate over five billion PID control loops in industry worldwide [41]. This is because PID controllers are relatively simple to use and provide satisfactory performance in many continuous process control tasks.

23.5.3 Supervisory Level

The next level in the control pyramid is known as the **supervisory level**. The term supervision indicates this layer's intention to oversee an entire unit operation and/or entire process with the aim of guiding the system in a defined direction. The objective is typically related to directly maintaining target values for intermediate and final CQAs despite the array of disturbances acting on the process. While process control at the regulatory layer is only interested in ensuring a constant environment to which the product is exposed, supervisory control typically aims at monitoring and controlling product characteristics directly. Other objectives of this layer are related to ensuring minimal energy and material consumption, ensuring production rate across unit operations, and avoiding over-recycle of product due to recirculation loops [42]. All those tasks are achieved by continuously updating the set points of the regulatory layer and thus the supervisory level is not directly manipulating process actuators. Hence, the occasional usage of the term set point controls for this layer of the control pyramid [43].

Product quality attributes as well as material and energy consumption depend on various process parameters. Hence, the problem tackled at the supervisory level is multivariable in nature and thus requires in-depth knowledge of the underlying complex process behavior. In the process control field, such a system is conceptualized as a multiple-input multiple-output (MIMO) process subject to a wide range of constraints. Indeed, the supervisory controller has to make sure that based on process- and product-related measurements, the MIMO process results in a desired behavior while meeting all constraints.

Control strategies that are abundantly referred to in the pharmaceutical industry rules, regulations, and guidelines are typically operated at this supervisory layer. For example, in the FDA Q8(R) Pharmaceutical Development Guidance for Industry, the different elements of a control strategy all relate to ensuring product quality, which is exactly the responsibility of the supervisory control layer. As the control pyramid indicates, to achieve a state of control at this level, a well-functioning regulatory control layer is mandatory to guarantee that the product is exposed to an exactly determined process environment.

Over the years different control strategies have been developed that specifically target the complex problems observed at the supervisory level. An important difference with the controllers found in the lower layer is the fact that supervisory control is not always automated. Manual analysis and manipulations of the process performance are very common in certain industrial applications. The reasoning behind this is related to the fact that the integrated multivariable processes being studied at this level are not always fully determined and work typically at a much slower pace as compared with the fast regulatory processes, leaving the possibility for thorough analysis and evaluation by a process operator or engineer before coming up with a new control action. An overview of different supervisory strategies in the pharmaceutical manufacturing industry is nicely illustrated in the work of Yu et al. [14]. In this work three levels of complexity are proposed depending on the complexity of the supervisory control strategy (Figure 23.7).

23.5.3.1 Level 3 The least advanced and still the most widely adopted within the pharmaceutical industry is **level 3 control**. This type of supervision is characterized by extensive end-product testing while over-restricting variations in material attributes and process parameters. Such a control strategy is typically in place when there is only a poor understanding of the underlying mechanisms governing the process, i.e. how CMAs and CPPs relate to the CQAs of the product. Clearly this is a labor-intensive control strategy with limited feedback to compensate for disturbances throughout process operation. That is why overprocessing of materials is a common strategy at this level, e.g. over-drying of wet granular material to ensure compliance to the residual moisture tolerance limits. In addition, analysis of CQAs is guaranteed through offline testing with long time delays. Indeed, a control strategy based on strict regulations and the quality-by-testing principle is not aligned with the objectives of the quality by design (QbD) [19, 34, 44] and PAT guidances for industry. At most the measurements taken over time are used for thorough statistical analysis to determine if the process is within specifications or starts drifting. Such a passive control strategy only touches the field of statistical process control (SPC). Indeed the application of such a level 3 control strategy is only applicable if the risks of failure and its accompanied losses are low. Since profitability through optimization has become a major concern in the industry, such a wasteful control strategy is not adequate for most pharmaceutical manufacturing lines. In fact, a transition toward continuous manufacturing of drug substances and drug products requires a faster and more flexible approach toward supervisory process control that does not depend on laborious manual analysis.

23.5.3.2 Level 2 **Level 2 supervisory control** is unquestionably more in line with the current QbD philosophy since

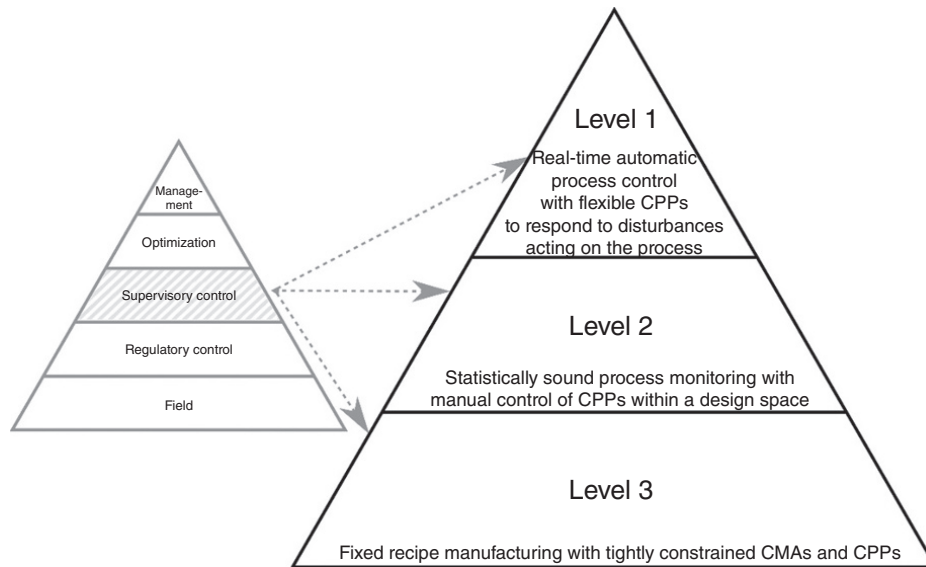


FIGURE 23.7 Supervisory control strategy levels. *Source:* Adapted from [14]. Copyright (2014) Elsevier.

it is based on knowledge gained through appropriately designed development studies, which means precise identification of the sources of process variability together with a clear understanding of their propagated effects in order to predict them accurately. Indeed, this is exactly what the regulators are pushing for with the concept of a design space of a process, i.e. a range in the multidimensional space of CMAs and CPPs for which it has been demonstrated that quality is ensured. For routine process operation, the normal operating range is chosen to fit well within this design space. Different tools exist to build a knowledge base upon which the design space can be founded, including experimental analysis using factorial design of experiments, data-driven modeling, and detailed mechanistic modeling [45].

An important scientific basis for level 2 control is found in the field of SPC. The goal of SPC is to detect the existence, magnitude, and timing of changes that might cause a process to deviate from its desired operation. At its heart, SPC is based on common cause variation data, probability distributions, statistical inference, and control charts. From these, metrics such as capability and performance to determine process robustness are derived [46]. Yet, these methodologies are aimed at monitoring rather than actively controlling the process; the term statistical process monitoring is suggested by some authors [47]. Figure 23.8 gives an illustration of a process average using an individual control chart (I-Chart) for a drug product continuous direct compression (CDC) line as provided by Almaya et al. [48]. This work exemplifies the role of SPC to monitor intra-batch variability as a result of start-up, steady-state intermittent pausing, and shutdown operations of a CDC line. An important aspect, clearly elaborated in this study, is the fact that although a process might be

in a state of statistical control, it might not conform to the specified quality limits. Hence, careful comparison of the measured results against the relevant requirements is always needed.

In classical SPC, where only the CQAs of the final product are monitored, it is up to process operators and engineers to find out a probable root cause for an out-of-control event. Clearly, this is based on their understanding of the process and a one-at-a-time, i.e. univariate, inspection of the different process variables. For complex multivariable batch processes, i.e. characterized by numerous interactions and correlations between the measured variables and even more for data-rich continuous processes, the expertise required by the process operator becomes more and more demanding. A valuable alternative is multivariate statistical process control (MSPC). Instead of using a complex data set as a whole, the variable space is compressed to a reduced space composed of latent variables (LVs). These LVs explain the main underlying phenomena governing the observed process behavior. An overview of the different multivariate modeling methodologies such as principal component analysis (PCA) and partial least squares (PLS) together with the possibilities they offer in a process context is given by MacGregor et al. [49]. An important feature is the ease at which information can be visualized using the few LVs or diagnostics such as the Hotelling T^2 statistic, the squared prediction error (SPE), or the distance to the model in the X-space (DModX) [50]. For these multivariate diagnostics, variable control charts such as Shewhart, cumulative sum (CUSUM), and exponentially weighted moving average (EWMA) are also applicable. In addition, whenever one of the considered diagnostics is outside its control limits, fault identification is made possible by tracing back the contribution of each original

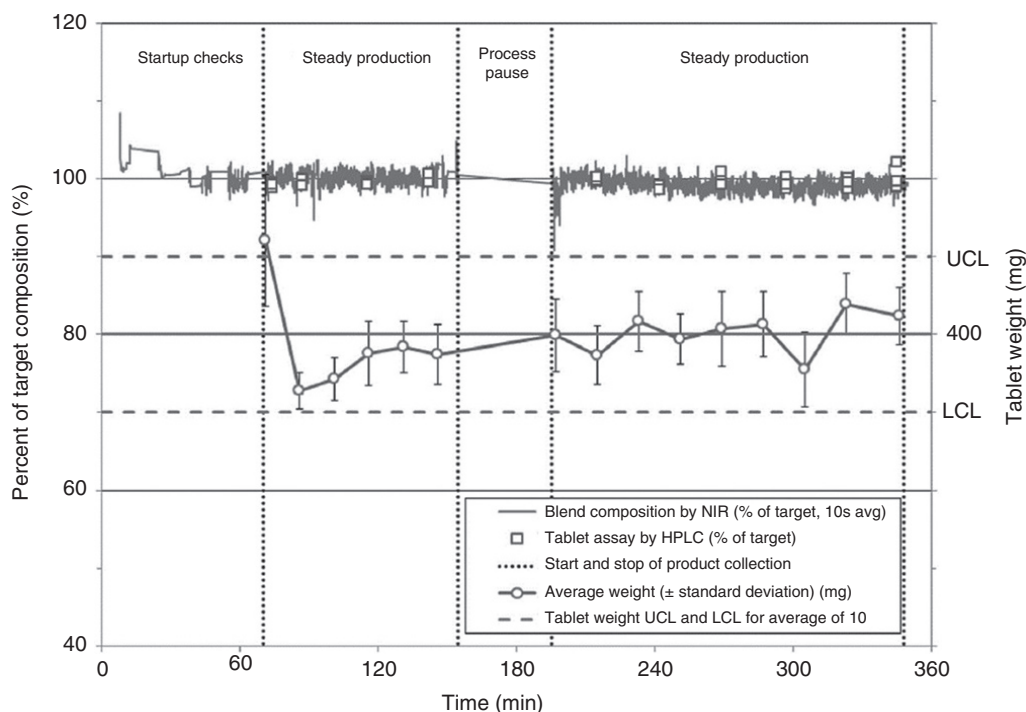


FIGURE 23.8 I-Chart for a drug product CDC line. *Source:* Reprinted with permissions from Almayra et al. [48]. Copyright (2017) Elsevier.

measured variable to an out-of-control event using contribution plots. This clearly indicates that MSPC is an important tool in the pursuit of controlled industrial processes. An example for a continuous twin-screw wet granulation line is given by Silva et al. [51]. In this study, different logged process parameters were used in an MSPC strategy to determine the state of the process while subjected to different artificial disturbances.

Initiated by the 2004 US FDA PAT guidance for industry, (near) real-time assessment of product quality is upcoming in the pharmaceutical industry. For this purpose multivariate data analysis is commonly used to relate multivariate measurements obtained from real-time process analyzers (e.g. NIR spectroscopy, Raman spectroscopy, etc.) to an estimate of quality in real time. This is commonly referred to as the act of measurement calibration. Different applications of these so-called PAT tools for the production of oral solid dosage forms are summarized in the work of Matero et al. [52] and Laske et al. [4]. A summary of the challenges and issues related to spectroscopic PAT tools is provided by Chen et al. [53], while the work of Goodwin et al. [54] presents a control strategy to ensure tablet content and content uniformity after compaction. To this end, compaction force-weight regulatory control of the tablet press is combined with at-line testing of mean and individual weight as well as tablet content testing using NIR spectroscopy. A similar level of control complexity is proposed in the work of Burggraeve et al. [55] by combining process measurements, real-time product

measurements, and MSPC to monitor a batch fluid bed granulation process, i.e. batch statistical process control (BSPC).

Although the usage of real-time analyzers is receiving considerable attention over the last couple of years, the capital and operational investments that come with such instruments can definitely be a barrier. For some process variables (especially those that are of interest in a solids handling process), real-time measurements do not guarantee robustness at commercial scale or do not yet exist. An attractive alternative is therefore the use of so-called soft sensors or inferential sensors that use real-time measurements of readily available secondary variables, i.e. process parameters such as temperature, humidity, mass flow rate, torque, etc. together with a model to estimate a desired primary variable, i.e. product quality attribute. Process models based on fundamental and empirical knowledge that are used in the early stages of process development as well as LV modeling are good starting points for such soft sensor strategies [56]. A systematic approach for the development of data-based soft sensors is given by Lin et al. [57]. The work of Rehr et al. [58] demonstrates a soft sensor to predict API concentrations after continuous extrusion, blending, and twin-screw wet granulation. Interestingly, a transfer function model of second order was applicable for all three applications. In batch operations such as fluid bed drying, soft sensors are extensively used for end-point detection. An example for a semicontinuous dryer is given by

Fonteyne et al. [59]. In this work a mechanistic model using mass and energy balances is compared against real-time NIR and Raman monitoring as well as the conventional offline Karl Fischer method.

Current practice in the pharmaceutical industry is mainly based on locking production settings, i.e. open-loop manufacturing with strict regulations according to level 3 control. Instead, future manufacturing should rely on flexible process settings to actively anticipate product variability [52, 60]. By using multivariate modeling in combination with information about the design space, active process manipulation to keep CQAs within specification limits is also made possible with a level 2 supervisory control strategy. This way, variability observed before processing (e.g. variability in raw materials, equipment status, and environmental conditions) as well as during processing (e.g. variability due to process operation, equipment status, environmental conditions, and utilities) can be compensated. A simple illustration is manual process adjustment with the help of an out-of-control action plan (OCAP). An OCAP constitutes a flowchart that describes a sequence of activities that must take place following an out-of-control event recorded with the help of a control chart [61]. Clearly, such an action plan is limited to the knowledge used to define the strategy – hence the importance of a fundamental process understanding. Note that the use of an OCAP is a curative measure that is only triggered whenever a process is considered out of control. Different studies for drug product manufacturing have however shown the use of preventive actions to comply with product specification limits. Burggraef et al. [62] uses a feed-forward control strategy based on multiple linear regression (MLR) models and PLS models to determine an optimal temperature during the drying phase of fluid bed granulation. In this way, the granule density at the end of the process falls within the desired range.

In literature, LV models are mostly used to understand the process and predict quality attributes using raw material attributes and process parameters. However, they can also be used for product and process design by using them in the inverse direction. This means using product quality specifications to estimate the required raw material properties or process parameters. This methodology, referred to as model inversion, requires numerically constrained optimization tools to come up with potential solutions [63, 64]. A neat example that uses model inversion in the context of pharmaceutical manufacturing is given by García-Muñoz and Mercado [65] and later on extended by García-Muñoz et al. [24]. In the latter paper, an optimal inventory selection system is proposed, which, based on a PLS regression model inversion, is capable of choosing which raw material lots should be blended according to their physical properties in order to cancel out undesired interactions across materials.

This way, the next best campaign in terms of solid dosage dissolution is calculated up front using the available materials of the inventory. This methodology can thus be seen as a level 2 supervisory control strategy that acts in a feed-forward manner to reduce variability in the CQAs of its process. Similar work is presented by Muteki et al. [66].

23.5.3.3 Level 1 While level 2 supervisory control has the advantage of increasing operational flexibility through a design space concept, it does not propose real-time process manipulations to automatically correct for deviations from the target. Hence, manual evaluation and interpretation is required from process operators and engineers. In this respect, operating procedures with sufficient detail help these practitioners in standardizing the operation of a process. However, manual interventions are prone to external factors and might not always take into account the complex underlying interactions of a manufacturing process. In addition, ensuring exact timing and intensity of a manual process intervention can be troublesome [67]. These issues are especially valid for continuous oral solids manufacturing lines that are usually characterized by a relatively small residence time and material holdup within each unit operation as well as small final dosage forms when compared with the material throughput. Hence, an automatic control strategy that actively manipulates the CPPs based on real-time measurements is paramount. Within the current regulatory framework of the pharmaceutical industry, this automatic system is sometimes referred to as PAT-based control (the term process analytically controlled technology [PACT] is also used) and is covered by **level 1 control** as proposed by Yu et al. [14].

A straightforward example of level 1 control for a CDC line using in-line Raman-based feedback control is given by Nagy et al. [68]. With the help of a PI controller, the concentration of API surrogate in the blended material stream could be kept on target despite systematic operational disturbances. A similar approach was presented in the work of Zhao et al. [69]. Because of its critical importance, API composition control is generally targeted first for a level 1 control strategy. However, other intermediate and final CQAs could also benefit from automatic process control. For example, an experimental study presented by Correa [70] illustrates closed-loop control of API size metrics by adjusting milling speed in response to measurements coming from an online particle size analyzer. Also, the study of Baronsky-Probst et al. [71] reports the application of a feedback mechanism to control the specific mechanical energy (SME) and strand diameter in a continuous twin-screw hot melt extrusion process.

As it appears from the previous examples, the underlying control principles used in the automatic process control of the regulatory layer, such as ON-OFF and PID control, can also be

used at the supervisory level for level 1 control. However, the difference lies in the fact that supervision typically relates directly to the automatic control of intermediate and final CQAs themselves, hence going beyond control of the CPPs that determine the environment to which a product is exposed. In a pharmaceutical manufacturing context, the application of such a level 1 control strategy can be referred to as **advanced process control** (APC). It should be noted that although the origin of APC is found in the chemical, petrochemical, and oil refining (CPOR) industries, the exact usage of the term in different contexts is subject to debate. Some authors in the field of pharmaceutical manufacturing refer to APC as all possible control strategies that make use of PAT, mathematical models, or other techniques to actively control intermediate or final CQAs [72]. But in chemical engineering, the definition is more strictly defined in terms of problem statement, i.e. APC is referred to as the control of MIMO processes that are subject to physical constraints with the goal of improved product quality while reducing operational costs [41, 43]. Although both definitions are very similar, in the former pharmaceutical approach, APC does not necessarily need to tackle a MIMO process and may include the usage of real-time process analyzers for automatic control purposes. However, as product quality is typically depending on multiple variables and their interactions within the process, the problems encountered in the pharmaceutical industry are typically of multivariable nature, i.e. MIMO systems. In addition, the fact that reducing operational costs in real time is not specifically mentioned in pharmaceutical APC could relate to the point that this technology still needs to prove its benefits at commercial scale. Hence, the focus of this chapter is mainly on ensuring product quality within drug product manufacturing.

Different automatic control strategies exist to ensure proper supervision. A straightforward solution to such a multivariable control problem is the application of multiple controllers, e.g. PID controllers, acting on the single-input single-output (SISO) subprocesses. This is also referred to as decentralized control [73]. An experimental example for a pilot-scale integrated continuous pharmaceutical manufacturing line is given by Lakerveld et al. [74], while Ierapetritou and Ramachandran [75] report a decentralized control CDC line using a tumbler mixer. The advantage of the decentralized control approach is related to its simplicity and ease at which it is maintained. Clearly, an important part of decentralized MIMO control is finding the optimal pairing between inputs and outputs. Different methodologies exist to quantify the degree of interaction, ranging from the commonly used relative gain array (RGA) to more elaborate measures such as Gramian-based interaction measures. A more heuristic approach is suggested by Lakerveld et al. [42] and Rehr et al. [76], which uses sensitivity analysis to evaluate input–output pairing for pharmaceutical manufacturing processes. In addition, the work of Ramachandran et al. [77]

provides a simulation study that compares the influence of different degrees of process interactions on the input–output pairing and the decentralized controller performance for a CDC line. The results in this work clearly illustrate that process interactions limit the performance of decentralized control strategies. Also, simulation studies provided by Ramachandran and Chaudhury [78] and Singh et al. [79] demonstrate the use of decentralized control for a single pharmaceutical unit operation and an entire continuous pharmaceutical production line, respectively. The latter focuses on continuous dry granulation using roller compaction technology. In this specific case, the strong interaction between different variables results in an oscillatory response when tracking the targeted CQA set points. When confronted with such high degree of interactions, valuable alternatives exist, for example, the application of interaction decoupling with the help of a pre-compensator, optimal control using a linear–quadratic–Gaussian (LQG) controller, and H ∞ robust control [80]. To date, these more complex multivariable controllers have not yet been reported in the framework of continuous oral solids manufacturing.

Instead, most modern constrained multivariable controllers for the process industry are based on receding horizon model-based predictive control algorithms. In a PAT context such a control strategy is sometimes referred to as PAT-enabled model-based control [81]. A vital element in those algorithms is the inclusion of a detailed conceptualization of the systems behavior. The reason for this requirement is related to the fact that a computerized system, instead of a process operator, needs to come up with the next best control action while taking into account the underlying variable interactions, constraints, and time delays. In the field of process systems engineering, this knowledge is provided in the form of a mathematical model. The model itself is used in real time within the model predictive control (MPC) software controller to predict the process output at future time instances given a combination of the past and future control actions. By including this knowledge, a set of optimal corrective actions can be calculated in real time to ensure product quality while under constant pressure of external disturbances. This concept also highlights the greatest drawback of model predictive control, i.e. an appropriate mathematical model of the process needs to be available. In practice developing such a model can be a laborious exercise for nonexperts.

MPC is not a single algorithm. Instead, it is a family of automatic process control algorithms. More specifically this framework is characterized by (i) the explicit real-time application of a process model and (ii) the calculation of an optimal control action through minimization of a cost function. Hence, all of the MPC algorithms share a common strategy: the future outputs $\hat{y}(t+k|t)$ (e.g. CQAs) of a continuous process are predicted over a determined horizon N , i.e. the

prediction horizon, at each instant $t+k$. To do this end, the process model is used in combination with the past values of the inputs (e.g. CMAs and CPPs) applied to the process and outputs measured from the process, as well as future control signals $u(t+k|t)$. This set of future control signals is calculated using a cost function and optimizer that aims at future process outputs that follow a reference defined by the process operator. While control actions for more than one step in the future are being calculated, only the action relevant to the current sample instance $u(t|t)$ is applied to the process at each sampling instance. When the next measurement $y(t+1)$ becomes available, the whole procedure is repeated – hence the concept of a receding horizon [82] (Figure 23.9).

From the control systems hierarchy, it is clear that the supervisory layer acts upon the regulatory layer. As mentioned before, the output of the supervisory layer is directly fed to the regulatory layer as new set points. An alternative approach, which is of particular interest for automatic level 1 supervision, is to merge the regulatory and supervisory layers so that the latter is directly manipulating the process parameters. While this may seem as a more straightforward approach, in most cases it is not the best choice. Difficulties in the identification and tuning are introduced, while the overall system becomes more sensitive to errors and failures. Hence, such a control strategy is avoided at the lowest control level [36].

Table 23.2 gives an overview of different simulation studies that successfully demonstrate the use of an MPC strategy in the context of continuous oral solids manufacturing. Although the starting point of these studies is a detailed mechanistic model, the model used for control action calculation is often reduced in complexity. Such models, typically linear in the variables, can be derived from simulation data or obtained

through model linearization around a nominal steady-state operating point. To investigate the performance of the proposed control strategies of Table 23.2, the process is simulated using the mechanistic model, while the control algorithms use its linearized version.

Successfully simulating the closed-loop behavior of a continuous manufacturing line proves to be very useful to determine the overall performance of a process controller beforehand. Indeed, tedious physical experiments are excluded, while the boundaries of the control system in terms of robustness under uncertainty are easily explored. However, reproducing these results on the actual application is another challenge one should not take too lightly.

In practice, not all variables of interest can be monitored in real time during manufacturing because instrumentation for real-time analysis might be lacking or instrumentation required to monitor certain CQAs might be too costly. Sometimes software problems can also prevent the integration of measurement systems, e.g. appropriate drivers that link hardware and third-party software products may not be readily available. Most industrial instrumentation also requires proper interfacing in the process stream to obtain qualitative measurements [91]. This last argument is especially valid for bulk solids manufacturing that is characterized by sticking, fouling, and flowability issues. Therefore, custom interfacing solutions are a necessity that does not disturb the process itself [31, 92–94]. Despite having a functioning interface, colored noise might affect the measurement to a greater extent in case of the real-time application. Indeed, mathematical models in the form of a soft sensor can be valid solutions. However, these also require in-depth understanding of the process behavior that might not be available for some unit operations of a continuous oral solids manufacturing line.

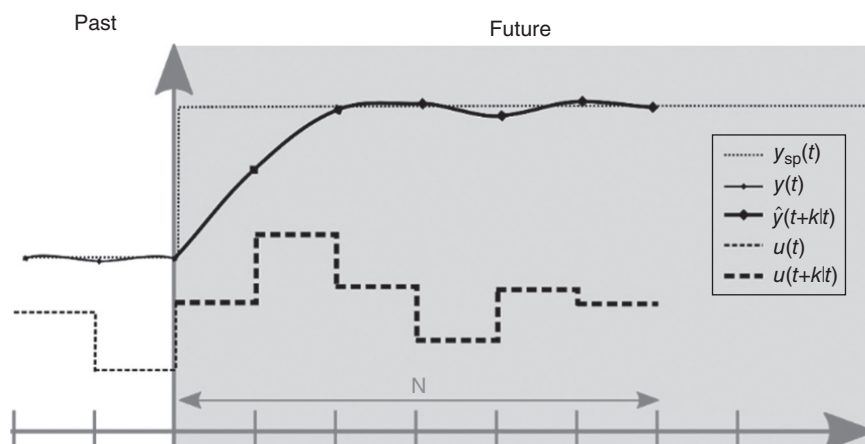


FIGURE 23.9 Illustration of the receding horizon concept used in model predictive control.

TABLE 23.2 MPC Simulation Studies in the Field of Continuous Pharmaceutical Oral Solid Dosage Manufacturing

Author	Unit Operation/ Continuous Line	Controlled Variables	MPC Algorithm	Models for Control
Valencia [83]	Tumbler mixer	Outlet mass flow rate, mixing variance	Constrained linear MPC	Transfer function models
Hsu et al. [84]	Roller compaction unit	Ribbon density, roll gap	Dynamic matrix control (DMC)	Finite step response models
Singh et al. [85]	Continuous direct compression line	Total flow rate, blend RSD, API composition, API-excipient ratio, tablet weight, tablet dissolution	Unspecified	Time-variant output-error models
Mesbah et al. [86]	Film drying unit	Film thickness, solvent weight fraction	Constrained nonlinear MPC	Nonlinear mechanistic model
Sen et al. [87]	Continuous API purification and processing line	Crystallization temperature, saturation concentration, drying gas temperature, API composition, holdup	Unspecified	First-order plus dead-time transfer function models (FOPDT)
Rehrl et al. [76]	Feeding blending unit	Outlet mass flow rate	Constrained linear MPC	State space model
Mesbah et al. [88]	Integrated continuous manufacturing line	API dosage, production rate, impurity content	Quadratic dynamic matrix control (QDMC)	Finite step response models
Haas [89]	Tablet press unit	Tablet weight, tablet hardness	Unspecified	Output-error models
Su et al. [90]	Feeding blending unit	API composition, API relative standard deviation, powder flow rate	Extended prediction self-adaptive control (EPSAC)	State space model

It is not only the measurement systems that pose important hurdles when implementing an APC solution. The low level behavior of process actuators can also result in erroneous APC solutions. Indeed, actuator dynamics are often neglected in mechanistic simulation studies because the dynamics are unknown or assumed to be insignificant. In addition, the IT infrastructure used to orchestrate APC is another important aspect. The presence of latency and jitter in the data communication might result in serious discrepancies between a simulation and a real-world validation experiment. Hence, these practicalities should be accounted for when aiming for a successful implementation of level 1 supervisory control.

Despite these challenges, both academic and industrial case studies have reported the successful use of MPC for automatic control of CQAs relevant to continuous oral solid dosage manufacturing (Table 23.3). Ultimately, the objective is to operate and evaluate the continuous process completely automatic, i.e. a production plant on autopilot.

The work presented in Table 23.3 clearly exemplifies the importance of a level 1 supervisory control strategy, i.e. mitigate process disturbances in real time by actively adapting the CPPs. Despite having such an automatic control system for a continuous manufacturing line, it is still possible that a severe process disturbance propagates through the entire

manufacturing line to cause OOS final drug products. While such a scenario for a traditional batch manufacturing process would result in the loss of an entire batch, continuous manufacturing demands for another approach. In this respect, the concept of a residence time distribution (RTD) is well investigated [100]. The RTD is a fundamental descriptor of a system as it gives distinctive clues on both the average time materials spent inside the system and the type of axial mixing occurring within a system, i.e. dispersion [101, 102]. Hence, identification of the multiple RTDs related to the unit operations and mechanical connections allows for accurate tracking of materials propagating in a continuous manufacturing line. Different studies in the field of continuous manufacturing have reported the use of RTDs to predict the impact of process disturbances and variations. Gao et al. [103] and Tian et al. [104] present experimental results obtained for continuous powder mixing. The work of Engisch and Muzzio [105] demonstrates the use of RTD modeling for a CDC line using simulations, whereas Van Snick et al. [106] present experimental findings for a similar manufacturing route. With respect to continuous granulation, an in-depth study is given by Kumar et al. [107] for a twin-screw wet granulation unit operation, while Mangal and Kleinebudde [108] and Kruisz et al. [109] report the application of RTD modeling for dry granulation units. Besides product tracking and scenario analysis for different disturbances, RTD models can also be used in the definition of a diversion strategy. In this way, OOS intermediate and end

TABLE 23.3 MPC Experimental Validation Studies in the Field of Continuous Pharmaceutical Oral Solid Dosage Manufacturing

Author	Unit Operation/ Continuous Line	Controlled Variables	MPC Algorithm	Model for Control
Blackwood [95]	Continuous direct compression	Holdup mass, production rate, API concentration	Unspecified	Unspecified
Singh et al. [96]	Continuous direct compression	API composition	Unspecified	First- and second-order plus dead- time transfer function models. Regressed using step response experiments
Bhaskar et al. [97]	Tablet press unit	Tablet weight, pre and main compression force	Unspecified	First- and second-order plus dead- time transfer function models. Regressed using step response experiments
Barros et al. [98]	Tablet press unit	Tablet weight, pre and main compression force	Unspecified	First-order plus dead-time transfer functions and a nonlinear polynomial. Regressed using step response experiments
Nicolai et al. [99]	Twin-screw wet granulation unit	L/S ratio	Extended prediction self-adaptive control (EPSAC)	Autoregressive moving average with external input (ARMAX)

products can be rejected, hence ensuring that only drug products that conform to specifications arrive to the patient. It is expected that a diversion strategy is especially important during start-up and shutdown phases of a continuous manufacturing line [16]. Figure 23.10 depicts a CDC line with a built-in diversion apparatus [110]. Note that in case of material diversion, a lack of materials might be introduced in the downstream unit operations. An adequate strategy that is either based on sufficiently large material buffers or a dynamic control system that matches the mass flow rate is an essential component of a continuous manufacturing line with intermittent material diversion.

23.5.3.4 Systematic Frameworks for Supervisory Control Implementation To date, an off-the-shelf APC software and hardware solution, which specifically targets the continuous solid dosage industry, is not yet commercially available. However, a few systematic approaches to process control design and implementation are proposed in literature [90, 111, 112]. These studies elaborate on in-depth findings for the practical realization of supervisory monitoring and control using multiple software and hardware solutions.

A key element in such realizations is a distributed plant-wide information system for advanced control purposes. Other than the process automation system, it captures only the relevant measurement data provided by univariate sensors and multivariate analyzers. But at the same time, such a system should allow for complex mathematical calculations, supervisory monitoring, and control as well as regulatory compliant data storage and reporting. In a PAT context, this system is referred to as a PAT data management platform

[113]. In fact, such a platform adds an additional IT layer to the automation architecture as it mainly consists of a computer network and accompanying software that shares and stores PAT-relevant data over its various nodes with the aim of enhancing process supervision. Commercial PAT data management platforms are available from different vendors, e.g. SIEMENS SIPAT, Optimal synTQ, ABB ADI and xPAT, Perceptive Engineering PharmaMV, OSIsoft PI System, Sartorius BioPAT MFCS, and Symbion Systems Inc. Symbion DX & RX. While some of these software products can be applied to a wide range of pharmaceutical manufacturing processes and related instrumentation, others are tailored for specific applications. It is important that the PAT data management platform is capable of interacting with different computerized systems and third-party software applications. This way, all relevant measurement data can be obtained (e.g. univariate process sensors and multivariate instrumentation) while at the same time allowing for more elaborate data analysis (e.g. multivariate regression using a calculation engine). Clearly, standardized communication protocols are proven to be crucial in such a context. For example, the object linking and embedding for process control unified architecture (OPC UA) protocol is extensively used to rapidly implement machine-to-machine communication.

23.5.4 Optimization Level

23.5.4.1 Optimal Operation As illustrated and discussed in the previous section, improved supervisory control offers a lot of potential to safeguard drug product quality for continuous pharmaceutical manufacturing. However, when

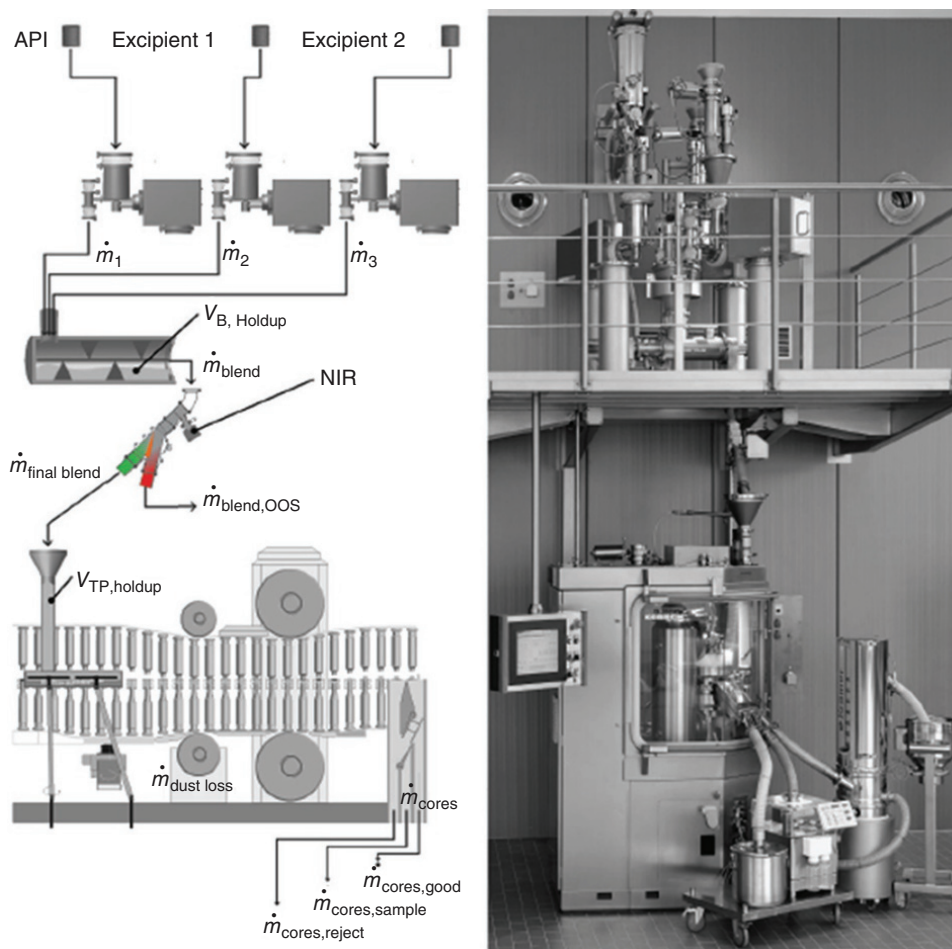


FIGURE 23.10 QbCon[®] CDC line with built-in diversion apparatus between continuous blender and tablet press. (Left) Schematic. (Right) Photograph [110]. *Source:* Image courtesy of L.B. BOHLE Maschinen + Verfahren GmbH, Ennigerloh, Germany, 2016.

operating a manufacturing line or production process, optimizing its performance on a dynamic basis is as important. Process performance encompasses different objectives such as profitability, efficiency, variability, capacity, and sustainability, to name a few. For example, guaranteeing a profitable process is achieved by evaluating different economic aspects of the process such as the time-varying market demand, energy consumption, raw material prices, utility prices, and product price on a frequent basis. Such aspects range from minute-based (e.g. energy markets) to too much slower characteristics (e.g. inventories) that only vary over the course of several days or weeks. To obtain maximum performance, process control can also offer solutions by steering the manufacturing process toward a performance optimum, given a set of constraints and market-defined economic considerations. Again different approaches exist to do so.

Initial work in the optimization field predominantly focused on the application of a hierarchically layered optimization strategy. Such an approach is characterized by an upper optimization layer, i.e. the real-time optimization

(RTO) layer, which establishes optimal operational points, i.e. set points or steady states for the CPPs and/or quality attributes, for individual process units or an entire plant. This is done with the help of real-time updated steady-state process models while taking into account the different aspects defined under the process performance definition. Next, the operating points suggested with the economic optimizer are imposed with the help of the underlying advanced controller, e.g. a model predictive controller, to maintain the process outputs at their respective set points. Hence, a cascade of information is passed on from the optimization layer to the supervisory layer from which it could be sent to the underlying regulatory control layer that is in direct control of the low level processes that make up a process plant (Figure 23.11).

For batch manufacturing, numerous studies have proven to increase the economic feasibility for different applications in a broad range of industrial sectors. An example relevant to the pharmaceutical drug product industry is reported in the study of Gagnon et al. [115]. This work demonstrates the

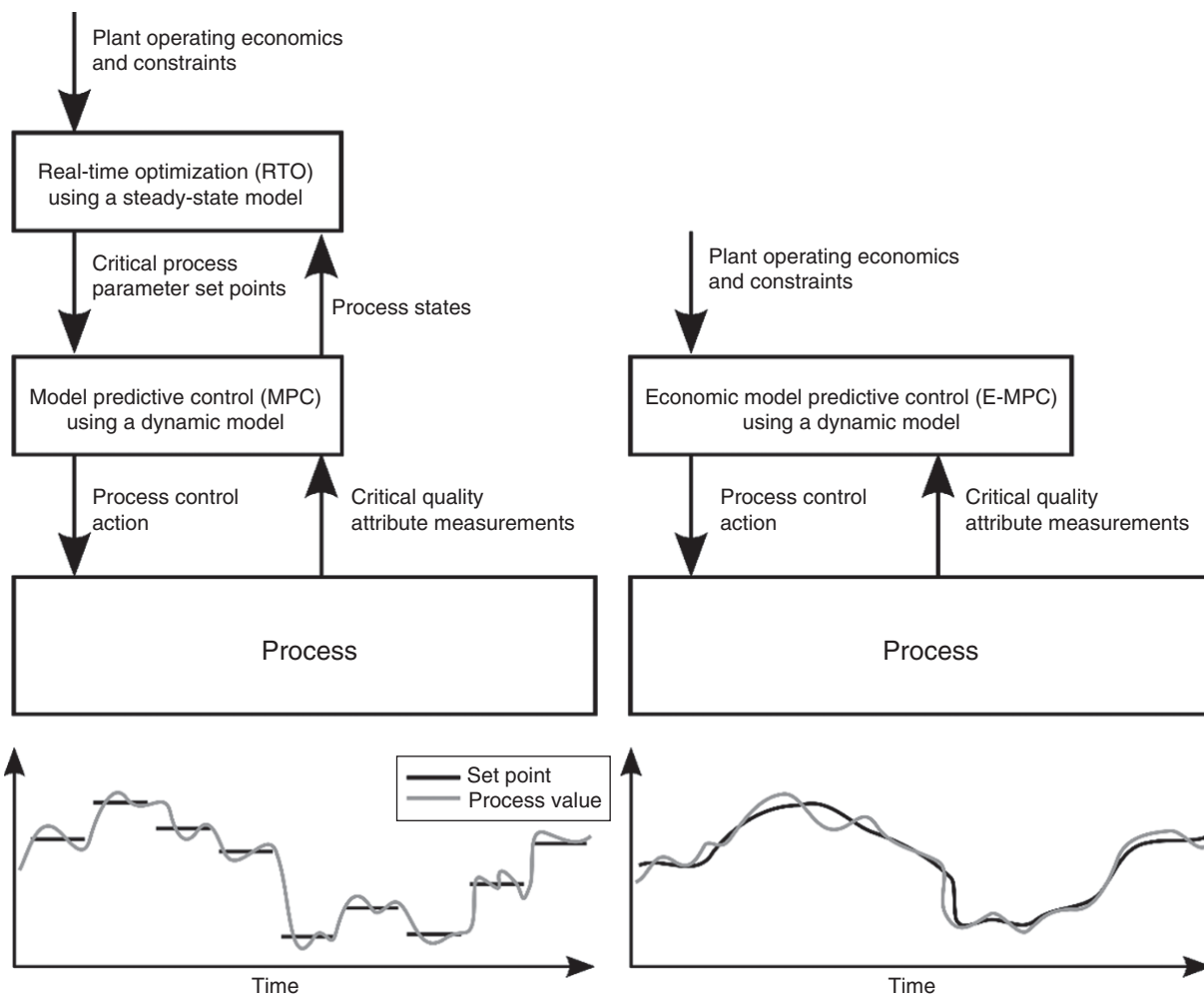


FIGURE 23.11 Visualization of two different approaches for process control at the optimization level. (Left) MPC with RTO. (Right) E-MPC. *Source:* Adapted from Petersen et al. [114]. Copyright (2017) Elsevier.

use of process control to minimize the energy consumption during batch fluidized bed drying. With respect to continuous manufacturing, a simulation example is given by Singh et al. [116] for a continuous direct compaction line. Here, a combination of MPC and moving-horizon real-time optimization (MH-RTO) was proposed. For this particular case, it could be shown that the MH-RTO strategy has the potential to increase revenue and profit by more than 10% when adapting the production rate, i.e. material flow rate according to market demands.

Over the last decade, there has been a tendency to include economic objectives directly in the supervisory controller itself. This way, process operation becomes even more flexible to highly dynamic market-driven operations that is not possible with a steady-state RTO approach. For example, several economic formulations of MPC (E-MPC) and nonlinear MPC (E-NMPC) have been reported [117] (Figure 23.11). A common thread is the fact that such controllers integrate

the economic cost function and constraints of the RTO layer as an additional soft constraint to the cost function of the MPC [118]. In fact the implementation strategy of E-MPC is very similar to the one of MPC. To date, E-MPC example studies have not yet been reported in the field of continuous oral solids manufacturing.

23.5.5 Management Level

Traditionally, information related to the process economic performance was rarely shared with the people who were actually operating the process. The other way around was also the case, i.e. insignificant diagnostics, alarms, and failures were not reported to the enterprise management. Moreover, reporting performance came with significant delays due to the weekly or monthly nature of the financial system. Today, such a business approach is unacceptable. Managing

a manufacturing process is much more real time. Therefore, business intelligence must be just as dynamic.

The evolution of information systems and industrial automation have made huge amounts of data available for analysis that could improve the overall process performance. Providing the proper information to everyone in the organization is today's challenge [84]. Commercial software packages exist to bridge the gap between management and operations. For example, real-time performance measures can be used to calculate key performance indices (KPIs) directly from process sensors and analyzers; manufacturing execution systems (MES) schedule production orders while keeping track of material and energy balances; enterprise resource planning (ERP) keeps track of all resources (financial, staff, raw materials, finished products, customer, etc.) within an enterprise. A more elaborate overview can be found elsewhere [119–121].

23.6 CONCLUSIONS AND PERSPECTIVES

Continuous manufacturing of oral solid dosage forms is gaining momentum in the pharmaceutical industry. Unlike traditional manufacturing, characterized by intermediate product analysis, fully integrated continuous processes require alternative strategies to guarantee the final drug product quality specifications. As a response, PAT and its underlying systems engineering tools are increasingly being adopted by industry, regulators, and academia.

This chapter focuses on the different aspects related to process control. The aim is to provide the reader with a clear overview of the different levels of complexity involved in an overall process control system while avoiding mathematical detail. Abundant literature references are provided to illustrate the state of the art in this field. In our opinion there is much needed to provide clear insights into the benefits offered by process control as well as the alignment of stakeholders' expectations.

REFERENCES

- Ess, S.M., Schneeweiss, S., and Szucs, T.D. (2003). European healthcare policies for controlling drug expenditure. *Pharmaco Economics* 21 (2): 89–103.
- Current and Resolved Drug Shortages and Discontinuations Reported to FDA (February 2018). <https://www.accessdata.fda.gov/scripts/drugshortages/default.cfm> (accessed 9 November 2018).
- Mukherjee, S. (January 2017). Why some of big pharma's best-selling drugs are in trouble this year. <http://fortune.com/2017/01/19/%20pharma-patents-expiration> (accessed 9 November 2018).
- Laske, S., Paudel, A., and Scheibelhofer, O. (2017). A review of PAT strategies in secondary solid oral dosage manufacturing of small molecules. *Journal of Pharmaceutical Sciences* 106 (3): 667–712.
- Lee, S.L., O'Connor, T.F., Yang, X. et al. (2015). Modernizing pharmaceutical manufacturing: from batch to continuous production. *Journal of Pharmaceutical Innovation* 10 (3): 191–199.
- Gaffney, A. (2014). Number of drug recalls surges at FDA, led by mid-level concerns. <https://www.raps.org/regulatory-focus%E2%84%A2/news-articles/2014/8/number-of-drug-recalls-surges-at-fda,-led-by-mid-level-concerns> (accessed 9 November 2018).
- Koba, M. (2015). Drug shortage in the U.S. May Pose deadly problem for patients. *Time* (January 2015). <http://www.time.com/3655660/usdrug-shortage>(accessed 9 November 2018).
- Steiner, R. and Jornitz, M. (2017). *Continuous Processing in the Pharmaceutical Industry: Status and Perspective, Continuous Manufacturing of Pharmaceuticals*. Wiley.
- Padmanabhan, B. (2017). True continuous manufacturing, pharmaceutical manufacturing. <https://www.pharmamanufacturing.com/articles/2017/true-continuous-manufacturing> (accessed 9 November 2018).
- Srai, J.S., Badman, C., Krumme, M. et al. (2015). Future supply chains enabled by continuous processing – opportunities challenges May 20–21 2014, continuous manufacturing symposium. *Journal of Pharmaceutical Sciences* 104 (3): 840–849.
- Vervae, C. and Remon, J.-P. (2016). Chapter 13. In: *Handbook of Pharmaceutical Granulation Technology*, 3e. CRC Press.
- FDA (2004). *Guidance for Industry: PAT—A Framework for Innovative Pharmaceutical Development, Manufacturing and Quality Assurance*. FDA.
- Rathore, A.S., Bhambure, R., and Ghare, V. (2010). Process analytical technology (PAT) for biopharmaceutical products. *Analytical and Bioanalytical Chemistry* 398 (1): 137–154.
- Yu, L.X., Amidon, G., Khan, M.A. et al. (2014). Understanding pharmaceutical quality by design. *The AAPS Journal* 16 (4): 771–783.
- Murphy, T., O' Mahony, N., Panduru, K., et al. (2016). Pharmaceutical manufacturing and the quality by design (QBD), process analytical technology (PAT) approach. 2016 27th Irish Signals and Systems Conference (ISSC) (21–22 June 2016), Londonderry, pp. 1–7
- Myerson, A.S., Krumme, M., Nasr, M. et al. (2015). Control systems engineering in continuous pharmaceutical manufacturing May 20–21, 2014, continuous manufacturing symposium. *Journal of Pharmaceutical Sciences* 104 (3): 832–839.
- Rantanen, J. and Khinast, J. (2015). The future of pharmaceutical manufacturing sciences. *Journal of Pharmaceutical Sciences* 104 (11): 3612–3638.
- Wade, H.L. (2004). *Basic and Advanced Regulatory Control: System Design and Application*, 2e. ISA.
- International Conference on Harmonisation of Technical Requirements of Registration of Pharmaceuticals for Human Use (2009). *Pharmaceutical Development Q8 (R2)*.
- Ishikawa, K. (1986). *Guide to Quality Control – Revised Edition*. Asian Productivity Organization.
- Hlinak, A.J., Kuriyan, K., Morris, K.R. et al. (2006). Understanding critical material properties for solid dosage form design. *Journal of Pharmaceutical Innovation* 1 (1): 12–17.

22. El Hagrasy, A.S., Hennenkamp, J.R., Burke, M.D. et al. (2013). Twin screw wet granulation: influence of formulation parameters on granule properties and growth behavior. *Powder Technology* 238: 108–115.
23. Fonteyne, M., Correia, A., De Plecker, S. et al. (2015). Impact of microcrystalline cellulose material attributes: a case study on continuous twin screw granulation. *International Journal of Pharmaceutics* 478 (2): 705–717.
24. García Muñoz, S., Padovani, V., and Mercado, J. (2014). A computer aided optimal inventory selection system for continuous quality improvement in drug product manufacture. *Computers & Chemical Engineering* 60: 396–402.
25. Engisch, W.E. and Muzzio, F.J. (2015). Feedrate deviations caused by hopper refill of loss-in-weight feeders. *Powder Technology* 283: 389–400.
26. Peeters, E., De Beer, T., Vervaet, C., and Remon, J.-P. (2015). Reduction of tablet weight variability by optimizing paddle speed in the forced feeder of a high-speed rotary tablet press. *Drug Development and Industrial Pharmacy* 41 (4): 530–539.
27. Vercruysse, J., Toiviainen, M., Fonteyne, M. et al. (2014). Visualization and understanding of the granulation liquid mixing and distribution during continuous twin screw granulation using NIR chemical imaging. *European Journal of Pharmaceutics and Biopharmaceutics* 86 (3): 383–392.
28. Seferlis, P. and Georgiadis, M.C. (2004). *The Integration of Process Design and Control*, 1e, vol. 17. Elsevier Science.
29. Vercruysse, J., Delaet, U., Van Assche, I. et al. (2013). Stability and repeatability of a continuous twin screw granulation and drying system. *European Journal of Pharmaceutics and Biopharmaceutics* 85 (3 Part B): 1031–1038.
30. Saleh, M.F., Dhenge, R.M., Cartwright, J.J. et al. (2015). Twin screw wet granulation: effect of process and formulation variables on powder caking during production. *International Journal of Pharmaceutics* 496 (2): 571–582.
31. Scheibelhofer, O., Hohl, R., Salar-Behzadi, S. et al. (2014). Automatic correction for window fouling of near infrared probes in fluidised systems. *Journal of Near Infrared Spectroscopy* 22 (3): 229–238.
32. Matthews, G., Hancock, P.A., and Desmond, P.A. (2017). *The Handbook of Operator Fatigue*. CRC Press.
33. Staniforth, J.N. (1982). Relationship between vibration produced during powder handling and segregation of pharmaceutical powder mixes. *International Journal of Pharmaceutics* 12: 199–207.
34. International Conference on Harmonisation of Technical Requirements of Registration of Pharmaceuticals for Human Use (2005). Quality Risk Management Q9.
35. Mikulak, R.J., McDermott, R., and Beauregard, M. (2008). *The Basics of FMEA*, 2e. Productivity Press.
36. Skogestad, S. (2004). Control structure design for complete chemical plants. *Computers & Chemical Engineering* 28 (1–2): 219–234.
37. Hale, G. and Martin, P.G. (2009). *Automation Made Easy: Everything You Wanted to Know About Automation – And Need to Ask*. ISA.
38. Dunn, W.C. (2005). *Introduction to Instrumentation, Sensors and Process Control*. Artech House Publishers.
39. Visioli, A. (2006). *Practical PID Control*. London: Springer-Verlag.
40. Ang, K.H., Chong, G., and Li, Y. (2005). PID control system analysis, design, and technology. *IEEE Transactions on Control Systems Technology* 13 (4): 559–576.
41. Nordh, P. (2016). Understand advanced process control. *Chemical Engineering Progress* 112 (6): 69–72.
42. Lakerveld, R., Benyahia, B., Braatz, R.D., and Barton, P.I. (2013). Model-based design of a plant-wide control strategy for a continuous pharmaceutical plant. *AIChE Journal* 59 (10): 3671–3685.
43. Tatjewski, P. (2007). *Advanced Control of Industrial Processes*. London: Springer-Verlag.
44. International Conference on Harmonisation of Technical Requirements of Registration of Pharmaceuticals for Human Use (2008). Pharmaceutical Quality System Q10.
45. Lepore, J. and Spavins, J. (2008). PQLI design space. *Journal of Pharmaceutical Innovation* 3 (2): 79–87.
46. Yu, L.X. (2008). Pharmaceutical quality by design: product and process development, understanding, and control. *Pharmaceutical Research* 25 (4): 781–791.
47. De Ketelaere, B., Hubert, M., and Schmitt, E. (2015). Overview of PCA-based statistical process-monitoring methods for time-dependent, high-dimensional data. *Journal of Quality Technology* 47 (4): 318–335.
48. Almaya, A., De Belder, L., Meyer, R. et al. (2017). Control strategies for drug product continuous direct compression – state of control, product collection strategies, and startup/shutdown operations for the production of clinical trial materials and commercial products. *Journal of Pharmaceutical Sciences* 106 (4): 930–943.
49. MacGregor, J.F., Yu, H., García Muñoz, S., and Flores-Cerrillo, J. (2005). Data-based latent variable methods for process analysis, monitoring and control. *Computers & Chemical Engineering* 29 (6): 1217–1223.
50. Dunn, K. (2015). *Process Improvement Using Data*, version: 294-34b8.
51. Silva, A.F., Sarraguça, M.C., Fonteyne, M. et al. (2017). Multivariate statistical process control of a continuous pharmaceutical twin-screw granulation and fluid bed drying process. *International Journal of Pharmaceutics* 528 (1–2): 242–252.
52. Matero, S., v. Den Berg, F., Poutiainen, S. et al. (2013). Towards better process understanding: chemometrics and multivariate measurements in manufacturing of solid dosage forms. *Journal of Pharmaceutical Sciences* 102 (5): 1385–1403.
53. Chen, Z., Lovett, D., and Morris, J. (2011). Process analytical technologies and real time process control a review of some spectroscopic issues and challenges. *Journal of Process Control* 21 (10): 1467–1482.
54. Goodwin, D.J., van den Ban, S., Denham, M., and Barylski, I. (2018). Real time release testing of tablet content and content uniformity. *International Journal of Pharmaceutics* 537 (1–2): 183–192.
55. Burggraeve, A., Van Den Kerkhof, T., Hellings, M. et al. (2011). Batch statistical process control of a fluid bed granulation process using in-line spatial filter velocimetry and product

- temperature measurements. *European Journal of Pharmaceutical Sciences* 42 (5): 584–592.
56. Fortuna, L., Graziani, S., Rizzo, A., and Xibilia, M.G. (2007). *Soft Sensors for Monitoring and Control of Industrial Processes*. London: Springer-Verlag.
 57. Lin, B., Recke, B., Renaudat, P. et al. (2005). A systematic approach for soft sensor development. In: *Computer Aided Chemical Engineering*, vol. 20 (ed. L. Puigjaner and A. Espuña), 1147–1152. Elsevier.
 58. Rehl, J., Karttunen, A.-P., Nicolai, N. et al. (2018). Control of three different continuous pharmaceutical manufacturing processes: use of soft sensors. *Journal of Pharmaceutical Innovation* <https://doi.org/10.1016/j.ijpharm.2018.03.027>.
 59. Fonteyne, M., Gildemyn, D., Peeters, E. et al. (2014). Moisture and drug solid-state monitoring during a continuous drying process using empirical and mass balance models. *European Journal of Pharmaceutics and Biopharmaceutics* 87 (3): 616–628.
 60. Oksanen, C.A. and García Muñoz, S. (2010). Process modeling and control in drug development and manufacturing. *Computers & Chemical Engineering* 34 (7): 1007–1008.
 61. Montgomery, D.C. (2009). *Introduction to Statistical Quality Control*, 6e. Wiley.
 62. Burggraeve, A., Silva, A.F.T., Van Den Kerkhof, T. et al. (2012). Development of a fluid bed granulation process control strategy based on real-time process and product measurements. *Talanta* 100: 293–302.
 63. Jaeckle, C.M. and MacGregor, J.F. (2000). Industrial applications of product design through the inversion of latent variable models. *Chemometrics and Intelligent Laboratory Systems* 50 (2): 199–210.
 64. Tomba, E., Facco, P., Bezzo, F., and Barolo, M. (2013). Latent variable modeling to assist the implementation of quality-by-design paradigms in pharmaceutical development and manufacturing: a review. *International Journal of Pharmaceutics* 457 (1): 283–297.
 65. García-Muñoz, S. and Mercado, J. (2013). Optimal selection of raw materials for pharmaceutical drug product design and manufacture using mixed integer nonlinear programming and multivariate latent variable regression models. *Industrial & Engineering Chemistry Research* 52 (17): 5934–5942.
 66. Muteki, K., Swaminathan, V., Sekulic, S.S., and Reid, G.L. (2012). Feed-forward process control strategy for pharmaceutical tablet manufacture using latent variable modeling and optimization technologies. 8th IFAC Symposium on Advanced Control of Chemical Processes (10–13 July 2012), Furama Riverfront, Singapore.
 67. Montague, G.A., Glassey, J., and Willis, M.J. (2003). French fry quality improvement using advanced control techniques. *Journal of Food Engineering* 57 (4): 357–365.
 68. Nagy, B., Farkas, A., Gyürkés, M. et al. (2017). In-line Raman spectroscopic monitoring and feedback control of a continuous twin-screw pharmaceutical powder blending and tableting process. *International Journal of Pharmaceutics* 530 (1–2): 21–29.
 69. Zhao, X.J., Gatamel, C., Dirion, J.L., et al. (2013). Implementation of a control loop for a continuous powder mixing process. AIChE Annual Meeting, San Francisco (6 November 2013).
 70. Correa, E.J. (2011). PAT particle size analyzer (PSA) application for API milling manufacturing process. ISPE Great Lakes Chapter 2011 Meeting. Doi:10.1021/acs.oprd.5b00088.
 71. Baronsky-Probst, J., Möltgen, C.-V., Kessler, W., and Kessler, R.W. (2016). Process design and control of a twin screw hot melt extrusion for continuous pharmaceutical tamper-resistant tablet production. *European Journal of Pharmaceutical Sciences* 87: 14–21.
 72. Moshgbar, M. and Hammond, S. (2010). Advanced process control. *Quality Assurance Journal* 13 (3–4): 62–66.
 73. Wang, Q.-G., Ye, Z., Hang, C.-C., and Cai, W.-J. (2008). *PID Control for Multivariable Processes*. Berlin Heidelberg: Springer-Verlag.
 74. Lakerveld, R., Benyahia, B., Heider, P.L. et al. (2015). The application of an automated control strategy for an integrated continuous pharmaceutical pilot plant. *Organic Process Research & Development* 19 (9): 1088–1100.
 75. Ierapetritou, M. and Ramachandran, R.G. (2016). *Process Simulation and Data Modeling in Solid Oral Drug Development and Manufacture*. Humana Press.
 76. Rehl, J., Krusz, J., Sacher, S. et al. (2016). Optimized continuous pharmaceutical manufacturing via model-predictive control. *International Journal of Pharmaceutics* 510 (1): 100–115.
 77. Ramachandran, R., Arjunan, J., Chaudhury, A., and Ierapetritou, M.G. (2011). Model-based control-loop performance of a continuous direct compaction process. *Journal of Pharmaceutical Innovation* 6 (4): 249–263.
 78. Ramachandran, R. and Chaudhury, A. (2012). Model-based design and control of a continuous drum granulation process. *Chemical Engineering Research and Design* 90 (8): 1063–1073.
 79. Singh, R., Ierapetritou, M., and Ramachandran, R. (2012). An engineering study on the enhanced control and operation of continuous manufacturing of pharmaceutical tablets via roller compaction. *International Journal of Pharmaceutics* 438 (1–2): 307–326.
 80. Skogestad, S. and Postlethwaite, I. (2005). *Multivariable Feedback Control: Analysis and Design*, 2e. Wiley.
 81. ABB (February 2018). Advanced process control in biopharmaceutical production. <http://new.abb.com/control-systems/industry-specific-solutions/pharmaceutical-and-life-sciences/apc-mpc-model-predictive-control-for-biopharmaceutical>
 82. Camacho, E.F. and Bordons Alba, C. (2007). *Model Predictive Control*. London: Springer-Verlag.
 83. Valencia, A.F. (2005). Modeling and control of continuous tumble mixing of granular materials. Master thesis. University of Puerto Rico.
 84. Hsu, S.-H., Reklaitis, G.V., and Venkatasubramania, V. (2010). Modeling and control of roller compaction for pharmaceutical manufacturing – part II: control system design. *Journal of Pharmaceutical Innovation* 5 (1–2): 24–36.
 85. Singh, R., Ierapetritou, M., and Ramachandran, R. (2013). System-wide hybrid MPC–PID control of a continuous pharmaceutical tablet manufacturing process via direct compaction. *European Journal of Pharmaceutics and Biopharmaceutics* 85 (3 Part B): 1164–1182.

86. Mesbah, A., Ford Versypt, A.N., Zhu, X., and Braatz, R.D. (2014). Nonlinear model-based control of thin-film drying for continuous pharmaceutical manufacturing. *Industrial & Engineering Chemistry Research* 53 (18): 7447–7460.
87. Sen, M., Singh, R., and Ramachandran, R. (2014). A hybrid MPC-PID control system design for the continuous purification and processing of active pharmaceutical ingredients. *Processes* 2 (2): 392–418.
88. Mesbah, A., Paulson, J.A., Lakerveld, R., and Braatz, R.D. (2017). Model predictive control of an integrated continuous pharmaceutical manufacturing pilot plant. *Organic Process Research & Development* 21 (6): 844–854.
89. Haas, N.T., Ierapetritou, M., and Singh, R. (2017). Advanced model predictive feedforward/feedback control of a tablet press. *Journal of Pharmaceutical Innovation* 12 (2): 110–123.
90. Su, Q., Moreno, M., Giridhar, A. et al. (2017). A systematic framework for process control design and risk analysis in continuous pharmaceutical solid-dosage manufacturing. *Journal of Pharmaceutical Innovation* 12 (4): 327–346.
91. Andersson, M., Svensson, O., Folestad, S. et al. (2005). NIR spectroscopy on moving solids using a scanning grating spectrometer – impact on multivariate process analysis. *Chemometrics and Intelligent Laboratory Systems* 75 (1): 1–11.
92. Fonteyne, M., Vercruyse, J., De Leersnyder, F. et al. (2016). Blend uniformity evaluation during continuous mixing in a twin screw granulator by in-line NIR using a moving F-test. *Analytica Chimica Acta* 935: 213–223.
93. Germer, K., Wolf, B., and Eckardt, G. (2011). Influence of the installed in-line spatial filter velocimetry (SFV) probe on the fluidized bed stability, Parsum.
94. Lothian, J., Hamilton, P., and O'Callaghan, C. (February 2018). Real time monitoring of twin screw granulation. <https://www.innopharmalabs.com/tech/applications-and-processes/real-time-monitoring-twin-screw-granulation> (accessed 9 November 2018).
95. Daniel, B. (2013). Application of an advanced process controller to a continuous mixing, direct compression process, AIChE Annual Meeting (6 November 2013), San Francisco, CA.
96. Singh, R., Sahay, A., Karry, K.M. et al. (2014). Implementation of an advanced hybrid MPC-PID control system using PAT tools into a direct compaction continuous pharmaceutical tablet manufacturing pilot plant. *International Journal of Pharmaceutics* 473 (1–2): 38–54.
97. Bhaskar, A., Barros, F.N., and Singh, R. (2017). Development and implementation of an advanced model predictive control system into continuous pharmaceutical tablet compaction process. *International Journal of Pharmaceutics* 534 (1–2): 159–178.
98. de Barros, F.N., Bhaskar, A., and Singh, R. (2017). A validated model for design and evaluation of control architectures for a continuous tablet compaction process. *Processes* 5 (4): 76.
99. Nicolai, N., De Leersnyder, F., Copot, D. et al. (2018). Liquid-to-solid ratio control as an advanced process control solution for continuous twin-screw wet granulation. *AIChE Journal* 64 (7): 2500–2514.
100. Gao, Y., Muzzio, F.J., and Ierapetritou, M.G. (2012). A review of the residence time distribution (RTD) applications in solid unit operations. *Powder Technology* 228: 416–423.
101. Scott Fogler, H. (2006). *Elements of Chemical Reaction Engineering*. Prentice Hall PTR.
102. Levenspiel, O. (1998). *Chemical Reaction Engineering*. Wiley.
103. Gao, Y., Vanarase, A., Muzzio, F., and Ierapetritou, M. (2011). Characterizing continuous powder mixing using residence time distribution. *Chemical Engineering Science* 66 (3): 417–425.
104. Tian, G., Lee, S.L., Yang, X. et al. (2017). A dimensionless analysis of residence time distributions for continuous powder mixing. *Powder Technology* 315: 332–338.
105. Engisch, W. and Muzzio, F. (2016). *Journal of Pharmaceutical Innovation* 11: 64.
106. Van Snick, B., Holman, J., Vanhoorne, V. et al. (2017). Development of a continuous direct compression platform for low-dose drug products. *International Journal of Pharmaceutics* 529 (1–2): 329–346.
107. Kumar, A., Vercruyse, J., Toivainen, M. et al. (2014). Mixing and transport during pharmaceutical twin-screw wet granulation: experimental analysis via chemical imaging. *European Journal of Pharmaceutics and Biopharmaceutics* 87 (2): 279–289.
108. Mangal, H. and Kleinebudde, P. (2017). Experimental determination of residence time distribution in continuous dry granulation. *International Journal of Pharmaceutics* 524 (1–2): 91–100.
109. Kruijs, J., Rehl, J., Sacher, S. et al. (2017). RTD modeling of a continuous dry granulation process for process control and materials diversion. *International Journal of Pharmaceutics* 528 (1–2): 334–344.
110. Rehbaum, H. and Groß-Weege, C. (2016). Automatisierung kontinuierlicher Herstellungsprozesse – Herausforderung bei der kontinuierlichen Produktion pharmazeutischer Feststoffe. *TechnoPharm* 6 (6): 336–345.
111. Markl, D., Wahl, P.R., Menezes, J.C. et al. (2013). Supervisory control system for monitoring a pharmaceutical hot melt extrusion process. *AAPS PharmSciTech* 14 (3): 1034–1044.
112. Singh, R., Sahay, A., Muzzio, F. et al. (2014). A systematic framework for onsite design and implementation of a control system in a continuous tablet manufacturing process. *Computers & Chemical Engineering* 66: 186–200.
113. Rehbaum, H. (2017). Successful introduction of PAT requires a strategy, *Pharmaceutical Processing*, Interphex Supplement, pp. 12–17.
114. Petersen, L.N., Poulsen, N.K., Niemann, H.H. et al. (2017). Comparison of three control strategies for optimization of spray dryer operation. *Journal of Process Control* 57: 1–14.
115. Gagnon, F., Desbiens, A., Poulin, É. et al. (2017). Nonlinear model predictive control of a batch fluidized bed dryer for pharmaceutical particles. *Control Engineering Practice* 64: 88–101.

116. Singh, R., Sen, M., Ierapetritou, M., and Ramachandran, R. (2015). Integrated moving horizon-based dynamic real-time optimization and hybrid MPC-PID control of a direct compaction continuous tablet manufacturing process. *Journal of Pharmaceutical Innovation* 10 (3): 233–253.
117. Ellis, M., Durand, H., and Christofides, P.D. (2014). A tutorial review of economic model predictive control methods. *Journal of Process Control* 24 (8): 1156–1178.
118. Godoy, J.L., González, A.H., and Normey-Rico, J.E. (2016). Constrained latent variable model predictive control for trajectory tracking and economic optimization in batch processes. *Journal of Process Control* 45: 1–11.
119. Grossmann, I. (2005). Enterprise-wide optimization: a new frontier in process systems engineering. *AIChE Journal* 51 (7): 1846–1857.
120. Kletti, J. (2007). *Manufacturing Execution System – MES*. Berlin Heidelberg: Springer-Verlag.
121. Vallespir, B. and Alix, T. (2009). Advances in production management systems: new challenges, new approaches. International IFIP WG 5.7 Conference APMS, Bordeaux, France (21–23 September 2009). Springer-Verlag Berlin Heidelberg.

PART IV

APPLIED STATISTICS AND REGULATORY ENVIRONMENT

MULTIVARIATE ANALYSIS FOR PHARMACEUTICAL AND MEDICAL DEVICE DEVELOPMENT

FREDERICK H. LONG

Spectroscopic Solutions, LLC, Randolph, NJ, USA

Multivariate analysis (MVA) is the statistical analysis of many variables at once. Many problems in the pharmaceutical industry are multivariate in nature. The importance of MVA has been recognized by the US FDA in the recent guidance on process analytical technology [1]. MVA has been made much easier with the development of inexpensive, fast computers and powerful analytical software. Chemometrics is the statistical analysis of chemical data, which is an important area of MVA. Spectral data from modern instruments is fundamentally multivariate in character. Typically pharmaceutical process monitoring requires more than one variable. Furthermore, the powerful statistical methods of chemometrics are essential for the analysis and application of spectral data including near IR (NIR) and Raman. In this chapter, we will review the subject of chemometrics and MVA and its application in the pharmaceutical industry.

With spectral data, it is not uncommon to measure several thousand variables at one time. However, it is often hard to conceptualize so many variables; therefore, we will begin our discussion of MVA with a few simple examples that illustrate important statistical concepts that are essential in chemometrics. The first problem is a set of pharmaceutical quality data. Measurements of density and assay have been measured for 43 lots of material. The data is shown in Table 24.1. Inspection of the data reveals that the density values are near 1.0, while the assay values are closer to 100. A goal of the data analysis is to understand the variation within the data set. It will be advantageous to have the two variables in the data set with similar magnitudes; therefore we will scale each of the two variables by its own standard

deviation. The standard deviation, σ , of a set of measurements (x_1, \dots, x_n) is given by

$$\sigma = \left(\frac{\sum (x_i - \bar{x})^2}{n-1} \right)^{1/2} \quad (24.1)$$

where

\bar{x} is the average value of the n measurements.

The denominator in equation one is $n-1$, because once the average is calculated, there are $n-1$ degrees of freedom. We note that the standard deviation has the same units as the variable of interest.

A plot of the scaled data is shown in Figure 24.1. The x -axis is the scaled density and the y -axis is the scaled assay values. Each point represents one of the 43 lots of material. From the plot in Figure 24.1, one data point is far away from all of the others. Statisticians call data points that do not belong to the data set outliers. Outliers are important to identify and remove from the analysis of the data set, because a single outlier can greatly influence the statistical analysis and obscure underlying trends in the data. We note that while outliers are often removed in a research and development environment during method development, great caution must be used in removing outliers during validation or use in actual production.

The scaled data are replotted in Figure 24.2, with the outlier point removed. The reader will also note that the origin of the graph has been moved to the center point of the data set.

TABLE 24.1 Pharmaceutical Quality Data Example

Density (g/cm ³)	Assay (mg)
0.801	121.410
0.824	127.700
0.841	129.200
0.816	131.800
0.840	135.100
0.842	131.500
0.820	126.700
0.802	115.100
0.828	130.800
0.819	124.600
0.826	118.310
0.802	114.200
0.810	120.300
0.802	115.700
0.832	117.510
0.796	109.810
0.759	109.100
0.770	115.100
0.759	118.310
0.772	112.600
0.806	116.200
0.803	118.000
0.845	131.000
0.822	125.700
0.971	126.100
0.816	125.800
0.836	125.500
0.815	127.800
0.822	130.500
0.822	127.900
0.843	123.900
0.824	124.100
0.788	120.800
0.782	107.400
0.795	120.700
0.805	121.910
0.836	122.310
0.788	110.600
0.772	103.510
0.776	110.710
0.758	113.800

This operation is called *mean centering*, when the average of the overall data set is subtracted from the data. As mentioned earlier, in MVA we are concerned with investigation of the variation within the data set. The average values of the data set are not of primary importance. Two arrows in the figure illustrate the two directions of variation within the data set. PC1 is the largest direction of variation and PC2 is the second direction of variation. It is important to note that PC1 and PC2 are perpendicular to each other. In MVA, PC1 and PC2 are the first and second principal components (PCs) of the data set, respectively.

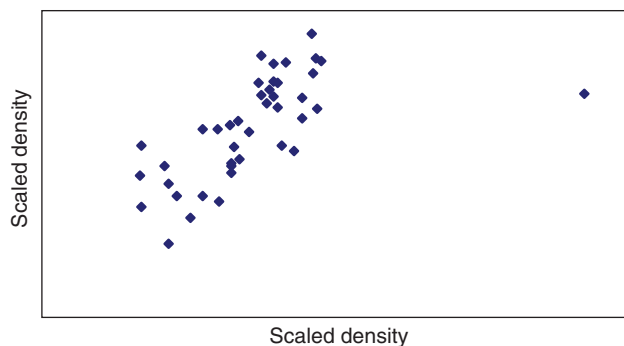


FIGURE 24.1 Scaled pharmaceutical quality data. Both the density and assay are scaled by the standard deviation of the data for each variable. Because the variables are scaled by the standard deviation, they are dimensionless.

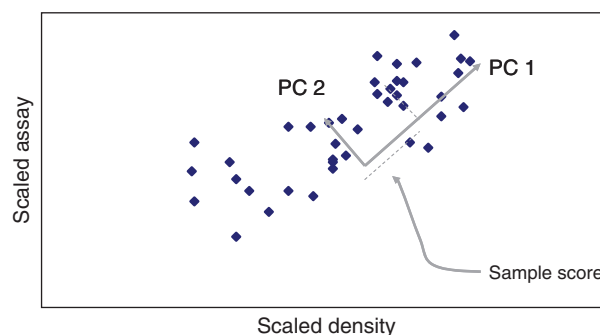


FIGURE 24.2 Scaled pharmaceutical quality data showing both the first and second principal components (PCs) for the data set. The first PC is the direction of the maximum variation within the data set. The second PC is perpendicular to the first PC. The scores for each sample point are given by the projection of the data point onto the PC vector.

For each one of the data points, the projection of the data point onto the PC1 or PC2 vector is called a score value. Plots of score values for different PCs, typically PC1 versus PC2, are called score plots. Score plots provide important information about how different samples are related to each other. PC plots, also called loading plots, provide information about how different variables are related to each other. Because we are working with scaled variables, the PCs and scores are dimensionless variables.

The mathematics of principal component analysis (PCA) can be clearly described using linear algebra [2]. An excellent discussion of linear algebra can be found in the references [3]. By convention, the data matrix, \mathbf{X} , has p columns and n rows, and each column represents another variable and new rows for each observation or sample. The average data matrix, $\bar{\mathbf{X}}$, is the average of each individual column (i.e. variable) in the data set. Mean centering is written as

$$(\mathbf{X} - \bar{\mathbf{X}}) \tag{24.2}$$

The covariance matrix is written as

$$\mathbf{C} = (\mathbf{X} - \bar{\mathbf{X}})^T (\mathbf{X} - \bar{\mathbf{X}}) \tag{24.3}$$

where

an superscript **T** represents a matrix transpose.

The covariance matrix is a square symmetric $p \times p$ matrix. It provides information about the relationship between different variables. For example, the ij element of the covariance matrix quantifies the relative change between the ij variables. If an element of the covariance matrix is zero, there is no relationship (correlation) between the two variables.

Related to the covariance matrix is the correlation matrix where all the variables have been scaled their standard deviations. The correlation matrix is useful when one or more of the variables have much higher numerical values than the other variables. The scaling of the variables means that all variables will contribute to the analysis in roughly the same way. Mathematically the correlation matrix, **R**, is written as

$$r_{ik} = \frac{s_{ik}}{\sqrt{s_{ii}\sqrt{s_{jj}}} = \frac{\sum_{i=1}^n (x_{ji} - \bar{x}_i)(x_{jk} - \bar{x}_k)}{\sqrt{(x_{ji} - \bar{x}_i)^2} \sqrt{(x_{jk} - \bar{x}_k)^2}}$$

$$\mathbf{R} = \begin{pmatrix} 1 & r_{1p} \\ & 1 \\ r_{p1} & & 1 \end{pmatrix} \tag{24.4}$$

where the elements of the correlation matrix are given by r_{ij} .

R is a square $p \times p$ matrix, where p is the number of variables. The diagonal elements of **R** are equal to 1.

PCA is the systematic analysis of the covariance or correlation matrix. It can be shown that the eigenvalues are positive and the eigenvectors are orthogonal for both matrices [4]. The eigenvector equation for **C** is

$$\mathbf{C}u_i = \lambda_i u_i \tag{24.5}$$

where

u_i is the i th eigenvector.

λ_i is the corresponding eigenvalue.

By convention the eigenvalues are placed in descending order, where λ_1 is the largest eigenvalue. In PCA the eigenvectors are also called PCs. It can be shown that the first PC represents the largest source of variance in the data set. The % variation explained by the i th PC is given by

$$100 \times \frac{\lambda_1}{\sum_i \lambda_i} \tag{24.6}$$

It is common with spectral data that the data set can be well approximated by a few PCs. As explained earlier, score values provide information about the relationship between different observations. The PCs form a basis set that can be used to approximate the original data set. For a single mean-centered observation, x_j ,

$$x_j = \sum_{i=1}^p t_{ji} PC_i = \sum_{i=1}^A t_{ji} PC_i + E \tag{24.7}$$

where

t_{ji} are the score values.

A is the number of PCs.

E is the error when the number of PCs is less than the number of variables.

Because the PCs are orthogonal, a direct expression for the score values can be given by the following equation:

$$t_{j,i} = (\mathbf{x}_j - \bar{\mathbf{X}}) \cdot PC_i \tag{24.8}$$

Equation (24.7) is derivable from Eq. (24.6) by taking a dot product of both sides and exploiting the orthogonality of the PCs. The previous example is somewhat trivial because only two variables were involved.

Let us now consider another example with more variables. In Table 24.2, a set of data describing the properties of 43 raw materials is shown. The variables that describe the raw materials are labeled QV1–QV8. The variables QV1–QV8 describe different properties of the raw material such as moisture, assay, and particle size. Using commercial software, we can do a PCA of the data set using the same approach that was used for the first data set, i.e. scaling by standard deviation and mean centering. A few of the critical results are shown in the figures below. The loading (PC) plot shows some results that are clearly interpretable (Figure 24.3). The PC plot shows how different variables relate to each other. In the plot the reader can observe that QV5 and QV8 are close to each other and therefore are well correlated to each other. QV1 and QV7 are also correlated. A plot of the score values for each one of the 43 raw materials is shown in Figure 24.4. The origin of the score plot corresponds to the average of the entire data set. The samples that are farther away from the origin are more likely to be possible outliers. The ellipse in Figure 24.4 is called the Hotelling T^2 ellipse and is showing the 95% probability level for outliers. The Hotelling T^2 ellipse is based on scaled squared score values [2]. The T^2 value for observation i is given below:

TABLE 24.2 Multivariable Quality Data Set

Primary ID	QV1	QV2	QV3	QV4	QV5	QV6	QV7	QV8
1	110	2	2	180	1.5	10.5	10	70
2	110	6	2	290	2	17	1	105
3	110	1	1	180	0	12	13	55
4	110	1	1	180	0	12	13	65
5	110	1	1	280	0	15	9	45
6	110	3	1	250	1.5	11.5	10	90
7	110	2	1	260	0	21	3	40
8	110	2	1	180	0	12	12	55
9	100	2	1	220	2	15	6	90
10	130	3	2	170	1.5	13.5	10	120
11	100	3	2	140	2.5	8	140	m
12	110	2	1	200	0	21	3	35
13	140	3	1	190	4	15	14	230
14	100	3	1	200	3	16	3	110
15	110	1	1	140	0	13	12	25
16	100	3	1	200	3	17	3	110
17	110	2	1	200	1	16	8	60
18	70	4	1	260	9	7	5	320
19	110	2	0	125	1	11	14	30
20	100	2	0	290	1	21	2	35
21	110	1	0	90	1	13	12	20
22	110	3	3	140	4	10	7	160
23	110	2	0	220	1	21	3	30
24	110	2	1	125	1	11	13	30
25	110	1	0	200	1	14	11	25
26	100	3	0	0	3	14	7	100
27	120	3	0	240	5	14	12	190
28	110	2	1	170	1	17	6	60
29	160	3	2	150	3	17	13	160
30	120	2	1	190	0	15	9	40
31	140	3	2	220	3	21	7	130
32	90	3	0	170	3	18	2	90
33	100	3	0	320	1	20	3	45
34	120	3	1	210	5	14	12	240
35	110	2	0	290	0	22	3	35
36	110	2	1	70	1	9	15	40
37	110	6	0	230	1	16	3	55
38	120	1	2	220	0	12	12	35
39	120	1	2	220	1	12	11	45
40	100	4	2	150	2	12	6	95
41	50	1	0	0	0	13	0	15
42	50	2	0	0	1	10	0	50
43	100	5	2	0	2.7	1	1	110

$$S_{ia}^2 = \frac{\sum_{i=1}^N t_{ia}^2}{N} \quad (24.9)$$

$$S_{ia}^2 = \frac{\sum_{i=1}^N t_{ia}^2}{N}$$

where

A is the number of PCs.

t_{ia} is the a th PC score value for the i th sample.

S_{ia}^2 is the variance of t_a , because the average of the score values is zero [2].

T^2 is closely related to the often used parameter Mahalanobis distance. An important property of the T^2 statistic is that it is directly proportional to an F value, which is a statistical parameter that is rigorously related to a probability value.¹ The numerical value of the F value is dependent on the number of samples, PCs, and probability level desired, α . Examination of Eq. (24.9) for two PCs shows that

$$F(\alpha) = C \left(\frac{t_1^2}{S_1^2} + \frac{t_2^2}{S_2^2} \right) \quad (24.10)$$

where

C is a constant.

Equation (24.10) is an equation for an ellipse in the t_1, t_2 space. By convention, the Hotelling T^2 ellipse is usually drawn at the 95% probability level.

PCA can be viewed as a method for approximating the original data set. The approximation is based on a linear combination of the PCs where the amplitude coefficients are the previously described scores. The approximation is exact when the number of PCs equals the number of variables in the data set. For most spectral data sets, a small number of PCs (also called factors) can be used to approximate the spectral data set very well. The determination of the correct number of factors can be done by a variety of numerical methods. Too many factors in the PCA model will overfit the data, and the model will not predict reliably. Most MVA software packages will suggest a suitable number of PCs. The suggested number is usually a good starting point; however, it is best practice to verify the optimum number of PCs with additional independent test data.

Classification is an important application of chemometrics. It is the sorting of data into different groups. These groups can be quite diverse such as different sources or different quality grades of the same raw material. Chemometric methods for raw material identification using NIR or Raman spectra are important but are relatively simple and are discussed elsewhere [5]. In this chapter we will discuss soft independent modeling of class analogies (SIMCA) [6], a method for classification of similar classes using MVA. PCA score plots sometimes show data sets to consist of several subgroups. For example Figure 24.5 shows the score plot

¹ $T_i^2 \frac{(N-A)N}{A(N^2-1)}$ is approximately F-distributed (see Ref. [2]).

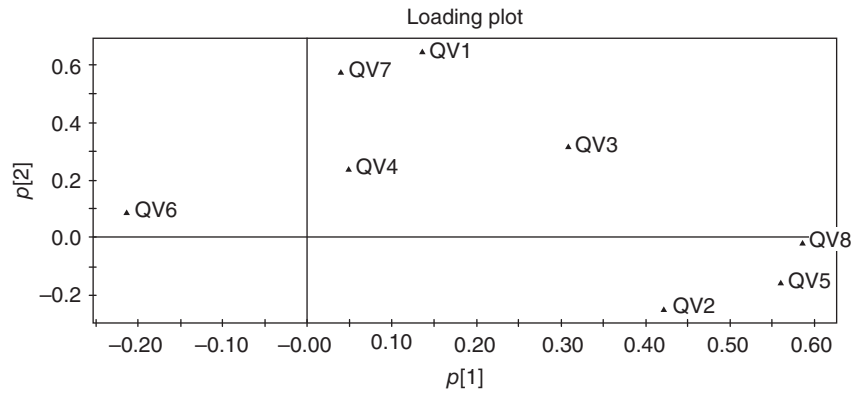


FIGURE 24.3 Loading plot for the data set in Table 24.2. The first principal component is plotted on the x-axis, and the second principal component is on the y-axis. Variables that are close to each other are highly correlated.

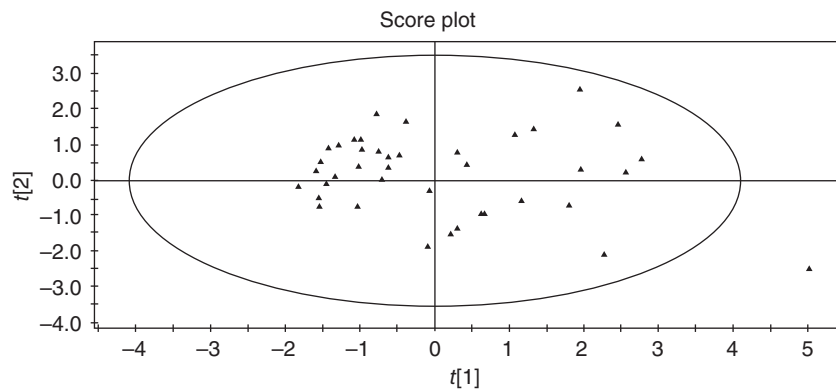


FIGURE 24.4 Score plot of the data set in Table 24.2. The ellipse is the Hotelling T^2 ellipse at 95% probability level. Samples outside the ellipse have a probability of greater than 95% of being statistical outliers.

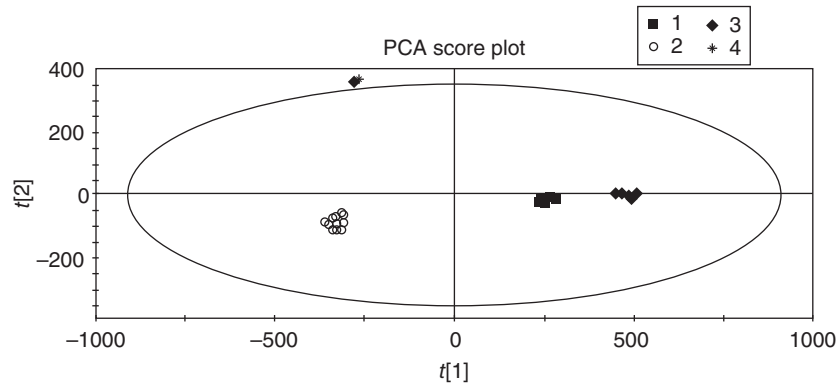


FIGURE 24.5 Score plot of mid-IR spectral data for a series of oils. *Group 1*, corn oil; *group 2*, olive oil; *group 3*, safflower oil; *group 4*, corn margarine.

for the mid-IR spectra for a series of oils. Grouping in the score plot clearly illustrates the differences between the four oils (olive, corn, safflower, and corn margarine).

SIMCA is designed to improve on this separation of classes by using the residuals from the PCA. Residuals are

the difference between the PCA model and the data. In the SIMCA analysis, a separate PCA model is built for each class in the training set. The average residual value for each class (S_0) is also calculated. Test or validation data are then fit to each PCA class model. The correct class is the class that

has the best fit to the PCA model. The comparison is quantified by the use of the scaled residual (s_i/S_0) (distance to model (DmodX)) values. The equations are given below:

$$s_i = \sqrt{\frac{\sum_{k=1}^K e_{ik}^2}{(K-A)}} \quad (24.11)$$

$$S_0 = \left(\frac{\sum_{i,j} e_{ij}^2}{(N-A-1)(K-A)} \right)^{1/2}$$

where

N is the number of samples.

A is the number of PCs.

K is the number of variables.

s_i is the root mean square residual value for the i th sample.

e_{ij} is the spectral residual, i.e. the difference between the spectra and the PCA model for observation i and variable j .

If the test sample residual is close to the average residual for the entire class, then the sample has a high probability of belonging to the class. The relationship to actual probability values is possible because the scaled residual values $(S_i/S_0)^2$ in Eq. (24.11) are described by an F-distribution. The results of an SIMCA analysis are often displayed in a Coomans plot. In a Coomans plot two classes are compared as shown in Figure 24.6.

A typical Coomans plot is shown below. PCA models for corn oil and olive oil are used to predict the classification of a set of test samples. The test samples include olive, corn, corn margarine, safflower, and walnut oils. The different classes are as shown in the legend. The x -axis on the Coomans plot

is the DmodX value for the corn oil; the y -axis is the same for olive oil. The red vertical line is the 5% probability level for the corn oil model; samples to the right of this line are probable outliers for the corn oil models. The red vertical line is the same for olive oil. Note most olive and corn oil test samples are correctly classified. Test samples from other classes are well separated from the oil and corn oil groups. At the end of this chapter, an application of SIMCA to medical device development is presented.

In many cases, spectral data requires mathematical transformations before MVA is performed [7]. The mathematical transformations are collectively referred to as spectral preprocessing. Derivative preprocessing is the most common form of spectral preprocessing with NIR spectra. Derivative preprocessing will eliminate or at least minimize the background variation associated with the NIR spectra of many pharmaceutical materials. The effects of a first derivative preprocessing on a typical NIR spectrum are shown in Figure 24.7 (top and bottom). The first derivative removes the slowly varying baseline typical of NIR spectra of powders; positive and negative peaks correspond to regions where the slope of the raw spectrum has a positive or negative value. There are several methods for the calculation of spectral derivatives; however, they all start with the definition of first or second derivative from elementary calculus:

$$f'(x) = \frac{f(x + \Delta x) - f(x)}{\Delta x}$$

$$f''(x) = \frac{f(x + 2\Delta x) - 2f(x + \Delta x) + f(x)}{(\Delta x)^2} \quad (24.12)$$

$$f''(x + \Delta x) \approx f''(x) + (\Delta x)f^{(3)}(x)$$

The two most common approaches for the calculation of derivatives are gap and Savitzky–Golay derivatives [8]. A gap derivative is based on the calculation of a running average for n points to the right and the left of a center point. The average values for the right- and left-hand side are then used

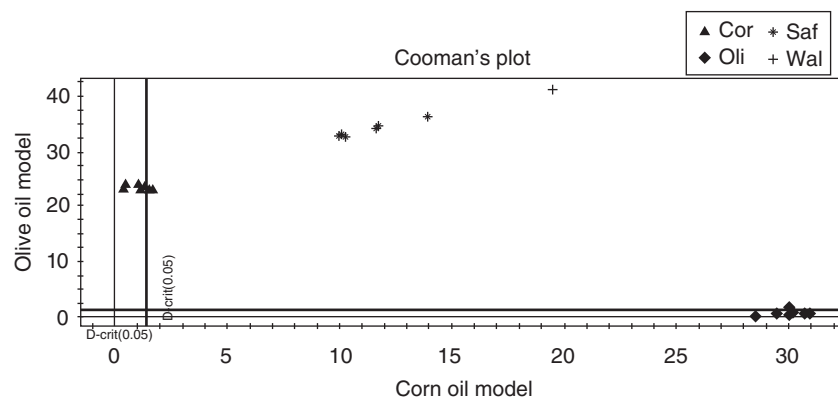


FIGURE 24.6 Coomans plot comparing the olive and corn oil classes using a test set. Triangle, corn oil; diamond, olive oil; asterisk, safflower oil; plus, walnut.

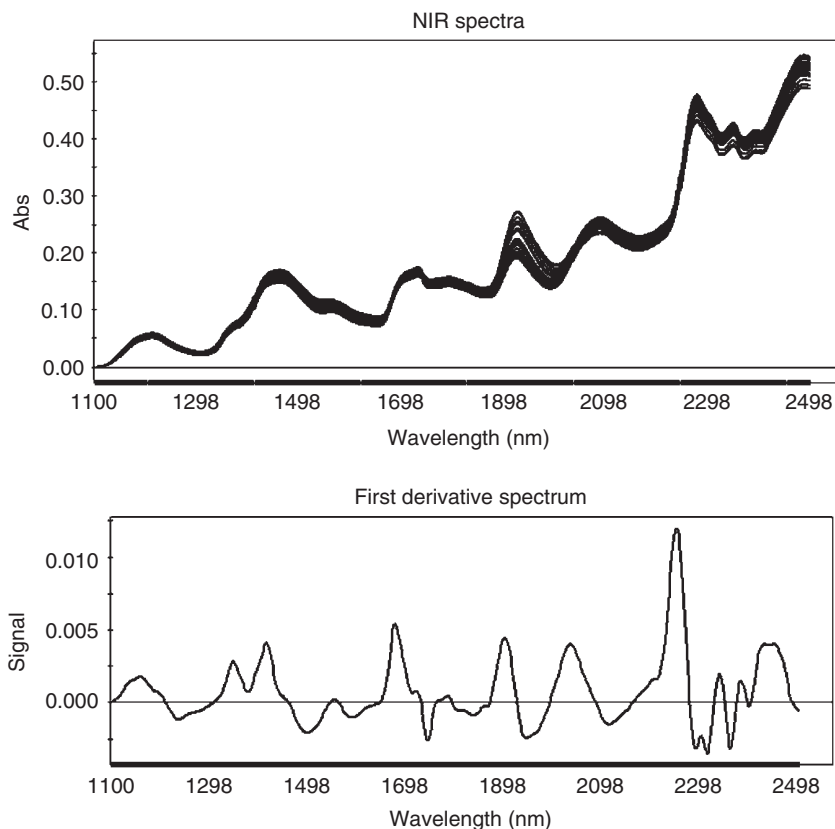


FIGURE 24.7 Effects of first derivative spectral preprocessing. (Top) Several raw NIR spectra. (Bottom) First derivative spectrum.

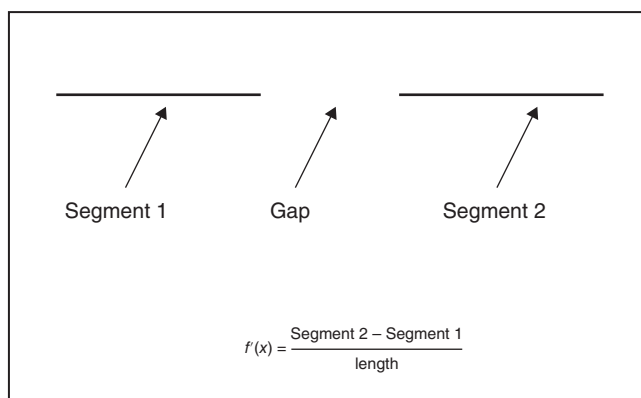


FIGURE 24.8 Illustration of gap derivative algorithm.

to calculate a finite difference derivative. There is an optional gap or separation between the right and left sides. Gap derivatives are described in Figure 24.8. A first-order gap derivative uses an n -point average to calculate a finite difference first derivative. Commonly the gap is set to zero in many applications.

Savitzky–Golay derivatives are based on fitting N points of the data to either a quadratic or cubic polynomial. The derivative is found by differentiation of the polynomial. For both methods of numerical differentiation, it is important to properly determine the number of points used in the

averaging. Too few points can compromise the signal to noise; too many points will filter out important high-frequency components of the data.

Standard normal variant (SNV) is a preprocessing method that is used to autoscale individual spectra [9]. The equation for SNV is given below:

$$\frac{x - \mu}{\sigma} \tag{24.13}$$

where

μ is the average value for the spectrum of interest.

σ is the standard deviation of the numbers that make up the spectrum.

During SNV preprocessing, the average value for each spectrum is subtracted, and then the spectrum is divided by the standard deviation for the sample spectrum. After SNV preprocessing, the range of each spectrum will be approximately -2 to 2 . SNV processing can be used to correct for laser intensity variation in Raman spectra and several kinds of path length variation in NIR spectra. The effects of SNV preprocessing are illustrated in Figure 24.9.

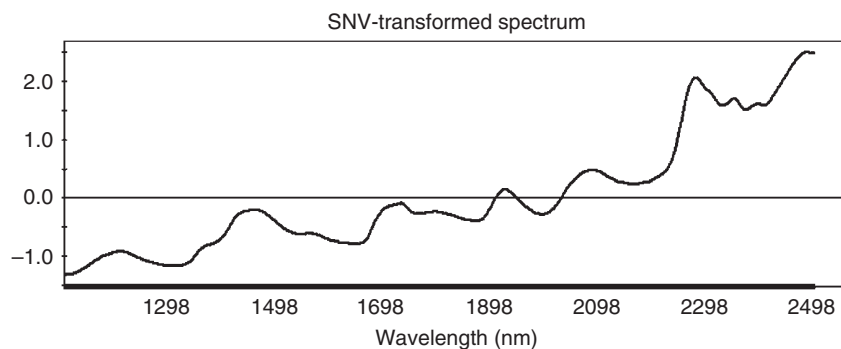


FIGURE 24.9 Effects of SNV preprocessing on spectrum from Figure 24.7.

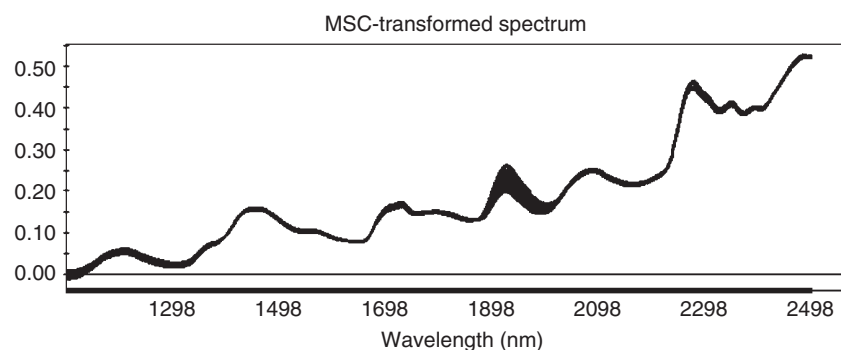


FIGURE 24.10 Effects of MSC preprocessing on spectrum from Figure 24.7. Note difference in y-axis from Figure 24.9.

Multiplicative scatter correction (MSC) is a preprocessing method designed to eliminate background variation in NIR spectra due to scattering [10]. The effects of MSC are similar to SNV in many real-world applications; however, it is a distinct method. MSC uses the average spectrum of the entire data set and not individual spectra. The sample spectra are then regressed against the average spectrum, producing slope and offset values at each wavelength for all samples in the data set. The slope and offset values are then used to correct the data set. Results of MSC preprocessing are illustrated in Figure 24.10. MSC preprocessing will remove the variation due to scattering in the data set but not change the average spectral value as SNV preprocessing does. The equations for MSC preprocessing are given below:

$$\begin{aligned} x_j &= a_j \mathbf{1} + b_j \boldsymbol{\mu} + \varepsilon_j \\ x'_j &= (x_j - a_j) / b_j \end{aligned} \quad (24.14)$$

where

$\boldsymbol{\mu}$ is the average spectrum.

b_j is the slope.

a_j are the offset values for each wavelength.

An important advantage of MSC preprocessing is that it can be used on filter wheel data, where the wavelength spacing is irregular and only a few wavelengths are typically measured.

There are many other preprocessing methods used in chemometrics such as wavelets, orthogonal signal correction, and extended MSC (EMSC) [3, 11]. In practice combinations of different preprocessing can also be used. However, the three methods discussed (derivatives, SNV, and MSC) are still the most commonly used preprocessed in real-world applications.

Partial least squares (PLS) is an extension of PCA where both the \mathbf{X} and \mathbf{Y} data are considered [12, 13]. In PCA only the \mathbf{X} data is considered. The goal of the PLS analysis is to build an equation that predicts \mathbf{Y} values (laboratory data) based on \mathbf{X} (spectral) data. The PLS equation or calibration is based on decomposing both the \mathbf{X} and \mathbf{Y} data into a set of scores and loadings similar to PCA. However, the scores for both the \mathbf{X} and \mathbf{Y} data are not selected based on the direction of maximum variation but are selected in order to maximize the correlation between the scores for both the \mathbf{X} and \mathbf{Y} variables. As with PCA, in the PLS regression development, the number of components or factors is an important practical consideration. A short description of the PLS algorithm is given below; a more detailed discussion of the PLS algorithm

can be found elsewhere [12, 13]. Commercial software can be used to construct and optimize both PCA and PLS calibration models.

PLS decomposition of both \mathbf{X} and \mathbf{Y} data into scores and loadings is given in the equation below:

$$\begin{aligned}\mathbf{X} &= \mathbf{TP}^T + \mathbf{E} \\ \mathbf{Y} &= \mathbf{UQ}^T + \mathbf{f}\end{aligned}\quad (24.15)$$

The score matrices for \mathbf{X} and \mathbf{Y} , i.e. \mathbf{T} and \mathbf{U} , are calculated together. This self-consistent approach allows for a set of scores and loadings that represent the variation in the \mathbf{Y} data set. Therefore, the scores and loadings are much better than PCA scores and loadings for quantitative prediction. The algorithm proceeds by mean centering the data and then finding the first loading spectrum and first component scores. The prediction of a PLS method is summarized in the regression vector or coefficient, \mathbf{B} . The predictions are related to the \mathbf{x} sample data by

$$y = \mathbf{B} \cdot \mathbf{x} \quad (24.16)$$

We will now consider an example of a PLS calibration using NIR data. NIR transmission spectra from 155 tablets have been measured [14]. The tablet calibration set included samples with a range of assay values and several lots of production samples in order to capture the typical variations seen

in the tablets. After scanning with the NIR instrument, the amount of active ingredient in each tablet was measured by HPLC. The weight of the tablet was about 800 mg, and the target value for the drug content was 200 mg. We will use chemometrics to develop a model for the amount of active. This model could be used to monitor the stability of tablets over time in a nondestructive manner. For brevity, we will only outline the analysis procedure. Typical NIR transmission spectra for the pharmaceutical tablet are shown in Figure 24.11. The broad overlapping spectra with a considerable background are typical of NIR spectra. Derivative preprocessing can be used to remove the unnecessary background and elucidate the underlying peaks in the spectra. A first derivative spectrum is shown in Figure 24.12.

A calibration curve showing the predictions of the PLS model versus the laboratory data is shown in Figure 24.13. The clear quality of the calibration curve is evident. The calibration curve can be evaluated by several methods including outlier detection and removal and optimization of the spectral range used for PLS calibration. A detailed discussion of these issues can be found in the references [12, 13]. Common examples of quantitative methods done with NIR data and PLS regression are moisture, particle size, and assay [7].

Method validation for NIR or Raman spectroscopic methods using chemometrics is outlined in USP chapter <1119> [15]. The criteria for method validation are the same as other quantitative analytical methods: accuracy, precision,

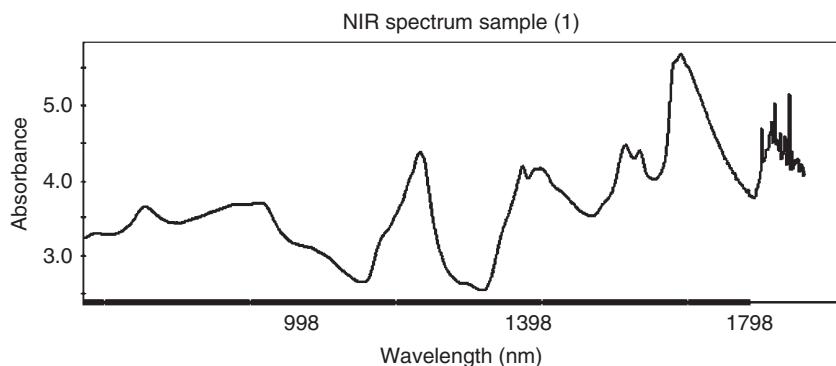


FIGURE 24.11 NIR transmission spectrum of a pharmaceutical tablet.

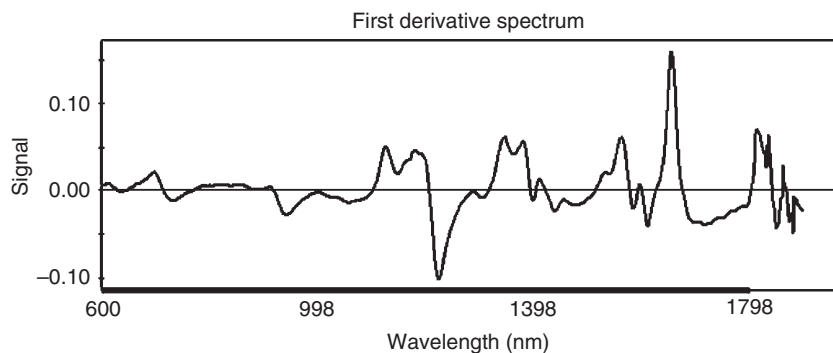


FIGURE 24.12 Spectrum from Figure 24.11 after first derivative preprocessing.

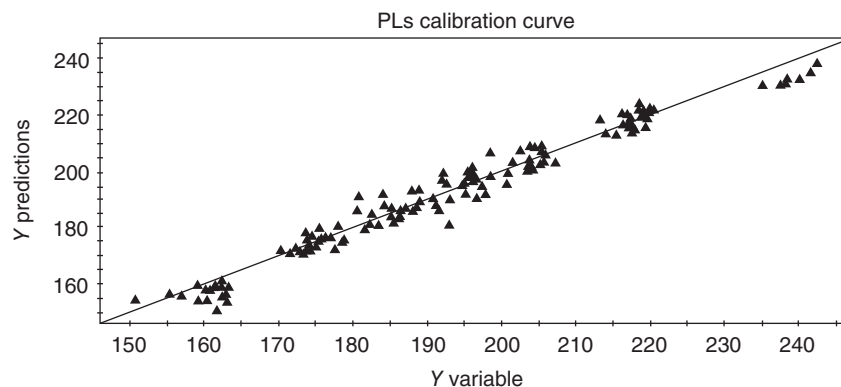


FIGURE 24.13 Calibration curve from PLS method for tablet assay value.

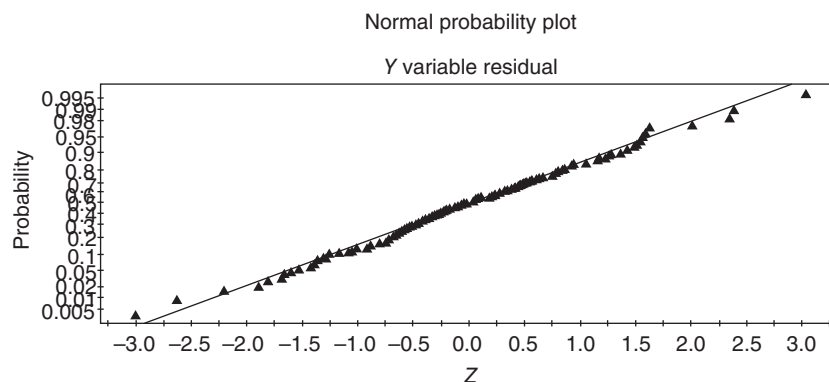


FIGURE 24.14 Normal probability plot for residuals. When the residuals fall on straight line, the calibration under consideration is linear.

intermediate precision, linearity, specificity, and robustness. However, because these methods are statistical in nature and are based on a previously validated analytical method, the validation of MVA methods is somewhat different than traditional analytical methods. In this chapter we will briefly discuss chemometric method validation; a more detailed discussion can be found elsewhere.

Accuracy of the MVA method refers to how closely the MVA method and the original laboratory method compare. The accuracy of a chemometric method is evaluated by comparing the predictions of the MVA model with the actual lab data for a set of validation samples. The validation samples should be from lots of material not used in the original calibration set. There are several mathematical ways to express the accuracy. The most commonly used approach is the standard error of prediction (SEP). The SEP is defined in Eq. (24.17):

$$\text{SEP} = \sqrt{\sum \frac{(\text{NIR} - \text{LAB})^2}{n}} \quad (24.17)$$

where

n is the number of validation samples.

The SEP value should be close to the actual error of the original laboratory method. The actual error of the laboratory method should include normal sources of variation such as different analysts, different instruments, and different lots of material analyzed on different days.

The *linearity* of a multivariate method is an important topic. Typically the linearity of a chromatographic method is evaluated by the R^2 (coefficient of determination) value of a recovery measurement. R^2 is the fraction of variation in the y variable explained by the linear fit; r , the correlation coefficient, quantifies the correlation between the x and y variables [16]. R^2 is often used in the analysis of chromatography recovery studies [16]. In contrast, R^2 is not a good statistical parameter for multivariate methods. The linearity of a multivariate method is evaluated by the inspection of the residual values, i.e. the difference between the predictions of the multivariate model and the actual laboratory data. A linear model will have residuals that are random, i.e. normally distributed. A nonlinear model will have residuals that are not normally distributed. The USP chapter <1119> states that the linearity should be evaluated by examination of the residuals, but no specific threshold or criteria are given. In the opinion of this author, visual inspection of the residuals using a normality plot is recommended. In Figure 24.14, a normality plot of residuals is shown. The data points in Figure 24.14

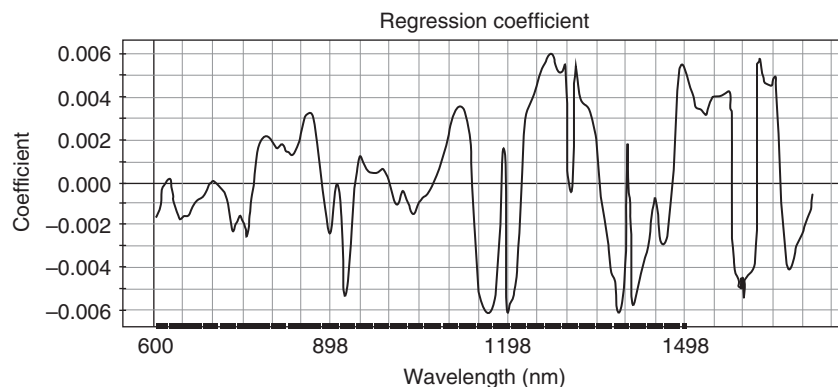


FIGURE 24.15 Regression coefficient from the PLS model for tablet assay described earlier. The regression coefficient is a method for documenting and examination of which wavelengths are most important for the PLS calibration.

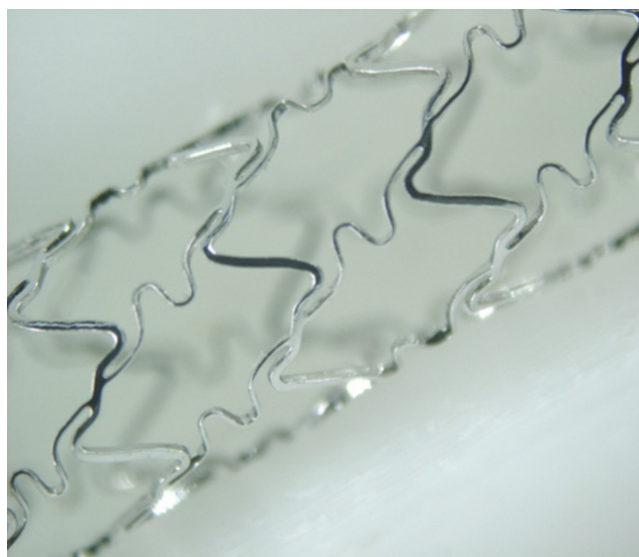


FIGURE 24.16 Photograph of a drug-eluting stent. *Source:* Reprinted with permission from Balss et al. [17]. Copyright 2008 American Chemical Society.

do follow a straight line, indicating a normal distribution of residuals, consistent with a linear model or calibration [13]. In some cases, the linearity of the model can be improved by removing some of points in the normality graph, which are probable outliers.

Method *specificity* is the extent to which the multivariate calibration is specific to the analyte of interest. With a PLS calibration, the specificity is documented by the regression coefficient of the calibration. The regression coefficient shows which wavelengths are most important for the PLS calibration. Important wavelengths may have either positive or negative regression coefficient values. The most important wavelengths should correspond to the absorption peaks of the analyte of interest. For example, the regression coefficient for a moisture model will have peaks at the known water

absorbance band locations. In practice, the regression coefficient is often documented in the method development report. A regression coefficient from the PLS calibration for tablet assay described earlier in this chapter is shown in Figure 24.15.

The *range* of a multivariate calibration method is determined by the range of laboratory values in the calibration and validation data sets. A method is validated over the range of lab values of the samples used in the independent validation set. The range of the validation samples can also depend upon the application of the method. For example, in-process testing or testing where a limited number of samples are available may require a fairly small range of values because samples outside of a small range are not available or do not exist.

An important application of the above principles is chemical imaging. In chemical imaging, spectral information about a sample of interest is collected at many physical points on the sample. With sufficient density of data collection, this can provide a detailed chemical map of the sample. Experimentally this can be done using numerous spectroscopies including Fourier transform IR (FTIR), Raman, and NIR. The promise of chemical imaging is to provide chemical imaging information about a sample at high spatial resolution. This is particularly important in pharmaceutical development, where many tablets or drug delivery systems are spatially inhomogeneous.

An example where chemical imaging has been used very successfully is drug-eluting stents (DES) [17]. DES consist of a metal stent with a thin coating. This coating contains a powerful immunosuppressant drug. The value of the drug is to control clotting near the stent so that the stent can keep the constricted artery open. A photo of a DES is shown in Figure 24.16. In recent work, confocal Raman microscopy was used to evaluate coatings on DES at high spatial resolution. The coating was a mixture of two polymers, namely, polyethylene vinyl acetate (PEVA) and polybutyl

methacrylate (PBMA), and the drug rapamycin. Three independent PLS models were constructed for the components of the mixture. Slightly different ranges and preprocessing were used for the polymers and the drug in order to obtain optimized calibration models. A mixture design was used to help construct the calibration set. This was critical because it allowed the entire mixture space to be spanned while independently varying all three components. Production samples were added to the calibration set to improve the calibration model. Because no reference method could provide spatially resolved data, spatially averaged images were used in the calibration set. Excellent results were obtained comparing the spatially averaged calibration models with independent validation samples. The PLS models were also found to exhibit mass balance. A typical image of the spatial drug distribution is shown in Figure 24.17. Further details about this work can be found in Ref. 17.

The second example of the use of MVA for medical device development is a classification problem. The data was analyzed using the SIMCA method, which was described in detail earlier in this chapter. During the development of DES, *in vivo* porcine studies were conducted. It was very important to understand the time scale of the porcine biological response to the stent. FTIR analysis was performed on the stent and the tissue around the stent at various time points. Data at different time points was collected: 0, 1, 14, 60, 75, 90, and 180 days. The SIMCA model had two groups, the original stent at 0 days and the pure porcine tissue. A Coomans plot of the results is shown in Figure 24.18 (Balss K., Long F., and Maryanoff C., unpublished results). The x - and y -axes are the scaled residual or DmodX distance of a given sample from either the porcine or stent model. SIMCA is a very sensitive classification method that can work with complex, highly dimensional data such as FTIR spectra. Samples below the

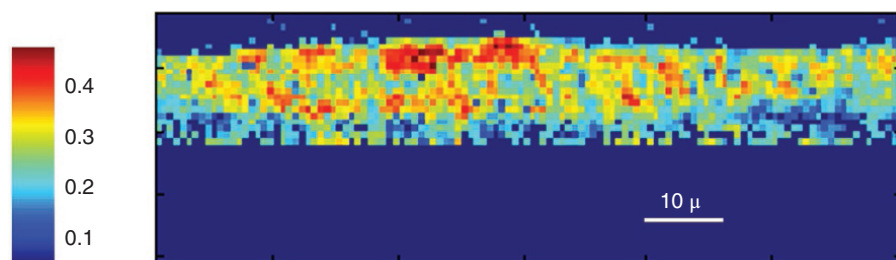


FIGURE 24.17 Predicted drug distribution in drug-eluting stent using a PLS calibration model. The thickness of the coating is approximately 10 μm . *Source:* Reprinted with permission from Balss et al. [17]. Copyright 2008 American Chemical Society.

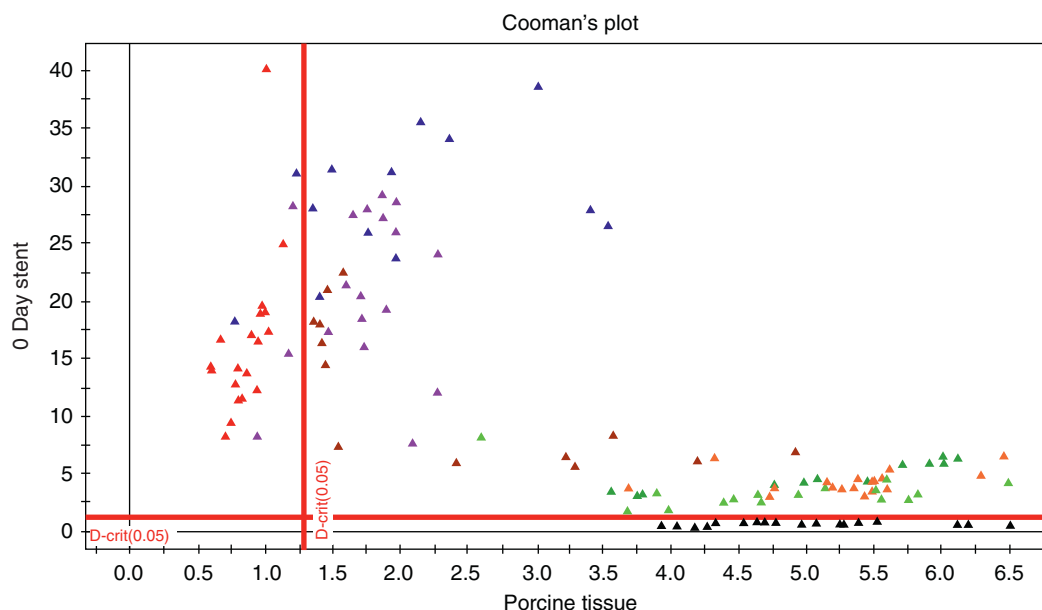


FIGURE 24.18 Coomans' plot based on FTIR spectra taken at different times, see text for details: black, 0 day; orange, 1 day; light green, 14 days; dark green, 60 days; brown, 75 days; purple, 90 days; blue, 180 days; red, porcine tissue. The 0 day (black) points are below the horizontal line and the porcine tissue (red) points are to the left of the vertical line.

red horizontal line are stent spectra, and samples to the left of the red vertical line are porcine tissue. The spectra of the stent and pig tissue are readily distinguished; furthermore the time evolution of the porcine response to the stent is clearly seen as the tissue around stent looks more and more like the normal porcine tissue. The SIMCA classification model was validated using several independent sets of data. This analysis is impossible by simple visual inspection of the spectra and demonstrates the power of MVA.

This chapter has briefly summarized the essential principles of chemometrics and their application to spectral data. Examples have included both pharmaceutical and medical device development. There are many applications of chemometrics that have not been discussed here due to space limitations. Batch monitoring involves the use of multivariate control charts based on score plots developed from a collection of good batches [18]. It can be used with spectral data, process data, or a combination of both. It has been used to monitor a variety of complex pharmaceutical products to improve yields and provide improved process understanding [2]. In particular, it can also be used in biopharmaceutical development. MVA can also be used in real-time monitoring of continuous manufacturing processes. In summary, chemometrics is a vital part of process analytical technology, quality by design, and the overall future of both pharmaceutical development and manufacturing.

REFERENCES

1. US Food and Drug Administration *Guidance for Industry Process Analytical Technology* (2004).
2. Eriksson, L., Johansson, E., Kettaneh-Wold, N. et al. (2006). *Multi- and Megavariate Data Analysis*. Umetrics AB.
3. Strang, G. (2007). *Computational Science and Engineering*. Wellesley-Cambridge Press.
4. Johnson, R.A. and Wichern, D.W. (2002). *Applied Multivariate Statistics*. Prentice Hall.
5. Long, F. (2008). *Am. Pharm. Rev.* Sept/Oct.
6. Wold, S. (1976). *Pattern Recogn.* 8: 127–139.
7. Siesler, H., Ozaki, Y., Kawata, Y., and Heise, H. (2002). *Near-Infrared Spectroscopy: Principles, Instruments, Applications*. Wiley.
8. Savitzky, A. and Golay, M.J.E. (1962). *Anal. Chem.* 36: 1627–1639.
9. Barnes, R.J., Dhanoa, M.S., and Lister, S. (1989). *J. Appl. Spectrosc.* 43: 772–777.
10. Martens, H., Jensen, S.A., and Geladi, P. (1983). *Proceedings from Nordic Symposium on Applied Statistics*, 205–233. Norway: Stokkand Forlag Publishers.
11. Martens, H. and Stark, E.J. (1991). *Pharm. Biomed. Anal.* 9: 625–635.
12. Wold, S. and Josefson, M. (2000). Multivariate analysis of analytical data. In: *Encyclopedia of Analytical Chemistry* (ed. R. Meyers), 9710–9736. John Wiley & Sons.
13. Esbensen, K. (2002). *Multivariate Data Analysis- In Practice*. CAMO Process AS.
14. Ritchie, G. (2002). Software shootout data. International Diffuse Reflection Conference, Chambersburg, PA (August 2002).
15. US Pharmacopoeia *Near IR Spectroscopy*, Chapter <1119> and *Raman Spectroscopy*, Chapter <1120>.
16. Miller, J. (2000). *Statistics and Chemometrics for Analytical Chemistry*, 137. Prentice Hall.
17. Balss, K., Long, F., Veselov, V. et al. (2008). *Anal. Chem.* 80: 4853–4859.
18. Wold, S., Cheney, J., Kettaneh, N., and McCready, C. (2006). *Chemom. Intell. Lab. Syst.* 84: 159–163.

PHARMACEUTICAL MANUFACTURING: THE ROLE OF MULTIVARIATE ANALYSIS IN DESIGN SPACE, CONTROL STRATEGY, PROCESS UNDERSTANDING, TROUBLESHOOTING, AND OPTIMIZATION

THEODORA KOURTI

Department Chemical Engineering, McMaster University, Hamilton, Ontario, Canada

25.1 INTRODUCTION

The goal of any industry (be it chemical, pharmaceutical, steel, pulp, and paper) is to produce a product satisfactory to the customer (that is, within prescribed quality specifications) under safety and environmental regulations and at a minimum cost. Quality control and regulatory specifications (safety, environmental) will help the manufacturer achieve the first three objectives but sometimes at the expense of cost. Understanding the process and monitoring and controlling process performance will help meet all four targets (quality product, safety constraints, environment constraints, minimum cost) simultaneously.

Process analysis and understanding, monitoring, and control have been practiced by several industries (notably petrochemical) for several decades. These industries adopted the above practices gradually. First they saw the need for real-time quality measurements and developed real-time analyzers; as an example, the first analytical and control instrument group of Universal Oil Products (UOP) was formed in 1959 with the mission to develop online analyzers for internal pilot plant applications. This first step made the industry capable of collecting real-time measurements of quality properties and of other process variables. The second step was the development of automatic process control techniques. This required some form of modeling. Attempts were made to understand the fundamental mechanisms of processes and build sophisticated mechanistic (first principles) or empirical

(data-driven) models. Later, in the 1990s, the industry made a third step, which was the use of multivariate statistical analysis methods. With these multivariate approaches it became possible to analyze and understand the process by looking at historical data containing hundreds of variables, detect abnormal situations, diagnose the sources of the abnormalities, and make appropriate modifications. Furthermore, by utilizing multivariate statistical process control (MSPC), it became possible to monitor the wellness of the process and product in real time by looking simultaneously at hundreds of variables as they are collected. As a result several industries managed not only to assure acceptable end-product quality but also to improve process performance and maintenance and to significantly reduce cost. The quick adoption of the methodologies and the benefits becomes evident from a very impressive set of applications presented by industry in 2003 in the symposium of “Abnormal situation detection and projection methods – Industrial applications” [1].

However, the picture in the pharmaceutical industry was different. In 2003, an article in the *Wall Street Journal* [2] stated that “The pharmaceutical industry has a little secret: Even as it invents futuristic new drugs, its manufacturing techniques lag far behind those of potato-chip and laundry-soap makers.” The article went on to explain that “in other industries, manufacturers constantly fiddle with their production lines to find improvements,” but “regulations leave drug-manufacturing processes virtually frozen in time.” Applications from the snack food industry (potato chip) that

were published that same year [3, 4] were illustrating the use of inferential sensors based on a digital imaging system that had been developed for monitoring and control of the amount of coating applied to the base food product and the distribution of the coating among the individual product pieces, with real-time results from the implementation of such imaging system on snack food production lines. The imaging system was used to monitor product quality variables and to detect and diagnose operational problems in the plants. It was also used to implement closed-loop feedback control over coating concentration. It was based on multivariate image analysis. Figure 25.1 shows the setup for feedback control based on image analysis. Images are collected by cameras from the process or from a stream exiting the process. This information then can be used in a feedback control loop.

The situation depicted in Figure 25.1 is the desired state in the pharmaceutical industry. That is, it is desired that we have the ability to measure or infer quality in real time, to assess the deviation from expected quality value, and to calculate a real-time control action for correction. Such ability stems from good models that provide process understanding, (different) models that convert spectral or other sensor data to quality, and (different) models that calculate the required control action. The word “different” was on purpose added in the previous phrase to indicate that in this endeavor different types and classes of models are employed and that there are several modeling activities required. The models required for all these classes of modeling may be first principles/mechanistic, or data based/empirical, or hybrid. Multivariate projection methods or latent variable methods play an integral part in empirical

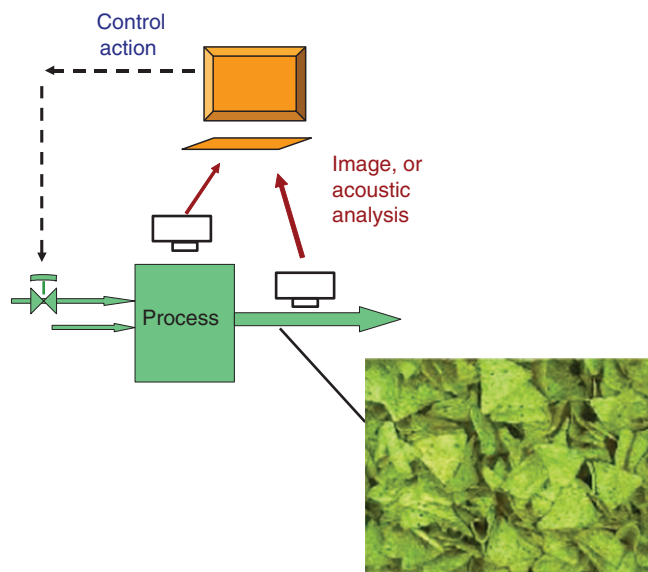


FIGURE 25.1 Monitoring and feedback control based on image and vibrational analysis. Example from snack food industry [3–5]. Source: Reprinted (adapted) with permission from Yu et al. [3]. Copyright (2003) American Chemical Society.

and hybrid modeling. Such models can be developed to relate final quality properties to raw material attributes and process parameters and can be used for process understanding, process monitoring, and troubleshooting. Multivariate models can also be developed and utilized for process control, scale-up, and site transfer. There is a wealth of literature describing the theoretical foundation of the latent variable or projection methods [6–8] as well as the experiences from practitioners in industry [1, 9].

A lot of changes have happened since 2003 in the pharmaceutical industry that has entered a new era. The introduction of concepts like quality by design (QbD), design space, control strategy, and real-time release (Table 25.1) and the adoption of continuous manufacturing by the pharmaceutical industry are examples of such changes. Multivariate methods are most suitable to address requirements associated with these concepts. The pharmaceutical industry can learn from existing methodologies and from experiences from other industries and utilize multivariate technologies for process and product design, process monitoring, and fast process and product improvements.

In this chapter the fundamentals of latent variable modeling will be presented briefly together with references for in-depth presentations of methodologies and their use for process understanding, troubleshooting, monitoring, and control. Case studies are shown for such applications or

TABLE 25.1 Terms Related to Quality by Design

Quality by design (QbD) is defined as a systematic approach to development that begins with predefined objectives and emphasizes product and process understanding and process control based on sound science and quality risk management
Design space is the multidimensional combination and interaction of input variables (e.g. material attributes) and process parameters that have been demonstrated to provide assurance of quality
Control strategy is a planned set of controls derived from current product and process understanding that ensures process performance and product quality. The controls can include parameters and attributes related to drug substance and drug product materials and components, facility and equipment operating conditions, in-process controls, finished product specifications, and the associated methods and frequency of monitoring and control
Real-time release testing is the ability to evaluate and ensure the quality of in-process and/or final product based on process data, which typically include a valid combination of measured material attributes and process controls.

These terms are defined by the International Conference on Harmonization of Technical Requirements for Registration of Pharmaceuticals for Human Use (ICH). ICH is unique in bringing together the regulatory authorities and pharmaceutical industry to discuss scientific and technical aspects of drug registration. ICH's mission is to achieve greater harmonization worldwide to ensure that safe, effective, and high-quality medicines are developed and registered in the most resource-efficient manner. <http://www.ich.org/home.html>.

are referenced. The chapter should be used by the reader as a guidance for the types of problems that can be solved utilizing the methodology; the reader will use the detailed references to seek in-depth analysis and detailed solutions to specific problems.

25.2 THE NATURE OF PROCESS AND QUALITY DATA

25.2.1 Multivariate Nature of Quality

Understanding the multivariate nature of quality is of great importance. Product quality is defined by the simultaneously correct values of all the measured properties; that is, product quality is a multivariate property. Most of the time the property variables are not independent of one another, and none of them adequately define product quality by itself; therefore it is not a good practice to separately monitor key properties of the final product using univariate control charts.

Figure 25.2 shows a classic illustration of the problem using separate control charts for two quality variables (y_1 , y_2). In this figure, the two variables are plotted against each other (upper left of the figure). The same observations are also plotted as individual (univariate) charts for y_1 (the horizontal plot) and y_2 (the vertical plot) with their corresponding upper and lower control limits. Suppose that when only common cause variation is present, y_1 and y_2 follow a multivariate normal distribution; the dots in the joint plot represent a set of observations from this distribution. Notice that y_1 and y_2 are correlated. The ellipse represents a $(1 - \alpha)\%$ joint confidence limit of the distribution (i.e. when the process is in control, $\alpha\%$ of the points will fall outside the ellipse).

The point indicated by the \otimes symbol is clearly outside the joint confidence region, and it is different from the normal in-control population of the product. However, neither of the univariate charts gives any indication of a problem for point \otimes ; it is within limits in both of the charts. The individual

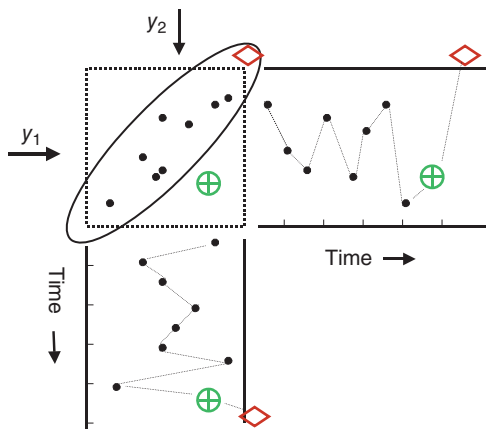


FIGURE 25.2 The multivariate nature of quality.

univariate charts effectively create a joint acceptance region shaped like a square (shown with the ellipse). This will lead to not only accepting wrong products as good (point \otimes) but also rejecting a good product as bad (point \diamond). The problem worsens as the number of variables increases. It is clear that an efficient fault detection scheme should look at the variables together. Multivariate charts are required in order to test quality [10] when it is described by many variables.

Recognizing the multivariate nature of quality, it should guide the procedures that will be used for the following cases:

- Raw material evaluation.
- Intermediate quality evaluation.
- Final quality evaluation.
- Process control for quality.
- Product transfer and scale-up (the multivariate nature of quality should be preserved for raw materials, intermediate qualities, and final qualities). This is a minimum requirement. Later we will discuss the requirements on the multivariate space of the process variable trajectories as well.

25.2.2 Real-Time Monitoring and Process Signature: The Need to Utilize Information from Process Data

It is well known in industry that the same measured quality properties can be achieved by taking different process paths. However, very frequently, the few properties measured on a product are insufficient to define entirely the product quality. In polymer industry, for example, if only the viscosity of a polymer is measured and kept within specifications, any variation in end-use application (downstream processability) that arises due to variation in chemical structure (branching, composition, end-group concentration) will not be captured.

Consider the example of Figure 25.3, where the trajectories of each one of three process variables are plotted for three different batch runs, A, B, C. The final product at time t (end-point quality) is determined from properties $y_{1,t}$ and $y_{2,t}$. These properties are plotted for each product produced by runs A, B, C against the desired confidence limits of acceptable multivariate quality defined by the ellipse. Suppose that for all three runs, the two “end properties” are on target in the multivariate control chart (points fall within the ellipse). However the process trajectories follow different paths for each run. Are these runs equivalent?

To achieve consistency in all the product properties (measured quality and ability to process down the stream), the process conditions (path to end point) must also be kept in statistical control. When this is not the case, although the measured product properties may be on target, the properties that determined the processability of the product may not be within acceptable limits. Therefore monitoring process data

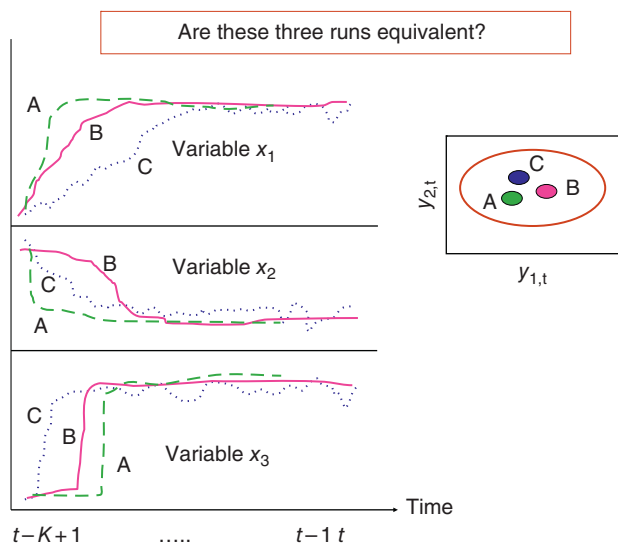


FIGURE 25.3 Different paths to the end point may result in similar values for some quality properties but may affect others in a different way. Consistent paths will assure that overall quality is consistent.

(temperatures, pressures, etc.) together with real-time analyzers will not only give valuable information about events with special causes that may affect the final quality but also give early warnings for potential equipment failure.

Another example that further corroborates this argument has been reported from the pharmaceutical industry [11]: “Conventional process control of drying of granulate in a fluidized bed drier would be to measure the loss on drying of a sample of powder, to determine water content. An advance on this may be to determine water content using an on-line NIR technology. However, true Process Understanding requires that the route by which you get to this end point be known and controlled. For example if the drying process is too vigorous, attrition may cause the granulate to generate an unacceptably high level of fine particles, which may cause downstream processing problems or dissolution issues; equally if drying is too slow, the potential for degradation of the drug molecule may exist.”

This “process path to the end point” is also discussed in the European regulatory perspective [12] where it is reported that “during discussions within the industry, the term *process signature* has been mentioned regularly.” To get a common understanding of this, the EU PAT Team had invited public comments on the following definition: “A collection of batch specific information that shows that a batch has been produced within a design space of the product.” The EU PAT Team mentions as examples of process signatures the amount of water added in relation to time (wet massing), air flow rate, and bed temperature during fall rate drying (fluidized bed drying). They concluded that their understanding is that there

is no unique process signature, but instead a family of process signatures with common characteristics (salient features).

The above observations point to the importance of monitoring the process together with the product quality. By monitoring only the quality variables (in a univariate or multivariate chart), one performs statistical quality control (SQC). Real-time measurements on temperatures, pressure, pH, rpm, etc., combined real-time measurements from analytical technology (spectroscopy, ultrasound, etc.) will lead to online *process* monitoring and make MSPC, fault detection and isolation, possible. Combining information from the process measurements with the information from the analytical tools gives a very powerful tool to monitor the process. These two sets of measurements are not independent from one another, but are interrelated. As a matter of fact, these measurements “confirm” each other. This is the reason that process variables are sometimes used to assess the reliability of real-time analyzers. It will also be pointed out later that sometimes real-time process measurements may eliminate the need of some real-time analyzers. Information about the process may also include the vessels used for a specific run, the operators that were on shift, suppliers of raw material, etc. Another advantage to using process measurements is that any abnormal events that occur will also have their fingerprints in the process data. Thus, once an abnormal situation is detected, it is easier to diagnose the source of the problem, as we are dealing directly with the process variables. For example, a pending equipment failure means that our production is not in control. However, there are situations that while there may be a pending equipment failure, real-time quality measurements may still be acceptable. By monitoring process variables we have a very high probability to detect a pending problem.

Process data can be utilized together with appropriate models to:

- Infer final product quality from process conditions during production.
- Ease process understanding and troubleshooting.
- Infer a quality in real time (soft sensors).
- Establish an overall “process signature” and monitor it.
- Monitor analyzer reliability.
- Check that the process is in a state of statistical process control (SPC).
- Decide on midcourse correction of variable trajectories to control final quality.
- Establish operational knowledge that can be used for product transfer and scale-up.

It should be emphasized here that although some relations between process operating conditions and final quality are known from the initial design of experiments (DOE), once

we are in production, these relations may be influenced by other factors and may change locally. By investigating production data we can uncover the true relationships between process conditions and quality under the closed-loop operations.

Finally, it is very frequently stated that critical process parameters should be identified and monitored together with critical quality attributes utilizing SPC. It should be emphasized here that the parameters that appear to be critical in the DOE are not necessarily the ones that will give information about the “wellness of the process” in SPC charts. The reason for this is that the parameters that are identified as important in DOE will be tightly controlled during production. SPC charts on routine data are noncausal; therefore things that were important in DOE will not be important in SPC, unless something goes really wrong (i.e. the controller fails and cannot keep the desired target). As an example, suppose that temperature is important to the yield, as determined by DOE. That means that during production, the desired temperature profile will be regulated by the controllers; in SPC monitoring, what is important (and in some types of processes will indicate the presence of excess impurities and other disturbances) is how much effort the controller is putting to maintain the temperature, that is, how much the valve to the cooling agent opened or closed during the reaction. Therefore monitoring the controller action will provide much more information about abnormal situations than monitoring the temperature, although temperature was identified as critical process parameter by DOE.

25.2.3 Multivariate Nature, Structure, and Other Characteristics of Process Data

Databases containing measurements collected during production may become very large in size. The data are noncausal in nature (unless they come from designed experiments). They consist of highly correlated variables with many missing measurements and low content of information in any one variable (due to the low signal-to-noise ratios).

Multivariate Structure of Data: The convention that will be used throughout this chapter in expressing data is that of Table 25.2. Other formats, which may appear in specific sections only, will be defined in their corresponding sections.

25.2.4 Process Analysis and Process Understanding

Sometimes the term “process analysis” is incorrectly being used as an equivalent term to process analytical chemistry. Collecting real-time measurements on a specific property may or may not reflect what is happening in the rest of process, or the state of the process, unless the state of the process is completely observable in that quality, as discussed in Section 25.2.2; therefore collecting a real-time quality measurement is not process analysis.

TABLE 25.2 Some Multivariate Process Data Formats

Matrix Symbol	Dimensions	Explanation
X	$(n \times k)$; two-way matrix n observations in time or n batches; k process variable measurements	Data from a continuous process, at given instant in time; summary data from a batch (max T , min T , length of batch run, etc.)
Y	$(n \times m)$; two-way matrix n observations in time or n batches; m product quality values	Quality data from a continuous process corresponding to the process measurements in X , properly lagged; quality data at the end of a batch
X	$(I \times J \times K)$; three-way matrix I batches; J process variables, measured at K time intervals, for each batch	Data collected from batch process at several time intervals during production
Z	$(n \times r)$; two-way matrix n observations in time or n batches; r other variable measurements	Raw material; total cycle times; length between processes; preprocessing information

A process can be defined as a series of physical and/or chemical operations that converts input to output. Process analysis is a systematic examination of a process to understand it in order to develop ideas to improve it. Improvement could translate to better quality, lower cost, more efficient energy consumption, less pollutants to environment, and safer operation. One can perform process analysis utilizing both offline and real-time measurements.

Process analysis leads to process understanding. Again there may be several definitions of process understanding. There is a widespread belief that one gains process understanding only when they can describe the process by first principles, that is, by a theoretical or mechanistic model. However, one can gain tremendous insight into the process from empirical models derived from databases. These empirical models can lead to fast improvements that in several situations would have been impossible if people had been waiting for the development of theoretical first principles

models. Empirical models based on process data can be extremely valuable to diagnose abnormal operations such as pending equipment failure.

So while process understanding may by some definitions mean uncovering the mechanisms and path of a chemical reaction, or modeling a fermentation process, it may also mean uncovering production problems like the examples reported by practitioners in several industries:

- Diagnosing that the production of abnormal batches followed a specific pattern and as a result uncovering an incorrect operator practice that led to an abnormal batch every time a routine maintenance task was taking place.
- Understanding why recent process data that were projected on a latent variable space seem formed two clusters indicating different operation practices and hence solving an important operation problem as a result. Examination revealed that the cooling agent valve was not capable of meeting the demands in hot days and the reactor temperature could not be controlled properly. The valve was resized.
- Assessing that an operational problem caused the readings of a specific thermocouple to be erroneous and appear as outliers; thermocouple was too close to entrance of cool reactant; erroneous readings fed to the controller alter inappropriately the reactor temperature.
- Understanding where is the maximum process variability; is this variability noise or is it assignable to a cause – can we reduce it?

25.3 LATENT VARIABLE METHODS FOR TWO-WAY MATRICES

Latent variables exploit the main characteristic of process databases, namely, that although they consist of measurements on a large number of variables (hundreds), these variables are highly correlated, and the effective dimension of the space in which they move is very small (usually <10 and often as low as 2). Typically only a few process disturbances or independent process changes routinely occur, and the hundreds of measurements on the process variables are only different reflections of these few underlying events. For a historical process dataset consisting of a ($n \times k$) matrix of process variable measurements \mathbf{X} and a corresponding ($n \times m$) matrix of product quality data \mathbf{Y} , for linear spaces, latent variable models have the following common framework [13]:

$$\mathbf{X} = \mathbf{T}\mathbf{P}^T + \mathbf{E} \quad (25.1)$$

$$\mathbf{Y} = \mathbf{T}\mathbf{Q}^T + \mathbf{F} \quad (25.2)$$

where

\mathbf{E} and \mathbf{F} are error terms.

\mathbf{T} is an ($n \times A$) matrix of latent variable scores.

\mathbf{P} ($k \times A$) and \mathbf{Q} ($m \times A$) are loading matrices that show how the latent variables are related to the original \mathbf{X} and \mathbf{Y} variables.

The dimension A of the latent variable space is often quite small and determined by cross-validation or some other procedure.

Latent variable models assume that the data spaces (\mathbf{X} , \mathbf{Y}) are effectively of very low dimension (i.e. non-full rank) and are observed with error. The dimension of the problem is reduced by these models through a projection of the high-dimensional \mathbf{X} and \mathbf{Y} spaces onto the low-dimensional latent variable space \mathbf{T} , which contains most of the important information. By working in this low-dimensional space of the latent variables (t_1, t_2, \dots, t_A), the problems of process analysis, monitoring, and optimization are greatly simplified. There are several latent variable methods. Principal component analysis (PCA) models only a single space (\mathbf{X} or \mathbf{Y}) by finding the latent variables that explain the maximum variance. Principal components can then be used in regression (principal component regression [PCR]). In PCR there appears to be a misconception that the principal components (PCs) with small eigenvalues will very rarely be of any use in regression. The author's personal experience is that these components can be as important as those with large variance. Projection to latent structures or partial least squares (PLS) maximizes the covariance of \mathbf{X} and \mathbf{Y} (i.e. the variance of \mathbf{X} and \mathbf{Y} explained plus correlation between \mathbf{X} and \mathbf{Y}). Reduced rank regression (RRR) maximizes the variance of \mathbf{Y} and the correlation between \mathbf{X} and \mathbf{Y} . Canonical variate analysis (CVA), or canonical correlation regression (CCR), maximizes only the correlation between \mathbf{X} and \mathbf{Y} . A discussion of these latent variable models can be found elsewhere [13]. The choice of method depends on the objectives of the problem; however, all of them lead to a great reduction in the dimension of the problem. Some of them (PCR and PLS) model the variation in the \mathbf{X} space as well as in the \mathbf{Y} space. This point is crucial in most of the applications related to PAT that are discussed in the following sections, as well as for the problem of treating missing data. The properties of PCA and PLS are discussed briefly below.

25.3.1 Principal Component Analysis

For a sample of mean centered and scaled measurements with n observations on k variables, \mathbf{X} , the PCs are derived as linear combinations $\mathbf{t}_i = \mathbf{X}\mathbf{p}_i$ in such a way that, subject to $|\mathbf{p}_i| = 1$, the first PC has the maximum variance, the second PC has the next greatest variance and is subject to the condition that it is

uncorrelated with (orthogonal to) the first PC, etc. Up to k PCs are similarly defined. The sample PC loading vectors \mathbf{p}_i are the eigenvectors of the covariance matrix of \mathbf{X} (in practice, for mean centered data the covariance matrix is estimated by $[n - 1]^{-1} \mathbf{X}^T \mathbf{X}$). The corresponding eigenvalues give the variance of the PCs (i.e. $\text{var.} [\mathbf{t}_i] = \lambda_i$). In practice, one rarely needs to compute all k eigenvectors, since most of the predictable variability in the data is captured in the first few PCs. By retaining only the first A PCs, the \mathbf{X} matrix is approximated by Eq. (25.1).

25.3.2 Partial Least Square (PLS)

PLS can extract latent variables that explain the high variation in the process data, \mathbf{X} , which is most predictive of the product quality data, \mathbf{Y} . In the most common version of PLS, the first PLS latent variable $\mathbf{t}_1 = \mathbf{X}\mathbf{w}_1$ is the linear combination of the x -variables that maximizes the covariance between \mathbf{t}_1 and the \mathbf{Y} space. The first PLS weight vector \mathbf{w}_1 is the first eigenvector of the sample covariance matrix $\mathbf{X}^T \mathbf{Y} \mathbf{Y}^T \mathbf{X}$. Once the scores for the first component have been computed, the columns of \mathbf{X} are regressed on \mathbf{t}_1 to give a regression vector, $\mathbf{p}_1 = \mathbf{X}\mathbf{t}_1 / \mathbf{t}_1^T \mathbf{t}_1$; the \mathbf{X} matrix is then deflated (the $\hat{\mathbf{X}}$ values predicted by the model formed by $\mathbf{p}_1, \mathbf{t}_1$, and \mathbf{w}_1 are subtracted from the original \mathbf{X} values) to give residuals $\mathbf{X}_2 = \mathbf{X} - \mathbf{t}_1 \mathbf{p}_1^T$. \mathbf{Q} are the loadings in the \mathbf{Y} space. In the so-called NIPALS algorithm, \mathbf{q}_1 is obtained by regressing \mathbf{t}_1 on \mathbf{Y} , and then \mathbf{Y} is deflated $\mathbf{Y}_2 = \mathbf{Y} - \mathbf{t}_1 \mathbf{q}_1^T$. The second latent variable is then computed from the residuals as $\mathbf{t}_2 = \mathbf{X}_2 \mathbf{w}_2$, where \mathbf{w}_2 is the first eigenvector of $\mathbf{X}_2^T \mathbf{Y}_2 \mathbf{Y}_2^T \mathbf{X}_2$, and so on. The new latent vectors or scores ($\mathbf{t}_1, \mathbf{t}_2, \dots$) and the weight vectors ($\mathbf{w}_1, \mathbf{w}_2, \dots$) are orthogonal. The final models for \mathbf{X} and \mathbf{Y} are given by Eqs. (25.1) and (25.2) [14, 15].

25.3.3 Latent Variables for Process Understanding

Latent variable methods are excellent tools for data exploration to identify periods of unusual/abnormal process behavior and to diagnose possible causes for such behavior (troubleshooting). The scores and loadings calculated by PCA and PLS and the weights by PLS can be utilized for this purpose. By plotting the latent variables (t_1, t_2, \dots, t_A) against each other, the behavior of the original data set (be it process, \mathbf{X} , or quality data, \mathbf{Y}) can be observed on the projection space. By examining the behavior in the projection spaces, regions of stable operation, sudden changes, or slow process drifts may be readily observed. Outlier and cluster detection becomes also easy, both for the process and for the quality space. An interpretation of the process movements in this reduced space can be found by examining the loading vectors ($\mathbf{p}_1, \mathbf{p}_2, \dots, \mathbf{p}_A$) or, ($\mathbf{w}_1, \mathbf{w}_2, \dots, \mathbf{w}_A$) in the case of PLS and the contribution plots in the case of PCA on \mathbf{X} or a PLS analysis on \mathbf{X} and \mathbf{Y} ; each point on a t_1 vs. t_2 plot is the summary of measurements on k variables.

Figure 25.4a gives a simplified schematic interpretation of the methods. Suppose that we have measurements from five variables in a process (here we plot the variable deviations from their nominal trajectories) during a time period. Suppose that variables x_1, x_3 , and x_4 are correlated with each other, while variable x_2 is correlated with x_5 . With the multivariate projection methods, new variables (latent variables) are calculated. In PCA, the first PC, t_1 , is a weighted average of x_1, x_3 , and x_4 , while the second PC, t_2 , is a weighted average of x_2 and x_5 . PCA can be seen as a classification of the main events that affect a process. The first PC corresponds to the event that affects the largest number of variables, the second to the event that affects the next number of variables, and so on. The description here is a simplified explanation. It may be the case that two or more events affect the same variable, in which case this variable will contribute to the values

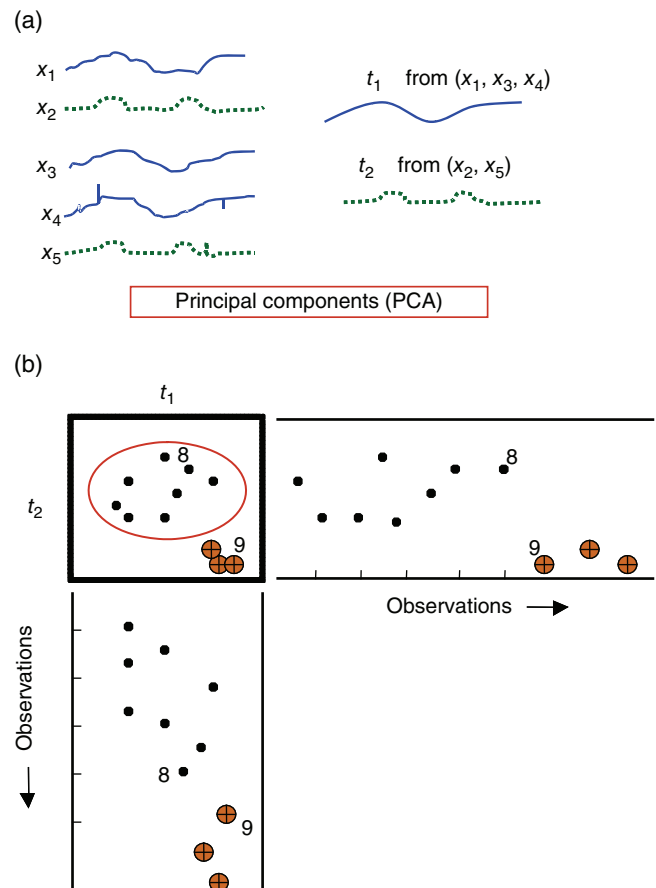


FIGURE 25.4 (a) Simple interpretation of PCA and dimensionality reduction. The principal components t_1 and t_2 use the correlation of five variables and break the process in two orthogonal events. The first principal component corresponds to the event that affects the largest number of variables, the second to the event that affects the next number of variables, and so on. (b) These components can be plotted against each other. A five-variable system is projected onto a two-dimensional plane.

of more than one component. We have reduced the number of the initial five raw variables to two PCs; we have *reduced the dimensionality* of the system. We can now plot these components against each other, as shown in Figure 25.4b. Each point on the plot summarizes the behavior of five raw variables. When the process is in statistical control, the points will be within the control limits, shown with an ellipse that is determined by statistical criteria (discussed later). If there is a problem in the process, the points will plot out of the ellipse. Notice that by using the latent variables, a five-variable system is *projected* onto a two-dimensional plane. This is why these methods are also called projection methods. Plotting PCs against each other is a good way to visualize process behavior and detect outliers and clusters. Typically a small number of components are required to describe the main events in a unit (usually <10 , and sometimes only 3–4).

25.3.3.1 Process Understanding Example 1: Relating Issues Across Unit Operations

The power of projection methods in exploring large databases is demonstrated with the following example, shown in Figure 25.5, where we plot projections of the raw material quality, micronized material quality, and final quality. More than 25 measurements of physical and chemical properties are collected per lot of raw material. These variables are projected on a space defined by latent variables, which allow us to visualize better the process behavior. In the particular example, raw material is

produced at three supplier locations. The raw material properties are within univariate specifications at all locations. Projected on a multivariate space, t_1 vs. t_2 however, they form three clusters: one cluster projects on the low part of the t_1 vs. t_2 plot (negative values of t_2), while the other two clusters at the upper part of the plot (positive values of t_2). One of these two clusters is marked by a small ellipse. This indicates that in a multivariate sense the material possesses slightly different characteristics depending on the location it was produced (covariance structure changes with location). A few of the batches of this raw material were subsequently used for a specific product, and we show its behavior after micronization and filling. The material properties after micronization are also projected on PCs, and it could be observed that the material corresponding to the batches in the small ellipse projects on a different location than the other batches. The filling performance of the material originating from the batches of the small ellipse is different than the rest of the material as well. The conclusion for this example is that the raw material differences propagate in the final quality. Given the large number of variables involved, it is clear that the projection methods provided a very quick diagnostic of the problem. This could have not been possible by dealing with univariate charts. A note here is that although the control ellipses shown are set by default in the vendor software, they are not interpretable when there is clustering; the assumptions for the calculation of these ellipses are for process monitoring

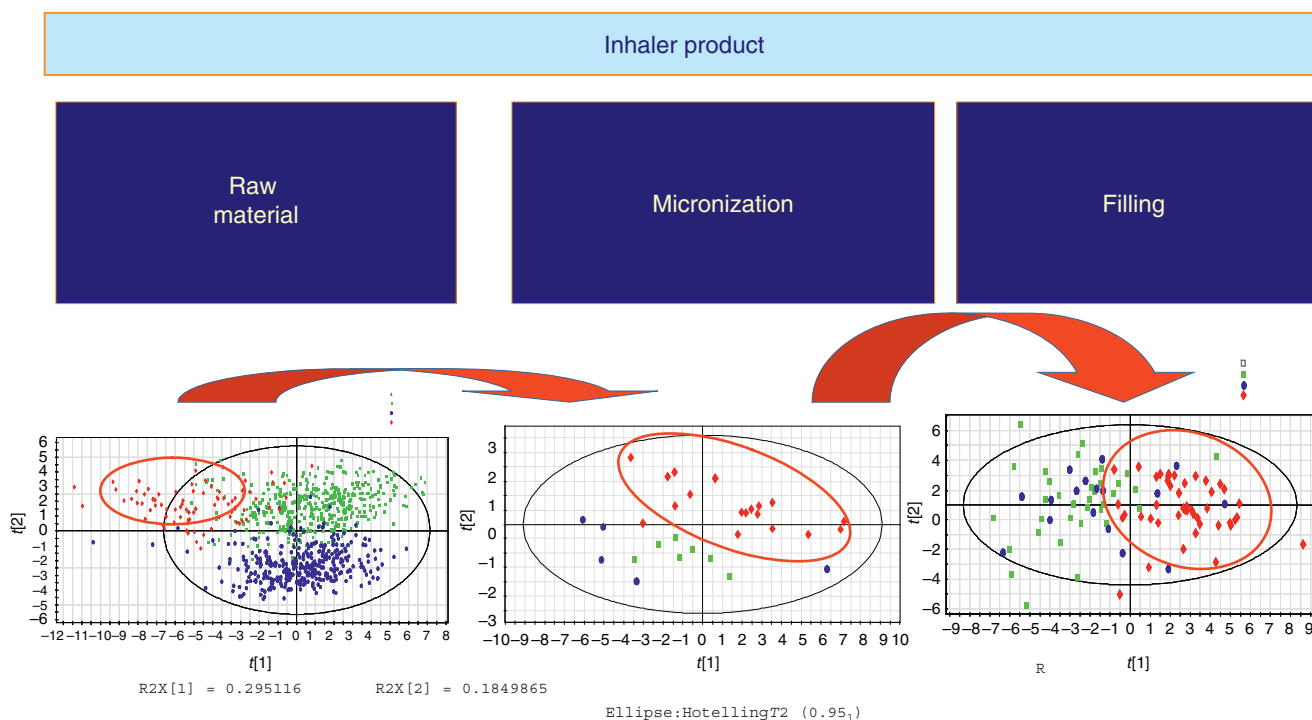


FIGURE 25.5 Representation utilizing projection methods. Raw material properties, micronization properties, and filling performance are projected on a latent variable space. Product batches produced from similar raw material (circled by the small ellipse) have similar filling performance.

and not for process exploration where there is intentional variation such that introduced by DOE.

25.3.3.2 Process Understanding Example 2: Quick Diagnosis of Effect of Raw Material Variability to Granule Characteristics The advantage of PLS is that many highly correlated quality responses can be analyzed simultaneously. The **Y** matrix can contain several parameters related to quality. In this example we wanted to see the correlation between certain active pharmaceutical ingredient (API) physical properties and the granule properties. We chose to use as **Y** the entire tapped density profile and the entire particle size distribution (PSD) profile. As seven variables corresponded to each property profile, we did not have to use any special scaling discussed in the multiblock section. First we use as **X** matrix the API properties and used PLS between **Y** and **X** to relate API variability to the variability of the granule properties. Figure 25.6 shows the variability explained for each variable in the tapped density profile and granule size profile in a cross-validated model. ■ is the direct fitting variability and ▨ the cross-validated variability. One can very clearly

visualize that the tapped granule density is very highly correlated with the specific API physical property variability (as a high percentage of variability of the tapped density profile is explained by the variability of the API property values).

25.4 MULTIVARIATE STATISTICAL PROCESS CONTROL

From routine operation we can establish acceptable limits of good process behavior. On a t_1 vs. t_2 plane, such limits will take the form of an ellipse. When the process is in statistical control, the points will be within the ellipse. If there is a problem in the process, the points will plot out of the ellipse. In Figure 25.4b the ellipse is calculated based on PCA on the data from good operation. Notice that while for raw correlated data the ellipse is tilted, indicating correlation (Figure 25.2), this is not the case when it is calculated for the PCs that are orthogonal.

To monitor the process in real time however, it would have become cumbersome to have to plot all combinations of PCs

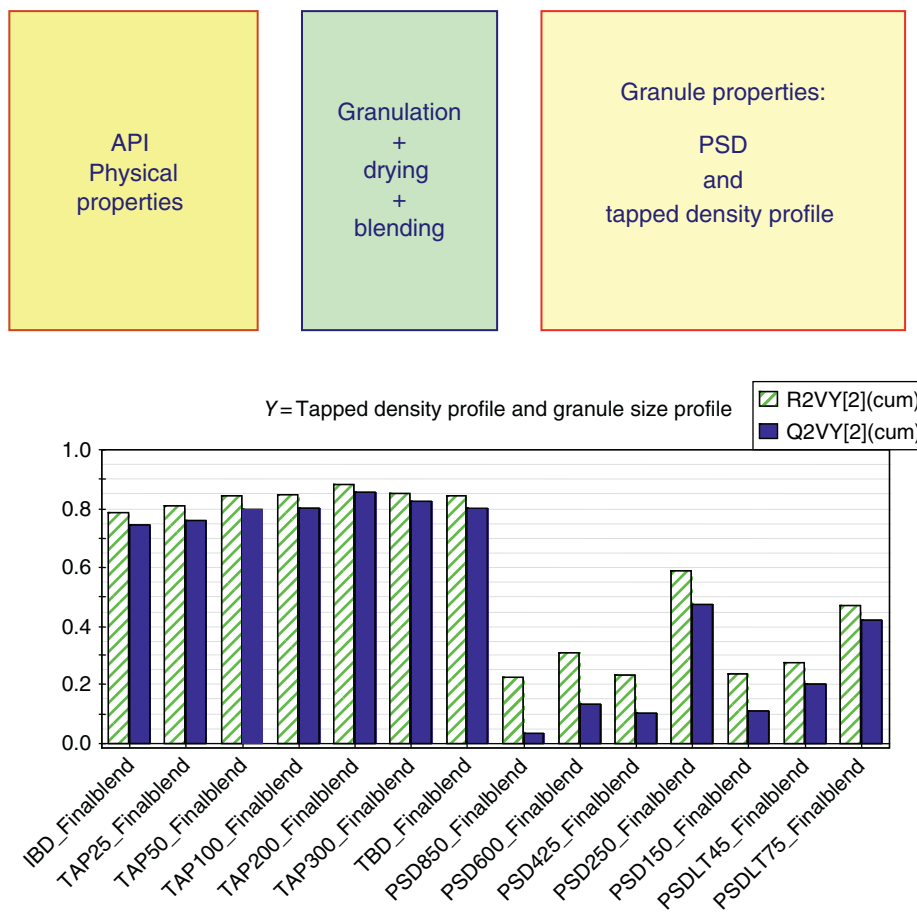


FIGURE 25.6 PLS is a powerful tool to help visualize multiple responses. Here we study the effect of API physical property characteristics on the tapped density profile and the granule size distribution.

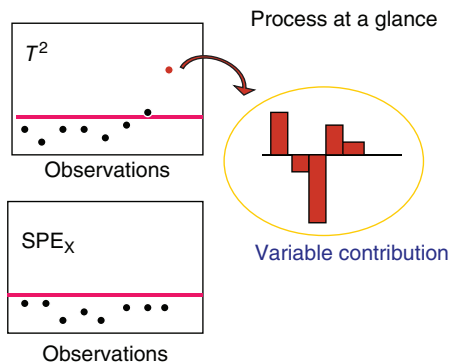


FIGURE 25.7 Two charts (T^2 and SPE) are required to give a picture of the wellness of the process at a glance. The hundreds of measurements collected from the process variables at each instant in real time are translated into one point for the T^2 chart and one point for the SPE chart.

(even if we had four components, we would need six charts). A statistic (Hotelling's T^2) can be calculated, and the overall variability of the main events of the system can be monitored with a single chart, like the one shown at the upper left corner of Figure 25.7. The red line corresponds to acceptable performance. For the case of two components, this solid line corresponds to the perimeter of the ellipse of Figure 25.4b. For three components it would correspond to the surface of an ellipsoid, and for four components the surface of a hyperellipsoid.

Hotelling's T^2 for scores is calculated as

$$T_A^2 = \sum_{i=1}^A \frac{t_i^2}{\lambda_i} = \sum_{i=1}^A \frac{t_i^2}{s_{t_i}^2} \quad (25.3)$$

where

$s_{t_i}^2$ is the estimated variance of the corresponding latent variable t_i .

This chart essentially checks if a new observation vector of measurements on k process variables projects on the hyperplane within the limits determined by the reference data.

As mentioned above, the A PCs explain the main variability of the system. The variability that cannot be explained forms the residuals (squared prediction error [SPE]). This residual variability is also monitored and a control limit for typical operation is being established. By monitoring the residuals (Figure 25.7, bottom left), we test that the unexplained disturbances of the system remain similar to the ones observed when we derived the model. For example, a model derived with data collected in the summer may not be valid in the winter when different disturbances affect the system (cooling water temperatures are different, equipment walls

are colder, valves may reach limits in capacity of providing heating agent, etc.). It is therefore important to check the validity of the model by checking the type of disturbances affecting the system. When the residual variability is out of limit, it is usually an indication that a new set of disturbances have entered the system; it is necessary to identify the reason for the deviation, and it may become necessary to change the model.

SPE_X is calculated as

$$SPE_X = \sum_{i=1}^k (x_{\text{new},i} - \hat{x}_{\text{new},i})^2 \quad (25.4)$$

where

\hat{x}_{new} is computed from the reference PLS or PCA model.

Notice that SPE_X is the sum over the squared elements of a row in matrix \mathbf{E} in Eq. (25.1). This latter plot will detect the occurrence of any new events that cause the process to move away from the hyperplane defined by the reference model.

25.4.1 Calculation of Chart Control Limits

For Hotelling's T^2 chart (either for PCA or for PLS), an upper control limit based on the A first PCs and derived from n observations is obtained using the F -distribution and given by

$$T_{A, \text{UCL}}^2 = \frac{(n^2 - 1)A}{n(n - A)} F_{\alpha(A, n - A)} \quad (25.5)$$

where

$F_{\alpha(A, n - A)}$ is the upper 100 $\alpha\%$ critical point of the F -distribution with $(A, n - A)$ degrees of freedom.

For the SPE_X chart limits can be computed using approximate results from the distribution of quadratic forms. The critical upper 100(1 - α)% confidence interval on SPE is given as

$$\theta_1 \left[\frac{z_\alpha \sqrt{2\theta_2 h_0^2}}{\theta_1} + \frac{\theta_2 h_0 (h_0 - 1)}{\theta_1^2} + 1 \right]^{1/h_0} \quad (25.6)$$

where

z_α is the unit normal deviate corresponding to the upper 100(1 - α)%.

α is the chance taken to incorrectly declare a fault because of the type I error:

$$\theta_i = \sum_{j=A+1}^m \lambda_j^i = \text{Tr}(\mathbf{E}^i) \text{ for } i=1,2,3, \quad (25.7)$$

where

λ_i is the i th eigenvalue referring to the covariance matrix and $h_0 = 1 - (2\theta_1\theta_3/3\theta_2^2)$.

Nomikos and MacGregor [16] used an approximation based on the weighted chi-squared distribution ($g\chi^2(h)$). They suggested a simple and fast way to estimate the g and h , which is based on matching moments between a $g\chi^2(h)$ distribution and the reference distribution of *SPE* at any time interval. The mean ($\mu = gh$) and the variance ($\sigma^2 = g^2(2h)$) of the distribution are equated to the sample mean (b) and variance (v) at each time interval. Therefore, the g and h are estimated from the equations $\hat{g} = v/2b$ and $\hat{h} = 2b^2/v$.

Hence, the upper control limit on the *SPE* at significance level α is given by

$$\frac{v}{2b} \chi_{\alpha}^2 \left(\frac{2b^2}{v} \right) \quad (25.8)$$

It should be emphasized that the models built for process monitoring model only common cause variation and not causal variation. The main concepts behind the development and use of these *MSPC* charts based on latent variables for monitoring continuous processes were laid out in early 1990s [10].

These two charts (T^2 and *SPE*) are two complementary indices; together they give a picture of the wellness of the system at a glance. With this methodology, the hundreds of measurements collected from the process variables at each instant in real time are translated into one point for the T^2 chart and one point for the *SPE* chart (these two points summarize the process at that instant). As long as the points are within their respective limits everything is in order. Once a point is detected out of limit, then the so-called *contribution plots* can be utilized, which give us a list of all the *process* variables that mainly contribute to the out of limit point and hence allow us to diagnose the process problem immediately. Contribution plots can be derived for out of limit points in both charts:

Contributions to SPE: When an out-of-control situation is detected on the *SPE* plot, the contribution of each variable of the original data set is simply given by $(x_{\text{new},j} - \hat{x}_{\text{new},j})^2$. Variables with high contributions are investigated.

Contributions to Hotelling's T^2 : Contributions to an out of limit value in Hotelling's T^2 chart are obtained as follows: a bar plot of the normalized scores (t_i/s_{t_i}) [2] is plotted, and

scores with high normalized values are further investigated by calculating variable contributions. A variable contribution plot indicates how each variable involved in the calculation of that score contributes to it. The contribution of each variable of the original data set to the score of component q is given by

$$\begin{aligned} c_j &= p_{q,j}(x_j - \bar{x}_j) \text{ for PCA} \\ \text{and} \\ c_j &= w_{q,j}(x_j - \bar{x}_j) \text{ for PLS} \end{aligned} \quad (25.9)$$

where

c_j is the contribution of the j th variable at the given observation.

$p_{q,j}$ is the loading.

$w_{q,j}$ is the weight of this variable to the score of the PC q . \bar{x}_j is its mean value (which is zero form mean centered data).

Variables on this plot that appear to have the largest contributions to it but also the same sign as the score should be investigated (contributions of the opposite sign will only make the score smaller). When there are K scores with high values, an "overall average contribution" per variable is calculated over all the K scores [7].

As an example consider Figure 25.4b, which illustrates that two clusters of points were observed on a t_1 vs. t_2 plot. The use of contribution plots may help to investigate which variables have contributed to the move from point 8 to 9. So Eq. (25.9) would give the contribution of variable j to the move of the score values between two observations (say, 8 and 9), for component q is calculated as [15] $p_{jq} \times (x_{j,9} - x_{j,8})$ for PCA.

Utilizing contribution plots, when an abnormal situation is detected, the source of the problem can be diagnosed such that corrective action is taken. Some actions can be taken immediately in real time. Others may require interventions to the process. One such example of an abnormal situation appeared in a reactor, in which the reactor temperature should be controlled in an exothermic reaction to 50 °C. On a very hot day, the charts indicated abnormalities. Contribution plots pointed to a break in the correlation of cooling water flow and reactor temperature. It turned out that although the cooling water valve was fully open, it could not cope with the demand, as the cooling water was warmer. The valve had to be resized. *MSPC* pointed to a problem that had to be corrected. Therefore the contribution plots are very important tools in understanding factors influencing the process during production and help in an "ongoing process understanding" philosophy.

25.4.2 How to Utilize Control Charts

Since latent variable-based control charts were introduced, their use in industry is increasing. The charts answer the need of process industries for a tool that allows them to utilize the massive amounts of data being collected on hundreds of process variables, as well as the spectral data collected from modern analyzers.

Latent variable control charts can be constructed to monitor either a group of response variables \mathbf{Y} (e.g. product quality variables) or a group of predictor variables \mathbf{X} (process variables). For example, multivariate charts can be constructed to assess the consistency of the multivariate quality of raw materials, \mathbf{Z} , as well as to test the final product, \mathbf{Y} , for consistent quality. If there is spectral analysis on some of the materials, then multiblock concepts, discussed later can be used.

A very important advantage of latent variables is that they can be used to monitor predictor variables taking into account their effect on the response variables. A model is built to relate \mathbf{X} and \mathbf{Y} using available historical or specially collected data. Monitoring charts are then constructed for future values of \mathbf{X} . This approach means that the process performance can be monitored even at times when the product quality measurements, \mathbf{Y} , are not available.

The main approach of SQC methods developed throughout the statistical literature has been to monitor only product quality data (\mathbf{Y}) and, in some cases, a few key process variables (\mathbf{X}). However, often hundreds of process variables are measured much more frequently (and usually more accurately) than the product quality data. So monitoring the process data is expected to supply much more information on the state of the process and supply this information more frequently. Furthermore, any special events that occur will also have their fingerprints in the process data. So, once a special event is detected, it is easier to diagnose the source of the problem as we are dealing directly with the process variables. On the contrary, control charts on the product variables only indicate that the product properties are no longer consistent with specification but they do not point to the process variables responsible for this.

Control charts on process variables are useful in multistep operations when quality data are not available between successive steps. For example, if a catalyst is conditioned in a batch process before being used for polymer production, the quality of the catalyst (success of conditioning) is assessed by its performance in the subsequent polymer production. It would be useful to know if the catalyst will produce good product before using it; monitoring the batch process variables with a latent variable chart would give early detection of poor quality product. Similarly, the few properties measured on a product sometimes are not sufficient to define product performance for several different customers. For example, if only viscosity of a polymer is measured, end-use applications that depend

on chemical structure (e.g. branching, composition, end-group concentration) are unlikely to receive good material. In these cases the process data may contain much more information about events with special causes that affect the hidden product quality variables.

The philosophy applied in developing MSPC procedures based on projection methods is the same as that used for the univariate or multivariate Shewhart charts. An appropriate reference set is chosen, which defines the normal operating conditions for a particular process. Future values are compared against this set. A PCA or PLS model is built based on data collected from periods of plant operation when performance was good. Periods containing variations due to special events are omitted at this stage. The choice and quality of this reference set is critical to the successful application of the procedure.

25.5 BATCH PROCESS MONITORING

Figure 25.3 shows schematically the nature of batch process trajectories, which are nonlinear and dynamic. Modeling batch operations requires taking into account their nonlinear dynamic nature. The methodology for developing multivariate control charts based on latent variables for batch process monitoring was initially presented by Nomikos and MacGregor [16–18] in a series of landmark papers. Any operation of finite duration, like batch granulation, batch drying, blending/mixing additives for a finite time, fermentation, batch distillation, drying, etc., can be modeled by the same methodology. This section will present the main issues that need to be addressed in batch empirical modeling and will also give references to publications where these issues are discussed in detail.

25.5.1 Modeling of Batch Process Data

Most of the processes in the pharmaceutical industry are batch processes. Collecting real-time data during a batch process generates very large data sets. The top of Figure 25.8 gives the possible measurements that could be collected for a batch process. Information may be collected, at different time intervals for the duration of the batch, for several process variables like agitation rate (rpm), pH, cooling agent flow (F), and temperatures in different locations in the reactor (T_1, T_2, T_3). Data may also be collected in the form of spectra from real-time analyzers, like near infrared (NIR). Finally information may be available on raw material analysis, recipe, other preprocessing data, even information on who was the operator on shift, and which vessel was used.

Historical data collected from a batch process had traditionally been represented by a three-dimensional data array $\underline{\mathbf{X}}$ where a matrix $\underline{\mathbf{X}}$ ($I \times J \times K$) indicates that J process

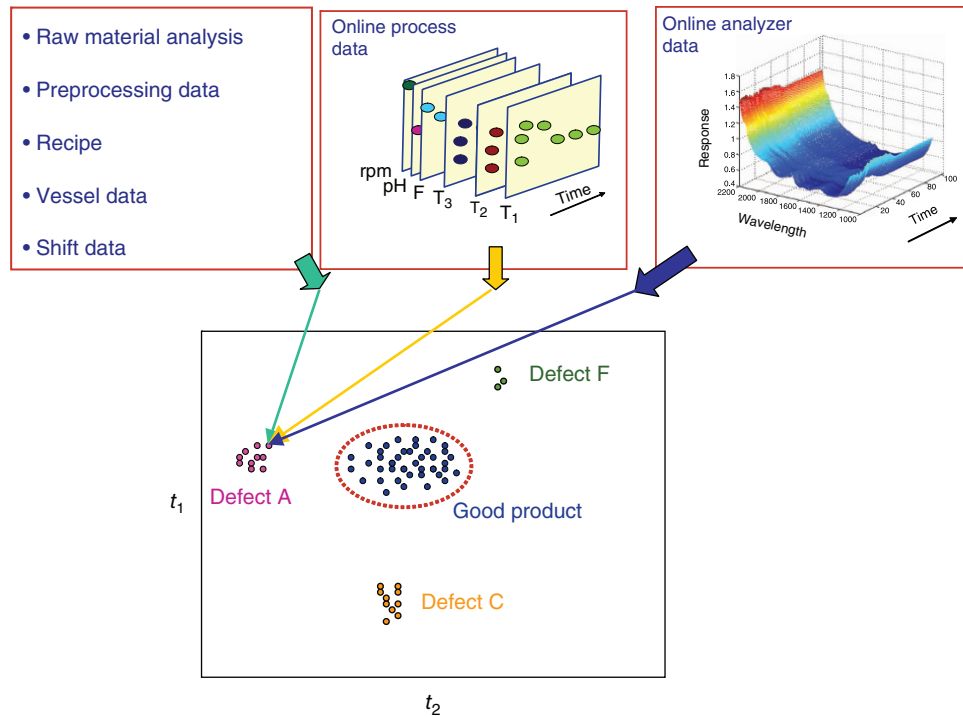


FIGURE 25.8 Data generated during batch runs are projected on a lower-dimensional space, defined by two principal components, t_1 and t_2 . Each point on the plane corresponds to one batch run, that is, each point is the summary of the hundreds of measurements taken during the run.

variables are measured at K time intervals, or K aligned observation numbers (AON), for each one of I batches. Kourti [19, 20] discussed that in practice it is not necessary that the same number of measurements are available for all the variables for the duration of the batch process. Some variables may not be present or measured for the full duration of the batch. Furthermore, the frequency of measurements may be different due to several reasons: (i) some variables may be measured more frequently than others (i.e. some every minute and others every 15 minutes), and (ii) certain phases in the process may be sampled more frequently to catch important phenomena (in emulsion polymerization, particle nucleation occurring at the very first few minutes in the reaction determines the number of particles and the PSD; one may need to capture this with more frequent sampling at those stages). Therefore, Kourti argued that the data set in such situations does not form a complete cube, but rather a cube where some columns are missing (Figure 25.9). Consequently the methods that are used to model batch processes should be capable of modeling the structure of this incomplete cube. There are several methods for modeling three-way data. The choice of the method depends on the use of the model (i.e. prediction of final quality, monitoring, process control) and the types of the data sets available. Critical discussions on batch process modeling procedures for robust process monitoring, fault detection, and control can be found in selected publications [16–23].

The method presented by Nomikos and MacGregor [16] is termed in the literature “batchwise unfolding” and is capable of modeling the incomplete cube structure. Furthermore, it is capable of modeling three-way structures generated when formulating the control problem of batch processes using latent variables, which is discussed later. The method unfolds the three-dimensional structure into a two-dimensional array. In this new array, different time slices are arranged next to each other; variables observed at a given time interval are grouped in one time slice; the number of variables in each time slice may vary. Figure 25.9 shows an example of an unfolded matrix where variable x_4 is not measured at time $t = 1$ and variable x_2 is not measured at times $t = 3–4$. Once the three-way structure is unfolded to a two-way matrix \mathbf{X} , Eqs. (25.1) and (25.2) can be used to model \mathbf{X} using PCA or \mathbf{X} and \mathbf{Y} using PLS.

Multivariate control charts (Hotelling’s T^2 and SPE) can be constructed for batch processes in a straightforward manner. Multivariate charts have superior detection capabilities to univariate charts for batch processes. In the words of a colleague from industry,

In most cases in practice, changes in the covariance structure precede detectable deviations from nominal trajectories. This was the problem that univariate monitoring approaches for batch processes could not address. In most process upsets it is the correlation among the monitored variables that

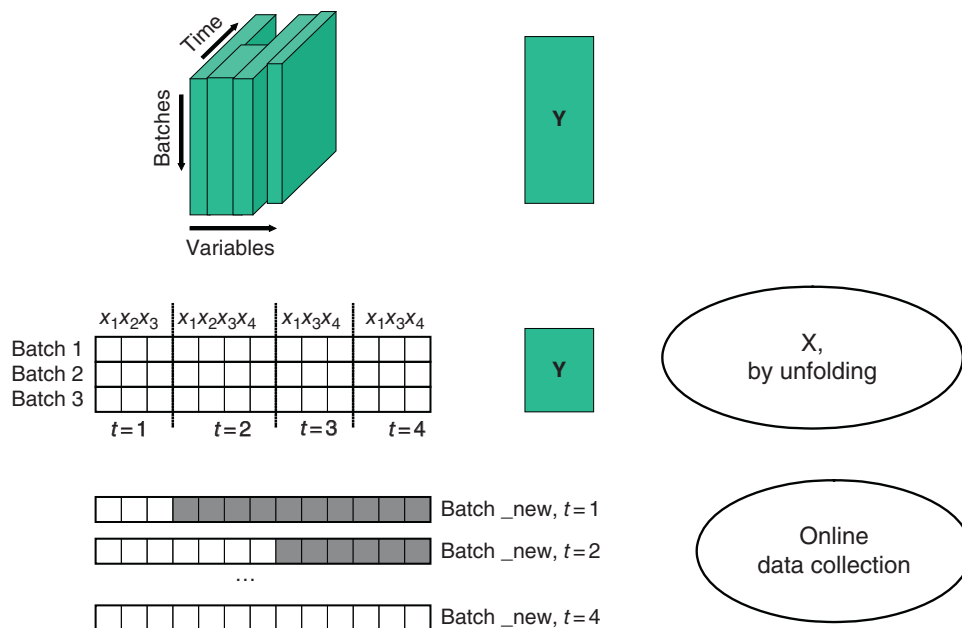


FIGURE 25.9 The structure of data in batch processes.

changes first, and later, when the problem becomes more pronounced, the monitored variables deviate significantly from their nominal trajectories. There are cases where a process upset will change dramatically only the correlation among the variables without causing any of the variables involved to deviate significantly from its nominal trajectory. These particular cases, although rare, can result to significant cost to a company since they can go unnoticed for long periods of time (usually they are detected from a customer complaint). (P. Nomikos 2002; personal communication)

25.5.2 Alignment of Batch Processes of Different Time Duration

Sometimes batches have different time duration. In other words, using the same recipe, it may take different time to achieve the same conversion. This is due to the fact that time is not the deciding factor for the completion of the batch. Sometimes the deciding factor may be the rate by which a certain reactant is added. One important stage before modeling batch process data is the alignment or synchronization of the data, such that they are expressed against the correct aligning factor (which may not be time). With alignment or synchronization we must achieve the following: (i) Establish common start points at different phases of the run. For example, we could define that the first observation in all the runs for phase one of the reaction will correspond to the start of the monomer feed, for the second phase to the initiator injection, etc. (ii) Match the shape of the trajectories of key variables. Once the shapes match, it is not necessary that the length of the batches match [20].

A critical discussion of the various synchronization approaches can be found in Kourti [20]. Attempts for batch data alignment involved the use of the cumulative amount of reactant added to the reactor as an indicator variable, where the variable trajectories were expressed as a function of the indicator variable, rather than time. The extent of the reaction was also used as an indicator variable [24]. Dynamic time warping based on speech recognition methods was also suggested for alignment procedures. Another approach was the use of total time as a variable in the Z matrix as extra information to describe the batch [25]. P.A. Taylor (Chemical Engineering, McMaster, 1998; personal communication) suggested to include the cumulative warping, up to a given warped observation, as a new variable trajectory; his argument was that this would provide much richer information on the state of the batch by comparing it with the “typical batch” and would provide it in real time, rather than waiting for the batch to finish so that we can calculate the total time; the cumulative time spent could be used as extra trajectory in the case of alignment with an indicator variable. This suggestion was used later with excellent results and also provided the basis to designing batches with desired duration. The cumulative warp can be used as an extra variable to take into account time effects on batch quality when batches were synchronized by dynamic time warping, while García-Muñoz et al. [26] used the cumulative time when batches were synchronized by the indicator variable approach. Provided that an indicator variable exists (or can be constructed by non-linear transformations from other variables and/or process knowledge), the indicator variable approach is usually

chosen as the simplest and most convenient for industrial applications.

25.5.3 Mean Centering and Scaling the Incomplete Cube

Mean centering the two-way matrix, formed by batchwise unfolding the three-way data, is equivalent to subtracting from each variable trajectory its average trajectory over the I batches, thus converting a nonlinear problem to one that can be tackled with linear methods such as PCA and PLS.

When the three-way data form a full cube, that is, an $\underline{\mathbf{X}}(I \times J \times K)$ matrix, it is common practice to autoscale the two-way matrix formed by unfolding $\underline{\mathbf{X}}$ (that is, divide each column by its standard deviation). This accomplishes two things: (i) gives an equal weight to all time periods and consequently does not give high weights to noisy phases or underweight a variable in tight control (after all a variable is in tight control either because it is important to the product quality or because of safety and/or environmental concerns) and (ii) for the case of complete cube gives an equal weight to all the variables considered. However, in the case where variables are sampled less frequently or are not present for the full run, the weights have to be adjusted accordingly, depending on the objective. To give equal weight to all variables for example, after autoscaling, each column corresponding to variable j must be divided by $\sqrt{K_j}$ where K_j is the number of times that variable j was sampled in the run. In the example of Figure 25.9, after autoscaling the two-way matrix, all columns corresponding to x_1 and x_3 must be divided by $\sqrt{4}$, to x_2 by $\sqrt{2}$, and to x_4 by $\sqrt{3}$.

25.5.4 Online Monitoring of Batch Processes

Each batch run has a finite duration and the process variables exhibit a dynamic behavior during the run. This means that not only the autocorrelation structure of each variable changes during the run but also the cross-correlation of the variables changes as well. Models utilizing batchwise unfolding take into account this changing covariance structure of variables and time for the batch duration. For online monitoring, it is the structure of each evolving batch that it is compared against the typical behavior, as modeled by the training set of batches. The procedure for online monitoring of batch processes is slightly more complicated than that for continuous process because of the following:

- For the online monitoring of continuous processes, at every time instant, we have a vector of new observations $\mathbf{X}_{\text{new}}^T$ that has a length equal to the number of columns in the model matrix \mathbf{X} (and occasionally some measurements may be missing due to sensor failure or other reasons).

- In batch process monitoring, we have a vector with a length equal to the number of columns in the unfolded \mathbf{X} *only* when the batch run has finished. At any other time, data are missing from this vector, simply because they have not been collected yet. In Figure 25.9, the new vector `batch_new` is shown for different time intervals; gray areas have not been collected yet. Of course, the part with the collected data may also have missing data due to sensor failure or other reasons.

Therefore the score calculations and the limits for the multivariate control charts have to be developed in such a way that they take these “incomplete” measurement vectors into account. The procedure for the development of the multivariate control charts for the duration of the batch was outlined by Nomikos and MacGregor [16]. To deal with the incomplete measurement vector, several approaches have been suggested. García-Muñoz et al. [27] recently investigated these approaches and demonstrated that using the *missing data* option and solving the score estimation problem with an appropriate method is equivalent to the use of an accurate forecast for the future samples over the shrinking horizon of the remainder of the batch. As PCA can model the covariance structure of the process variables, it facilitates handling of missing data. In batch processes a PCA model describes the variance–covariance structure between variables over the entire batch; in other words there exists information over all combinations of x_{jk} , j being the variable number and k the time interval (for example, one can find how variable 4 at time 6 is related to itself at time 15, but also to variable 8 at time 35). Because of the tremendous structural information built into these multivariate PCA models for batch processes, the missing data option for predicting the future trajectory is shown to yield the best performance by all measures, even from the beginning of the batch.

Provided that there are no faults, for the prediction of the future process variable trajectories, the final scores, and the product quality, these missing data estimation methods are very powerful. They have also been proven critical to the success of the control methods using latent variables. However, for process monitoring and online detection of process faults, all the alternative “filling-in” methods give similar results. When a fault occurs, the model structure is not valid anymore. In that case the differences among the trajectory estimation methods appear to be much less critical since the control charts used in each case are tailored to the filling-in mechanism employed. All the approaches appear to provide powerful charting methods for monitoring the progress of batch processes.

The calculation of monitoring charts and their limits for batch processes is discussed in Nomikos and MacGregor [16], where Hotelling’s T^2 statistic for the analysis of batch process data (called D statistic) is calculated as

$$D = \mathbf{t}_R^T \mathbf{S}^{-1} \mathbf{t}_R \mathbf{I} / (\mathbf{I} - 1)^2 \quad (25.10)$$

where

\mathbf{t}_R is the vector containing the \mathbf{R} retained components of the model.

\mathbf{S} represents the covariance matrix of the \mathbf{R} retained score vectors.

It is mentioned that \mathbf{S} is a diagonal matrix due to the orthogonality of the scores, which is true for the final score estimate (i.e. when the batch run is complete). When computing this statistic for the online monitoring of batches, one should consider that the covariance of the scores changes with time and the scores might become non-orthogonal; therefore Hotelling's statistic should be computed using the correct and complete variance-covariance matrix that corresponds to each time sample. This time-varying variance-covariance is computed using the reference set of batches. Therefore, the estimate of Hotelling's statistic at time k for batch i (D_{ki}) is a function of the estimate of the score vector for the \mathbf{R} retained components at time k ($\hat{\mathbf{t}}_{Rki}$) for batch i and the covariance matrix of the scores at time k (\mathbf{S}_k). Notice that D_{ki} will change depending on the method used to solve the missing data problem (or the option selected to "fill in"), since it is a function of $\hat{\mathbf{t}}_{Rki}$, which has been shown [27] to differ from method to method and option to option. Using this corrected version of Hotelling's statistic dramatically improves abnormality detection.

25.5.5 Industrial Practice

Industrial applications for batch analysis, monitoring, and fault diagnosis have been reported [26–31]. It should be noted here that several companies choose to use the methodology not for real time monitoring but as a tool to *real-time release* of the batch product in the following way: the batch is not monitored as it evolves; rather, immediately after the batch finishes, the process data are passed through the model, and the scores for the complete batch are investigated. If they are within control limits, the product is released. If there is a problem, the product is sent for analysis in the laboratory. This procedure saves the company time and money. The batch run may last two to three hours, but the product analysis may take many more hours. That means that they do not have to waste batches while they are waiting for the results from the laboratory. By checking the process data as soon as the batch is complete, they can detect problems before starting a new batch. Other applications of multiway methods to batch analysis, optimization, and control have been reported and will be discussed later in the corresponding sections. Multiway methods and DOE can be used [32] to determine

optimal process variable trajectories in a batch process in order to obtain a desired quality property.

There is a great potential for applications of multivariate batch analysis in the pharmaceutical industry. It is a superb tool for achieving process understanding. Using it to analyze past historical data will provide the user with summaries of the process history like the one shown in Figure 25.8. The figure shows the projection of several batch runs on a plane, defined by two PCs, t_1 and t_2 . Each point corresponds to one batch run, that is, each point is the summary of the hundreds of measurements taken during the run. For example, for a batch run that lasts 16 hours and where 5 minute averages are collected on 6 process variable trajectories, each point is the summary of 1152 process measurements plus all the spectra scans at all time intervals plus other information on recipe. In the figure one can immediately detect not only clusters of good operation but also clusters corresponding to specific product problems. Notice that the problems in product quality are observable by the projection of the process data and of the measurements taken by real-time analyzers; that is, the problem in the quality is observable without laboratory information. Utilizing contribution plots, one can interrogate the multivariate model and determine combinations of variables and periods of operation that will drive a process away from producing a good product to producing defect A. Such an example is discussed in the literature [21] where information from the process and the raw materials was incorporated in the analysis. Based on this analysis it was possible to determine raw material combinations and processing conditions that will result in a bad product.

In an industrial example, by monitoring the batch process variables, a pending equipment failure was detected [31]. A critical discussion on other batch process modeling and monitoring procedures and other issues related to batch process analysis can be found in two recent papers [19, 20]. A discussion of the various approaches to synchronize runs of different duration can be found in Refs. [14, 19, 20].

25.6 MULTISTAGE OPERATIONS: MULTIBLOCK ANALYSIS

There are multiple steps in pharmaceutical manufacturing and each step may involve multiple unit operations. Having a control chart for each unit rather than one for the whole process could be helpful to operators. However, building a model for each unit operation separately does not consider interactions between unit operations. Such cases can also be addressed with latent variable models. Rather than building a model for each unit, one can build a model for the full process that will take into account the interactions between units and their relative importance to the final product quality by weighting them differently. Then, from this model, individual charts per unit operation can be derived. This way,

interactions between unit operations are preserved. This is the approach of multiblock PLS (MB-PLS).

In the MB-PLS approach, large sets of process variables (\mathbf{X}) are broken into meaningful blocks, with each block usually corresponding to a process unit or a section of a unit. MB-PLS is not simply a PLS between each \mathbf{X} block and \mathbf{Y} . The blocks are weighted in such a way that their combination is most predictive of \mathbf{Y} . Several algorithms have been reported for multiblock modeling, and for a good review, it is suggested that the reader consult [33–35]

Multivariate monitoring charts for important subsections of the process, as well as for the entire process, can then be constructed, and contribution plots are used for fault diagnosis as before. In a multiblock analysis of a batch process, for example, one could have the combination of three blocks (\mathbf{Z} , \mathbf{X} , and \mathbf{Y}); block \mathbf{Z} could include information available on recipes, preprocessing times, and hold times, as well as information of the shifts (which operator was in charge) or the vessels used (i.e. which reactor was utilized); \mathbf{X} would include process variable trajectories; and \mathbf{Y} would be quality. Analysis of this type of data could even point to different ways the operators operate the units and relate product quality to operator, or different process behavior of vessels and identify faulty vessels, etc. The reader is referred to the work of Munoz et al. [21] for detailed examples where the multiblock analysis is utilized in batch processes for troubleshooting.

Several alternative ways to perform multiblock appear in commercial software. One approach that is being frequently used to deal with a data structure of several blocks involves two stages: PCA is performed for each one of the \mathbf{Z} and \mathbf{X} blocks, and then the scores and/or residuals derived from these initial models are related to \mathbf{Y} with a PLS. In an alternative version, PLS is performed between \mathbf{Z} and \mathbf{Y} or between \mathbf{X} and \mathbf{Y} , and the resulting scores are related to \mathbf{Y} . The users should exercise caution because these approaches may fail to take into account combinations of variables from different blocks that are most predictive of \mathbf{Y} . For example, in situations where process parameters in \mathbf{X} are modified to account for variability of raw material properties in \mathbf{Z} (i.e. when \mathbf{X} settings are calculated as a feed-forward control to deviations of \mathbf{Z}), a PLS between \mathbf{Z} and \mathbf{Y} will show that \mathbf{Z} is not predictive of \mathbf{Y} variability; similarly a PLS between \mathbf{X} and \mathbf{Y} will show that \mathbf{X} is not predictive of \mathbf{Y} ; a MB-PLS of $[\mathbf{Z}, \mathbf{X}]$ and \mathbf{Y} will identify the correct model. Finally MB-PLS handles missing data in a very effective way.

As might be expected in multistage continuous processes, there can be significant time delays between the moment an event occurs in one unit (and therefore affects the variables of that unit) and the moment its effect will become obvious on a product variable at the end of the process. These delays significantly affect the interaction and correlation structures of the process variables and need to be handled by lagged variables created from the original process variables. Data can be time-shifted to accommodate time delays between process units.

In some multistage operations, the path of the product through the various process units can be traced easily, and eventually one can relate a specific lot number to several process stages (via a MB-PLS). In such cases the process conditions of these units can be used to predict the quality of the product. There are situations, however, where a product (or the composition of the effluent stream of a process) is a result of a multistage operation but its path cannot be traced clearly, due to mixing of streams from several parallel units in one vessel and then splitting to a number of other vessels. A discussion on monitoring difficult multistage operations can be found in Kourti [20]. In those cases the best alternative in order to achieve consistent operation is to monitor each unit, separately, by a PCA model. By assuring a consistent operation per unit, one hopes for a consistent product. Once an unusual event is detected in one unit, one may decide not to mix the product further or investigate lab quality before proceeding to the next stage.

25.7 PROCESS CONTROL TO ACHIEVE DESIRED PRODUCT QUALITY

The term “control” appeared in the pharmaceutical literature to describe a variety of concepts such as end-point determination, feedback control, SPC, or simply monitoring. Process control refers to a system of measurements and actions within a process intended to ensure the output of the process conforms with pertinent specifications.

In this work, we use some terms related to process control with the following definition:

- Feedback control: To indicate that we are reactive, that is, the corrective action is taken on the process based on information from the process output (as, for example, measurements on product or intermediate quality at given time).
- Feed-forward control: To indicate that we are proactive, that is, the process conditions are adjusted based on measured deviations of the input to the process (as, for example, information on raw material).

25.7.1 Feed-Forward Estimation of Process Conditions

The concept of adjusting the process conditions of a unit based on measured disturbances (feed-forward control) is a concept well known to the process systems engineering community for several decades. The methodology is also used in multistep (multiunit) processes where the process conditions of a unit are adjusted based on information of the intermediate quality achieved by the previous unit (or based on raw material information).

An example of a feed-forward control scheme in the pharmaceutical industry where multivariate analysis (MVA) was involved is described by Westerhuis et al. [36]. The authors related crushing strength, disintegration time, and ejection force of the tablets to process variables from both the wet granulation and tableting steps and the composition variables of the powder mixture. They also included physical properties of the intermediate granules. The granule properties may differ from batch to batch due to uncontrolled sources as for example humidity or temperature. This model is then used for each new granulation batch. A feed-forward control scheme was devised that can adjust the variables of the tableting step of the process based on the intermediate properties to achieve desirable final properties of the tablets.

To the author's knowledge, there are several unpublished examples in the chemical and other industries where information on the raw data \mathbf{Z} is used to determine the process conditions \mathbf{X} or $\underline{\mathbf{X}}$ in order to achieve the desired quality \mathbf{Y} utilizing projection methods. Sometimes such information from \mathbf{Z} may simply be used to determine the length of the run, while in other cases it may be a multivariate sophisticated scheme that determines a multivariate combination of trajectories for the manipulated variables. To achieve this, historical databases can be used to develop multiblock models \mathbf{Z} , \mathbf{X} (or $\underline{\mathbf{X}}$), and \mathbf{Y} .

25.7.2 End-Point Determination

There have been reports in the literature where real-time analyzers are used for "end-point detection" or "end-point control." In most of these situations, a desired target concentration is sought, for instance, % moisture in drying operations.

An example is described by Findlay et al. [37] where NIR spectroscopy is used to determine granulation end point. The moisture content and particle size determined by the NIR monitor correlates well with offline moisture content and particle size measurements. Given a known formulation, with predefined parameters for peak moisture content, final moisture content, and final granule size, the NIR monitoring system can be used to control a fluidized bed granulation by determining when binder addition should be stopped and when drying of the granules is complete.

25.7.3 Multivariate Manipulation of Process Variables

It was discussed in Section 25.2.2 that regulating only the final value of a property (or even several properties) is not sufficient. In other words, end-point control may not be sufficient. The process signatures are equally important. These process signatures should be regulated in a correct multivariate way, not simply on a univariate basis. It is possible that two batch runs produce products with different quality, even if the trajectory (path to end point) of one quality variable

follows the same desired path in both of the runs. This will happen if the covariance structure of the trajectory of this variable with the trajectories of the rest of the process variables (temperatures, agitation rate, reactant addition) is different for these two batches. This concept is very important both in control and in scale-up. Latent variable methodology allows for taking into consideration the process variable trajectories in a multivariate way.

Control of batch product quality requires the online adjustment of several manipulated variable trajectories. Traditional approaches based on detailed theoretical models are either based on nonlinear differential geometric control or online optimization. Many of the schemes suggested in the literature require substantial model knowledge or are computationally intensive and therefore difficult to implement in practice. Empirical modeling offers the advantage of easy model building.

Lately latent variable methods have found their way to control batch product quality and have been applied in industrial problems. Latent variable methodology allows for taking into consideration the process signatures in a multivariate way for end-point detection problems. Marjanovic et al. [38] describe a preliminary investigation into the development of a real-time monitoring system for a batch process. The process shares many similarities with other batch processes in that cycle times can vary considerably, instrumentation is limited, and inefficient laboratory assays are required to determine the end point of each batch. The aim of the work conducted in this study was to develop a data-based system able to accurately identify the end point of the batch. This information can then be used to reduce the overall cycle time of the process. Novel approaches based upon multivariate statistical techniques are shown to provide a soft sensor able to estimate the product quality throughout the batch and a prediction model able to provide a long-term estimate of the likely cycle time. This system has been implemented online, and initial results indicate that it offers the potential to reduce operating costs.

In another application [39], latent variable methodology was used for soft sensor development that could be used to provide fault detection and isolation capabilities and that it can be integrated within a standard model predictive control framework to regulate the growth of biomass within a fermenter. This model predictive controller is shown to provide its own monitoring capabilities that can be used to identify faults within the process and also within the controller itself. Finally it is demonstrated that the performance of the controller can be maintained in the presence of fault conditions within the process.

Work has also been reported for complicated control problems where adjustments are required for the full manipulated variable trajectories [40]. Control through complete trajectory manipulation using empirical models is possible by controlling the process in the reduced space (scores) of a latent variable model rather than in the real space of the manipulated

variables. Model inversion and trajectory reconstruction are achieved by exploiting the correlation structure in the manipulated variable trajectories. Novel multivariate empirical model predictive control strategy (LV-MPC) for trajectory tracking and disturbance rejection for batch processes based on dynamic PCA models of the batch processes has been presented. The method presented by Nomikos and MacGregor [16] is capable of modeling three-way structures generated when formulating the control problem of batch processes using latent variables.

25.7.4 Setting Raw Material Multivariate Specifications as a Means to Control Quality

Dushesne and MacGregor [41] presented a methodology for establishing multivariate specification regions on raw/incoming materials or components. The thought process here is that if the process remains fixed, we should control the incoming material variability. PLS is used to extract information from databases and to relate the properties of the raw materials supplied to the plant and the process variables at the plant to the quality measures of the product exiting the plant. The specification regions are multivariate in nature and are defined in the latent variable space of the PLS model. The authors emphasize that although it is usually assumed that the raw material quality can be assessed univariately, that is, by setting specification limits on each variable separately, this is valid only when the raw material properties of interest are independent from one another. However, most of the times, the properties of products are highly correlated. In other words, treating the raw material properties in a univariate way for two properties, it would mean that (referring to Figure 25.1) while we can process only material that falls in the ellipse, we agree to buy material from the supplier with the specifications set in the square; that is, we agree to use material that, we know in advance, it will not perform well.

To develop models to address the problem, MB-PLS is used for \mathbf{Z} , \mathbf{X} , and \mathbf{Y} ; \mathbf{Z} contains measurements on N lots of raw material data from the past; \mathbf{X} contains the steady-state processing conditions used to process each one of the N lots; \mathbf{Y} contains final product quality for these N lots. The methodology could be easily extended to batch process \mathbf{X} .

It should become one of the priorities in industries to express the raw material orders as a multivariate request to the supplier.

25.8 OTHER APPLICATIONS OF LATENT VARIABLE METHODS

25.8.1 Exploiting Databases for Causal Information

Recently there has been a lot of interest in exploiting historical databases to derive empirical models (using tools such as neural network regression or PLS) and use them for process

optimization. The idea is to use already available data rather than collecting new through a DOE. The problem is that for process optimization, causal information must be extracted from the data so that a change in the operating variables can be made that will lead to a better quality product or higher productivity and profit. However, databases obtained from routine operation contain mostly noncausal information. Inconsistent data, range of variables limited by control, noncausal relations, spurious relations due to feedback control, and dynamic relations are some of the problems the user will face using such happenstance data. These are discussed in detail in the section “Hazards of fitting regression equations to happenstance data” in Ref. [42] where the advantage of experimental designs as a means of obtaining causal information is emphasized. In fact in a humorous way, the authors warn the young scientists that they need a strong character to resist the suggestion of their boss to use data from past plant operation every time they suggest performing designed experiments to collect data.

In spite of this several authors have proposed approaches to optimization and control based on interpolating historical bases. However, in all these cases, their success was based on making strong assumptions that allowed the database to be reorganized and causal information to be extracted. One approach was referred to as “similarity optimization,” which combined multivariate statistical methods for reconstructing unmeasured disturbances and nearest neighbor methods for finding similar conditions with better performance. However, it too was shown to fail for many of the same reasons. In general, it was concluded that one can only optimize the process if there exist manipulated variables that change independently of the disturbances and if disturbances are piecewise constant, a situation that would be rare in historical process operations.

The reader should therefore exercise caution in using historical databases when it comes to retrieving causal information. However, databases obtained from routine operation are a great source of data for building monitoring schemes.

25.8.2 Product Design

Given the reservations about the use of historical databases, one area where some success has been achieved is in identifying a range of process operating conditions for a new grade of product with a desired set of quality properties and in matching two different production plants to produce the same grade of product. If fundamental models of the process exist, then these problems are easily handled as constrained optimization problems. If not, optimization procedures based on response surface methodology can be used. However, even before one performs experiments, there exists information within the historical database on past operating conditions for a range of existing product grades.

In this case the historical data used are selected from different grades and therefore contain information on variables

for several levels of past operation (i.e. there is intentional variation in them and are not happenstance data). The key elements in this empirical model approach is the use of latent variable models that both reduce the space of \mathbf{X} and \mathbf{Y} to a lower-dimensional orthogonal set of latent variables and provide a model for \mathbf{X} as well as \mathbf{Y} . This is essential in providing solutions that are consistent with past operating policies. In this sense, PCR and PLS are acceptable approaches, while MLR, neural networks, and RRR are not.

The major limitation of this approach is that one is restricted to finding solutions within the space and bounds of the process space \mathbf{X} defined by previously produced grades. There may indeed be equivalent or better conditions in other regions where the process has never been operated before and hence where no data exists. Fundamental models or more experimentation would be needed if one hopes to find such novel conditions.

A very good discussion on these issues can be found in Garcia et al. [26]. The authors illustrate a methodology with an industrial batch emulsion polymerization process where the batch trajectories are designed to satisfy certain customer requirements in the final properties of the polymer while using the minimal amount of time for the batch run. The cumulative time, or used time, is added as an extra variable trajectory after the alignment of the batches.

25.8.3 Site Transfer and Scale-Up

Product transfer to different sites and scale-up fall into the same class of problems: one needs to estimate the process operating conditions of plant B in order to produce the same product that is currently produced in plant A.

Attempts have been made to solve such problems with latent variable methods utilizing historical data from both locations from transferring other products.

The main points to keep in mind when addressing such a problem are the following:

- The quality properties of the product should always be checked within a multivariate context because univariate charts may be deceiving. The multivariate quality space for both sites should be the same. Correct product transfer cannot be achieved by comparing end-point quality on univariate charts from the two sites (or from pilot scale and manufacturing). The product quality has to be mapped from site to site in a multivariate way (the products in both sites have to project on the same multivariate space).
- The end-point quality may not be sufficient to characterize a product. The path to end product is important. Whenever full mechanistic models exist, these models describe the phenomena that are important for the process and therefore determine this path. When changing

sites, the full mechanistic model will describe the desired path in the new site taking into account size, mass and energy balances, and/or other phenomena related to the process. When mechanistic models do not exist, this mapping of the “desired process paths” or “process signatures” has to happen with empirical data.

A methodology has been developed for product transfer and scale-up based on latent variables [43]. The methodology utilizes databases with information on previous products and their corresponding process conditions from both sites. The two sites may differ in equipment, number of process variables, locations of sensors, and history of products produced.

25.9 QUALITY BY DESIGN

25.9.1 Design Space: Expressing Quality as a Function of Input Material Attributes and Process Parameters

ICH has adopted the concept of design space. The definition of design space as “the multidimensional combination and interaction of input variables (e.g., material attributes) and process parameters that have been demonstrated to provide assurance of quality” [44] stems from a well-known fact that if the variability in the raw material is not compensated by the process, it will be transferred to quality.

The effect of the raw material attributes in the process performance, if the process operating conditions remain fixed, is clear in the example depicted in Figure 25.5. Recall that in that example the raw material is characterized by more than 25 physical and chemical properties and that these variables are projected on a space defined by latent variables, which allow us to visualize better the process behavior. Raw material is produced at three supplier locations. Although the raw material properties are within univariate specifications, at all locations, the projection in three clusters indicates that in a multivariate sense the material possesses slightly different characteristics depending on the location it was produced (covariance structure changes with location). Some batches of that raw material were used for a specific product. The material properties after micronization are projected on principal components. It can be observed that after micronization the batches from a specific supplier location (circled by the small ellipse) project on a different location from the rest. The filling performance of the material originating from this location is also different from the rest of the material. *The raw material variability is transferred to quality if the process remains fixed.* A note here is that although the control ellipses (large ellipses) shown are set by default in the vendor software, they are not interpretable when there is clustering; the assumptions for the calculation of these ellipses are for

process monitoring and not for process exploration where there is intentional variation such that introduced by DOE.

The concept of the design space can be easily understood with the example below, depicted in Figures 25.10 and 25.11. In the figures we have a process where the raw material is described by attributes z_1 and z_2 , quality is described by q_1 and q_2 , and unit operations described by process parameters x_1 and x_2 for unit 1 and xx_1 and xx_2 for unit N . Two attributes, two quality, and two process parameters per unit are used for illustration purposes, but this does not affect generalization of the following discussion to more. Each circle represents the values of these parameters for one batch. Figure 25.10 shows what happens when a fixed process is considered, depicted by dark grey circles. Suppose that we run the traditional three batches at a selected range of z_1 – z_2 and selected range of process parameters and we achieve the target quality (all dark grey circles fall on a multivariate target). The light grey circles represent raw material from, say, a different manufacturer, with attribute values different than the range initially examined. If we process the light grey material on the fixed process conditions used for the previous three batches of different raw material, chances are that the final quality will differ. Figure 25.11 illustrates that if we carefully choose to operate at appropriate different process conditions for each different material, then we can have quality on target. In other

words, there is a multidimensional combination of raw material and process parameters that assures quality.

These appropriate process conditions (depicted by the paths that relate raw material and process parameters with quality) are the solutions to the equations of the model that relates raw material and process conditions to quality, and these solutions are obtained when we solve for the values of process conditions given the values of the raw material properties, such that quality falls in a desirable range. Sometimes optimization can be used to introduce constraints, such that the solution takes into account cost, time duration of a process, etc. The model may be theoretical or empirical or a hybrid.

25.9.2 Design Space Modeling

The design space can be established as a model that relates input material and process parameters to quality. The model may be theoretical (based on first principles) or empirical, derived from DOE, or a hybrid. Together with the model one has to specify the range of parameters for which the model has been verified. The model may cover one unit operation or a series of unit operations.

The model will express quality as a function of the raw material attributes and process parameters:

$$\text{Quality} = f(\text{raw material, process parameters})$$

or more specifically as

$$[q_1, q_2, \dots, q_N] = f(z_1, z_2, \dots, z_K, x_1, x_2, \dots, x_M, \dots, xx_1, xx_2, \dots, xx_P)$$

and then solve for the combination of process parameters that will result in a desired q_1, q_2, q_N given the values of z_1, z_2, z_N . The function may be linear or nonlinear, and more than one model will in general be required to describe the behavior of a multiunit plant if we wish to be able to predicts intermediate quality as well (i.e. granule properties). Multivariate projection methods can be used for empirical modeling. That is, the design space is a collection of models that relate (i) the final quality to all previous units, raw material, and intermediate quality and (ii) intermediate quality to previous unit operations and raw material. The design space consists of models (relationships/paths) plus the range of parameters for which the models have been verified. Random combinations in the range of parameters will not work in general (that is, in Figure 25.11, dark grey raw material operated in light grey conditions may not result in acceptable quality). Therefore the range without the paths or multidimensional combinations (the model) cannot describe the design space unless it is the range selected with traditional approaches; this range may however be only one of the acceptable solutions and therefore restricts our flexibility in dealing with a wide

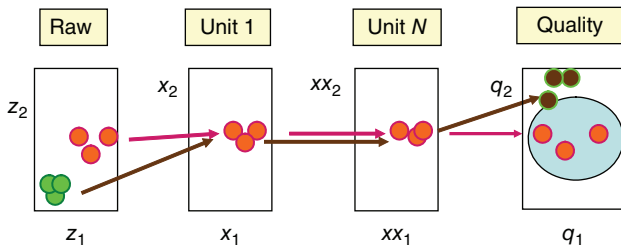


FIGURE 25.10 By maintaining fixed process conditions, we propagate raw material variability to quality.

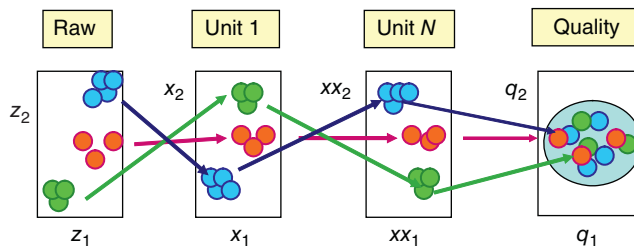


FIGURE 25.11 By taking a feed-forward approach where the process conditions are flexible to account for raw material variability, we can maintain quality on target.

variety of raw materials and/or dealing with disturbances that affect the process.

The above function uses more than two attributes and process parameters in the multidimensional relationships to reflect a general case. Multivariate projection methods or latent variables are proven very useful to describe and solve for these relationships where many variables are present.

The effect of raw material on the quality as it propagates through different unit operations is shown for a tableting process in Figure 25.12. When the raw material properties have certain characteristics (marked with a small ellipse), the material projects on a different area. The properties of granules produced from raw material with such characteristics (black) are different from the rest, and the final quality also shows differences. The difference in the quality can be theoretically explained based on the physical phenomena that govern the whole process. The idea of the design space is to express these phenomena by a model.

Recognizing the continuum in drug production that spans from the drug substance to the drug product will help create a more versatile and robust design space. The final product that delivers the API to a patient is indeed the result of a multidimensional combination of raw material attributes and process parameters that span several unit operations including those of the drug substance production (like reaction and crystallization), the ones from drug product production (like granulation and compression), and packaging. Each one of them has an impact

on one or more final quality characteristics, as well as on stability.

Incorporating stability into the QbD framework has been discussed; it involves including stability time profiles in the model for the design space [5].

Therefore, if the design space is addressed in a holistic form, then the quality of the final product, including its stability, should be expressed as a function of raw material characteristics and process parameters. Also by treating the design space in a holistic way, it would provide the manufacturer with the most cost-efficient operation and guarantee high yield and low operating costs, because problems at later unit operations will be anticipated and corrected in earlier operations. In other words the control strategy will be part of the design, such that it can be implemented in the most cost-effective way.

A model that describes the design space for the entire tableting process can be derived by relating quality to the raw material properties as well as to the process parameters of the unit operations (Figure 25.13). One row in the database depicted in Figure 25.13 would include the process conditions and quality experienced by the product as it is processed through the units. The empirical models derived are causal and based on carefully designed experiments (DOE). Some DOEs will also be necessary to estimate parameters even if mechanistic models are used.

The level of detail in the models varies depending on the objective of the model and the depth of process understanding

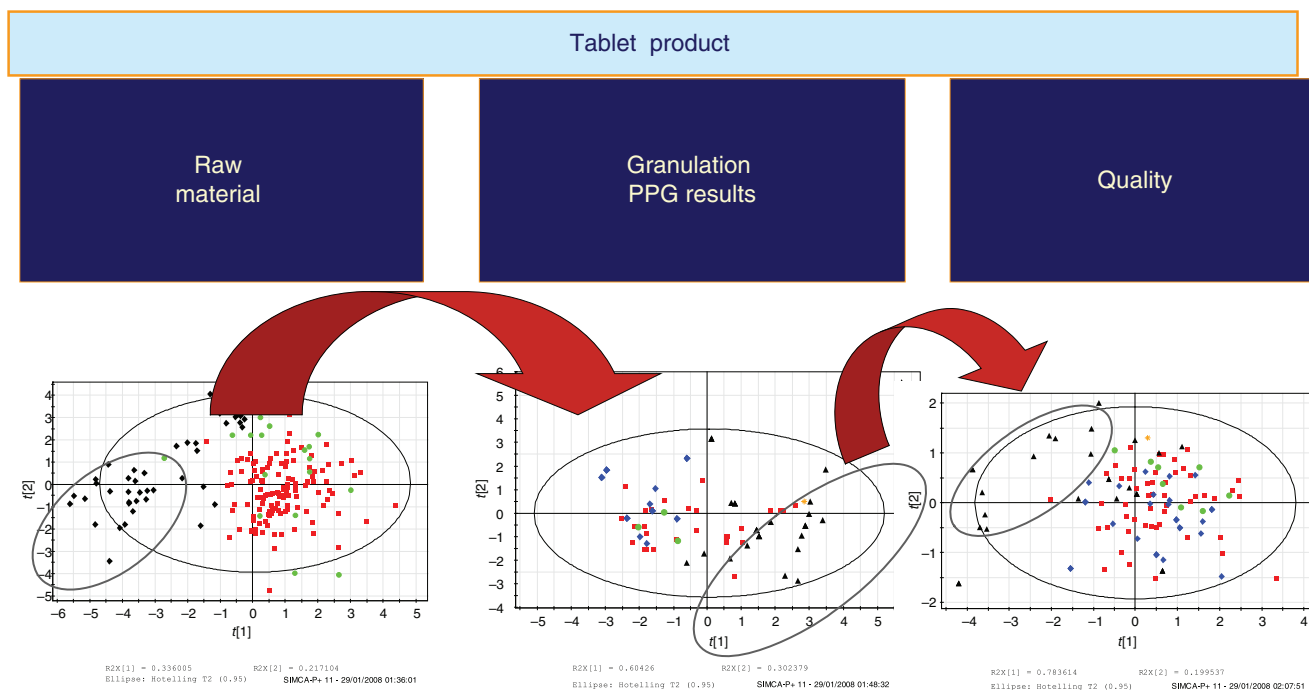


FIGURE 25.12 Projection space representation for a tablet product. Batches produced from raw material with similar characteristics have similar final quality.

one wishes to achieve. For example, the variable trajectories of a granulation may be described by carefully selected summary data or by the full variable trajectories aligned against time or another indicator variable.

25.9.3 Control Strategy

Based on the process understanding gained from the design space modeling, the control strategy can be derived to assure final quality. There are several ways of controlling a process, as discussed in Section 25.7. If we decide to keep the process fixed, we may apply a control strategy for the incoming material to reduce raw material variability (see Section 25.7.4).

If we wish to apply the principles depicted in Figure 25.11, then feed-forward control should be applied (Section 25.7.1).

Figure 25.14 depicts action in feed-forward control that would apply in case of Figure 25.12. When a different raw material enters the process, we have the choice to adjust

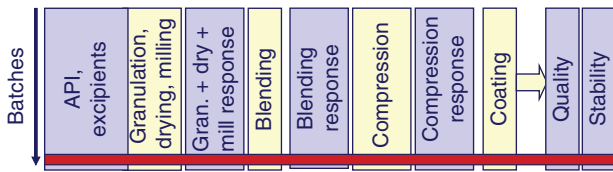


FIGURE 25.13 Both quality and stability profiles can be modeled as a function of input material, process variables, and intermediate attributes.

granulation conditions in a feed-forward manner. However, we also have the choice to adjust compression, that is to apply the feed-forward action at a later stage. When a deviation in the granules is detected that may result in quality different than typically observed if the compression operates at certain conditions, we may bring the quality on target by altering the compression settings. The choice of the process conditions at the unit operation at which we will perform the action will be dictated by a model that takes into account the value of the properties of the input to that unit and calculates process conditions such that quality is on target. When the model is empirical, MVA can be used.

Figure 25.15 illustrates such a case. The example here illustrates a feed-forward control scheme for unit N based on input information on the “state-of-the-intermediate product” from unit $N - 1$. The settings are calculated and adjusted such that the target value for quality Y is met. A multivariate model was built (from batch data) to relate product quality to the process parameters of unit N and the “state-of-the-intermediate product” from unit $N - 1$ (i.e. input to unit N). From this model, a quantitative understanding was developed showing how process parameters in N and the state-of-the-intermediate product from $N - 1$ interact to affect quality. Using MVA assures that the multivariate nature of quality is respected. In this case, the input to unit N is such that the five batches that project in an area within the small circle (two dark batches and three grey) have the same state of intermediate product – meaning that up to that time the five batches experienced same raw material and processing

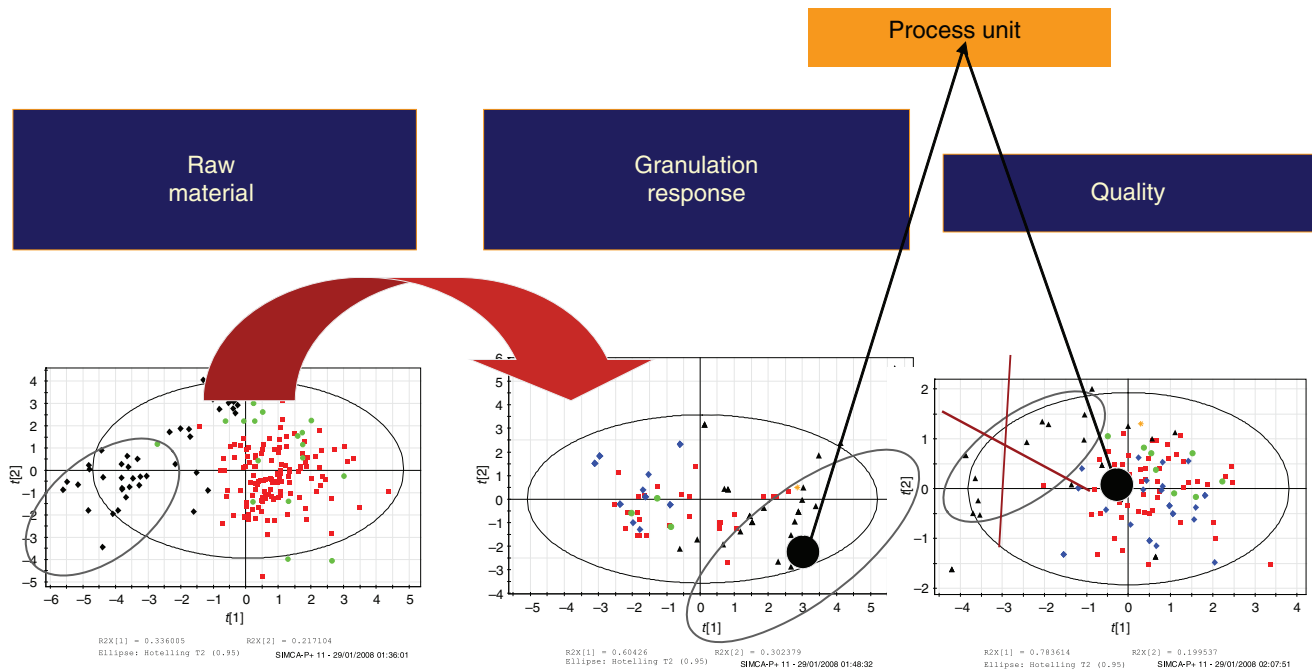


FIGURE 25.14 Feed-forward control. When a deviation in the granules is detected that may result in quality different than typically observed if the next process operates at given conditions, we may bring the quality on target by altering the process conditions.

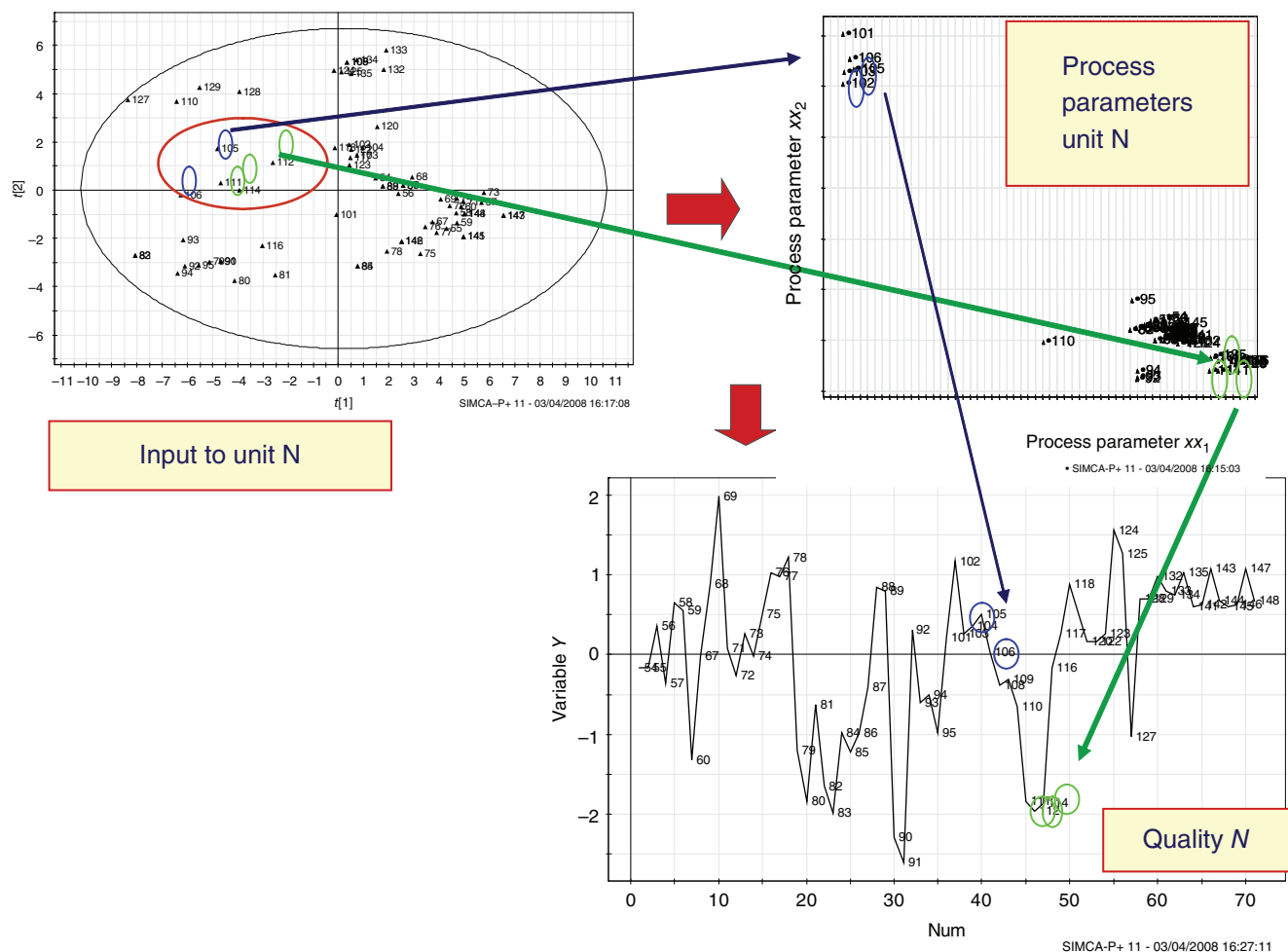


FIGURE 25.15 Control strategy using projection space.

conditions. The grey batches when processed with typical operating conditions in unit N , marked grey, resulted in quality below average. By taking a feed-forward action and processing the dark batches with different operating conditions, in unit N , the quality improves with values above average.

For real-time monitoring and control of an individual unit operation, as, for example, batch granulation, the principles described in batch process monitoring (Section 25.5.4) and process control by manipulating multivariate trajectories (Section 25.7.3) apply.

25.10 DESIGN SPACE MANAGEMENT

It is accepted that the design space will evolve after the initial submission, and therefore design space management is very important in the product life cycle. There are several issues to consider with design space management beyond the obvious

ones (that is, beyond managing the design space at the current site to address issues not considered because of limited data available by the initial submission). These include situations where a larger scale is considered at the same site, as well as when there is site transfer. Other issues would be situations where there are different suppliers of API or excipients and when the raw material characteristics are altered slightly within the same supplier. Production changes like opportunities to use soft sensors instead of real-time analyzers or to expand the current process analytical technology (PAT) capabilities should also be considered. Solutions to such problems can be addressed under the framework of design space management.

25.10.1 Quality by Design in Analytical Methods

The methodology described for design space can be applied in analytical methods. Chromatography is not only a laboratory method but also a unit operation in biopharmaceuticals.

Process transfer ideas can be also applied in method transfer ideas; in other words, method transfer and site transfer could be treated with similar principles.

MSPC for analytical methods has been reported. Multivariate monitoring of a chromatographic system has been carried out using a check sample containing five analytes to test column performance [45]. A T^2 chart and an SPE_X chart were used to monitor analyte peak area % of the five analytes. The results indicated that false alarms, which would have occurred with univariate charts, were avoided and points out of control due to change in correlation could be detected (impossible with univariate charts).

A recent application of QbD in analytical method utilizing MVA can be found in Ref. [46].

25.11 REGULATORY CONSIDERATIONS FOR MULTIVARIATE ANALYSIS MODELS

Regulatory agencies have responded to the need of clarifications related to the new concepts (i.e. QbD, PAT, real-time release, design space, continuous manufacturing) and the tools (i.e. modeling) required to achieving them. Regulatory guidances as well as other publications prepared jointly with industry can be utilized for clarifications.

Modeling: Latent variable models are just a fraction of the models used in the pharmaceutical industry under the new initiatives. There are many considerations in the development, validation, and maintenance of models depending on their use.

ICH published points to consider related to modeling [47] and clarified one important point: within each implementation mode, for the purpose of regulatory consideration, an important factor to consider is the model's contribution in assuring the quality of the product. In that context, models can be classified as high, medium, or low impact:

- a. *High impact models:* A model can be considered high impact if prediction from the model is a significant determinant of quality of the product (e.g. a chemometric model for product assay, a surrogate model for dissolution).
- b. *Medium impact models:* Such models can be useful in assuring quality of the product but are not the sole determinant of product quality (e.g. most design space models, many in-process controls).
- c. *Low impact models:* These models are typically used to support process and product development and design efforts (e.g. formulation or reaction optimization).

A series of articles [48, 49] jointly authored by regulatory agencies and industry provide points for consideration for the building and use of models in the regulated

pharmaceutical industry. It offers examples of how models can play a part in the QbD framework, how these models can be developed, and how model information can be utilized as a part of the control strategy. Multivariate models are empirical models. Changes in raw materials, instrumentation, and equipment will affect the models and impact the robustness of the model during the life cycle of the related product. One could not overemphasize the importance of life cycle considerations (model maintenance). These papers outline points to consider for a model's life cycle in a regulated environment taking into account the impact of the model according to ICH [47].

MVA and real-time release. Workshops were run where industry and regulators presented jointly case studies. Recordings of such presentations are available on MVA and real-time release [50] and on MSPC [51].

PAT. The use of models in PAT is also clarified in other guidances by EMA [52, 53] and FDA [54].

Continuous manufacturing. In recent years the pharmaceutical industry is adopting continuous manufacturing. Joint industry–regulatory workshops took place to address emerging issues in the field including the use of multivariate modeling for PAT and process monitoring and control. Several publications appeared with joint authors from industry and regulatory agencies. Some publications relevant to modeling are Refs. [55–57].

25.12 RECENT APPLICATIONS OF MULTIVARIATE ANALYSIS IN PHARMACEUTICAL INDUSTRY

MVA has been applied for QbD and real-time release in new products [50] as well as for QbD for legacy products [58, 59] and process understanding [60] utilizing models that involved stability as described in Figure 25.13.

In pharmaceutical tablet manufacturing processes, a major source of disturbance affecting drug product quality is the lot-to-lot variability of the incoming raw materials, as depicted in Figure 25.10. As suggested earlier the process conditions can be adjusted to account for this variability such that the quality remains on target. A number of practitioners have adopted such practices [61–63]. The approach described in Section 25.7.4 where the process conditions remain constant but specifications are applied on incoming materials was also applied [64].

The concept of product design described in Section 25.8.2 was applied [65, 66] and further developed [67, 68]. MVA was utilized for applications involving process monitoring and control [69–71] including control of continuous reactors [72]. Multivariate applications were not limited to pharmaceutical processes. A remarkable activity was observed for applications to biopharmaceutical processes [73–78]. Further

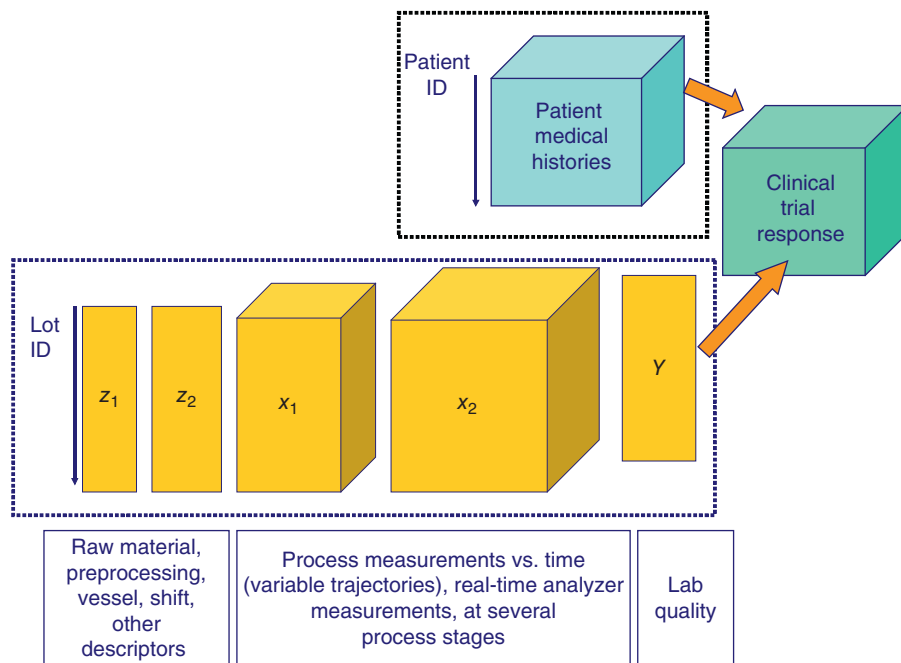


FIGURE 25.16 Examples of complex data structures emerging in industry that can be mined for a wealth of information.

applications and considerations for MVA in pharmaceutical industry are reported in Refs. [79, 80].

25.13 CONCLUDING REMARKS

As more complex structures of data are being generated, MVA offers great opportunities for information integration and analysis. Manufacturing data as well as patient histories can be integrated and then incorporate into design space the clinical trial responses.

Figure 25.16 shows an example of the possibilities that can be explored. Quality in product Y can be related to past information of raw materials, preprocessing and holding times, the type of the vessel used, the operator that run the process, and other recipe information as well as process measurement trajectories and analyzer information. The quality Y (and details of manufacturing) as well as the patient medical histories and clinical responses can be used to establish a better understanding of the design space.

REFERENCES

- Kourtí, T. (2005). Symposium report on abnormal situation detection and projection methods. *Chemom. Intell. Lab. Syst.* 76: 215–220.
- Abboud, L. and Hensley, S. (2003). Factory shift: new prescription for drug makers: update the plants; after years of neglect, industry focuses on manufacturing; FDA acts as a catalyst; the three-story blender. *Wall Street J.*, Eastern edition, New York, NY, September 3 A.1.
- Yu, H., MacGregor, J.F., Haarsma, G., and Bourg, W. (2003). Digital imaging for on-line monitoring and control of industrial snack food processes. *Ind. Eng. Chem. Res.* 42: 3036–3044.
- Yu, H. and MacGregor, J.F. (2003). Multivariate image analysis and regression for prediction of coating content and distribution in the production of snack foods. *Chemom. Intell. Lab. Syst.* 67: 125–144.
- Kourtí, T., Gonzalez, S., Balaguer, P., et al. (2010). Stability in the QbD framework. Presented in IFPAC 2010, Baltimore, MD (2–4 February 2010).
- Kourtí, T. (2006). Process analytical technology beyond real-time analyzers: the role of multivariate analysis critical reviews in analytical chemistry. *Crit. Rev. Anal. Chem.* 36: 257–278.
- Kourtí, T. (2009). Multivariate statistical process control and process control, using latent variables. In: *Comprehensive Chemometrics*, vol. 4 (ed. S. Brown, R. Tauler and R. Walczak), 21–54. Oxford: Elsevier.
- Brown, S., Tauler, R., and Walczak, R. (eds.) (2009). *Comprehensive Chemometrics*, vol. 1–4. Oxford: Elsevier.
- Miletic, I., Boudreau, F., Dudzic, M. et al. (2008). Experiences in applying data-driven modelling technology to steelmaking processes. *Can. J. Chem. Eng.* 86: 937–946.
- Kourtí, T. and MacGregor, J.F. (1995). Process analysis, monitoring and diagnosis using multivariate projection methods – a tutorial. *Chemom. Intell. Lab. Syst.* 28: 3–21.

11. Muirhead, G.T. (2005). Process analytical technology at GlaxoSmithKline. Presented at the 19th IFPAC, International Forum Process Analytical Technology, Arlington, VA (10–13 January 2005).
12. Graffner, C. (2005). PAT – European regulatory perspective. *J. Process Anal. Technol.* 2 (6): 8–11.
13. Burnham, A.J., Viveros, R., and MacGregor, J.F. (1996). Frameworks for latent variable multivariate regression. *J. Chemom.* 10: 31–45.
14. Kourti, T. (2005). Application of latent variable methods to process control and multivariate statistical process control in industry. *Int. J. Adapt. Control Signal Process.* 19: 213–246.
15. Kourti, T. (2002). Process analysis and abnormal situation detection: from theory to practice. *IEEE Control. Syst.* 22 (5): 10–25.
16. Nomikos, P. and MacGregor, J.F. (1995). Multivariate SPC charts for monitoring batch processes. *Technometrics* 37 (1): 41–59.
17. Nomikos, P. and MacGregor, J.F. (1994). Monitoring of batch processes using multi-way principal component analysis. *AIChE J* 40 (8): 1361–1375.
18. Nomikos, P. and MacGregor, J.F. (1995). Multiway partial least squares in monitoring batch processes. *Chemom. Intell. Lab. Syst.* 30: 97–108.
19. Kourti, T. (2003). Multivariate dynamic data modelling for analysis and statistical process control of batch processes, start-ups and grade transitions. *J. Chemom.* 17: 93–109.
20. Kourti, T. (2003). Abnormal situation detection, three way data and projection methods – robust data archiving and modeling for industrial applications. *Annu. Rev. Control* 27 (2): 131–138.
21. García-Muñoz, S., Kourti, T., MacGregor, J.F. et al. (2003). Troubleshooting of an industrial batch process using multivariate methods. *Ind. Eng. Chem. Res.* 42: 3592–3601.
22. Camacho, J., Picó, J., and Ferrer, A. (2008). Bilinear modelling of batch processes. Part I: theoretical discussion. *J. Chemom.* 22 (5): 299–308.
23. Camacho, J., Picó, J., and Ferrer, A. (2008). Bilinear modelling of batch processes. Part II: a comparison of PLS soft-sensors. *J. Chemom.* 22 (10): 533–547.
24. Neogi, D. and Schlags, C.E. (1998). Multivariate statistical analysis of an emulsion batch process. *Ind. Eng. Chem. Res.* 37: 3971–3979.
25. Kassidas, A., MacGregor, J.F., and Taylor, P.A. (1998). Synchronization of batch trajectories using dynamic time warping. *AIChE J* 44: 864–875.
26. García-Muñoz, S., MacGregor, J.F., Neogi, D. et al. (2008). Optimization of batch operating policies. Part II. Incorporating process constraints and industrial applications. *Ind. Eng. Chem. Res.* 47 (12): 4202–4208.
27. García-Muñoz, S., MacGregor, J.F., and Kourti, T. (2004). Model predictive monitoring for batch processes with multivariate methods. *Ind. Eng. Chem. Res.* 43: 5929–5941.
28. Kosanovich, K.A., Piovoso, M.J., and Dahl, K.S. (1995). Multi-way PCA applied to an industrial batch process. *Ind. Chem. Res.* 35: 138–146.
29. Lennox, B., Montague, G.A., Hiden, H.G. et al. (2001). Process monitoring of an industrial fed-batch fermentation. *Biotechnol. Bioeng.* 74 (2): 125–135.
30. Kourti, T., Nomikos, P., and MacGregor, J.F. (1995). Analysis, monitoring and fault diagnosis of batch processes using multiblock and multiway PLS. *J. Process Control* 5: 277–284.
31. Nomikos, P. (1996). Detection and diagnosis of abnormal batch operations based on multiway principal component analysis. *ISA Trans.* 35: 259–267.
32. Duchesne, C. and MacGregor, J.F. (2000). Multivariate analysis and optimization of process variable trajectories for batch processes. *Chemom. Intell. Lab. Syst.* 51: 125–137.
33. Westerhuis, J., Kourti, T., and MacGregor, J.F. (1998). Analysis of multiblock and hierarchical PCA and PLS models. *J. Chemom.* 12: 301–321.
34. Qin, J.S., Valle, S., and Piovoso, M.J. (2001). On unifying multiblock analysis with application to decentralized process monitoring. *J. Chemom.* 15: 715–742.
35. Höskuldsson, A. (2008). Multi-block and path modelling procedures. *J. Chemom.* 22 (11–12): 571–579.
36. Westerhuis, J.A., Coenegracht, P.M.J., and Coenraad, F.L. (1997). Multivariate modelling of the tablet manufacturing process with wet granulation for tablet optimization and in-process control. *Int. J. Pharm.* 156: 109–117.
37. Findlay, P., Morris, K., and Kildsig, D. (2003). PAT in fluid bed granulation. Presented at AIChE, San Francisco, CA (2003).
38. Marjanovic, O., Lennox, B., Sandoz, D., et al. (2006). Real time monitoring of an industrial batch process. Presented at CPC7: Chemical Process Control, Lake Louise, Alberta, Canada (8–13 January 2006).
39. Zhang, H. and Lennox, B. (2004). Integrated condition monitoring and control of fed-batch fermentation processes. *J. Process Control* 14: 41–50.
40. Flores-Cerrillo, J. and MacGregor, J.F. (2004). Control of batch product quality by trajectory manipulation using latent variable models. *J. Process Control* 14: 539–553.
41. Duchesne, C. and MacGregor, J.F. (2004). Establishing multivariate specification regions for incoming materials. *J. Qual. Technol.* 36: 78–94.
42. Box, G.E.P., Hunter, W.G., and Hunter, J.S. (1978). *Statistics for Experimenters. An Introduction to Design, Data Analysis and Model Building*, Wiley Series in Probability and Mathematical Statistics. New York: Wiley.
43. García-Muñoz, S., MacGregor, J.F., and Kourti, T. (2005). Product transfer between sites using joint Y_PLS. *Chemom. Intell. Lab. Syst.* 79: 101–114.
44. International Conference on Harmonization (2009). ICH Draft Step 4, Q8(R2) Pharmaceutical development August 2009. http://www.ich.org/fileadmin/Public_Web_Site/ICH_Products/Guidelines/Quality/Q8_R1/Step4/Q8_R2_Guideline.pdf.
45. Nijhuis, A., de Jong, S., and Vandeginste, B.G.M. (1997). Multivariate statistical process control in chromatography. *Chemom. Intell. Lab. Syst.* 38: 51–62.

46. Muteki, K., Morgado, J.E., Reid, G.L. et al. (2013). Quantitative structure retention relationship models in an analytical quality by design framework: simultaneously accounting for compound properties, Mobile-phase conditions, and stationary-phase properties. *Ind. Eng. Chem. Res.* 52: 12269–12284.
47. ICH Quality Implementation Working Group (QIWG) (December 2011). Points to Consider (R2), ICH-Endorsed Guide for ICH Q8/Q9/Q10 Implementation. http://www.ich.org/fileadmin/Public_Web_Site/ICH_Products/Guidelines/Quality/Q8_9_10_QAs/PtC/Quality_IWG_PtCR2_6dec2011.pdf.
48. Kourti, T., Lepore, J., Liesum, L. et al. (May/June 2015). Scientific and regulatory considerations for implementing mathematical models in the quality by design (QbD) framework. *Pharm. Eng.* 35 (3): 80–88.
49. Kourti, T., Lepore, J., Liesum, L. et al. (August 2015). Scientific and regulatory considerations for implementing mathematical models in the quality by design (QbD) framework, part 2. *Pharm. Eng.* 35 (4): 38–51.
50. Jørgensen, G.H. and Kourti, T. 2014. *Challenges in the implementation of model based and PAT based RTRT for a new Product*, Joint European Medicines Agency/Parenteral Drug Association Quality-by-Design Workshop, European Medicines Agency, London (28–29 January 2014). http://www.ema.europa.eu/docs/en_GB/document_library/Presentation/2014/02/WC500162149.pdf and http://www.ema.europa.eu/ema/index.jsp?curl=pages/news_and_events/events/2013/12/event_detail_000808.jsp&mid=WC0b01ac058004d5c3
51. Lorenz, L. and Sargi, L. (2014). Applying QbD for a legacy product and achieving real time release testing by a design space approach with supportive PAT and soft sensor based models: challenges in the Implementations. Joint European Medicines Agency/Parenteral Drug Association Quality-by-Design Workshop, European Medicines Agency, London (28–29 January 2014). http://www.ema.europa.eu/docs/en_GB/document_library/Presentation/2014/03/WC500162194.pdf and http://www.ema.europa.eu/ema/index.jsp?curl=pages/news_and_events/events/2013/12/event_detail_000808.jsp&mid=WC0b01ac058004d5c3
52. EMA (2014). Guideline on the Use of Near Infrared Spectroscopy by the Pharmaceutical Industry and the Data Requirements for New Submissions and Variations, 27 January 2014, EMEA/CHMP/CVMP/QWP/17760/2009 Rev2, Committee for Human Medicinal Products (CHMP), Committee for Veterinary Medicinal Products (CVMP).
53. EMA (2014). Addendum to EMA/CHMP/CVMP/ QWP/ 17760/2009 Rev 2: Defining the Scope of an NIRS Procedure, 5 June 2014, EMA/CHMP/CVMP/ QWP/63699/2014, Committee for Human Medicinal Products (CHMP), Committee for Veterinary Medicinal Products (CVMP).
54. FDA (2015). Development and Submission of Near Infrared Analytical Procedures Guidance for Industry. DRAFT GUIDANCE. U.S. Department of Health and Human Services Food and Drug Administration, Center for Drug Evaluation and Research (CDER). March 2015 Pharmaceutical Quality/CMC.
55. Allison, G., Cain, Y.T., Cooney, C. et al. (2015). Regulatory and quality considerations for continuous manufacturing. May 20–21, 2014 continuous manufacturing symposium. *J. Pharm. Sci.* 803–812.
56. Kourti, T., Madurawe, R., Brorson, K. et al. (2017). ISPE 2016 continuous manufacturing conference highlights. *Pharm. Eng.* 37 (3): 36–42.
57. Regulatory perspectives on continuous pharmaceutical manufacturing: moving from theory to practice. 2nd International Symposium on Continuous Manufacturing of Pharmaceuticals: Implementation, Technology & Regulatory, MIT (26–27 September 2016). <https://iscmp2016.mit.edu/regulatory-white-paper>
58. Lourenc,o, V., Lochmann, D., Reich, G. et al. (2012). A quality by design study applied to an industrial pharmaceutical fluid bed granulation. *Eur. J. Pharm. Biopharm.* 81: 438–447.
59. Yacoub, F., Lautens, J., Lucisano, L., and Banh, W. (2011). Application of quality by design principles to legacy drug products. *J. Pharm. Innov.* 6: 61–68.
60. Huang, J., Goolcharan, C., and Ghosh, K. (2011). A quality by design approach to investigate tablet dissolution shift upon accelerated stability by multivariate methods. *Eur. J. Pharm. Biopharm.* 78: 141–150.
61. Muteki, K., Swaminathan, V., Sekulic, S.S., and Reid, G.L. (2011). De-risking pharmaceutical tablet manufacture through process understanding, latent variable modeling, and optimization technologies. *AAPS Pharm. Sci. Technol.* 12: 1324–1334.
62. Muteki, K., Swaminathan, V., Sekulic, SS, and Reid, G.L. (2012). Feed-forward process control strategy for pharmaceutical tablet manufacture using latent variable modeling and optimization technologies. Preprints of the 8th IFAC Symposium on Advanced Control of Chemical Processes the International Federation of Automatic Control, Furama Riverfront, Singapore (10–13 July 2012).
63. Liu, Z., Bruwer, M.-J., MacGregor, J.F. et al. (2011). Modeling and optimization of a tablet manufacturing line. *J. Pharm. Innov.* 6: 170–180.
64. Liu, Z., Bruwer, M.-J., MacGregor, J.F. et al. (2013). Setting up simultaneous specifications on multiple raw materials to ensure product quality and minimize risk. Presented at IFPAC-2013, Baltimore, MA (January 2013).
65. Torres, M.T., Nichols, E., MacGregor, J.F., and Hoare, T. (2014). Designing multi-responsive polymers using latent variable methods <https://doi.org/10.1016/j.polymer.2013.12.041>. *Polymer* 55.
66. Yu, W., Muteki, K., Zhang, L., and Kim, G. (January 2011). Prediction of bulk powder flow performance using comprehensive particle size and particle shape distributions. *J. Pharm. Sci.* 100 (1).
67. Tomba, E., Barolo, M., and García-Muñoz, S. (2012). General framework for latent variable model inversion for the design and manufacturing of new products. *Ind. Eng. Chem. Res.* 51: 12886–12900.
68. García-Muñoz, S. (2013). Two novel methods to analyze the combined effect of multiple raw-materials and processing conditions on the product's final attributes: JRPLS and TPLS. *Chemom. Intell. Lab. Syst.* 133: 49–62.
69. Machin, M., Liesum, L., and Peinado, A. (2011). Implementation of modelling approaches in the QbD framework: examples from the Novartis experience. *Eur. Pharm. Rev.* 16: 5–8.
70. Ewan C.W. Mercer, John Mack, David Lovett and Furqan Tahir Application of multivariate process modelling for monitoring and control applications in continuous pharmaceutical

- manufacturing. In Ferreira, A. P., Menezes, J. C., & Tobyn, M. (Eds.). (2018). *Multivariate Analysis in the Pharmaceutical Industry* (1st Ed.). London: Academic Press.
71. McCready, C. (2018). Recent advances in the integration of mechanistic models with data driven approaches, for enhanced monitoring, forecasting, optimization and real time control of pharmaceutical processes. IFPAC 2018 Conference, North Bethesda (11–14 February 2018).
72. Tahir, F., Krzemieniewska-Nandwani, K., Mack, J. et al. (2017). Advanced control of a continuous oscillatory flow crystallizer. *Control. Eng. Pract.* 67: 64–75.
73. Ündey, C., Ertunc, S., Mistretta, T., and Looze, B. (2010). Applied advanced process analytics in biopharmaceutical manufacturing: challenges and prospects in real-time monitoring and control. *J. Process Control* 20: 1009–1018.
74. Mercier, S.M., Diepenbroek, B., Dalm, M.C.F. et al. (2013). Multivariate data analysis as a PAT tool for early bioprocess development data. *J. Biotechnol.* 167: 262–270.
75. Kourti, T. (2015). Multivariate analysis for process understanding, monitoring, control and optimization in Lyophilization processes. In: *Biopharmaceutical Drug Product Development and Quality by Design*, AAPS Advances in the Pharmaceutical Sciences Series (ed. S. Hershenson, F. Jameel, M.A. Khan and S. Martin-Moe). Springer. AAPS Press.
76. Ganguly, J., Obando, L., Higgins, J.P., and Yoon, S. (2011). Multivariate data analysis in biopharmaceuticals. In: *PAT Applied in Biopharmaceutical Process Development and Manufacturing: An Enabling Tool for Quality-by-Design*, Biotechnology and Bioprocessing Series (ed. C. Ündey, D. Low, J.M.C. de Menezes and M. Koch). New York: CRC Press.
77. Ündey, C., Wang, T., Looze, B. et al. (2015). Predictive monitoring and control approaches in biopharmaceutical manufacturing. *Eur. Pharm. Rev.* 20 (4): 63–69.
78. Roy, K., Ündey, C., Mistretta, T., and Sodhi, M.S. (March 2014). Multivariate statistical monitoring as applied to clean-in-place (CIP) and steam-in-place (SIP) operations in biopharmaceutical manufacturing. *Biotechnol. Prog.* 30 (2).
79. Ferreira, A.P. and Tobyn, M. (2015). Multivariate analysis in the pharmaceutical industry: enabling process understanding and improvement in the PAT and QbD era. *Pharm. Dev. Technol.* 20 (5): 513–527.
80. Ferreira, A. P., Menezes, J. C., & Tobyn, M. (Eds.). (2018). *Multivariate Analysis in the Pharmaceutical Industry* (1st ed.). London: Academic Press.

QUALITY BY DESIGN: PILOT TO REALITY-THE HONEYMOON PHASE TO THE STORMY YEARS

MARY T. AM ENDE*

Process Modeling & Engineering Technology Group, Pfizer Inc., Groton, CT, USA

CHRISTINE B. SEYMOUR

Global Regulatory Affairs CMC, Pfizer Inc., Groton, CT, USA

26.1 INTRODUCTION

Over a decade ago, the International Conference on Harmonization (ICH) issued new regulatory guidances for the pharmaceutical industry in the United States, Europe, and Japan [1–3]. These guidances, Q8–Q10, set new expectations for designing and developing drug products using a science- and risk-based approach. In the United States, the Food and Drug Administration (FDA) initiated the Chemistry, Manufacturing, and Controls (CMC) Quality by Design (QbD) Pilot Program to encourage pharmaceutical companies to share the science- and risk-based approaches used to design and develop the drug candidate. A total of 12 pharmaceutical companies participated in the CMC QbD Pilot Program. The FDA division overseeing this was the Office of New Drug Quality Assessment (ONDQA) in the Center for Drug Evaluation and Research (CDER). A summary of the status by Director Moheb Nasr as of February 2007 [4] stated the following:

- Nine original NDAs and two supplements were accepted.
- Eight approved and three under review (31 July 2008) [5].
- Succeeding in integrating concepts of QbD into filings.
- Flexibility granted dependent on the level of knowledge and understanding demonstrated in filings.

*Current address: Lyndra Therapeutics, Watertown, MA, USA

The Pilot Program completed shortly afterward in March 2007 [5].

Chantix[®] was Pfizer's first product filed using the QbD approach and was Pfizer's first product to participate in FDA's Pilot Program. The drug product was approved on 11 May 2006. Chantix is available as 0.5 and 1.0 mg immediate release tablets for the treatment of nicotine addiction to aid in smoking cessation and is referred to as Pilot A. With a successful launch campaign, the product demand exceeded the original forecasts. The Pfizer Global Manufacturing (PGM) site in Illertissen, Germany, was challenged to streamline processing to maximize production rates. Through application of QbD and life cycle efficiency improvements, PGM–Illertissen Right First Time team set out to do just that. During the “Honeymoon Phase” of the QbD, the flexibility and benefits fostered by the QbD approach will be demonstrated.

Pfizer was fortunate to have a second drug product, Selzentry (also referred to as Maraviroc or Pilot B), participate in FDA's QbD Pilot Program. Selzentry[®] was introduced as a novel treatment for HIV available as immediate release tablets in 150 and 300 mg dosage strengths. FDA approval of Selzentry was granted on 6 August 2007. These two pilot programs allowed an alternative approach for describing the risk- and science-based approaches to the regulators during review, referred to here as “the Honeymoon Phase.” This was a valuable opportunity to explore the best method for sharing this information since there was no clear guidance of expectations for a QbD submission.

Chemical Engineering in the Pharmaceutical Industry: Drug Product Design, Development, and Modeling, Second Edition.

Edited by Mary T. am Ende and David J. am Ende.

© 2019 John Wiley & Sons, Inc. Published 2019 by John Wiley & Sons, Inc.

The manufacturing process description and metrics summarizing changes that were used to streamline the approach will be contrasted to recent experience of three QbD products filed seven years later during the “the Stormy Years.”

26.2 PREQUEL: THE ALLURE OF QUALITY BY DESIGN (QBD) IN 2005

In 2005, the regulatory environment was transforming from testing for quality to designing for quality and from prescriptive, “checkbox” submissions to science-based submissions and regulatory reviews (Figure 26.1). The benefits of the emerging science- and risk-based development paradigm were compelling: improvements in quality and process understanding, application of modeling and technology (e.g. PAT, computational modeling, continuous manufacturing), capitalizing on experience, development efficiency, regulatory flexibility, and global harmonization. In addition, other anticipated benefits were reductions in cost, regulatory burden, recalls and manufacturing failures, uncertainty and risk, process validation, and regulatory documentation for US new drug applications (NDA) and EU market authorization applications (MAA). Most of these anticipated benefits were expected to be derived from the closer collaboration between research and manufacturing, as well as industry and regulatory agencies. In this chapter, the reality of the situation is examined through a comparison of two products filed during the QbD Pilot Program to three products filed seven years later.

26.3 THE HONEYMOON PHASE

The agency was requesting the industry to share much more development information in NDAs. This was a huge shift in the level of information in the pharmaceutical development sections of the NDA focused on the final commercial

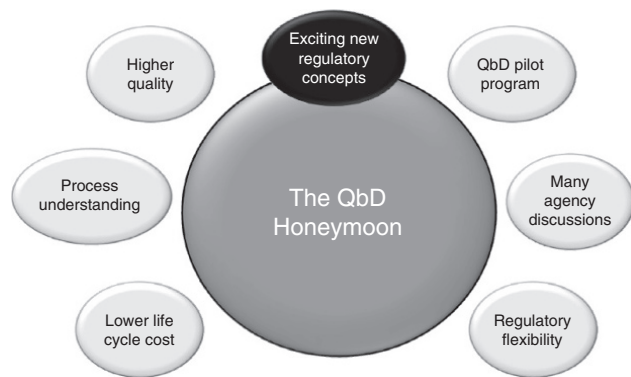


FIGURE 26.1 The “Honey Moon Phase” of FDA’s Quality by Design Pilot Program.

formulation and process to disclosing the risks and scientific understanding, as well as describing issues or edges of failure. The ICH guidances helped the industry prepare for the change and understand the evolving expectations in concept, but translating that into practice would take support to manage that change. To aid in this transition, the FDA created a QbD Pilot Program from July 2005 to March 2007 as a means for pharmaceutical companies to gain experience with the emerging changes toward QbD through face-to-face interactions with the agency during preparation and review phases. For Chantix and Selzentry, the two drug products submitted to the agency through the QbD Pilot Program, there were 2–3 face-to-face meetings and 0–2 teleconferences during the review to foster that closer collaboration during the review process (Table 26.1).

These additional CMC interactions during the regulatory review were an important component of the smooth transition to industry embracing and implementing the QbD concept. There were five new regulatory concepts that were addressed during these pilot product filings, including:

1. Design space and design space depiction.
2. Criticality and number of categories.
3. Japan Pharmaceutical Affairs Law (JPAL).
4. Change control matrix.
5. Regulatory agreement.

The first four concepts are exemplified during the Honeymoon Phase; however the fifth concept of creating a regulatory agreement never developed into a practice for reasons other than scientific.

26.3.1 Design Space and Design Space Depiction

In the first case, the overall operating region was depicted as the “knowledge space” for Chantix (Figure 26.2). This schematic representation of the operating regions across the unit operations captured the overall knowledge of the drug product, including non-critical parameters and attributes. Each unit operation is shown as a column in sequential order from left to right. The upper and lower portions of the figure capture the parameters and attributes, respectively. Those parameters impacting the quality attributes are shown by connecting lines. The criticality level (discussed in more

TABLE 26.1 The Number of FDA Face-to-Face and Teleconference Interactions During the QbD Pilot Program for the Chantix and Selzentry New Drug Applications (2005–2007)

Interactions	Chantix	Selzentry
F2F meetings	3	2
Teleconferences	2	0

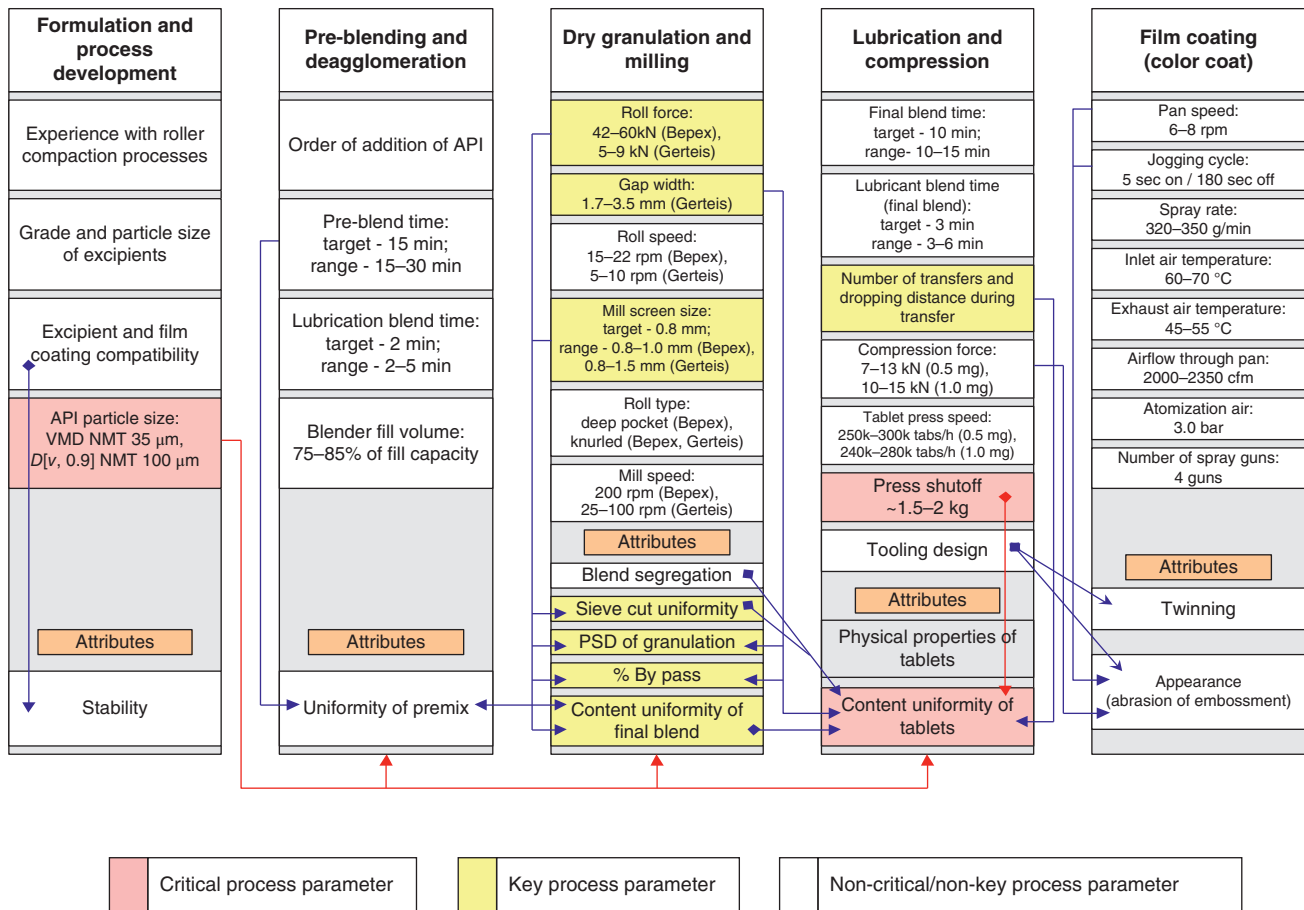


FIGURE 26.2 New concepts translated into practice during the QbD Pilot Program. Schematic showing the “knowledge space” and three-level criticality for Chantix immediate release tablets (Pilot A).

detail in the next section) is color-coded light gray for “key” and dark gray for “critical,” while non-critical is not highlighted. The overall operating regions, or knowledge space, depicted in Figure 26.2 are condensed into the design space (Figure 26.3) by removing the non-critical components. This first approach appeals more to spatial-temporal reasoning prominent among visual thinkers.

In the second case, a tabular format was used to depict knowledge and design spaces for Selzentry tablets (Table 26.2). The unit operations were listed in the first column, followed by the parameters and attributes in the second column. The criticality designation was shown in the third column, while the design space and normal operating ranges (NOR)/targets were provided in the last two columns. In this approach, the unit operations flow vertically in sequence versus horizontally in the previous case. This approach appeals more to mathematical thinkers who think in patterns and systems. Both approaches were understood and well received from the CMC reviewers at the regulatory agency during the QbD Pilot Program.

26.3.2 Criticality and Number of Categories

During FDA’s QbD Pilot Program, there was no clear guidance on criticality and the number of levels that would be acceptable. The major benefit of designing the products for global regulatory submissions is the augmented learning from each experience for one product that helped develop an internal best practice at the time.

During the preparation of the pilot QbD dossiers, intensive internal debates revealed the benefits of three levels of criticality including non-critical, key, and critical especially for harmonization with the three major markets of the United States, Europe, and Japan. The “key” category of criticality indicated that although the process parameter (KPP) had an effect on the critical quality attribute (CQA), it did not result in an edge of failure or had no relevance to the overall quality of the product. There were two critical process parameters (CPPs) in the initial product design space including the active pharmaceutical ingredient (API) particle size and tablet press shutoff that assured product quality for the only CQA,

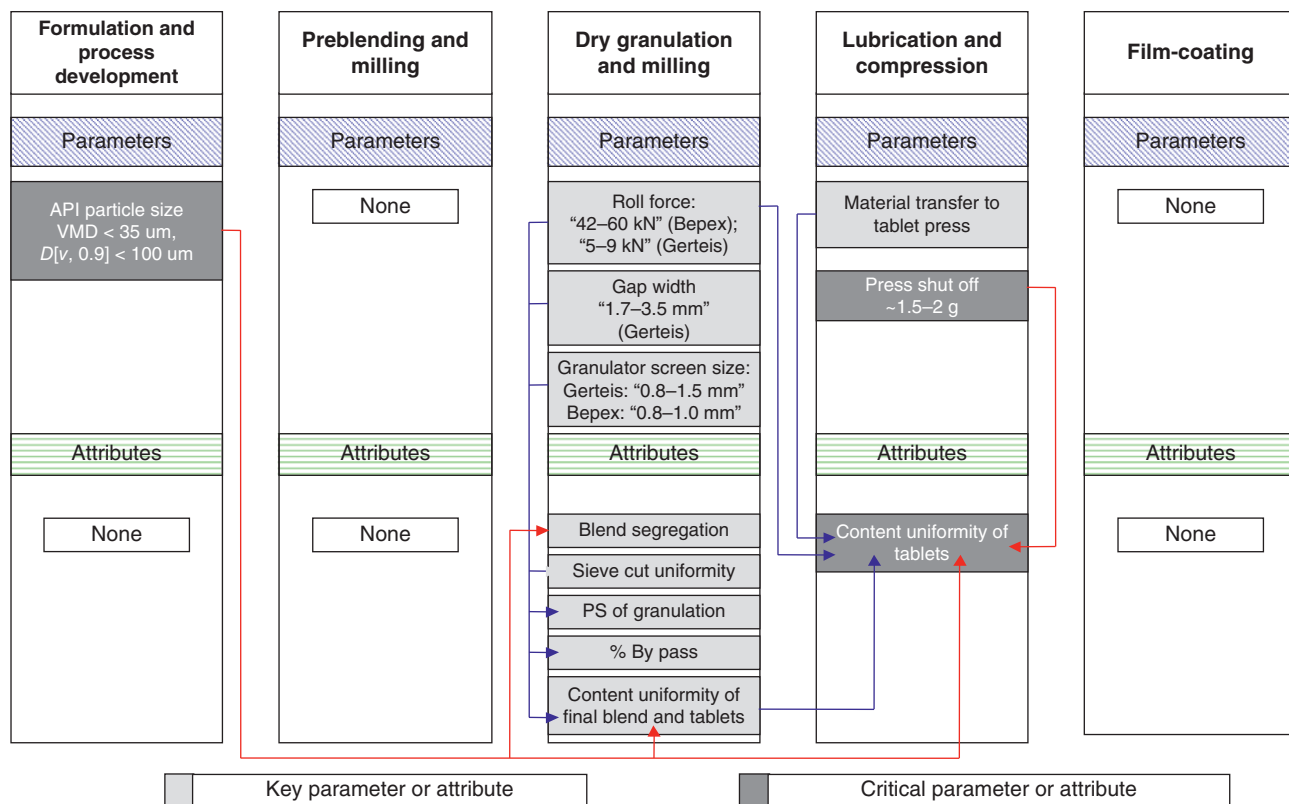


FIGURE 26.3 New concepts translated into practice during the QbD Pilot Program. "Design space" and three-level criticality for Chantix immediate release tablets (Pilot A).

content uniformity of tablets (see Figure 26.4). The API and DP co-development teams collaborated to ensure these CPPs were thoroughly understood and controlled, as described in more detail by McCurdy et al. [6].

Within a few short years of this initial pilot program for small molecules, further developments of criticality emerged to consider the "critical" category as a continuum rather than discrete [7, 8]. With that expansion of criticality interpretation, the "key" category became unnecessary.

26.3.3 Japan Pharmaceutical Affairs Law (JPAL)

During these early years of QbD, Japan issued the Pharmaceutical Affairs Law (PAL) intended to control and regulate the manufacturing, importation, and sale of drugs and medical devices¹. The JPAL provides guidance to pharmaceutical companies on how to translate their QbD control strategy. In preparing the Japan NDA, the Chantix team realized that the JPAL nomenclature for indicating major and minor parameters aligned well with the three levels of criticality initially used in early QbD filings for non-critical, key, and critical process parameters.

¹<http://www.jouhoukoukai.com/repositories/source/pal.htm>.

An example of the method of manufacture description and the depiction of KPP and CPP are shared below using the JPAL conventions. The parameters are highlighted in light gray to indicate KPP or dark gray to indicate a CPP impacting the CQA (Table 26.3).

26.3.4 Change Control Matrix

An important consideration for both the regulators and the pharmaceutical companies during the earlier phase of QbD involved defining how future product process changes would be approved, notified, or informed to the regulatory agencies. Table 26.4 provides an example of type of future changes that could occur, placing them into three main categories: (i) staying within the NOR, (ii) staying within the design space or proven acceptable range (PAR), and (iii) changing beyond the design space. For example, to redefine the design space, then a prior approval supplement is warranted. The next most important change is to a CPP. If a CPP change was made within the NOR, no agency notification would be required. However, if a CPP was changed within the PAR/design space, an annual report might be warranted if it was not previously planned through a comparability protocol. Making a change to the design space for a CPP might require a change being effected (CBE) or annual report.

TABLE 26.2 New Concepts Translated into Practice During the QbD Pilot Program. “Knowledge Space” and Three-Level Criticality for Selzentry Immediate Release Tablets

Identifier	Parameter/Attribute	Designation	Design Space Range	Normal Operating Range/Set Point
Drug substance	Particle size distribution	Non-critical	$D[v, 0.9] < 750 \mu\text{m}$	$D[v, 0.9] < 382 \mu\text{m}$
Magnesium stearate	Quantity added intragranular	Non-critical	0–0.5% w/w	0% w/w
Step 1: Blend	Blend time	Non-critical	5–20 minutes	10 minutes
	Blender speed	Non-critical	10–20 rpm	12 rpm
	Blender fill	Non-critical	40–64%	58–64%
Step 2: Screen	Mill type	Non-critical	Impact or screening mill	Screening mill
	Screen aperture	Non-critical	0.5–1.0 mm	0.8 mm
Step 3: Blend	Blend fill	Non-critical	40–64%	58–64%
	Blender speed	Non-critical	10–20 rpm	12 rpm
	Blender time	Non-critical	5–20 minutes	5 minutes
Step 4: Lubrication blend	Blend fill	Non-critical	40–64%	Optional step
	Blender speed	Non-critical	12 rpm	
	Blender time	Non-critical	5 minutes	
Step 5: Roller compaction	Roll force	Key	2 to 6 kN	5 kN
	Roll gap	Key	2.0–4.0 mm	2.5 mm
Step 6: Mill	Mill screen aperture size	Key	0.8–1.25 mm	1.0 mm
	Granulator mill speed	Non-critical	10–100 rpm	35 rpm
Step 7: Lubrication blend	Blend time	Non-critical	5–20 minutes	5 minutes
	Blender speed	Non-critical	10 to 20 rpm	12 rpm
Step 8: Compression	Compression force	Non-critical	10–40 kN (150 mg tablet) 20–52 kN (300 mg tablet)	15–35 kN (150 mg tablet) 30–50 kN (300 mg tablet)
Step 9: Film coating	Coating suspension concentration	Non-critical	15–20% w/w aqueous suspension	15% w/w aqueous suspension
	Spray rate	Non-critical	500–700 g/min	500–700 g/min
	Inlet air temperature	Non-critical	50–70 °C	50–70 °C

In the cases where the designation is changed from CPP to KPP, or KPP to CPP, an annual report may be sufficient means to communicate that in all categories, unless a comparability protocol was not included in the submission. Changes to KPP or non-critical parameters within the NOR and PAR would require no additional approval or notification except in the case of moving beyond the design space established in the regulatory submission. These changes to “key” parameters are not relevant to the current criticality continuum approach, but are included here for the early phase.

26.3.5 Regulatory Agreement

Of the five new regulatory concepts introduced during the CMC QbD Pilot Program, the regulatory agreement was the only one that was not successfully progressed. Several alternatives had been developed and shared with the regulatory reviewers and the legal teams for both industry and government; however this concept was addressed more readily by the change control matrix, which defined potential future changes and how that would be shared with the agency, as discussed in detail in the previous section.

26.3.6 QbD Honeymoon Benefits

The major benefits from the QbD paradigm shift from an industry perspective during the early phase of implementation were threefold – increased scientific debate was encouraged and fostered through frequent dialog, production was operated with a “lean manufacturing” concept more readily, and cost savings were achieved. To demonstrate those benefits, three specific examples are shared below. In addition, QbD facilitated the implementation of new technology such as continuous processing; the implementation of new technologies is not covered in this chapter.

26.3.6.1 Scientific Debate Both drug products in the QbD Pilot Program used particle size theoretical modeling to support the actual batch results to aid in specification setting. In one case, the particle size was a non-critical parameter with no significant impact on the product quality attributes. In the second case, the particle size was a CPP from the API input to the drug project based on its impact on content uniformity of tablets.

For the latter case, the initial proposal for a two-point particle size distribution specification was proposed to the

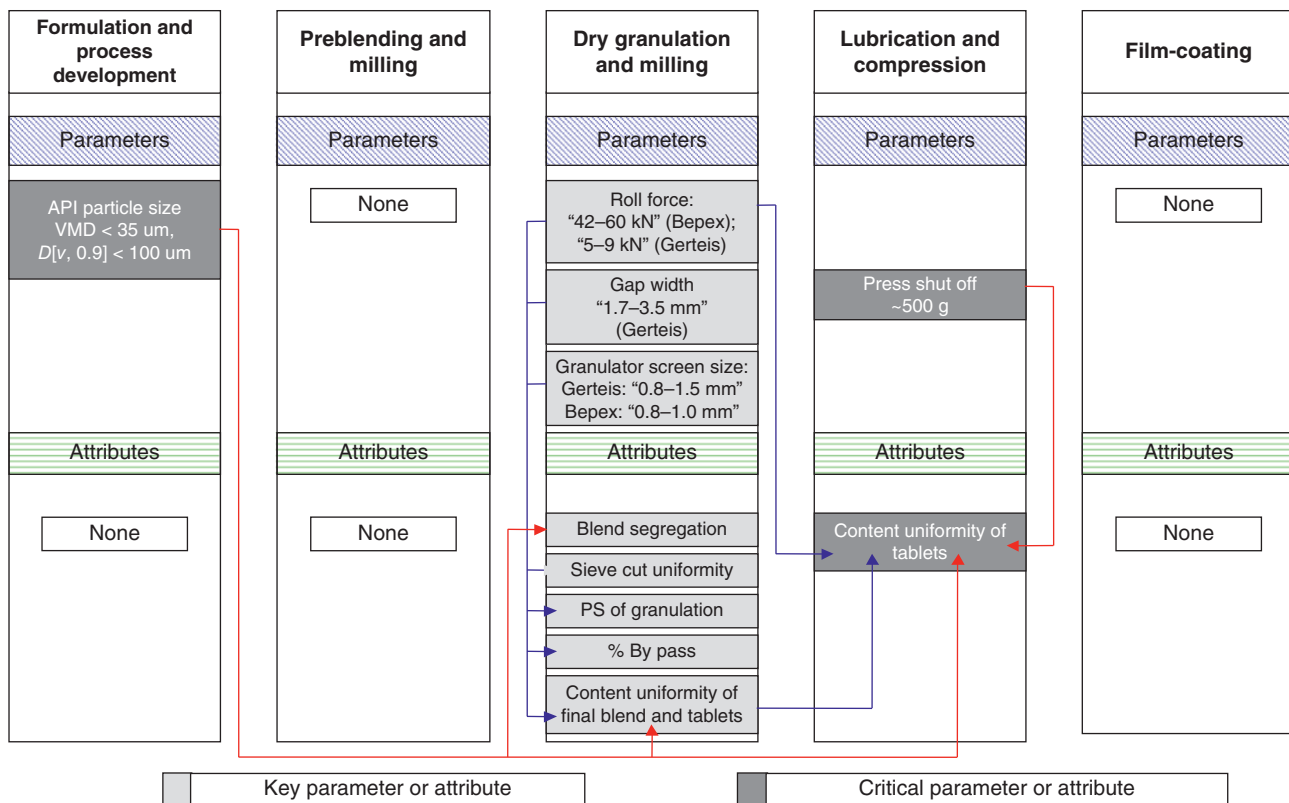


FIGURE 26.4 After further process understanding of the press shutoff, the design space was expanded for a critical process parameter from approximately 1.5–2.0 kg to approximately 500 g blend remaining in press for Chantix immediate release tablets.

TABLE 26.3 Example Description of Production Process for a Representative Batch Using the Expanded Design Space (Where *a–e* Indicate the Batch-Specific Mass Values)

Step	Description	
1	Charge a tumble blender with < <i>a</i> grams > drug, < <i>b</i> grams > excipients except lubricant. Blend the components	
2	Mill the blend from step 1 using a screening mill	
3	Charge a tumble blender with < <i>c</i> grams > intragranular lubricant into the blender from step 2. Blend the contents	
4	Dry granulate the blend from step 3 using a dry granulator under the following conditions:	← KPP
	Roll force: Bepex roller compactor: “40–60 kN” Gerteis roller compactor: “5–9 kN”	
	Gap width: Gerteis roller compactor: “1.7–3.5 mm”	
5	Mill the compacted blend from step 4 using a screening mill fitted with the following screen sizes:	← KPP
	Screen size: Bepex roller compactor: “0.8–1.0 mm” Gerteis roller compactor: “0.8–1.5 mm”	
6	Charge the granulation from step 5 into a tumble blender and blend	
7	Charge the remaining < <i>d</i> grams > extragranular lubricant to the blend from step 6, and blend	
8	Compress the blend from step 7 on a tablet press to target mg until not less than 500 g of blend remains	← CPP
9	Prepare the film coating suspension	
10	Charge the tablet cores from step 8 into a perforated coating pan. Apply the film coating suspension prepared in step 9 until the tablet weight increases to coated target, <i>e</i> mg	

TABLE 26.4 New Concepts Translated into Practice During the QbD Pilot Program Defining the Proposed Regulatory Requirements Incurred Due to Future Changes (i) Within the Normal Operating Range, (ii) Within the Design Space, and (iii) Beyond the Design Space

Category	Change Within Control Space (NOR)	Change Within Design Space (PAR)	Change Beyond Design Space
Redefinition of design space	Not applicable	Not applicable	Prior approval supplement
Critical process parameter	None	Annual report or none ^a	Change being effected (CBE) or annual report ^a
Critical process parameter changed to key	Annual report	Annual report	CBE or annual report ^a
Key process parameter changed to critical	Annual report	Annual report	CBE or annual report ^a
Key process parameter	None	None	Annual report or none ^a
Non-critical/non-key process parameter	None	None	None

^aLower reporting level used in conjunction with an approved comparability protocol.

regulators, with the $D[4,3] < 35 \mu\text{m}$ and the $D(v,0.9) < 100 \mu\text{m}$. During the QbD review, the FDA query requested a three-point specification for API particle size distribution with the recommendation to include two-sided acceptance criteria on $D[4,3]$ and a lower limit on $D(v,0.1)$. Upon meeting face-to-face with the FDA reviewers, a greater level of constraints was being requested moving to a five-point particle size distribution specification, including a lower and upper limit on $D(v,0.1)$, lower and upper limit on $D[4,3]$, and an upper limit on $D(v,0.9)$. Fortunately, there were additional opportunities to discuss this request by teleconference, and in the end, scientific and statistical knowledge succeeded in convincing the reviewers that the two-sided specifications on $D(v,0.1)$ and $D[4,3]$ were not warranted because of the following:

1. Significantly reducing the particle size of API by jet milling had no detrimental effect on final blend content uniformity or stratified tablet content uniformity.
2. $D[4,3]$ of the jet-milled lot was in the range of the $D(v,0.1)$ values for commercial milling process.
3. API particle size distribution parameters were highly correlated to each other.

Therefore, the drug substance particle size distribution was fully characterized by $D[4,3]$ and $D(v,0.9)$, and the final outcome was that the original two-point particle size distribution specification was accepted prior to approval.

26.3.6.2 Lean Drug Product Manufacturing As anticipated by the visionaries that conceived of the science- and risk-based QbD approach, the initial “honeymoon” phase resulted in higher quality products and improvements in process robustness. These results in the commercial manufacturing arena reaped the benefits of deeper and more thorough knowledge that allowed operating according to the lean

manufacturing principles. For instance, the Chantix product had the fortunate situation of marked increase in demand following launch exceeding 400% (Figure 26.5). The commercial scale was targeted to suit the forecasted demand. However, this rapid rise in product demand created the opportune time to implement efficiency improvements from 1 batch/day to 3 batch/day. The process time was reduced by 67%, while the non-value-added time was reduced by 28% for an overall 41% reduction.

Demonstration of the lean manufacturing impact on inventory can be seen by the stock levels at six months’ post-launch.

• Starter dose pack	3–6 weeks’ inventory
• 0.5 mg bottles	8 weeks’ inventory
• 1.0 mg bottles	< 5 days’ inventory

The most common dose of 1 mg was manufactured with only five days of inventory, unheard of before the QbD paradigm was implemented.

26.3.6.3 Cost Savings The overall value of the flexibility gained by submitting design space and the CMC QbD regulatory filings is compared to the previous method using traditional scale-up and post-approval changes (SUPAC) in Table 26.5. The four major changes implemented for the Pilot A product included (i) transferring the drug product manufacturing facility with adjacent sites, (ii) increasing drug product manufacturing scale twice, (iii) expanding the design space boundary for the CPPs of tablet shutoff, and (iv) changing the drug product water test from Karl Fischer to NIR.

In the first post-approval change in manufacturing facility, the QbD criteria required annual report submission of

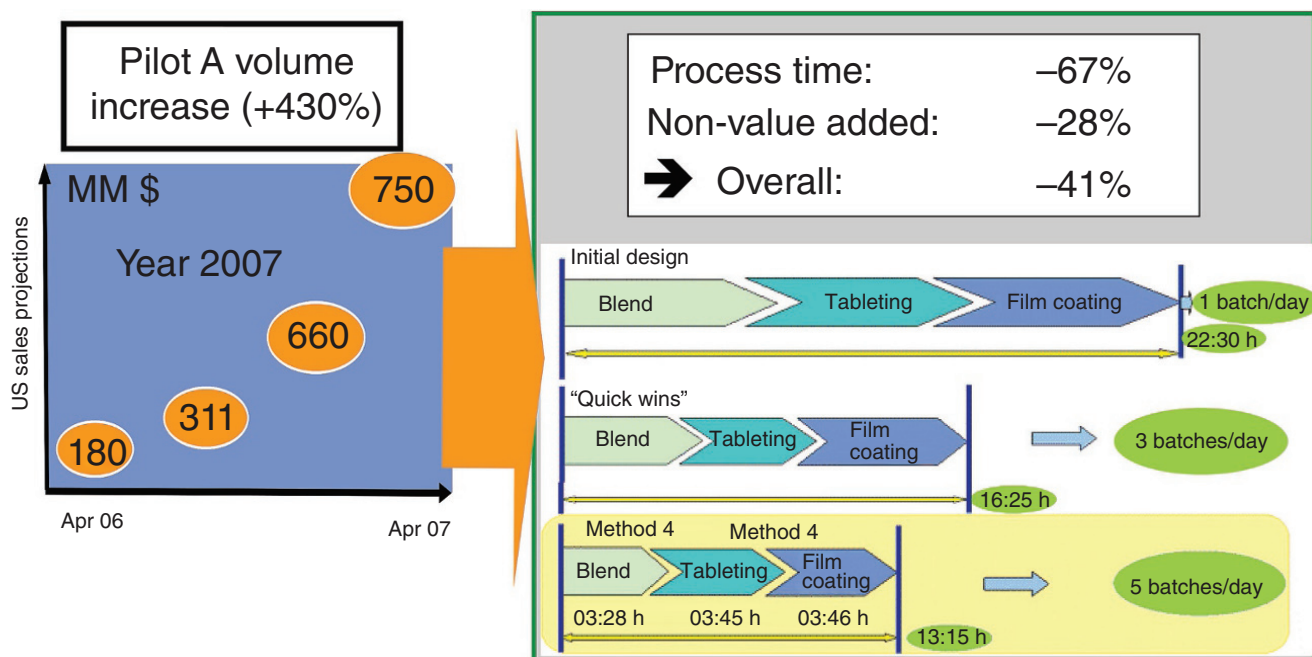


FIGURE 26.5 Important benefits from QbD development of Chantix allowed lean manufacturing and operational efficiency gains.

TABLE 26.5 Regulatory Flexibility During the “Honeymoon Phase” of QbD

Post-approval Change	Traditional SUPAC Criteria	QbD Criteria	Value
DP Mfg. site: Transfer to two adjacent facilities	SUPAC Level 1 – AR • Revised process description • Updated batch record • 1 batch LT stability	AR • Revised process description	Increase Mfg. flexibility and meet market demand
DP Mfg. scale: Increase scale from 115 to 320 kg and 720 kg	SUPAC Level 1 – AR • Revised process description • Updated batch record • 1 batch LT stability	AR • Revised process description	Meet market demand
Expansion of CPP boundary: Press shutoff	CBE or prior approval • Revised process description • Updated batch record • 1 batch LT stability • Change justification • Dissolution profiles	CBE-0 • Revised process description • Change justification	Increase yield/reduced cost
DP test method: Replace KF with NIR	Prior approval • NIR method summary • Comparative data of methods • NIR method validation	CBE-0 • NIR method summary • Comparative data of methods • NIR method validation	Reduced testing time/cost

the revised process description, but not the updated batch record or placing a batch on stability. The benefit of implementing this change was an increase in the manufacturing flexibility and meeting the market demand, which increased significantly.

The drug product manufacturing scale was increased from 115 to 320 kg and 720 kg to meet the market demand, with similar means for notifying as the first case.

For the situations of expanding the design space and changing a method, the SUPAC criteria would typically require prior approval, where any inventory manufactured with these changes would be stored until approval was received. QbD criteria required a CBE-0, allowing notification with immediate use, not pending approval. The value of these changes was an increase in yield, a reduction in cost, and a reduction in testing time.

The overall estimated direct and indirect cost savings from these changes were \$11 MM and \$200 MM, respectively.

These examples of increased scientific debate, enabling lean manufacturing, and corresponding cost benefits made it easy to argue for implementation of the QbD principles and embracing the associated regulatory filings in the future.

The “Honeymoon Phase” of QbD Pilot Program created significant value to products especially those that had a significant increase in product demand. The science- and risk-based approach of QbD further fostered development of new concepts to build for the future, including real-time release, analytical, stability, and packaging QbD (Figure 26.6).

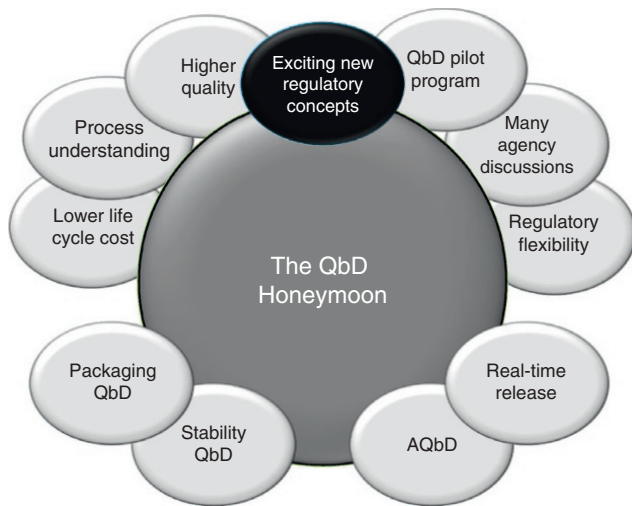


FIGURE 26.6 The “Honeymoon Phase” of Quality by Design Pilot Program creates new building blocks for the future.

26.4 THE STORMY YEARS

There were many benefits realized for the Chantix and Selzentry products realized during the Pilot Program that resulted in flexibility, including facilitating Chantix to stay on the market during the increased demand. In the Stormy Years, the experience of Chantix and Selzentry is compared with the experience with commercial applications several years later.

Seven years later, the business environment had greatly changed. Many of the pharmaceutical companies were reducing headcount (Table 26.6) during the Stormy Years and closing sites. Industry-wide, 18 research and development sites were closed between 2009 and 2013. In addition, companies were filing applications quickly around the world. Figure 26.7 shows the simultaneously global applications for a product; each line is a commercial application submission and approval timeline. Many of the applications are filed at the same time and had overlapping query timelines. More than

TABLE 26.6 Pressures in the Pharmaceutical Industry During the Stormy Years

Company	Work Force Reductions
AstraZeneca	17% reduction in 2014
Merck	15% reduction in 2009, additional 8500 reduction in 2013
Pfizer	15% reduction in 2009, additional 5% in 2013
Eli Lilly	14% reduction in 2007
Teva	5000 reduction in 2013

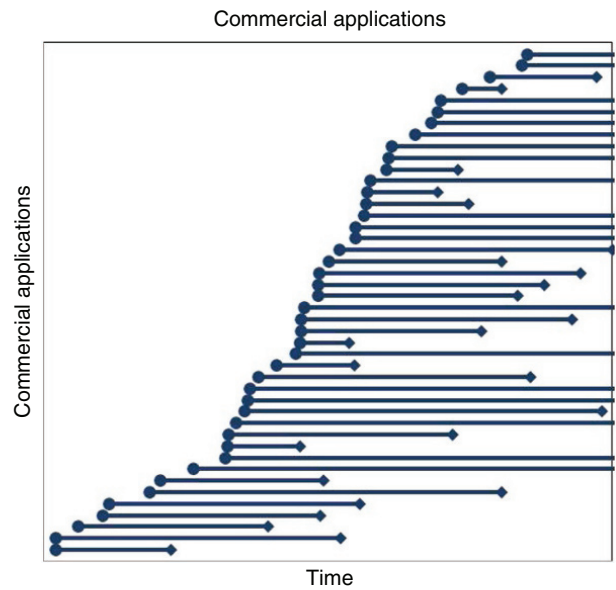


FIGURE 26.7 Global commercial regulatory applications for a product.

25 commercial applications for this product had been filed, and many of them had been approved in less than two years.

26.5 DOSSIER REVIEWS IN THE STORMY YEARS (THE HURRICANES)

The dossier reviews of Pfizer applications in the Stormy Years were substantially different than during the Pilot Program. Figure 26.8 shows the number of queries in the Honeymoon Phase and in the Stormy Years. Also shown in the figure is a non-QbD application during the Stormy Years (the middle bar). The bars for each application represent applications in two major markets. While the figure does not provide the actual number of queries, the dossiers in the Stormy Years received a substantially more queries than the dossiers in the Honeymoon Phase, and one of the markets issues a significantly more queries. In particular, QbD Product X had greater than 300% increase in the number of queries from Market 2. In average, Market 1 had a 140% increase in the number of queries, and Market 2 had a 280% increase in the number of queries when comparing the Honeymoon Phase and the Stormy Years. Also, Market 2 in average issues twice as many queries when examining all six products.

Figure 26.9 shows the total number of pages including the query responses and section updates in the Honeymoon Phase, for the non-QbD dossier, and for the QbD dossiers in the Stormy Years. The initial Pilot Program project Chantix had an extensive number of pages for query responses and section updates for Market 1. Market 2 had a 580% increase in the number of pages between the Honeymoon Phase and the Stormy Years. Also, Market 2 has 240% more pages than Market 1 in the Stormy Years.

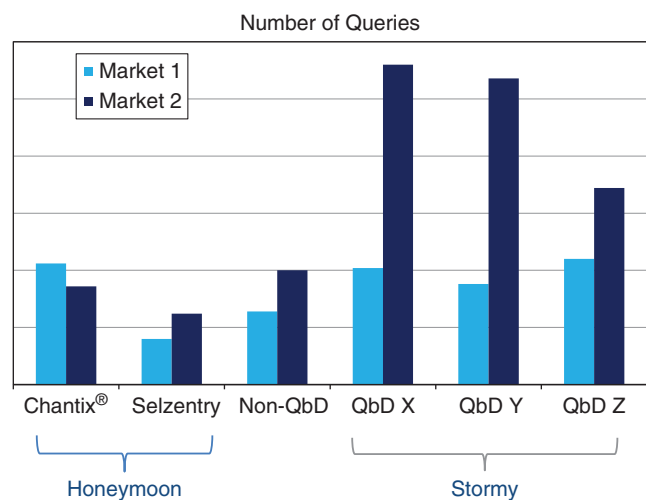


FIGURE 26.8 Comparison of the number of queries during the Pilot Program versus during the Stormy Years.

Figure 26.10 shows the number of pages in sections 3.2.S.2.6 and 3.2.P.2 at approval. In the Stormy Years, QbD X had an exceedingly long 3.2.S.2.6 in one market, and it was 600% longer than in the other market.

26.6 RESULTING DOSSIER CONTENT IN THE STORMY YEARS

There is a substantial difference in resulting dossier content in the Stormy Years; in particular, there is a large increase in the

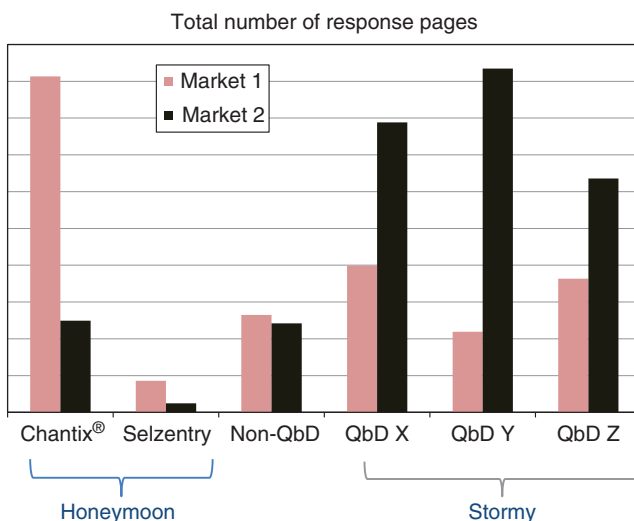


FIGURE 26.9 Comparison of the number of query response pages during the Pilot Program versus non-QbD and the Stormy Years.

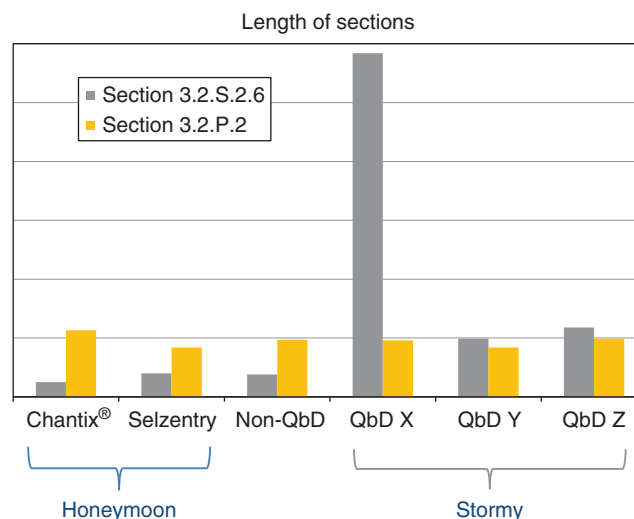


FIGURE 26.10 Comparison of the regulatory filing development sections for API (section 3.2.S.2.6) and drug product (section 3.2.P.2) during the Pilot Program versus non-QbD and the Stormy Years.

Honeymoon example

The CP-X solution in X is adjusted to temperatures not more than [redacted]. X acid [redacted] molar equivalents per equivalent CP-X) is added, followed by X [redacted] molar equivalents per equivalent CP-X) and the reaction stirred for a minimum of [redacted] at temperatures up to reflux. The reaction is cooled and combined with water and the organic product layer is separated. The water layer is washed with X and the organic product layers are combined. The product rich organic layer is washed with water, concentrated by distillation and X is displaced with X. The resulting slurry is concentrated by distillation, stirred for a minimum of [redacted] at temperatures up to reflux. The slurry is then cooled, stirred for a minimum of [redacted], filtered, washed with X and may be dried at temperatures up to [redacted].

Non-“QbD” example

PF-X (limiting reagent) X [redacted] to [redacted] equiv), X (PF-X; [redacted] to [redacted] equiv) and X [redacted] to [redacted] L/kg of PF-X), are combined and maintained at temperatures up to reflux. The reaction mixture is cooled, mixed and the solids are separated and washed. The filtrate is subjected to extractive work-up as follows.

The filtrate is treated with X until a pH of [redacted] is achieved. The phases are separated, and the aqueous phase may be washed [redacted]. The aqueous phase is then mixed with X and treated [redacted] until a pH of [redacted] is achieved. The layers are separated, and the organic layer is washed with water. The organic layer is mixed with X and concentrated. This solution of PF-X is used in Step 2.

Storm example

PF-X, X (up to [redacted] molar equivalents with respect to PF-X) and X [redacted] to [redacted] molar equivalents based on PF-X) are combined and mixed with the aqueous solution of PF-X from step 3 [redacted] molar equivalents with respect to PF-X) in a mixture of X and X (total solvent volume of [redacted] liters per kilogram of PF-X, ratio [redacted]) at temperatures up to [redacted] and are reacted for a time not longer than the upper contour line of Figure X.

Following the reaction, the mixture may be cooled, the phases are separated, the organic layer is extracted into aqueous hydrochloric acid ([redacted] liters per kilogram of PF-X) and X is added [redacted] molar equivalents with respect to PF-X). The aqueous product solution is heated to temperatures not less than [redacted] for [redacted] hours and then cooled, held at room temperature for not less than [redacted] hours and filtered through X, washing the X with either X or X [redacted] liters per kilogram of PF-X). The combined filtrate and washings are basified with X and extracted into a mixture of X and X (total solvent volume not less than [redacted] liters per kilogram of PF-X, ratio [redacted] to [redacted]). The organic phase is separated and washed with X. Additional X is added and the solution is concentrated and seeded. The resulting slurry may be concentrated further (to between [redacted] liters per kilogram of PF-X, and final temperature [redacted]) and then cooled to temperatures not less than [redacted]. The solids are isolated and washed with X and/or X and/or mixtures thereof (not less than [redacted] per kilogram of PF-X).

FIGURE 26.11 Comparison of the regulatory chemistry process description for the API (section 3.2.S.2.2) during the Pilot Program versus non-QbD and the Stormy Years.

number of CPP commitments in the manufacturing description sections.

Figure 26.11 shows a chemistry process description (section 3.2.S.2.2) from the Honeymoon Phase, the non-QbD dossier, and an example from the Stormy Phase. In each example, the highlight indicates the CPPs that are the regulatory commitments. The Pilot Program example has only seven CPPs, the non-QbD example has five CPPs, and the Stormy Phase example has approximately 20 CPPs.

Figure 26.12 is a similar figure for the drug product process description. The Pilot Program example only has two CPPs shown in dark gray highlights and two KPP shown in light gray highlights; the pilot example has KPP that had a lower reporting requirement than CPPs. The non-QbD example has 5 CPPs and the Stormy Phase example has approximately 24 CPPs.

26.6.1 Weathering the Storm

In addition to more queries, longer queries, and longer dossier sections in the Stormy Phase, some of the concepts that were accepted in the Honeymoon Phase were not accepted in the Stormy Phase, and new concepts were required during the Stormy Phase dossier reviews. Figure 26.13 illustrates the changes from the Honeymoon Phase to the Stormy Years.

Some of the changes between the Honeymoon Phase and the Stormy Phase were as follows:

- In the Honeymoon Phase, three levels of critically were accepted: critical, key, and non-critical process parameters. During the Stormy Phase dossier reviews, KPP were not accepted, and KPP were converted to CPPs with higher reporting requirements.
- There was less communication with the agencies during the Stormy Phase than in the Honeymoon Phase. In the Pilot Phase, there were many agency meetings as shown above. In the Stormy Phase, there were less meetings/teleconferences with very limited discussions; when there was a meeting/teleconference, it was only for critical details for approval of the dossier. The limited meetings made the topics described below a challenge to resolve.
- In the Stormy Phase as described above, the required level of detail in the dossier was far greater, and there is therefore less flexibility. In the Stormy Phase, many of the individual query responses were greater than 50 pages long, and for one Stormy Phase commercial application dossier review, the full risk assessment (> 500 pages) was required. There were many more CPPs required in the Stormy Phase than in the Honeymoon Phase.
- In the Stormy Phase, new regulatory requirements were raised during the dossier reviews. For example, the validation of the design space and the movement within the design space were imposed. It can be

Honeymoon example

Method of Manufacture Registration Document		Page 1 of 2
CP-X	COATED TABLETS - 0.5 MG	
CP-X		
DESCRIPTION OF PRODUCTION PROCESS FOR A REPRESENTATIVE BATCH		
Step 1	Add CP-X	silica, microcrystalline, calcium hydrogen phosphate, anhydrous, and silica colloidal, anhydrous to a suitable blender and blend.
Step 2	Pass the blend through an appropriate screen.	
Step 3	Charge magnesium stearate (approximately two-thirds of the total) into the blend from Step 2. Tumble blend contents.	
Step 4	Dry granulate the blend from Step 3 using a dry granulator.	
	Roll Force: 40-60 kN	
	Chute: 5-9 MPa	
	Gap Width: 1.7-2.5 mm	
Step 5	Mill the compacted blend from Step 4 using a screening mill fitted with a screen using sufficient Mill Speed.	
	Mill Screen Size: 0.8-1.0 mm	
	Chute: 0.8-1.2 mm	
Step 6	Change the granulation from Step 5 into a suitable bin and blend.	
Step 7	Change the remaining magnesium stearate to the blend from Step 6 and blend.	
Step 8	Compress the blend on a tablet press.	
	The tablet press is to be shut off (signifying the end of the compression run) with approximately 1.5-2 kg of blend remaining.	
Step 9	Prepare the film-coating suspension using Opady® 3046 (YS-1-1822-A) purified water and film-coating solution using Opady® Clear (YS-2-1914-A) purified water.	
Step 10	Film-coat the tablet cores from Step 8 with the suspension prepared in Step 9. Opady® White (YS-1-1822-A), followed by the solution prepared in Step 9. Opady® Clear (YS-2-1914-A), and submit samples of the finished product (one sent to Quality Operations for final release).	

Current Non-“QbD” example

1. Preparation of the Granulating Solution
1. Dissolve the X and the X in the purified water.

2.2. Granulation and Drying (Intra-granular / Wet Granulation)
1. Add the X, X and X (intra granular portion) to the high shear granulator and mix for a minimum of minutes.
The dry ingredients may be passed through a screen, if necessary, to de-lump prior to preblending.
2. While mixing, add the granulating solution to the high shear granulator and mix for a further minutes to achieve a suitable granulation.
3. To facilitate the drying process, pass the wet mass through a mm screen into a fluid bed dryer.
4. Dry the granulation in a fluid bed dryer. The in-process control for granulation is listed in Table 1.
Table 1. In-Process Controls for granulation
Test Target Value
Loss on drying
1. Pass the dried granulation through a mill fitted with a aperture screen.

Current “QbD” example

DESCRIPTION OF MANUFACTURING PROCESS, INCLUDING PROCESS CONTROL POINTS:

Step 1. Charge X drug substance and X into a suitable liter bin and blend for min (NOR) Range min to rpm.

Step 2. De-agglomerate drug substance and X blend from Step 1 using a suitable mill at rpm +/- rpm; Range = min (NOR) to mm (NOR); mm (NOR) screen size. Collect the material into an appropriately sized liter bin.

Step 3. Charge X, X and X into a suitable liter bin and blend for min (NOR); Range = min (NOR) at rpm.

Step 4. Pass the excipient blend through the mill at rpm; Range = min (NOR) to mm (NOR); mm (NOR) screen size and collect into the liter bin used in Step 2.

Step 5. Blend the drug substance pre-blend 1 and the excipient pre-blend 2 for min (NOR); Range = min (NOR) at rpm +/- rpm in a liter bin.

Step 6. Add X to the liter bin and blend for in (NOR); Range = min (NOR) at rpm.

Step 7. Dry granulate the blend from Step 6 using a roller compactor at a target roll force of (NOR Target), kN/cm (NOR Range), (NOR Range); gap width mm (NOR Target), Range = mm (NOR Range); mm (NOR) mill the compacted blend using a suitable screening mill with a mm (NOR Target), mm (NOR Range) aperture screen into a suitable liter bin.

Step 8. Add X to the liter bin used to collect the granulate from Step 7 and blend for min (NOR); Range = min (NOR) at rpm.

FIGURE 26.12 Comparison of the regulatory drug product process description (section 3.2.P.3.3) during the Pilot Program versus non-QbD and the Stormy Years.

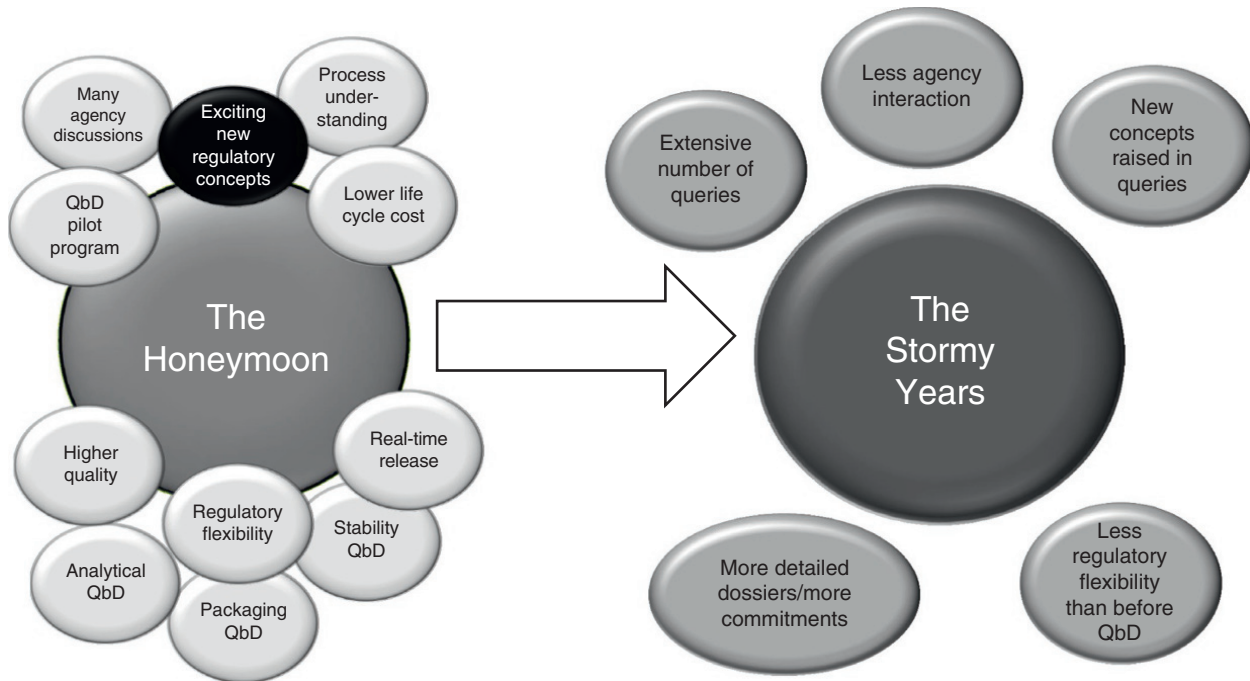


FIGURE 26.13 Quality by design – pilot to reality: the “Honeymoon Phase” and the “Stormy Years.”

difficult for a sponsor company to debate new regulatory strategies during a commercial application review as alignment between many functions is required and obtaining an application approval is highly desired.

- In the Stormy Phase, it was difficult to discuss and implement new regulatory strategies; analytical QbD was proposed in the Stormy Phase dossiers, and with the complexity of the concepts and the lack of meetings, analytical QbD was withdrawn from the dossiers.

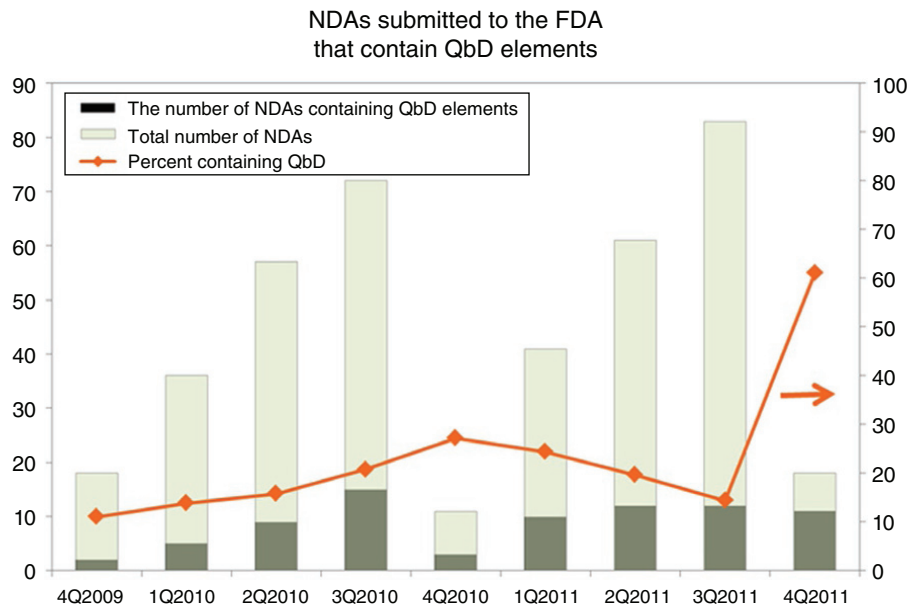


FIGURE 26.14 Number of NDAs containing QbD elements.

26.6.2 Number of NDAs Submitted to the FDA Containing QbD Elements

The FDA provided metrics in 2009 through 2011 for the number of NDAs that contain QbD elements and the resulting information². As shown in Figure 26.14, approximately 35 NDAs per year, or 20% of all NDAs submitted, contained QbD elements. The FDA has not provided QbD metrics since 2011.

26.7 SUMMARY

When QbD started being developed in around 2005, there were many allures of QbD, and many of these concepts came to fruition. Below are the authors' perspective of the allures of QbD and the status of these concepts.

The regulatory environment was transforming:

- From testing for quality to designing for quality. **YES, the regulatory culture is changing from testing for quality designing for quality.**
- From prescriptive and “checkbox” submissions/regulatory to science-based submissions/regulatory reviews. **YES, the regulatory culture is changing from “checkbox” submissions/review to a science- and risk-based submissions/review, and there are different levels of implementations in markets.**

²<https://www.accessdata.fda.gov/scripts/fdatrack/view/track.cfm?program=cder&id=CDER-ONDQA-NDAs-for-NMEs-and-non-NMEs-Containing-QbD-Elements>.

The allure of QbD was to improve:

- Quality. **YES, the systematic approach for design and development is improving quality.**
- Process understanding. **YES, the systematic approach for design and development is improving process understanding.**
- Regulatory flexibility. **A QbD dossier typically has more process knowledge that allows for regulatory flexibility, and also more critical process parameters tend to be required.**
- Application of modern technology, e.g. continuous process. **YES, QbD has made it possible to file modern technology, and the industry had begun to file continuous processes.**
- Capitalize on experience. **YES, platform experience and understanding can be utilized in QbD dossiers.**
- Development efficiency. **YES and NO, the systematic approach for design and development can improve development efficiency, and the extensive queries in the Stormy Phase increased development requirements.**
- Global harmonization. **NO, resulting dossier content differed in the three Stormy Phase examples.**

The allure of QbD was to reduce:

- Cost. **YES and NO, the systematic approach for design and development can decrease cost, and the extensive queries in the Stormy Phase required increased efforts.**

- Regulatory burden. *NO, more critical process parameters were required in the three Stormy Phase examples.*
- Uncertainty and risk. *YES, the systematic approach for design and development can decrease uncertainty and risk.*
- NDA documentation. *NO, typically a QbD dossier has longer sections and receives more queries.*

ACKNOWLEDGMENTS

The authors extend their gratitude to the following colleagues for their contributions: Georg Bernhard, Veronika Lubczyk, Ulrich Dressler, Thomas Katzschner, Andrew Prpich, Holger Weyhers, Emmanuel Ruppert, Ralph Moessner, Yang Ying, Klaus-Peter Kreth, Jurgen Werani, Karl Redl, Chris Neu, Dan Blackwood, Mark Jenkins, Alastair Coupe, Ron Oglivie, Roger Nosal, Timothy Watson, and Thomas Garcia.

The contents of this chapter were shared during two invited talks at the American Institute of Chemical Engineers Annual Meeting [9, 10].

REFERENCES

1. FDA (2009). International Conference on Harmonisation: guidance on Q8(R1) Pharmaceutical Development. *Federal Register* 74 (109): 27325–27326.
2. FDA (2006). International Conference on Harmonisation: guidance on Q9 Quality Risk Management. *Federal Register* 71 (106): 32105–32106.
3. FDA (2009). International Conference on Harmonisation: guidance on Q10 Pharmaceutical Quality System. *Federal Register* 74 (66): 15990–15991.
4. Pickett, J. (2007). *Inspection Monitor*, “CMC pilot program succeeding in integrating concepts of QbD: Nasr,” February 2007 (posted by Entrepreneur website).
5. Morris, W. and Hough, E. (2008). *Parenteral Drug Association (PDA) Newsletter*, “QbD pilot program for biotech seeks similar answers as one for small molecules,” November–December 2008.
6. McCurdy, V., am Ende, M.T., Busch, F.R. et al. (July/August 2010). Quality by design using an integrated active pharmaceutical ingredient-drug product approach to development. *Pharmaceutical Engineering* 30 (4): 1–16.
7. Garcia, T., Cook, G., and Nosal, R. (2008). PQLI key topics – criticality, design space, and control strategy. *Journal of Pharmaceutical Innovation* 3: 60–68.
8. Nosal, R. and Schultz, T. (2008). PQLI definition of criticality. *Journal of Pharmaceutical Innovation* 3: 69–78.
9. am Ende, M.T. and Seymour, C. (2013). QbD pilot to reality: the honeymoon phase. AICHE Meeting, San Francisco, CA (4 November 2013).
10. Seymour, C. and am Ende, M.T. (2013). QbD pilot to reality: the stormy years. AICHE Meeting, San Francisco, CA (4 November 2013).

INDEX

- absorption, drug, 383, 388, 405
active pharmaceutical ingredients (API), 3, 4, 8–17, 451, 452, 454–456, 467, 633–637, 640, 641, 643
actuator dynamics, 576
actuators, 567, 570, 576
additive manufacturing, 14, 16, 17
adhesion (adhesive), 227, 228, 234, 235, 238
advanced process control (APC), 573, 574, 576, 577
advanced regulatory control, 570
aerosol test, 364
aggregation, 175–178, 180
agitation, 148–150, 154, 163
agitator, 452–456
amorphous solid dispersion, 267–272, 274–277, 281, 287, 288
angular momentum, 423
annealing, 297–299, 454, 455, 463, 466
Antoine's equation, 418
arching, 58–60, 64–66, 69, 74, 196, 197
aseptic processing, 324, 347, 370–374
asymmetric membrane, 389, 401, 402
atomization, 154, 273, 275, 278, 282–288
 airflow, 420, 421, 431
attrition, 147, 148, 160
auger, 505–507
autoclaves, 316–318
automation, 562, 567, 577, 580
AVL code coupling interface (ACCI), 424, 425
axial dispersion effect in continuous sterilization, 337–339
- bacteria, 312
batch process monitoring, 612, 624
batch sterilization, 316, 318, 320–322, 331–335
BIER apparatus, 328, 329
bimodality, 176–178
- bimodal size distribution, 129, 130
bin, 489, 500–504
 design, 57, 58, 61, 68, 71, 72, 74–76
 recommended mass flow hopper angles, 68, 69, 73
 recommended outlet to prevent arching, 65, 66, 74
binder, 82, 89, 104
bioburden, 322–324, 339, 371–374, 377
biological indicators (BI), 318
Biologics Licensing Application (BLA), 8, 9
biopharmaceutical, 21–27, 33
biopharmaceutical classification system (BCS), 26, 27, 29, 34–36, 50
birth event (PBM), 447
blending, 33, 35, 45, 46, 48, 53–57, 62, 73, 74, 547, 555, 560
Bond number (Bog), 192–194, 196, 197
boundary condition, 97, 99, 102
boundary layer thickness, 257–259
bowl, 188, 189, 197, 198
breakage, 147, 148, 160–164, 167, 173, 175, 176, 178, 180
breakage function, 128, 129, 131, 132, 143
breakage zone, 163
brittle, 228, 236, 240
bubble point test, 362, 366–368
bulk density (ρ_{bulk}), 187, 191, 193, 196–198
bumping flow, 165–167
business intelligence, 580
- cake filters, 356
capillary, 148–150, 161
capillary number (Ca), 155, 160
capsules, 187, 188, 199
cell average technique, 176
cellulose acetate (CA), 248

- Center for Biologics Evaluation and Research (CBER), 4
Center for Drug Evaluation and Research (CDER), 4
central processing unit (CPU), 424, 425, 431, 441
Cesium-137, 342
CFD-DEM, 421, 422, 424, 425, 427–433, 436, 438, 441, 442, 444, 445, 447, 448, 503
CFD-TFM (two fluid model), 421, 422, 425–428, 447
Chantix, 634, 636–640
characterization, 23, 27, 28
chemical imaging, 597
chemical potential, 387
chemical sterilization, 314, 345–356, 373
chlorine dioxide, 350
chopper, 155, 156, 163
 speed, 156
circulation time, 154
clinical scale, 452
coalescence, 148, 151–153, 157–160
coating deposition, 413, 414
coating growth, 160
coating uniformity, 410, 413, 414, 416, 419–421, 433, 435–437, 442, 445, 447, 448
Cobalt-60, 342
coefficient of proportionality, 259
coefficient of restitution, 423, 424, 426, 430
coefficient of variation (CoV), 435–437, 441, 445, 446, 448
cohesion, 227, 228, 231
cohesive, 189, 190, 192, 193, 195–197, 199
cohesive strength, 59, 60, 62, 64–66, 68, 71, 72, 74, 77
collapse, 295
collapse temperature, 296, 297, 304
collision, 156, 158
collision velocity, 155, 156, 158–164, 168
commercial scale, 452
compaction force, 234
compaction simulation, 204, 208, 209
compaction simulator, 235, 236
compartment model, 419, 420, 428, 437, 438, 441–448
compatibility, 452
compound annual growth rate (CAGR), 4, 6, 7
compressibility test, 69
compression, 188–191, 195–199, 203–209, 211–222
 distribution, 233
 force, 204, 207, 208, 219, 220, 222, 227, 230, 233, 234, 238, 240
computational fluid dynamics (CFD), 267, 282, 284, 287, 290, 419, 421, 422, 424–433, 436, 438, 441, 442, 444, 447, 460, 461, 490, 491, 499, 503
computational predictive models, 490, 506
congealing, 454–456, 463
conservation equations, 280
Consigna25, 174
consolidation, 147, 148, 151, 155–157, 159, 163, 165, 167
constraints, 567–570, 574, 578, 579
contact angle, 148
contact model (DEM), 192, 196, 423, 424, 431
continued process verification, 536, 545
continuous control, 567, 568
continuous direct compression, 547
continuous manufacturing (CM), 15, 16, 173, 529–531, 533, 543, 545
continuous mixing technology (CMT), 547–554, 556, 557, 560
continuous production, 529
continuous sterilization, 318–320
controlled release, 248
 equations for, 384, 390–400
control strategy, 4, 318, 375, 535, 536, 539, 541, 542, 544, 545, 601–602, 619, 623–625
control systems hierarchy, 567, 575
control volume, 489–491, 515–517, 519
cooling, 452–454
 HME, 472–474, 477, 478, 480
Coomans plot, 592, 598
crack propagation, 160
critical attributes, 410, 413, 416
criticality, 632–635
critical material attributes (CMA), 562, 563, 570, 571, 575
critical process parameters (CPP), 43–46, 49, 421, 454, 562
critical quality attributes (CQA), 33, 36–38, 42–47, 49, 370, 561–564, 570, 571, 573–576, 633, 634
crushing force, 220, 221
current Good Manufacturing Practice (cGMP), 4, 10, 370, 373, 376
cycle time, 420, 431, 433
cyclone, 273, 274, 276, 280, 283, 288
damping parameter (DEM), 423
death event (PBM), 447
decimal reduction time *see* D-value
deformation, 155, 157–161, 231
degradation, 452, 454, 456
delayed release, 409–411, 413, 416
dense discrete particle model (DDPM), 425
densification, 147
density, 247, 248, 250, 253, 257, 261–265, 278, 281–283, 285–288
depth filters, 355, 356
depyrogenation, 321, 339
design of experiments (DoE), 29, 42, 47, 174, 175, 472, 481, 484
design space, 3, 469, 472, 480, 481, 486, 509, 520, 522, 571, 573, 601–602, 604, 620–626
destructive nucleation, 163, 164, 168
die-wall pressure, 231
differential scanning calorimetry (DSC), 275, 277, 464, 466
diffusion, 383–387, 390, 391, 393, 398–405
 test, 361–364, 367–369
discrete control, 567, 568
discrete element method (DEM), 190–192, 196, 197, 199, 211, 216–219, 419, 421–423, 447, 448, 491, 498–504, 506
discrete particle model (DPM), 284, 425
discrete sensors, 567, 568
discrete-time Markov chain (DTMC), 438, 439
disinfection, 315, 341, 342, 353
disintegration, 232
dissolution, 21–25, 34–39, 44, 45, 47, 49, 50, 151, 156, 167, 232, 268–270, 272, 280, 456

- profile, 411, 412
- rate coefficient, 400, 404, 405
- solid solution, 470, 479, 480
- distribution nucleation, 148
- distributor plate (Wurster coating), 420, 421, 428–431, 433
- dosage form, 381, 383–386, 389, 390, 395, 397, 399
 - development, 409, 411, 413, 416
- dosator, 187–199
- drag coefficient, 259, 422, 423, 426
- drag force, 422, 424, 426
- drop controlled regime, 154
- drop/droplet, 148, 150, 151, 153, 165
- droplet size, 245, 247, 250, 252–255, 257, 260–263, 265, 414
- droplet transport, 245, 246, 255
- drop penetration time, 148–150, 153, 154, 163
- Drucker–Prager cap model, 212, 213, 215, 216, 513, 514
- drug delivery, 381–384, 386, 387, 389
- drug eluting stent (DES), 597, 598
- drug product, 3, 12, 13
- drug release, 409–412, 414
 - mechanism, 384
- dry granulation, 31–34
- dry heat sterilization, 314, 320, 321, 339, 373
- drying, 245–247, 252
- drying kinetics, 268, 276, 283–286
- drying rate, 414, 416
- dry layer resistance, 300, 303, 304, 306
- ductile, 229
- duration of release, 411, 412
- D-Value, 327, 328, 346, 347
- dwelt time, 207, 230, 231, 234, 238, 240
- dynamic pressure ratio, 259
- DY-value, 345, 346

- ease of spreading, 463, 466
- economic model predictive control (E-MPC), 579
- ejection, 234
- ejection force, 208, 233, 235, 236, 238
- elastic, 228
 - deformation, 231
 - recovery, 231
- electron beam (E-beam) sterilization, 341–343, 346
- electrostatic, 238
- encapsulation, 490, 503
- endospores, 312
- energy balance, 267, 280
- energy transfer, 492
- enthalpy, 475, 478
- equivalent sterilization cycles, 334–339
- erosion, 384–387, 395, 396
- ethylene oxide, 347, 348, 353–355
- EU, 371, 373, 374, 376
- eukaryotes, 312
- EU market authorization applications (MAA), 632
- European Medicines Agency (EMA), 3, 370–373
- excipients, 27–29, 31–36, 39–41, 46, 50, 82, 89, 90, 95, 101, 383, 403, 476
- extended particle system (XPS), 424
- extragranular, 119, 120, 139

- feed rate, 93, 106, 109, 113
- feed screw, 83, 84, 96, 97, 104, 109
- field level, 567
- filling, 187–199
- fill level, 165, 166
- fill weight, 187–190, 195, 197, 207
 - variability, 188–190, 207, 219
- film coating, 389, 514–516, 518–522
 - modeling, 491, 514, 515, 517–520
- filter, 478, 483
- filtration sterilization, 315, 355–370
- finer, 157, 158
- finite element analysis (FEA), 512–514
- finite element method (FEM), 211, 212, 216, 512
- first in, first out, 529
- flexible batch size, 545
- flowability, 53–59, 66, 119, 127, 134, 135, 138–142
- flow function, 64–66
- flow patterns, 53, 58–60, 71, 74, 77
- flow rate, 150, 151, 165, 167
- fluid bed coating, 409, 410, 413–417
- fluid distribution, 150
- fluidised bed, 148, 150, 151, 155, 159, 160, 163, 165, 166, 168
- fluidization, 413, 414, 416
 - airflow, 420, 421, 428–432, 445, 446
 - bubbling, 512
 - segregation, 500, 504
 - turbulent, 512
- fluidized bed(s), 419–421, 423, 425–428, 431, 437, 438, 441, 447
- Food and Drug Administration (FDA), 3, 4, 8, 9, 14, 16, 17, 633, 637, 643
- forces, 188, 190–193
- formulation, 153, 154, 166, 381–383, 388, 389, 395, 401, 451, 452, 454, 467
- fracture, 147
- fragmentation, 231
- free energy diagram, 475
- free-flowing, 189, 192, 193, 196, 199
- freeze drying, 293–296, 299–306
- freezing, 293–299, 303–305
- friction, 157
 - DEM, 422, 424–426, 430
 - solid solution, 474, 475
- frictional forces, 234
- Froude number, 166, 460, 491, 509–511
- functional coating, 409–411, 413
- funnel flow, 60, 61, 66–69, 71, 72, 74–77
- F-Value, 328, 330

- gap, 188, 190, 195–198
- Gaussian noise, 178
- Geobacillus stearothermophilus*, 318, 322, 325
- Gibilaro model, 426
- Gidaspow model, 422, 423, 426, 427
- glass transition, 234

- glass transition temperature (T_g), 269–272, 277, 281, 282, 290, 297
- glass transition temperature of frozen concentrate (T_g'), 296–299
- glidant, 237, 240
- global impact, 4
- global optimisation, 178
- global sensitivity analysis (GSA), 262, 263
- Gordon–Taylor equation, 271
- granular temperature equation, 427
- granulation, 147, 149, 154, 235, 237
- granulator scale, 153
- granule, 119, 120, 127–129, 131–137, 139–142
 - density / porosity, 155–158, 163, 166
 - growth, 147, 148, 155, 157–159, 163–167
 - size, 92–94, 99, 100, 106, 107, 163, 164, 166–168
 - strength, 155, 156, 159–164
 - structure, 156, 166
- granule size distribution (GSD), 173–184
- graphical processing unit (GPU), 424, 425, 431
- gravimetric feeder, 547–550, 552–554, 557, 560
- hardness, 209, 220, 221
- heating, 452–454
- heat transfer, 283, 298–306, 414, 489, 512, 517
 - coefficient, 300, 301, 303–306
- heat transfer conductive, 474–476
- Heckel analysis, 209–211, 219
- HEPA filters, 320, 321, 355, 360, 364
- high shear mixer, 151, 156, 158, 159, 166, 167
- homogeneous, 470, 478, 479
 - Markov process, 438, 439
- homogenizer, 452, 454–456
- Honeymoon phase, 631, 632, 637–642
- Hook's law, 413
- hopper, 491, 500, 503–508
- hot melt extrusion (HME), 469, 472, 486
- hydrogen peroxide, 348
- hydrostatic pressure, 411, 412, 416, 417
- hypromellose (HPMC), 247, 248
- ice nucleation, 296–298
- ICH, 631, 632
- immediate release, 247, 248, 409–411
- immersion nucleation, 148
- impact and spreading, 245, 246, 251, 253, 254
- impact velocity, 155, 156, 158–164, 168
- impeller, 155, 160, 162–166
 - design, 489, 492, 505–511
 - speed, 155, 156, 162–167
 - tip speed, 162, 165, 166
- induction growth, 157–159, 166
- inertial growth, 160
- inhomogeneities, 189, 190
- inhomogeneous, 478, 479
- in-line blender, 533, 541
- in-line (powder) mixer, 547, 548, 553–556, 560
- integrity testing of sterilizing filters, 361–367
- intermediate bulk containers, 53, 55–59
- internal friction, 89, 92, 113
- International Council on Harmonization (ICH), 4
- interphase momentum exchange (CFD-TFM), 425, 426
- intragranular, 120
- ion-exchange, 386, 387, 396, 397
- ionizing radiations, 341, 342
- irradiation sterilization, 314
- irreducibility (Markov process), 438, 439
- irreversible, 463
- Japan New Drug Application (JNDA), 3
- Japan Pharmaceutical Affairs Law (JPAL), 4, 632, 634
- Kawakita equation, 210, 211
- Kelvin–Helmholtz (K-H), 257, 258
- key, 633–637, 641
- kinetic energy, 159, 160
- kinetic theory of granular flows (KTGF), 425–427
- kneading zone, 175, 176, 178, 180–182
- knowledge space, 632, 633, 635
- lag time, 410–412
- laminar flow, solid solution, 475
- Langmuir isotherm, 397
- Laplace's law, 412
- latent variable, 571
- latent variable model (LVM), 490–492, 494
- latent variables, 602, 606–608, 610–612, 615, 616, 618, 620, 622
- layering, 160
- lean manufacturing, 15, 16
- lethal rate, 333, 370
- level 1 control, 571, 573
- level 2 control, 570, 571
- level 3 control, 570, 571, 573
- LIGGGHTS, 191, 192
- liquid addition, 150, 157, 165
- liquid binder addition, 150, 157, 165
- liquid bridge, 156
- liquid distribution /dispersion, 147, 153, 154, 163–165, 167, 174–177, 180
- liquid level, 155, 165, 167
- liquid saturation, 148, 156–159, 161, 163, 166–168
- liquid-to-solid ratio (L/S), 174–176, 178, 180, 183
- local sensitivity analysis, 263
- loss in compactability, 133, 134, 140
- lubrication, 237, 238
- lubricity, 233
- lurication, 119, 120, 138, 140, 142
- lyophilization, 293, 295, 298, 299, 301–305
- macrovoid, 149, 150
- management level, 567, 579
- manometric temperature measurement (MTM), 303, 304
- Markov chain, 420, 428, 437–444, 447, 448
- Markov process, 438
- mass flow, 60, 61, 65–69, 71–78, 298, 299, 305
- massflux ratio, 259
- mass gain (coating), 435–437
- mass transfer, 283, 295, 299, 300, 302–306, 409, 414, 415, 489, 492

- Mastersizer, 175, 176
 material balance, 516, 517, 519
 material traceability, 542
 mathematical modeling, 247, 257
 MATLAB, 263
 matrix, 382, 384–387, 389, 390, 392–396, 399, 400, 403–405
 mean residence time (MRT), 534, 541
 mechanical dispersion regime, 154, 163, 165, 168
 mechanical properties, 23, 27, 28, 33, 34, 50
 mechanism, 383–385, 387, 389, 398, 399, 401, 405
 mechanism of release, 410–412, 416
 membrane filters, 356–358
 membrane-reservoir, 384, 385, 390, 392, 395, 402
 mesh (CFD-DEM), 424, 428–431
 microbial challenge test, 361
 microbial contamination, 312, 322–324, 356, 358
 see also Bioburden
 microbial death (microbial inactivation), 313–316, 321, 322, 325–335, 339, 341, 345–348, 353–354, 370–374
 microbial thermal death kinetics, 325
 milling, 119, 120, 127, 129, 131–135, 140, 167, 168, 490–492
 mill model, 127, 128, 135, 136, 138
 mini-batch basis, 536
 Ministry of Health, Labor and Welfare (MHLW), 4
 mixer granulation/mixer granulator, 151, 156, 158, 159, 166, 167
 mixing, 147, 452, 454–456, 472–476, 478, 480, 481, 491, 500, 502, 503, 508
 modeling, 96, 99, 101, 102, 104, 188, 190–193, 196, 197, 199, 293, 295, 298–306, 489–493, 495–500, 502, 503, 506, 507, 509, 511, 512, 520–522, 602, 606, 612–614, 616, 618, 619, 621, 625
 model inversion, 573
 model predictive control (MPC), 574–579
 modernizing pharmaceutical manufacturing, 529
 modulus of elasticity, 413, 416
 moist heat sterilization, 314, 339, 373, 375 *see also* steam sterilization
 moment of inertia, 423
 momentum conservation (CFD-TFM), 425
 Monte Carlo, 420, 428, 433–437, 441, 447, 448
 multiblock analysis, 616, 617
 multiparticulate, 409–413, 416, 417
 multiple-input multiple-output (MIMO) system, 570, 574
 multiplicative scatter correction (MSC), 594
 multiscale modeling, 446
 multivariate analysis (MVA), 486, 587, 596, 598, 599
 multivariate data analysis, 275
 multivariate process data formats, 605
 multivariate statistical process control, 571, 601, 609

 Navier–Stokes equation, 421, 422, 425
 near-infrared spectroscopy (NIR), 275, 587, 590, 592–597
 new chemical entities (NCE), 4, 6, 8, 10
 New Drug Application (NDA), 3, 8, 9, 16, 631, 632, 634, 643
 new molecular entities (NME), 4, 6, 8
 Newton's equation of motion (CFD-DEM), 422, 423
 nip, 82–84, 88, 90, 95–99, 102–105, 109, 113
 nip angle, 90, 98, 99, 103, 104, 106, 113
 nitrogen dioxide, 349, 350

 non-inertial growth, 159
 nonionizing radiations, 342–344
 non-Newtonian, 247, 258
 normal stress, 97–99, 104, 113
 Noyes and Whitney dissolution model, 456
 nozzle, 148, 153, 165, 187–193, 195–197, 199, 273–276, 283–288
 nucleation, 147, 150, 153, 154, 159, 165
 nucleation number, 152
 nucleation ratio, 151, 152
 nucleation regime, 148, 153, 154, 165
 nucleation regime map, 148, 153, 154
 nuclei/nucleus, 147, 148, 150, 151, 157, 159
 nuclei size distribution, 147, 148, 151, 153
 Nutsche filters, 357

 objective function, 178, 180, 184
 Ohnesorge number, 248, 258, 259
 ointment, 451–457, 463–467
 ON-OFF control, 569, 574
 Opadry, 248, 250, 256, 263
 open-loop manufacturing, 573
 optical coherence tomography (OCT), 421, 446
 optimization, 556
 optimization level, 577, 579
 osmotic delivery (or release), 389, 399, 401, 402
 osmotic pressure, 411, 417
 osmotic rupturing, rupturing multiparticulate, 409–413, 416, 418
 out-of-specification product, 562, 576, 577
 over-compression, 223
 over-wetting, 247, 251, 252
 Ozone, 348, 349

 parametric release, 374, 375
 partial least squares (PLS), 594–598, 606, 607, 609–613, 615, 617, 619, 620
 particle circulation, 413, 414, 416
 particle formation, 270, 284
 particle morphology, 277, 278, 284
 particle properties, 277, 279, 287
 particle size, 81, 82, 88, 92, 93, 95, 96, 101, 104, 106–108, 110, 112, 113, 155, 156, 162, 163, 190–193, 195, 197
 particle size distribution (PSD), 273–275, 277–279, 412, 414, 416
 PAT data management platform, 577
 patent, 7–10, 14
 pattern air, 250–253, 255, 257, 260–262, 265
 peak flow stress, 155–157, 161
 PEPT, 163, 165
 permeability test, 70
 personalized medicine, 16
 Pharmaceuticals and Medical Devices Agency (PMDA), 3, 8, 370
 phase Doppler anemometry (PDA), 254
 phase Doppler particle analysis (PDPA), 254, 265
 physical stability, 268–272, 274, 276
 physicochemical properties, 451, 452, 467
 PID control, 569, 570, 573, 574
 plasma sterilization, 348, 350
 plastic, 228, 236, 240
 deformation, 160, 161, 164, 229, 231, 234

- plug, 187–191, 195, 197–199
 plug flow, 535, 541
 Poisson distribution, 152
 Poisson's ratio, 208, 212, 214, 215
 polyethylene glycol (PEG), 248, 250, 254, 451, 452, 454, 455, 463–467
 polymer, 267–273, 276, 277, 290
 polymorph, 23–25, 466
 polyvinyl alcohol (PVA), 247, 248, 250, 254, 260, 261
 poorly soluble drugs, 268, 269, 272
 population balance equations, 127
 population balance model (PBM), 129, 131, 138, 173–178, 180, 183, 184, 446, 447, 459
 calibration, 174, 177, 178, 180, 183, 184
 compartmentalization, 176
 general, 173–175
 kernel, 176–178, 184
 numerical solution, 176
 pores/porosity, 147–151, 155–157, 159, 166
 porosity, 204, 209, 211, 219–221, 414–416
 portable continuous miniature modular (PCMM), 547, 548, 552
 powder, 187–199
 powder agitation, 154
 powder bed, 188–199
 height, 188, 190, 191, 193, 195–197
 powder flow, 84, 90, 92, 96, 97, 102, 105, 107, 109, 110, 498, 501, 503, 506
 powder velocity, 150, 153, 154, 165, 166
 precipitation, 268, 269
 pre-compression, 204–207, 230
 predictive modeling, 39, 41
 pressure, 191–199
 hold test, 364, 369, 370
 primary drying, 293–296, 298–305
 principal component, 588, 591
 principal component analysis (PCA), 496, 497, 588–592, 595, 606, 607, 609–613, 615, 619
 probability matrix, 438–444, 446, 448
 process, 187–191, 193, 195, 197, 199
 analytically controlled technology, 573
 capability, 113
 control, 561–571, 577–580, 601–604, 609, 617, 625
 disturbances, 563, 576
 flow, 454
 map, 415, 416
 modeling, drug product, 490, 491, 498, 522
 parameter, 188, 190, 191, 195, 199, 633, 634, 636, 637, 641, 643
 performance, 578, 580
 scale-up, 282, 286, 287
 signature, 604, 618, 620
 space, 415, 416
 understanding, 601, 602, 604–609, 611, 616, 623, 625
 process analytical technology (PAT), 3, 16, 17, 274–276, 469, 472, 482, 484–486, 561, 562, 570, 572–574, 577, 580, 587, 599
 product tracing, 576
 programmable logic controller, 568
 projected area, 259
 prokaryotes, 312
 property model, 132, 135, 136, 138
 proven acceptable range (PAR), 634, 635, 637, 642
 psychometric chart, 416, 418
 pulsatile release, 410
 punch coating, 233–236, 238, 240
 punch cup volume, 232
 punch geometry, 233–235, 240
 punch picking, 227
 punch sticking, 227, 231, 233–237, 239

 QbD Pilot Program, 631–635, 637, 639, 642
 Qicpic, 174, 176
 quality attributes, 188, 199, 204, 207
 quality by design (QbD), 3, 4, 15, 409, 529, 545, 602, 620, 624, 631–635, 637–643
 quality target product profile (QTPP), 452

 radiation sterilization, 341–345
 Raman spectra, 587, 590, 593, 595, 597
 rapid growth, 159
 rate constant, 396, 405
 ratholing, 58–61, 64–66, 71, 72
 Rayleigh–Taylor (R-T), 257
 real-time optimisation (RTO), 578, 579
 real-time release, 602, 625
 rebound, 160
 receding horizon, 574, 575
 recirculation, 452–454
 regression analysis, 30, 41
 regulatory considerations, 625
 regulatory filing strategy, 3
 regulatory level, 567–570
 relative density (ρ_R), 193, 195–199, 209, 211, 212, 214, 215, 218, 219
 relative saturation, 280–282, 287, 288, 290
 release profile, 409–412
 research & development investments, 4, 6
 residence probability, 442, 443
 residence time, 175, 176, 420, 431, 432, 434, 436, 442, 443
 residence time distribution (RTD), 431–434, 436, 447, 476, 478, 480, 481, 533, 541, 547, 553–560, 576
 residuals, 591, 592, 596–598
 revenue, 4–7, 9, 14
 Reynolds number, 258–260, 456, 457
 rheological properties, 463, 465
 ribbon, 119, 120, 122–124, 126–129, 131–138, 140, 143, 227
 strength, 90, 94, 106
 thickness, 93, 94, 105, 106, 109, 112, 113
 risk assessment, 469, 472, 477, 480, 564
 roller compaction model, 81–84, 87–96, 99–107, 109, 111, 113, 119, 120–126, 129, 132–136, 138, 141, 143
 roll(er) gap, 92, 93, 98, 104, 107, 109–111, 113
 roll pressure, 90, 91, 101, 103, 105, 109–111, 113
 roping flow, 162, 165–167

 saturation, 148, 156–159, 163, 165–168
 Savitzky Golay, 592, 593
 scale down, 470

- scale-up, 23, 32, 33, 35, 41–46, 48, 49, 81, 105–109, 111, 113, 119, 122, 126, 138, 143, 163, 165–167, 247, 262, 263, 293–295, 299, 300, 303–306
 for dissolution, 453, 456, 463–465
 Schmidt number, 456, 457
 scraper, 452–456
 scraper pressure, 234
 screw speed, 174, 175
 secondary drying, 273–275, 293–295, 298, 301–304
 seed core, 410, 413, 416
 segregation, 228, 236
 dusting segregation, 61, 63, 78
 fluidization segregation, 61–63, 71
 sifting segregation, 61, 62, 71
 Selzentry, 631–633, 635, 639, 640
 semi-brittle failure, 160, 161, 163, 164
 semipermeable (layer, membrane, polymer), 410, 411, 413, 417
 semipermeable membrane, 387–389, 397, 398
 sensitivity analysis (SA), 262–264
 sensitivity indices, 263–265
 sensory evaluation, 466
 set point, 563, 568–570, 574, 575, 578, 579
 shear, 453, 454
 shear rate, 248, 249, 258, 261–263
 shear stress, 148, 151, 154, 160, 162, 164, 167, 168, 472, 474, 475, 480, 481
 shear test, 65, 67
 shelf life, 21, 30, 34, 35, 39–41
 Sherwood number, 456, 457
 simulation, 190–193, 197–199
 single droplet (or particle) drying, 267, 282, 290
 single-input single-output (SISO) system, 573, 574
 size distribution, 147, 148, 151, 153, 160
 size reduction, 147, 156
 soft independent modeling of class analogies (SIMCA), 590–592, 598, 599
 soft sensor, 572
 solid bridge, 156
 solid dosage form formulations, 504, 509–511, 515, 516, 518, 521
 solid fraction, 232
 solid oral dosage form, 22, 31, 34, 46
 solid oral drug product, 490–492, 498
 solids bulk viscosity (CFD-TFM), 425, 426
 solids pressure (CFD-TFM), 425, 426
 solids shear viscosity (CFD-TFM), 426
 solubility, 451, 452, 454–458, 463, 467
 aqueous, 469, 480
 drug, 470
 enhancement, 268, 280
 solvent, 267–275, 280
 spectroscopy, 275, 277, 279
 spore-forming bacteria, 312
 sporulation cycle, 313
 spray area, 153, 154
 spray dryer, 470, 471
 spray drying, 267–270, 272–278, 280
 spray flux, 148, 151–154, 165
 spraying, 147
 spray plume, 285–288
 spray rate, 150, 151, 165, 167, 421, 442, 444–447
 spray zone, 148, 150, 151, 153, 155, 165
 spreading, 152
 spring stiffness (DEM), 423, 424
 stability, 21, 23–27, 29, 30, 32, 38–41, 45, 50
 modeling, 39–41
 standard error of prediction (SEP), 596
 standard normal variant (SNV), 593, 594
 standard round concave, 233
 state of control, 529, 532, 544
 stationary distribution (Markov process), 438–440
 statistical process control (SPC), 570–572
 Stdef, 156, 158, 159, 161, 162, 165
 steady growth, 157–159
 steady state, 415
 granulation, 166, 167
 steam sterilization, 316, 318, 321, 328–331, 339
 sterile filtration, 315, 324, 355, 356, 360–362, 365, 371–373 *see also*
 filtration sterilization
 sterility, 311
 sterility assurance level (SAL), 323, 324, 326, 336, 342, 370–372
 sterilization, 311
 of culture media, 318–320, 336–339
 of solids, 312–322, 339–341
 sterilizing chemical agents, 345–347, 353
 sterilizing filters, 355, 358–361
 Stokes deformation number (Stdef), 156, 158, 159, 161, 164, 165
 Stormy years, 631, 632, 639–642
 strain rate, 155, 160
 supervisory level, 570, 574
 surface area, 157
 surface tension, 155, 247–250, 253, 254, 258, 259, 261–265
 surfactant, 268–270, 272, 273, 279
 sweller layer, 410, 412, 413, 416
 swelling, 384, 386, 387, 389, 396, 399, 400
 swept volume, 166
 Syamlal–O’Brien model, 426
 system model, 121, 127, 135–138, 141, 143
 tablet, 156
 compression, 490, 492, 512, 513
 disintegration rate, 220
 formulation, 21, 23, 29, 32, 33, 36
 hardness, 113
 mixing, 246
 press-rotary, 204–206
 press-single station, 204
 punches, 204, 206
 tensile strength, 132–144
 take-off force, 234, 236
 tangential forces (DEM), 424
 Target Product Profile (TPP), 21, 22, 33, 34, 42, 45, 46
 tensile strength, 203, 204, 207, 219–223, 229, 232, 234
 terminal sterilization, 324, 349, 370–375
 thermal degradation of product, 330, 335, 336
 thermal sterilization, 314, 315, 331–335
 thermodynamic model, 280, 287, 491, 515, 518–521
 3D printing, 14, 16, 17
 throughput, 174, 175

- time delay, 569, 570, 574
tracer granule, 160, 163, 166
triboelectrification, 234
tribology, 229
trouble-shooting, 53, 56, 71
 flow problems, 71–74
 segregation problems, 77, 78
true density (ρ_T), 193, 196, 198, 209, 219
twin-screw granulator (TSG), 173–175, 177, 180, 183, 184, 540, 541
twin-screw wet granulation (TSWG), 173, 174, 183
two-fluid atomizer, 413
two-fluid coaxial atomizers, 245, 247, 250, 254, 265
two-phase flow effects, 60, 71, 75–77

US Food and Drug Administration (FDA), 370–374
US Pharmacopeia (USP), 372, 373
UV radiation, 342

vacuum, 452, 454–456
validation, 318, 371–373, 376
variability reduction ratio, 533
vegetative state of bacteria, 312
velocity, 245, 247, 248, 253–256, 258, 260–263
viral challenge test, 361
virtual experimental design, 139, 143
viruses, 312
viscosity, 149, 154–156, 162, 247–250, 253, 258, 259, 261–265, 454, 455, 463–466
viscous dissipation, 157, 159–161
viscous Stokes number (Stv), 159

viscous stress tensor, 422
voidage, 149
volumetric, 187–189, 193, 195, 196

wall friction test, 67, 68, 71, 72, 74, 75
water intrusion test, 364
Wen and Yu model, 422, 426
wet granulation, 25, 29, 30, 32–34, 44, 147
 scale-up, 490, 491, 497, 507–512
wet massing, 167
wetting, 147, 148, 150, 154
 saturation, 150, 151
 zone, 175–180, 183
World Health Organization (WHO), 311, 312, 319, 322, 326, 354, 368, 370, 373
Wurster coating process, 419, 420, 424, 428, 439, 440, 447, 448
Wurster fluid bed coating, 413, 414

x-ray powder diffraction (XRPD), 275–277
x-ray tomography, 156

yield pressure, 204, 209, 210, 213
yield strength, 156, 158, 159
yield stress, 164
Young's modulus, 192, 193

zero-order controlled release, 383, 384, 386, 387, 397, 398, 400, 402, 403
Z-value, 328, 329