Modern Pharmaceutics

Fourth Edition, Revised and Expanded

edited by

Gilbert S. Banker

University of Iowa

Iowa City, Iowa

Christopher T. Rhodes

University of Rhode Island

Kingston, Rhode Island



Marcel Dekker, Inc. New York · Basel

Copyright © 2002 by Marcel Dekker, Inc. All Rights Reserved.

Copyright © 2002 Marcel Dekker, Inc.

ISBN: 0-8247-0674-9

This book is printed on acid-free paper.

Headquarters
Marcel Dekker, Inc.
270 Madison Avenue, New York, NY 10016
tel: 212-696-9000; fax: 212-685-4540

Eastern Hemisphere Distribution
Marcel Dekker AG
Hutgasse 4, Postfach 812, CH-4001 Basel, Switzerland
tel: 41-61-261-8482; fax: 41-61-261-8896

World Wide Web
http://www.dekker.com

The publisher offers discounts on this book when ordered in bulk quantities. For more information, write to Special Sales/ Professional Marketing at the headquarters address above.

Copyright © 2002 by Marcel Dekker, Inc. All Rights Reserved.
Neither this book nor any part may be reproduced or transmitted in any form or by any means, electronic or mechanical, including photocopying, microfilming, and recording, or by any information storage and retrieval system, without permission in writing from the publisher.

Current printing (last digit):
10 9 8 7 6 5 4 3 2 1

PRINTED IN THE UNITED STATES OF AMERICA

Copyright © 2002 Marcel Dekker, Inc.

DRUGS AND THE PHARMACEUTICAL SCIENCES

Executive Editor

James Swarbrick
PharmaceuTech, Inc.
Pinehurst, North Carolina

Advisory Board

Larry L. Augsburger
University of Maryland
Baltimore, Maryland

David E. Nichols
Purdue University
West Lafayette, Indiana

Douwe D. Breimer
Gorlaeus Laboratories
Leiden, The Netherlands

Stephen G. Schulman
University of Florida
Gainesville, Florida

Trevor M. Jones
The Association of the
British Pharmaceutical Industry
London, United Kingdom

Jerome P. Skelly
Alexandria, Virginia

Hans E. Junginger
Leiden/Amsterdam Center
for Drug Research
Leiden, The Netherlands

Felix Theeuwes
Alza Corporation
Palo Alto, California

Vincent H. L. Lee
University of Southern California
Los Angeles, California

Geoffrey T. Tucker
University of Sheffield
Royal Hallamshire Hospital
Sheffield, United Kingdom

Peter G. Welling
Institut de Recherche Jouveinal
Fresnes, France

Copyright © 2002 Marcel Dekker, Inc.

Preface

The first edition of *Modern Pharmaceutics* was published in 1980, the second in 1989, and the third in 1995. During the more than 20 years since we developed the concept of this text we have been privileged to work with many of the most notable pharmaceutical scientists as chapter authors. Some of our original first edition team are still with us. Others, for a variety of reasons—including change of career focus, retirement, and death—are no longer with us or able to cooperate in this endeavor. We place on record our gratitude to all our colleagues who have so unselfishly assisted us in any or all of our four editions.

We are also grateful to all those readers of our book who have provided comments on their perceptions of the value of *Modern Pharmaceutics* and suggested to us ways in which the book might be improved. We have most carefully considered all such ideas and some of the changes that we have implemented in this edition derive from advice given to us by industrial pharmaceutical scientists, university faculty, and students.

The fourth edition of *Modern Pharmaceutics* follows the same format and has the same goals as previous editions. Chapter 1 sets the stage by reviewing the role of drugs and drug products in treating and preventing disease, while also summarizing their primary quality features. Chapters 2 through 6 provide background that is fundamental to an understanding of drug action and the design of drug products. Chapter 7, on preformulation, is another fundamentals chapter, describing the manner in which drugs are characterized for their physical, chemical, and pharmacokinetic properties to provide a rational and scientific basis for drug product design.

Chapters 8 through 16 describe drug products and dosage forms, together with the routes of administration by which they are given. Chapters 17 through 28 also treat topics critical to drug product quality such as packaging, optimization, and food and drug laws, in addition to examining more specialized product classes and introducing several new chapters. As in previous editions, this book once again ends with a view to the future.

In the planning of this edition, we identified certain areas of recent growth that seemed to merit increased attention in this text, while attempting to further recognize pharmacy's international character. Also, we gave attention to topics that are of relatively more importance than previously. Overall the book has grown in size because there are a number of critical areas that justify significant additional coverage. We have been fortunate in obtaining the services of a number of distinguished individuals to cover topics not allocated whole chapters in previous editions.

The growing importance of botanicals and other natural products has been recognized by devoting one of our new chapters to this topic. Also, we have a new chapter on managed care since this is an area of great importance and impacts all facets of pharmacy. Further, we expect this topic to gain additional recognition in parts of the world where at present the implications of this discipline are not fully recognized. Also, appreciating the growing importance of generic products, especially in North America and the European Union, we have included in this edition

Copyright © 2002 Marcel Dekker, Inc.

a chapter focused on bioequivalence. Finally, recognizing the impact of the information revolution, catalyzed by computers and the Internet, we have included a new chapter on drug information.

With one exception, all the chapters from the third edition of *Modern Pharmaceutics* that appear in the fourth edition have been revised and updated. Many chapters were extensively updated, and some, such as the first and last chapters, were extensively rewritten. Due to the illness of Dr. Robinson, the chapter on sustained and controlled release drug delivery systems was updated with the assistance of Gil Banker, with Dr. Robinson's approval.

Although *Modern Pharmaceutics* continues to evolve, our basic goals remain those which we developed in the 1970s when we first delineated the concept of this comprehensive and integrated treatment of pharmaceutics, with a focus on drug product quality and performance. We are committed to producing an up-to-date, authoritative, multiauthored treatise on pharmaceutics, which can be used by both students and practitioners.

Gilbert S. Banker
Christopher T. Rhodes

Copyright © 2002 Marcel Dekker, Inc.

Contents

Preface
Contributors

1. Drug Products: Their Role in the Treatment of Disease, Their Quality, and Their Status and Future as Drug-Delivery Systems
   Gilbert S. Banker

2. Principles of Drug Absorption
   Michael Mayersohn

3. Pharmacokinetics
   David W. A. Bourne

4. Factors Influencing Drug Absorption and Drug Availability
   Betty-ann Hoener and Leslie Z. Benet

5. The Effect of Route of Administration and Distribution on Drug Action
   Svein Øie and Leslie Z. Benet

6. Chemical Kinetics and Drug Stability
   J. Keith Guillory and Rolland I. Poust

7. Preformulation
   Jens T. Carstensen

8. Cutaneous and Transdermal Delivery—Processes and Systems of Delivery
   Gordon L. Flynn

9. Disperse Systems
   Wandee Im-Emsap, Ornlaksana Paeratakul, and Juergen Siepmann

Copyright © 2002 Marcel Dekker, Inc.

10. Tablet Dosage Form
Mary J. Kottke and Edward M. Rudnic

11. Hard and Soft Shell Capsules
Larry L. Augsburger

12. Parenteral Products
James C. Boylan and Steven L. Nail

13. Design and Evaluation of Ophthalmic Pharmaceutical Products
John C. Lang, Robert E. Roehrs, Denise P. Rodeheaver, Paul J. Missel, Rajni Jani, and Masood A. Chowhan

14. Delivery of Drugs by the Pulmonary Route
Anthony J. Hickey

15. Sustained- and Controlled-Release Drug Delivery Systems
Gwen M. Jantzen and Joseph R. Robinson

16. Target-Oriented Drug-Delivery Systems
Vijay Kumar and Gilbert S. Banker

17. Packaging of Pharmaceutical Dosage Forms
Thomas J. Ambrosio

18. Optimization Techniques in Pharmaceutical Formulation and Processing
Joseph B. Schwartz, Robert E. O'Connor, and Roger L. Schnaare

19. Food and Drug Laws that Affect Drug Product Design, Manufacture, and Distribution
Garnet E. Peck and Roland Poust

20. European Aspects of the Regulation of Drug Products with Particular Reference to Development Pharmaceutics
Brian R. Matthews

21. Pediatric and Geriatric Aspects of Pharmaceutics
Michelle Danish and Mary Kathryn Kottke

22. Biotechnology-Based Pharmaceuticals
Paul R. Dal Monte, S. Kathy Edmond Rouan, and Narendra B. Bam

23. The Pharmacist and Veterinary Pharmaceutical Dosage Forms
J. Patrick McDonnell and Lisa Blair Banker

24. Dietary Supplements
Teresa Bailey Klepser

25. Bioequivalency
Christopher T. Rhodes

Copyright © 2002 Marcel Dekker, Inc.

26. Drug Information
Hazel H. Seaba

27. Managed Care and Pharmacotherapy Management
Julie M. Ganther and William R. Doucette

28. A View to the Future
Gilbert S. Banker and Christopher T. Rhodes

Copyright © 2002 Marcel Dekker, Inc.

Contributors

Thomas J. Ambrosio, Ph.D. Development Fellow, Package Development, Schering-Plough Research Institute, Kenilworth, New Jersey

Larry L. Augsburger, Ph.D. Shangraw Professor of Industrial Pharmacy and Pharmaceutics, Department of Pharmaceutical Sciences, School of Pharmacy, University of Maryland, Baltimore, Maryland

Narendra B. Barn, Ph.D. Project Management and R&D Strategy, GlaxoSmithKline, Collegeville, Pennsylvania

Gilbert S. Banker, Ph.D. Dean Emeritus and John L. Lach Distinguished Professor of Drug Delivery Emeritus, College of Pharmacy, University of Iowa, Iowa City, Iowa

Lisa Blair Banker, D.V.M. Blair Animal Clinic, West Lafayette, Indiana

Leslie Z. Benet, Ph.D. Professor of Biopharmaceutical Sciences, School of Pharmacy, University of California, San Francisco, California

David W. A. Bourne, Ph.D. Professor of Pharmacy, College of Pharmacy, University of Oklahoma, Oklahoma City, Oklahoma

James C. Boylan, Ph.D. Pharmaceutical Consultant, Gurnee, Illinois

Jens T. Carstensen, Ph.D. Professor Emeritus, School of Pharmacy, University of Wisconsin, Madison, Wisconsin

Masood A. Chowhan, Ph.D. Director of Consumer Products, Alcon Research Ltd., Fort Worth, Texas

Paul R. Dal Monte, Ph.D. Project Management and R&D Strategy, GlaxoSmithKline, Collegeville, Pennsylvania

Michele Danish, Pharm.D. Clinical Manager, Department of Pharmacy, St. Joseph Health Services, Providence, Rhode Island

Copyright © 2002 Marcel Dekker, Inc.

William R. Doucette, Ph.D. Associate Professor, Division of Clinical and Administrative Pharmacy, University of Iowa, Iowa City, Iowa

Gordon L. Flynn, Ph.D. Professor of Pharmaceutical Science, Department of Pharmaceutical Sciences, College of Pharmacy, The University of Michigan, Ann Arbor, Michigan

Julie M. Ganther, Ph.D. Assistant Professor, Department of Clinical and Administrative Pharmacy, University of Iowa, Iowa City, Iowa

J. Keith Guillory. Ph.D. Professor Emeritus, Division of Pharmaceutics, College of Pharmacy, University of Iowa, Iowa City, Iowa

Anthony J. Hickey, Ph.D. Professor, School of Pharmacy, University of North Carolina, Chapel Hill, North Carolina

Betty-ann Hoener, Ph.D. Professor, Department of Biopharmaceutical Sciences, School of Pharmacy, University of California, San Francisco, California

Wandee Im-Emsap College of Pharmacy, Freie Universität Berlin, Berlin, Germany

Rajni Jani, Ph.D. Senior Director, Department of Pharmaceutics, Alcon Research Ltd., Fort Worth, Texas

Gwen M. Jantzen, Ph.D. School of Pharmacy, University of Wisconsin, Madison, Wisconsin

Teresa Bailey Klepser, Pharm.D. Associate Professor, College of Pharmacy, Ferris State University, Big Rapids, Michigan

Mary Kathryn Kottke, Ph.D. Department of Regulatory Affairs, Cubist Pharmaceuticals, Inc., Lexington, Massachusetts

Vijay Kumar, Ph.D. Assistant Professor, Division of Pharmaceutics, College of Pharmacy, University of Iowa, Iowa City, Iowa

John C. Lang, Ph.D. Director of Emerging Technologies, Consumer Products Research and Development, Alcon Research Ltd., Fort Worth, Texas

Brian R. Matthews, Ph.D. Senior Director of EC Registration, Alcon Laboratories (UK) Ltd., Hemel Hempstead, United Kingdom

Michael Mayersohn, Ph.D. Professor of Pharmaceutical Sciences, College of Pharmacy, University of Arizona, Tucson, Arizona

J. Patrick McDonnell, B.S.Ph. Senior Compliance Auditor, Department of Biologics Quality Assurance, Fort Dodge Animal Health, Charles City, Iowa

Paul J. Missel, Ph.D. Principal Scientist, Department of Drug Delivery, Alcon Research Ltd., Fort Worth, Texas

Steven L. Nail, Ph.D. Associate Professor, Department of Industrial and Physical Pharmacy, Purdue University, West Lafayette, Indiana

Robert E. O'Connor, Ph.D. Pharmaceutical Sourcing Group Americas, a division of Ortho-McNeil Pharmaceutical, Bridgewater, New Jersey

Copyright © 2002 Marcel Dekker, Inc.

Steven Øie, Ph.D. Professor, Department of Biopharmaceutical Sciences, School of Pharmacy, University of California, San Francisco, California

Ornlaksana Paeratakul, Ph.D. Assistant Professor, Pharmaceutical Technology, Faculty of Pharmacy, Srinakharinwirot University, Nakhonnayok, Thailand

Garnet E. Peck, Ph.D. Professor and Director of the Industrial Pharmacy Laboratory, Department of Industrial and Physical Pharmacy, School of Pharmacy, Purdue University, West Lafayette, Indiana

Rolland Poust, Ph.D. Professor, Pharmaceutical Services Division, College of Pharmacy, University of Iowa, Iowa City, Iowa

Christopher T. Rhodes, Ph.D. Professor, Department of Applied Pharmaceutical Sciences, University of Rhode Island, Kingston, Rhode Island

Joseph R. Robinson, Ph.D. Professor of Pharmacy, School of Pharmacy, University of Wisconsin, Madison, Wisconsin

Denise P. Rodeheaver, Ph.D., D.A.B.T. Assistant Director, Department of Toxicology, Alcon Research Ltd., Fort Worth, Texas

Robert E. Roehrs, Ph.D.* Vice President, Department of Drug Regulatory Affairs, Alcon Research Ltd., Fort Worth, Texas

S. Kathy Edmond Rouan, Ph.D. Vice President, Cardiovascular and Urology Project Team Leadership and Management, Project Management and R&D Strategy, GlaxoSmithKline, Collegeville, Pennsylvania

Edward M. Rudnic, Ph.D. Advancis Pharmaceutical Corp., Gaithersburg, Maryland

Roger L. Schnaare, Ph.D. Philadelphia College of Pharmacy, Philadelphia, Pennsylvania

Joseph B. Schwartz, Ph.D. Philadelphia College of Pharmacy, Philadelphia, Pennsylvania

Hazel H. Seaba, M.S. Professor (Clinical) and Director, Division of Drug Information Service, College of Pharmacy, University of Iowa, Iowa City, Iowa

Juergen Siepmann, Ph.D. Assistant Professor, College of Pharmacy, Freie Universität Berlin, Berlin, Germany

*Retired.

Copyright © 2002 Marcel Dekker, Inc.

Chapter 1

Drug Products: Their Role in the Treatment of Disease, Their Quality, and Their Status and Future as Drug-Delivery Systems

Gilbert S. Banker

University of Iowa, Iowa City, Iowa

I. ROLE OF DRUGS AND DRUG PRODUCTS IN THE TREATMENT AND PREVENTION OF DISEASE

The methods of treating illness and disease as we enter the twenty-first century include the use of the following forms of therapy: (a) surgery, including organ transplantation; (b) psychotherapy; (c) physical therapy; (d) radiation, and (e) chemo or pharmacotherapy. Of these various methods, pharmacotherapy (treatment with drugs) is the most frequently used technique for treating disease, has the broadest range of application over the greatest variety of disease states, and is usually the most cost-effective and preferred treatment method. Although surgery is the preferred method of treating some ailments or disease states, when alternative methods are available, these methods (usually pharmacotherapy) will be employed first if feasible, in the initial attempt to secure satisfactory relief or control of the condition or a complete cure. As pharmacotherapy continues to improve, it is replacing other forms of treatment as the preferred method of therapy. Pharmacotherapy is, for example, increasingly becoming the treatment of choice in treating various forms of cancer, including breast cancer, replacing the use of radical surgery. Pharmacotherapy is now an effective option to surgery in the treatment of some forms of prostate disease. When cure rates or reliability of disease control by pharmacotherapy can match surgical treatment (e.g., prostate surgery or radical mastectomy), most patients will strongly prefer the chemotherapeutic approach or the use of chemotherapy combined with less radical surgical approaches.

In some surgical procedures, such as organ transplantation, the success of that procedure will be only as great as the course of pharmacotherapy that follows. Organ transplant recipients are required to continue drug therapy for the balance of their lives for control of their immune systems and to prevent organ rejection.

Pharmacotherapy is also very important in the prevention of disease, since vaccines and other immunizing agents are drug products. The impact of vaccines in eliminating or reducing the incidence of six diseases is shown in Fig. 1. Some diseases that previously killed or crippled tens of millions of people worldwide, often reaching epidemic proportions, are now virtually unknown in most of the world. Table 1 shows the average number of deaths in the United States per million people as a result of various diseases in a nonepidemic situation over the last century. The table lists diseases that have been obliterated or nearly obliterated in the United States through drug immunology as shown in Fig. 1, together with diseases that have now been brought largely under control or greatly reduced by the discovery and effective use of anti-infective drugs than can combat bacterial infections. Examples of diseases in this latter category are

Copyright © 2002 Marcel Dekker, Inc.

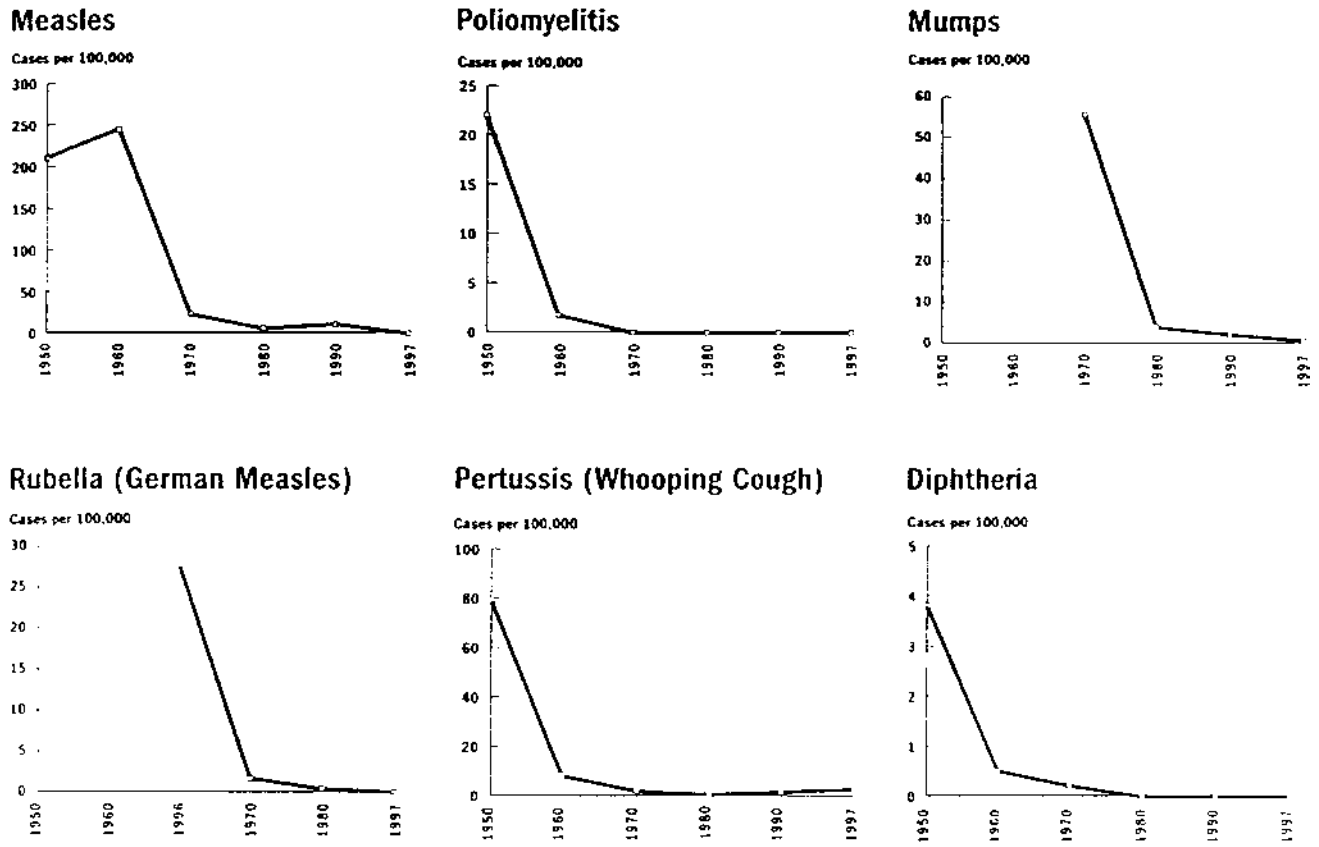


Fig. 1 Impact of vaccines in reducing incidence of diseases.

Table 1 Causes of Death in the United States per Million Population in Nonepidemic Years from Infectious and Other Identifiable Diseases

	Year							
Disease	1900	1920	1940	1960	1975	1985	1990	1999
Influenza and pneumonia	2030	2080	700	310	370	250	3313	311
Tuberculosis (all forms)	2020	1150	460	220	50	8	7	5
Diarrheas and intestinal	1330	540	100	50	40	1	0	0
Kidney diseases	890	890	820	210	110	85	83	95
Bronchitis	460	130	30	20	30	16	1	1
Diphtheria	430	260	10	3	0	0	0	0
Typhoid and paratyphoid	360	80	10	0	0	0	0	0
Syphilis	20	160	140	50	20	0	0	0
Measles	120	90	5	3	0	0	0	0
Whooping cough	120	120	20	7	0	0	0	0
Appendicitis	100	130	100	20	0	2	1	0
Scarlet fever	100	50	5	0	0	0	0	0
Malaria	80	40	10	0	0	0	0	0
Smallpox	20	6	0	0	0	0	0	0

Source: Official Statistics, U.S. Census Bureau.

Copyright © 2002 Marcel Dekker, Inc.

pneumonia, tuberculosis, certain diarrheal and intestinal disorders, and bronchitis.

Yet other diseases have been largely controlled by improved sanitation and public health procedures, alone or in combination with pharmacotherapy. Bubonic plague (the "Black Death"), malaria, and typhus are in this category. Largely through pharmacotherapy, including the use of immunological agents, epidemics of life-threatening diseases have been greatly reduced and, until recently, were thought to have been eliminated in all but the least developed regions of the world. We now know that this is not the case.

In past centuries epidemics swept entire countries and continents. In one 8-year period in the fourteenth century when bubonic plague was epidemic throughout Europe, two thirds of the population were infected; half of these died, totaling 25 million deaths. In 1918–1919 an influenza epidemic swept most of the world, causing 20 million deaths, with more than a half-million deaths occurring in the United States. Twenty or 30 years ago we thought such pandemics could be prevented. A pandemic occurs when a disease occurs over a wide geographic area and affects an exceptionally high proportion of some populations. As we enter the third millennium we are in the midst of a pandemic that is worse than anything the world has known before, based on numbers of human casualties.

Acquired immunodeficiency syndrome (AIDS) is a virally transmitted disease caused by HIV, a human immunodeficiency retrovirus, whose genetic material is RNA. AIDS is an especially nasty, progressive, and costly disease for which no cure currently exists. When a cure is found it will be a drug cure. The virus uses the enzyme reverse transcriptase to incorporate its genetic material into the infected host cell's genome. The primary host target cells are T_4 (T-helper) cells, which, after a significant level of destruction, compromises the body's immune system. A whole range of opportunistic infections then attack, often in concert, until death inevitably ensues.

Table 2 lists the distribution of HIV/AIDS in various regions of the world. The heavy infection level in sub-Saharan Africa is the result of several factors. First, the disease probably originated there. The oldest HIV sample came from the Republic of the Congo in 1959, but most researchers believe AIDS had been killing people many decades before then. Second, most of the countries in sub-Saharan Africa are underdeveloped and have lacked the resources to fight the disease or implement effective prevention plans. This region of the world now has 70% of the global total of HIV-infected people but only 10% of the world's population. The epidemic in Africa has exploded in the last 20 years. In 1982 only one African country had an HIV prevalence rate in adults above 2%. By 1998 it was estimated that more than 7% of the adults living in 21 African countries were HIV-positive or had AIDS. In two populations 25% of the adults were infected. Life expectancy has fallen to 30 years or less in some of these countries, making HIV an unprecedented catastrophe in the world's history [1]. In addition to the nearly 35 million people estimated to have HIV/AIDS going into the new millennium, to date AIDS has killed 20 million people [2]. It is clearly the most lethal pandemic the world has ever known and represents one of the biggest challenges to public health and pharmacotherapy today.

Table 2 Distribution of HIV/AIDS in Various Regions of the World

Region	Number infected[a]
Sub-Saharan Africa	22,500,000
Southeast Asia	6,700,000
Latin America	1,400,000
North America	890,000
Eastern Europe and Central Asia	560,000
Western Europe	500,000
Caribbean	330,000
Northern Africa and Middle East	210,000
Australia and New Zealand	12,000
Total	34,400,000

[a]As of the end of 1998.
Source: Ref. 1.

Another challenge facing pharmacotherapy as we enter the twenty-first century are diseases known as zoonoses. Zoonoses are animal-borne diseases, which are transmitted to humans. They often develop in a relatively small region of the world initially and then spread. They are often viral in origin and thus produce a far greater challenge to humankind than did many of the bacterial diseases of the past. Although controversy exists, there is compelling evidence that HIV is a variant of viruses that infect nonhuman primates in central Africa and is thus a zoonose disease.

Another recent example of a zoonose infection is the Nipah virus, named after the town in Malaysia where its first known victim lived. The animal vector of this disease has been identified as several species of bats. The Nipah virus has destroyed Malaysia's pig industry and it killed 105 people in 1999. The virus produces a severe form of encephalitis, and about 40% of infected individuals die.

Copyright © 2002 Marcel Dekker, Inc.

In the United States in recent years several diseases with animal hosts are now infecting humans. The Arena virus has as its host certain desert rodents. Rodent droppings when dry and airborne as dust particles infect humans, often with lethal consequences. The West Nile virus appeared in the United States in 1999. Birds carry this virus, and mosquitoes are the vector to humans. In 1999 the virus was localized in the New York City area. By 2000 it has spread across New York State and into New Jersey and Massachusetts. The West Nile virus is especially lethal when it infects children, the elderly, or those with a compromised immune system.

The subject of zoonoses was extensively discussed at the International Conference on Emerging Infectious Diseases, July 16–19, 2000 in Atlanta, Georgia [3]. A group of researchers from the University of Edinburgh reported that humanity is currently challenged by 1709 known pathogens, including viruses, bacteria, fungi, protozoa, and worms. Of these, 832, or 49%, are zoonotic. Of even greater concern is the fact that of the 156 disease that are considered to be new or "emerging," 114 are zoonoses, a staggering 73%. The diseases produced by these zoonoses often must primarily be challenged by pharmacotherapy. It is clear that much will remain to be done by pharmaceutical scientists in the years ahead, in addition to attacking the leading causes of death and disability.

II. ROLE OF PHARMACOTHERAY IN MORTALITY RATES, LIFE EXPECTANCY, AND QUALITY OF LIFE

Another impact of pharmacotherapy that is not generally recognized has been its effect on life expectancy and on the health of the newborn, infants, and children. Some of the earliest valid historical statistics on the death rates of children are found in parish records of London, England. One parish, St. Botolph's, "which was bordered by the city wall and eastern gate, the Tower of London and the Thames," has detailed records from 1558 to 1626, which have survived the years intact [4]. The population of this region probably enjoyed better medical care than did their rural neighbors. The stillborn death rate in the parish fluctuated between 40.8 and 133.4, averaging 71.6/1000 births. The current overall average stillborn death rate in the United States is about 20/1000 births. Infants dying in the first month in Shakespeare's day in London were known as chrisoms. The average chrisom death rate from 1584 to 1598 was more than 162/1000 (two to four times the average stillborn rate). Thus, about one child out of every three to four was stillborn or failed to survive the first month. By contrast, in the United States today fewer than 10 deaths per 1000 live births occur during the first year of life. Survival rates of infants and older children were equally grim in Shakespeare's England. Of every 100 children born in the late sixteenth century, only about 70 survived to the first birthday, about 48 to their fifth, and only 27–30 survived to their fifteenth birthday.

The death rate statistics of newborns, infants, and older children have greatly improved from the sixteenth, or even the twentieth century, to the opening of the twenty-first century. An interesting exercise when next you visit an old cemetery would be to read tombstones that predate 1900, or even 1940. You will find that one grave marker out of every two or three is for a newborn, infant, or a young child. This was largely related to the inability of the medical and pharmaceutical professions of that day to effectively combat infectious and children's diseases. Many of these diseases affected children almost exclusively among their fatality victims (see Table 1), including measles, scarlet fever, polio, and whooping cough. Added to the infectious "children's diseases" problem was the fact that these diseases often left their young survivors with permanent physiological damage, such as scarred heart valves, brain damage, poorly developed limbs or paralysis, and other defects that remained for the balance of life. The dramatic improvement in the infant mortality rate in the United States over the past 60 years is shown in Fig. 2 [5]. Although great advances have been made, a number of countries in Western Europe have rates that are substantially better than those of the United States.

The increasing life expectancy and the growing number of the elderly in our citizenry over the last century, and in recent decades, is well known (Figs. 3 and 4) [5]. A person born in 1920 could expect to live only 54.1 years. Today we can expect to live over 76 years (about 73 years for men and nearly 80 years for women). The life expectancy for women helps explain the fact that octogenarians (persons over 80) are the fastest-growing segment of any age group in our population. Another interesting way to look at life expectancy is shown in Fig. 5. Here we see the average additional life expectancy a person may have based on their current age. These figures, at all age levels, also show a steady upward progression over the last 27 years.

Figure 3 also shows what many experts predict. Life expectancy is not expected to plateau, but by 2030–2040 the life expectancy for women will be at least 90, and for men will be approaching 85. However, these

Copyright © 2002 Marcel Dekker, Inc.

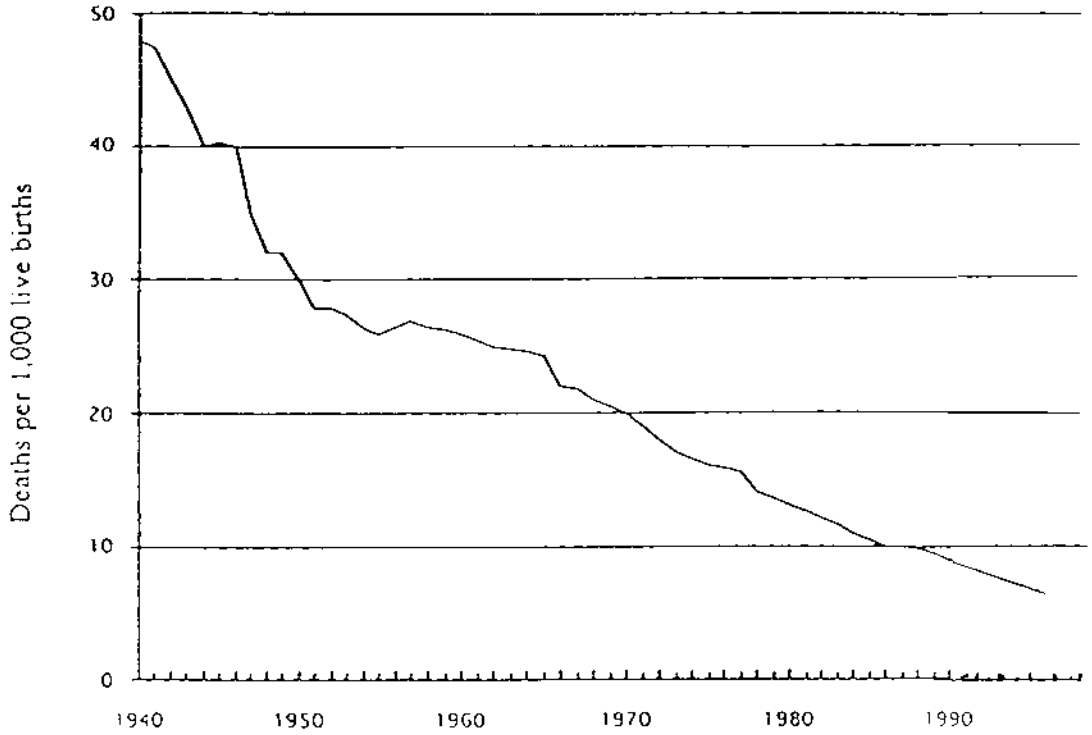


Fig. 2 Infant (under 1 year) mortality rates in the United States, 1940–2000.

numbers could be negatively impacted in the United States by the current AIDS epidemic. Figure 4 shows another result of our increasing life expectancy—the very rapid growth in the over-65 population as a percentage of the U.S. population from 1960 to the year 2000. In a report of the National Institute on Aging [6] it was noted that there are 35 million people in the United States age 65 or older. This number will double by 2030, continuing the dramatic shift to a much larger population of older Americans. Women currently account for about 60% of those over 65, and nearly half of these older women live alone. The percentage of older Americans living in poverty has dropped dramatically in the last 40 years, from 35% to 11%, but there is a disparity between races (8.2% for whites vs. 26.4% for blacks). This dramatic shift in our elderly population will have a large impact on social programs such as Social Security, Medicare, and Medicaid, as well as on the rapidly expanding growth and need for pharmaceutical services and pharmacy care in the years ahead.

The exact contribution of modern pharmaceuticals to our increased longevity can be only estimated and weighed in comparison with improved diet and lifestyles, sanitation, housing, and generally improved public health. However, advances in chemotherapy have certainly been the major factor in extending our life expectancy. Similarly, the contribution of pharmacotherapy to an improved quality of life in recent decades can be only estimated—but it has been a very major factor. It is clearly a leading factor in the well-being of the elderly, allowing them to remain active and essentially healthy through more years and over a greater fraction of their total life span. The role of pharmacotherapy in improving the quality of life of the mentally ill is also clearly evident. Hundreds of thousands of mentally ill patients in the United States alone who are currently being treated as outpatients can live essentially normal lives and can remain with their families through the use of drugs to control their illness. Without the availability of effective psychotherapeutic drug agents, many of these patients would require institutionalization or at least short- to midterm hospitalization.

Other diseases that formerly required long-term hospitalization or complete isolation include tuberculosis and the dreaded leprosy. Only a generation or two ago, for patients to be told that they had such diseases

Copyright © 2002 Marcel Dekker, Inc.

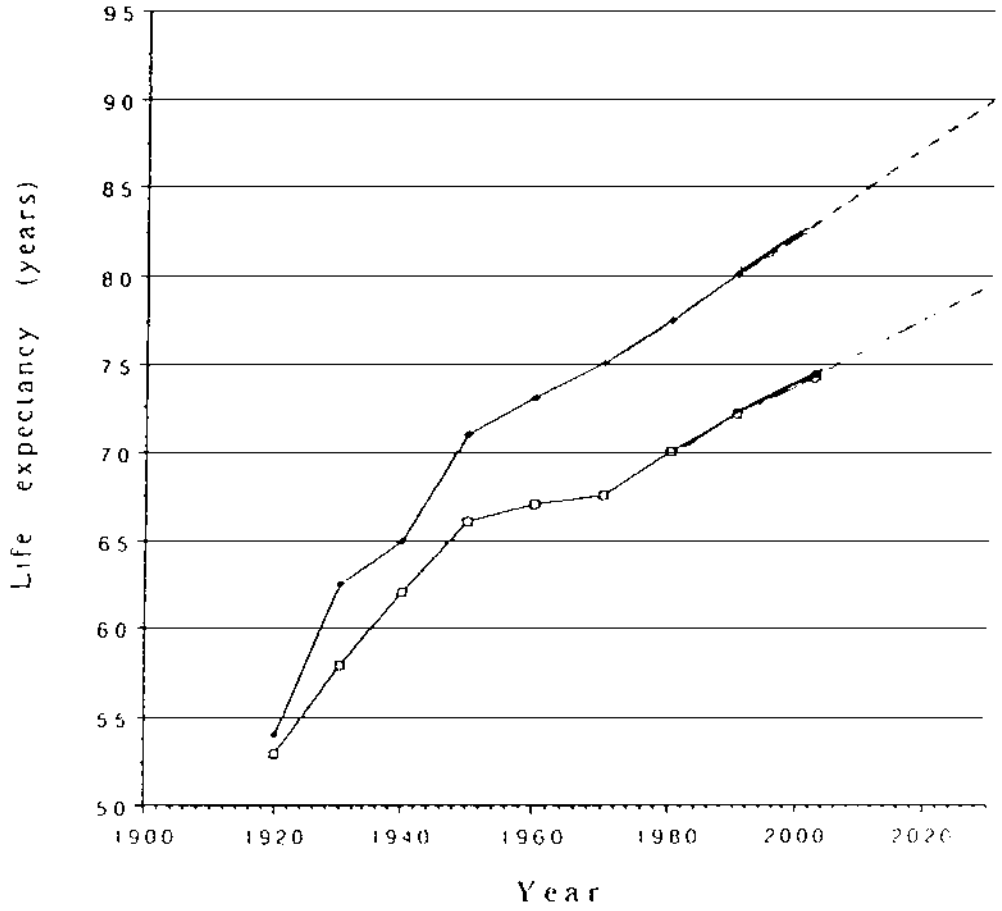


Fig. 3 Expectation of life in years at birth in the United States.

was equivalent to receiving a death sentence, or worse. These diseases are totally curable today by means of chemotherapy, and the patient no longer needs to be isolated in a sanitarium. Other diseases, such as rheumatoid arthritis, frequently drove patients to suicide. Today, even though we still lack cures for some of these diseases, we can contain and control them, permitting patients to lead nearly normal lives.

Great strides have been made in chemotherapy since World War II (1939–1945) and in the decades following the war. Antibiotics and other anti-infective drugs, steroids, psychotherapeutic agents, many new immunizing agents, important cardiovascular agents, antineoplastic agents, and numerous other drug classes and agents have appeared in the last four to five decades. Given the rapid advances in biotechnology, new drug innovation is entering another period of revolutionary growth. Nevertheless pharmaceutical scientists have no cause for complacency. We cannot yet cure the debilitating diseases of cystic fibrosis or muscular dystrophy. Many forms of cancer are treatable with only low to moderate success if detected early; that battle is far from won. Over one-half million Americans are dying each year from cancer, the number 2 cause of death in this country. Although death rates are consistently continuing to decline (Fig. 6), we must continue the battle to effectively combat many degenerative diseases affecting our growing elderly population, notably heart disease, the number 1 killer, which kills three-quarters of a million Americans a year. The growing challenge posed by cardiovascular disease, as Americans continue to age, is shown in Fig. 7. The current third leading cause of death also heavily affecting the elderly, is cerebrovascular disease, which kills about 150,000 Americans a year. This prevalence and estimated annual economic cost of some common diseases as shown in Table 3, further documents some of our remaining challenges in pharmacotherapy and

Copyright © 2002 Marcel Dekker, Inc.

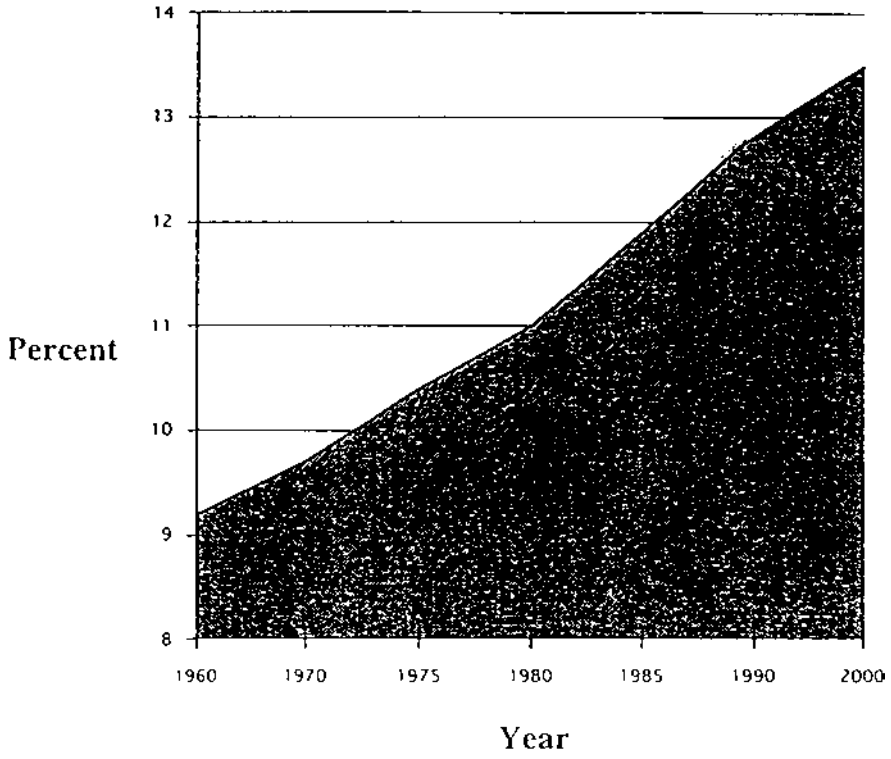


Fig. 4 Percentage of the total U.S. population older than 65 years, 1960–2000. (From Ref. 5.)

health care. The challenge of the AIDS pandemic and other emerging diseases was discussed in the previous section. Accidents are the fourth leading cause of death at 90,000–95,000/year. Pharmacotherapy is expected to provide major answers to most, if not all of these and many other disease challenges in the years to come.

III. DRUG AND DRUG PRODUCT QUALITY AND ITS EVALUATION

A. Reasons for the Drug Product Quality Question

The quality of drugs, which used to be an important discussion topic only for pharmaceutical manufacturers and experts in education, compendial standards, and regulatory enforcement, has now been placed in the spotlight of public attention. The reasons for the broadbased interest in drug product quality are based on the following factors, at least in part: (a) a clear realization that drug products are different from other consumer products; (b) rapidly increasing health care and drug product costs over the last several decades; (c) increasing public advertising of prescription drug products, with identification of price differentials among chemically equivalent generic products; (d) increasing payment of health care costs by third parties; (e) promulgation of the federal "maximum allowable cost" (MAC) regulations; and (f) effects of health care reform and increasing pressure to provide a prescription drug benefit under Medicare.

The greatest difference between drug products and other consumer products is that the principle of *caveat emptor* ("let the buyer beware") cannot operate in the usual way when the layperson acquires prescription drugs. In addition, laypeople lack the skills and sophistication to evaluate the quality and appropriateness of their prescription drug products, whereas they do have some ability to evaluate most other consumer products. Furthermore, consumers do not select prescription products as they do nearly every other product they purchase or use. A consumer is often compelled to buy a specific drug product, whereas he or she has more freedom of choice about when or if they buy other products. In the process, the layperson must usually trust the physician who prescribes the medication and the pharmacist who selects it (if a generic drug) and dispenses it. Other differences are that drug products are

Copyright © 2002 Marcel Dekker, Inc.

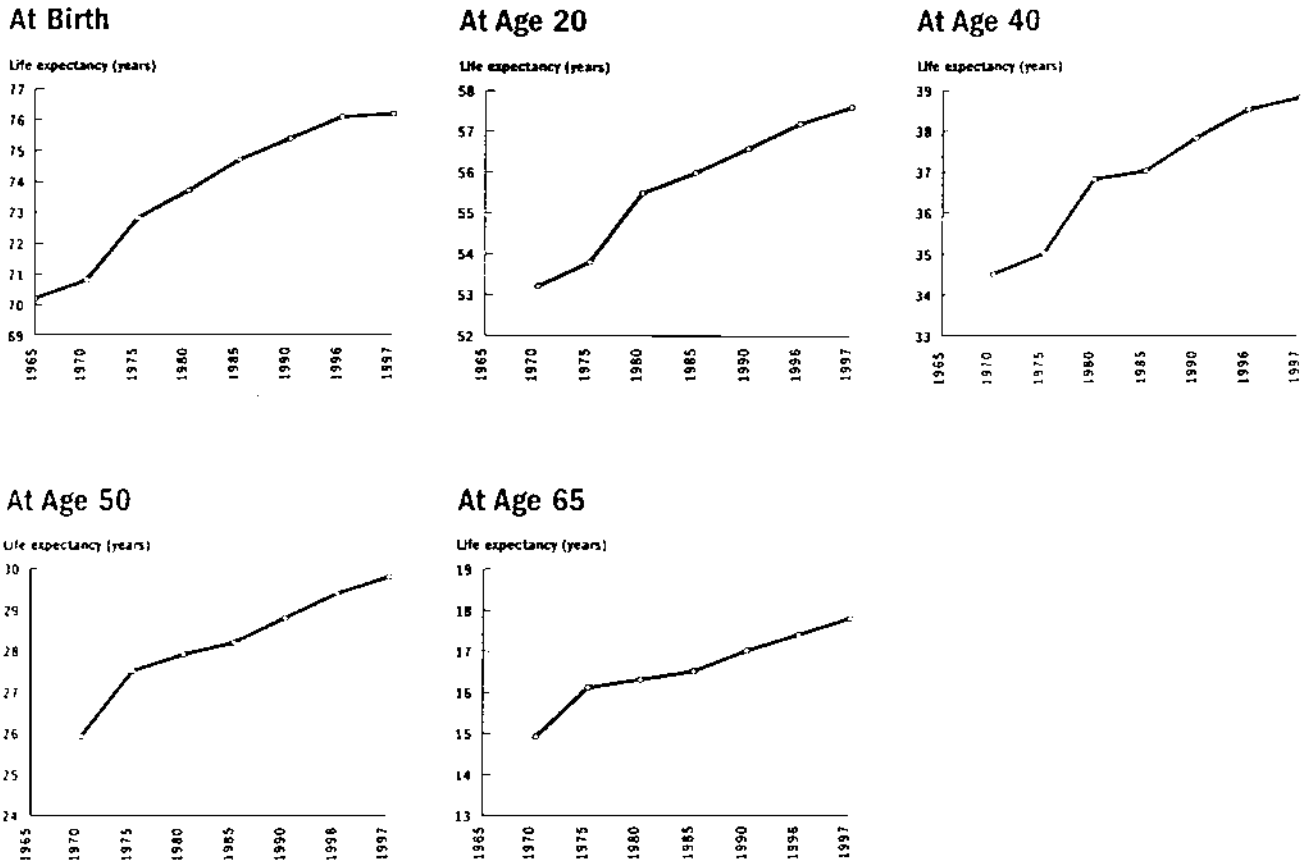


Fig. 5 Additional years of life expectancy at various ages, 1965–1970–1997.

typically more critical to the consumer's well-being than are other products. A poor quality drug product can have more serious consequences than a poor quality consumer product of nearly any other category. Drug products are also more complex than nearly any other class of consumer product. These last two factors are why drug products are subjected to many more tests and controls than are other types of products.

In the last four to five decades national health care expenditures for all types of health-related transactions, including dental, medical, hospital, prescription, and over-the-counter (OTC) drugs, has grown from about 5% of the country's gross domestic product (GDP) (national expenditures on all goods and services) to nearly 14% of the GDP today. Total health expenditures in the United States are now well over $1 trillion annually.

There are many reasons for the cost increases in health care, in addition to inflation, including the following: (a) there are more and more older people, requiring more care; (b) new and additional types of care and treatment are available and they tend to be more expensive, and (c) the quality of treatment has been improving. Aggregate expenditures in the health sector, have been rising at 10% or more a year over the last 15–20 years. While per capita health care expenditures for Americans are the highest in the world, a recent report from the World Health Organization (WHO) ranks the United States as only 37th out of 191 nations in terms of the health services provided to our citizens [7]. In this first ever ranking of United Nations member countries' health systems, some of the factors considered were life expectancy, health quality based on infant/child survival rates, system responsiveness, system cost per population served, fairness of financial contributions, and overall health system performance. The top ranked countries were France, Italy, San Marino, Andorra, Malta, Singapore, Spain, Oman, Austria, and Japan. The bottom ranked countries were all from sub-Saharan Africa plus Burma.

Copyright © 2002 Marcel Dekker, Inc.

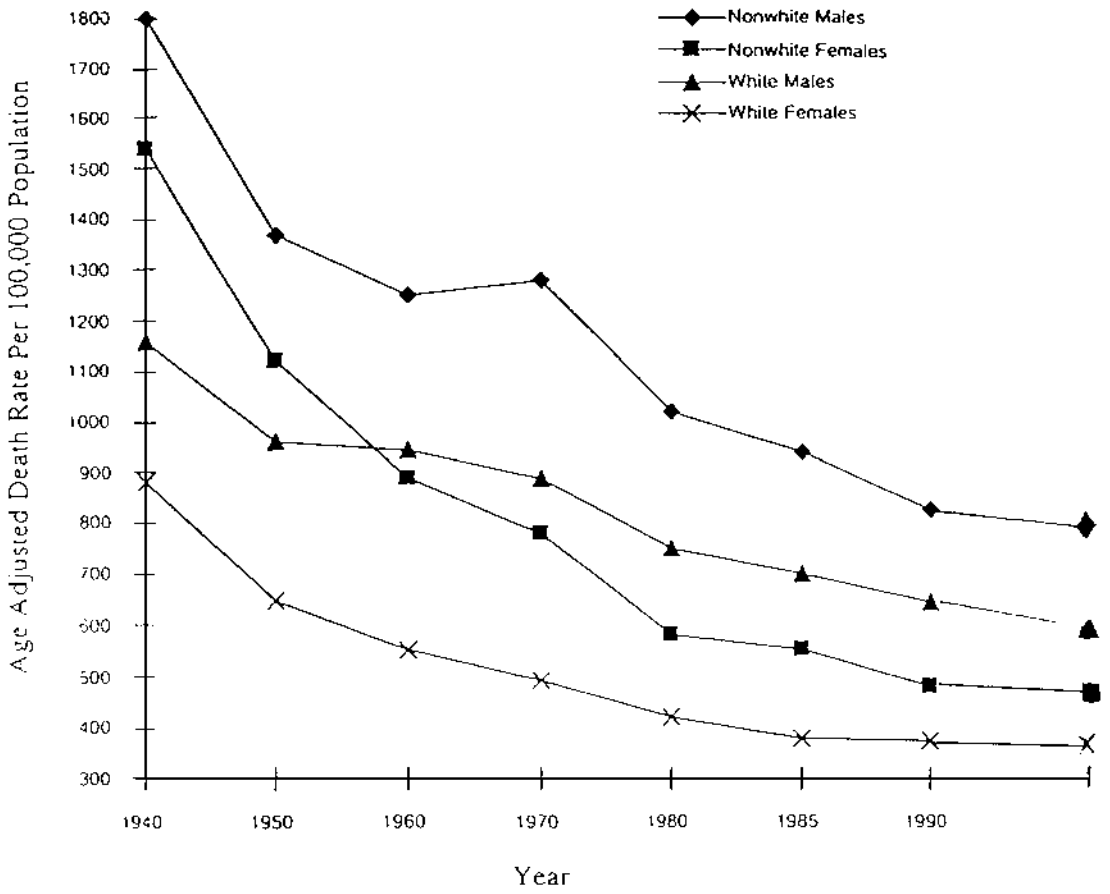


Fig. 6 Age-adjusted death rates in the United States from all causes by sex and by race (white and all others), 1940–1996. (Courtesy of National Center for Health Statistics.)

The WHO report noted that many countries are falling far short of their potential and that expenditures per se do not necessarily produce the best system. This is exemplified by the fact that the United States, which is ranked 37th, spends 13.7% of its gross domestic product on health care, while the United Kingdom, which is ranked 18th, spends only 5.8%.

While rankings are only as meaningful as the criteria on which they are based and the accuracy of their assessments, there is undoubtedly some validity to the WHO study and its rankings. One analysis of the relatively low U.S. ranking noted the large number of Americans who have no health insurance or other coverage and the lower level of healthy life expectancy in the United States compared to other industrialized nations. The WHO analysis also commented on the high rates of heart disease and tobacco-related cancers in the United States, together with the "extremely poor health" of minority groups, notably Native Americans and rural African Americans, as contributing to the relatively low ranking of U.S. health care.

The great majority of health care expenditures are for services—medical services, nursing services, or hospital services. While pharmacy (prescription products and services) constitutes a relatively small percentage of total health care expenditures (9–10% of the total), the pharmacy component has received a great deal of attention in recent years. There are several reasons for this. A product is involved in the pharmacy component, and it is much easier to analyze and attempt to minimize product costs compared to medical or hospital services. Second, a well-defined private industry is involved with the drug component—the pharmaceutical industry. Based in part on the relatively high profits shown by this industry it is a target for criticism by government and politicians. The pharmaceutical industry, its relationships to pharmacy, and the public are discussed in chapter 28. Lastly, the

Copyright © 2002 Marcel Dekker, Inc.

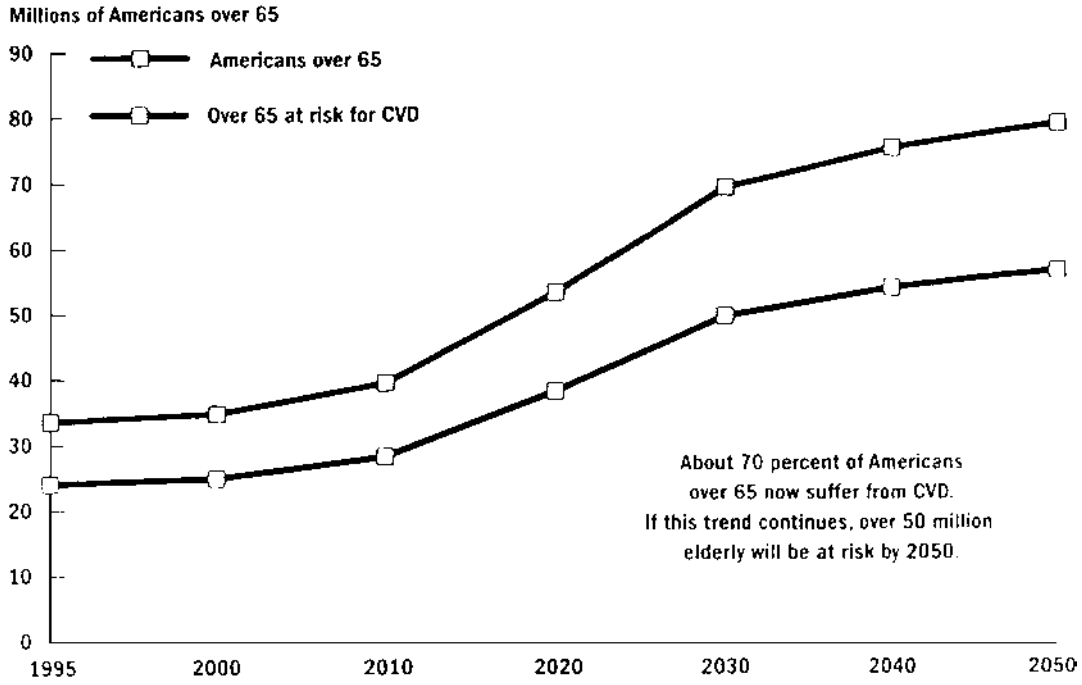


Fig. 7 Aging of baby boomers will dramatically increase population potentially at risk from cardiovascular disease (CVD).

Table 3 Prevalence, Cost, and Medicines in Development for Selected Major Diseases in the United States

Uncured disease	Approximate prevalence	Approximate annual economic cost ($billions)	Number[a] of medicines in development	Source
Alzheimer's disease	4,000,000	100.0	24	National Institute on Aging
Arthritis	43,000,000	54.6	28	Arthritis Foundation
Asthma	14,000,000	6.2	17	National Heart Lung and Blood Institute
Cancer	8,000,000	107.0	354	American Cancer Society
Congestive heart failure	4,900,000	20.2	17	American Heart Association
Coronary heart disease	13,900,000	95.6	38	American Heart Association
Depression	17,600,000	53.0	17	National Institute on Mental Health
Diabetes	15,700,000	98.2	26	National Institute of Diabetes and Digestive and Kidney Diseases
Hypertensive disease	50,000,000	31.7	10[b]	American Heart Association
Osteoporosis	10,000,000	13.8	19	National Osteoporosis Foundation
Schizophrenia	1,500,000	23.0	12	National Institute of Mental Health
Stroke	4,000,000	43.3	19	American Heart Association

Note: HIV/AIDS discussed separately in Chapter 1.
[a]PhRMA data.
[b]Hypertension medicines.
Source: PhRMA, 2000.

pharmaceutical component has been the fastest growing cost component of all health care expenditures, growing since 1993 at a steady 13.3 compound rate.

The U.S. prescription drug market in 1999 increased by 19%, totaling $2.7 billion, representing 2.7 billion prescriptions, a 9% increase over the prior year [8].

Copyright © 2002 Marcel Dekker, Inc.

Of this total prescription market the federal government paid about 13% (Medicaid), third parties paid 68%, and the individual paid about 19%[8]. The rapid growth of the prescription market in recent years is the result of numerous factors, but it certainly includes growing managed care utilization which includes some level of drug coverage, the continuing aging of the U.S. population, new product approvals and introductions (which tend to be more expensive), direct to consumer advertising by drug companies, inflation of drug product costs, and a strong economy. The aging-of-America effect is documented by following statistics [9]: the 50-plus age group consume 74% of all prescription drugs, 51% of all over-the-counter (OTC) drug products, and represent 65% of all hospital bed days. By 2005 30% of Americans will be 50 or older. The impact of the rapidly growing drug cost share of total operating expenses for HMOs is discussed in Chapter 27.

B. The Pharmacist's Responsibility and Role in Drug Product Selection

The pharmacist bears a heavy responsibility for the quality and appropriateness of the drug products that she or he dispenses. The Code of Ethics of the American Pharmaceutical Association states (in Section 2): "The pharmacist should never condone the dispensing of drugs and medications which are not of good quality or which do not meet standards required by law." Pharmacists are frequently obliged to make judgments concerning the quality of individual drug products and the various dosage forms and possible presentations available for individual drugs. This occurs as pharmacists serve on formulary committees, therapeutics committees, and in other "official" roles, as well as on a day-to-day basis, in evaluating manufacturers' and other data to select drug products for their patients that are not only cost-competitive, but are also safe and effective. The pharmacist's role in drug and drug product selection has increased dramatically in recent decades with the evolution of clinical pharmacy and pharmaceutical care and with the growth of third-party payments for prescription drugs. Fortunately, the pharmacist is the most knowledgeable expert on available drugs and drug products and criteria affecting drug product quality. Pharmacists are also better qualified and more knowledgeable than other health care professionals when it comes to drug product forms available, relative merits of different dosage forms, even within a given route of administration, and possible or most likely side effects within a given population group such as children or the elderly. Furthermore, pharmacists are well acquainted with the storage requirements for various drugs and drug products and with the physical signs by which deterioration may be detected. The pharmacist is the health care professional in the best position today to know the various drugs a given patient is taking and thus is best able to avert adverse therapeutic effects that can arise from many sources. This is because many patients see more than one physician, and they are increasingly taking a broader range of OTC products, including herbs, natural products, and nutriceuticals.

The questions of drug quality, drug cost, drug selection and effectiveness, and proper drug utilization have become a matter of widespread public interest. It is thus very important for pharmacists to be highly knowledgeable concerning all aspects of drug product quality and optimal drug utilization on an individual patient basis.

C. History and Evolution of Drugs and Drug Products

Every pharmacist and pharmacy student should read one of the many available books on the history of drugs and the drug industry [10–14]. A pictorial history such as *Drugs* [10] in the Life Science Library Series or one of the several illustrated books on the patent remedy era [11,12] is entertaining as well as educational reading. Several U.S. drug companies, on reaching the century mark, have also written interesting histories that also document U.S. pharmacy history [15–17]. Several pharmacy trade associations have written interesting histories directed to pharmacy practice [18,19]. A history of the U.S. Public Health Service also contains important bench marks for pharmacy [20]. These references are only a few of the many interesting historical books on pharmacy that could be cited.

When the history of drugs and drug products is expressed as a time continuum, as shown in Table 4, it is very apparent that most drug and drug product advancements have occurred over the last 50 years. Before the early twentieth century, the only purified organic chemical substances used in chemotherapy were aspirin, quinine, and morphine. For the preceding 4000 years, drugs changed relatively little. Early Egyptian physicians recorded over 70 "drugs" in 800 remedies administered in 14 different forms from pills to ointments and salves and poultices. The drugs were all from natural sources, ranging from spider webs, to animal excretions, to packs of mud, to poppy seeds. Between the eighth and thirteenth centuries, Arabic alchemists greatly advanced pharmaceutical art by

Copyright © 2002 Marcel Dekker, Inc.

Table 4 Periods and Notable Events in the History of the Development of Drugs and Drug Products

2000 B.C.	First drug records	
Ancient times to Middle Ages	Ancient to medieval pharmacy and medicine	Witch doctors Religious healers
1700		Natural products development
1800		First compendium (U.S.)
1850		First U.S. drug companies
1900	Patent remedy era	Development of analytical standards
1906		Wiley Act—first food and drug law
1938		Second major FDA legislation
1945–1965	Golden age of discovery	Development of sulfa anti-infectives (antibiotics, steroids, etc.)
1962		Kefauver-Harris Amendment
	New regulations	Accelerating development of cancer chemotherapy New drug-delivery systems
1977		Additional drug regulatory legislation—phase IV testing, etc.
1980s	Pharmaceutics advances	Full implementation of bioavailability standards Additional new drug-delivery capabilities
1980s–2000s		Mathematical optimization of drug product safety, effectiveness, and reliability
	Biotechnology era beginning	First recombinant DNA products Human insulin Human growth hormone Interferons, etc. Monoclonal antibodies Nucleotide blockage Growth in use of natural products and neutraceuticals
2000		Pharmaceutical company megamergers continue Sequencing of the human gerome Gene therapy developing Use of human tissues in medicine High throughput screening of potential drugs Combinatorial chemistry, biotechnology, and computers advance rational drug design

introducing extraction and distillation processes to concentrate and purify natural products. However, Arabic drugs did not reach Christian Europe until the late Middle Ages (thirteenth to fifteenth centuries). Pharmaceutical art in colonial America some 200 years ago was little better than that of the Arabic alchemists 1000 years earlier. The patent remedy era that flourished in the second half of the nineteenth century was a colorful period in pharmaceutical history when "cure-alls" were marketed by pitchmen from the backs of horse-drawn wagons, often as a follow-up to some free entertainment used to gather the locals. As depicted in the advertised indications from the label of one popular patent remedy of the day (Fig. 8), the only limit on claims appeared to be the imagination of the label writers [10]. As noted on the advertising copy of Fig. 8, these remedies were expensive in the days when an average wage earner made 15–20 dollars a week. Grandmother Pinkham used pictures of her grandchildren on her advertising and label copy; others, such as the medicinal syrups millionaire G. G. Green, placed on their advertising and labeling illustrations of the mansions that their high-profit "cure-alls" brought them. Most patent remedies contained common plant

Copyright © 2002 Marcel Dekker, Inc.



Fig. 8 Example of the advertised indication (cures) from the labeling of a widely used patent remedy, before Enactment of the Food, Drug and Cosmetic Act of 1906. (From Ref. 10.)

Copyright © 2002 Marcel Dekker, Inc.

extracts, such as taraxicum (common dandelion weed), at least 15% alcohol, and occasionally opium or other narcotic or addictive substances to help assure repeat sales. It is interesting to speculate on the number of persons in the temperance era who kept all those around them "dry" and then enjoyed their evening toddy of Lydia Pinkhams or stronger higher proof products before retiring. The Food, Drug, and Cosmetic Act of 1906 was intended to combat the abuses of the patent remedy era, at least in part by requiring labeling of the active ingredients contained in all pharmaceuticals and by broadly limiting fraudulent practices. The 1938 act went further, requiring that at least some principles of rational therapeutics be applied to all products and their claims. The 1962 act required many more proofs of safety and effectiveness. (See Chapter 19 for a more detailed description of the evolution of drug laws.)

One interesting way of examining drug and drug product quality is to analyze the changes in drugs and drug products over the centuries, especially the very rapid changes in the last half-century (see Table 4). Anyone who imagines that current drug products are optimal as to quality features or that we have reached the ultimate in chemotherapeutic capabilities is assuming that the history of drugs is now standing still as well as ignoring the science of pharmaceutics as we currently know it. Although the Middle Ages ended around 1450, historically speaking, drugs and drug products did not progress substantially above the quality and knowledge level of the medieval period until the late nineteenth and early twentieth centuries (Table 4). Drug and drug product advancements during the last 50–60 years have surpassed the total advancement in the field over the entire 4000-year history of drug development (see Chapter 28). As we enter the twenty-first century and the biotechnology scientific era, there is every likelihood that the rate of advancement in pharmacotherapy will overshadow everything we have known in the past.

D. Criteria for Drug and Drug Product Quality

Compendial standards and government regulations require that all drug products, whether ethical prescription or OTC products, meet strict standards of identity, potency, and purity. From about 1900 (see Table 4) until recent decades, standards of identity (the product is what it is actually labeled to be), potency (the active ingredient is present in the labeled amount), and purity [basically limiting nondrug materials as well as describing the amount of active ingredient(s) in natural substances] were thought to define drug quality adequately and were enforced under evolving law and U.S. Food and Drug Administration (FDA) regulation. The addition of a few physical tests, such as weight variation and disintegration time to compendial products such as tablets and capsules, was thought to accurately define the quality of these products. We now know that drug products require very careful evaluation to accurately reflect their quality and performance in clinical roles and that earlier concepts of evaluation required expansion. The designation "quality," applied to a drug product, according to a modern definition, requires that the product:

Contain the quantity of each active ingredient claimed on its label, within the applicable limits of its specifications
Contain the same quantity of active ingredient from one dosage unit to the next
Be free from extraneous substances
Maintain its potency, therapeutic availability, and appearance until used
Upon administration, release the active ingredient for full biological availability

In the contemporary definition of quality we see that the concepts of identity, potency, and purity are retained in the first three criteria, but that bioavailability potency maintenance (including maintenance of pharmaceutical elegance and therapeutic availability or full biological availability) are added. The definition recognizes that drug products may undergo changes with time that result in a loss of biological and therapeutic activity, even though the product complies completely with the original potency and purity standards and no significant drug decomposition has occurred. Such losses in therapeutic activity, without any chemical potency change, may occur as a result of a variety of causes, including:

1. Physical changes in the dosage form (moisture loss or gain, crystal changes in excipients, tablet hardening, loss of disintegration/deaggregation properties, etc.)
2. Physical changes in the drug (conversion of a more stable, less readily soluble polymorph, etc.)
3. Chemical changes or interactions involving excipients (such as esterification of coatings, rendering them less polar and less soluble)

As noted in chapters that follow, in vitro tests may not themselves be adequate to assure that a product possesses adequate or full bioavailability and

Copyright © 2002 Marcel Dekker, Inc.

therapeutic activity. Whether or not in vitro tests are adequate quality-control indications depends on their sensitivity to pick-up aging or environmental exposure effects that produce the type of changes noted in the preceding paragraphs or other changes that are of consequence to therapy. The effects on drug potency of environmental stress conditions (e.g., increased temperature, increased humidity, combinations of the two, or temperature cycling) are known for most drug products. The effects of such stress conditions on bioavailability or therapeutic activity are much less well known currently, even for many commercial products.

A drug product can be of no higher quality than the quality of the drug(s) and excipients (nondrug additives) from which the product is made. It is possible, however, for a product to be of much lower quality than its components, since quality and clinical performance are also related to:

1. The rationale of the dosage form design (e.g., if a drug is rapidly destroyed in solution at a pH of 2 or below, the design of an oral product must enable the drug to get through the stomach without substantial dissolution)
2. The method(s) of product manufacture
3. In-process and final quality-control procedures to assure that the quality designed and manufactured into the product is actually there
4. Reasonable convenience and ease of product use to assure patient compliance with prescribed dosages

Some of the factors and considerations in the design of high-quality drug products are shown in Table 5. Various physicochemical and pharmacokinetic properties of drugs that affect dosage form design, and their influence on design of high-quality products, are de-

Table 5 Factors in the Design and Production of High-Quality Drug Products

Input factors	Output factors
A sound development/design manufacturing base	Effectiveness
The preformulation research database Physicochemical properties of the drug Pharmacokinetic characterization of the drug	Safety Reliability
A rational dosage form design Formulation of a stable, reliable system	Stability Physical Chemical
Objective preclinical and clinical testing	Microbiological Bioavailability
A precise, reproducible manufacturing process Well-controlled manufacturing steps Coordinated manufacturing sequences Efficient, sanitary operation Modern plant and equipment Knowledgeable, well-qualified workers	Pharmaceutical Appearance Organoleptic properties
	Convenience Ease of use Dosing frequency Consumer acceptance
A sensitive product control system Raw material control Processing controls Final product controls Chemical control standards Physical properties standards Biological and microbiological standards	
Informed, qualified, and responsible personnel Management Research and development Quality control Production Services	

Copyright © 2002 Marcel Dekker, Inc.

scribed in Chapter 7 on preformulation. The criteria and properties defining the quality of various dosage forms are discussed in the relevant chapters on dosage forms. The features of an optimized drug are discussed in the relevant chapters on dosage forms. The features of an optimized drug product and the concepts of true drug-delivery systems are described in the next section.

IV. THE DRUG PRODUCT AS A DELIVERY SYSTEM

A. Drug Products, Drug-Delivery Systems, and Therapeutic Systems

Drug substances in their purified state usually exist as crystalline or amorphous powders or as viscous liquids. The majority of drug substances exist as white or light-colored crystalline powders. Although drugs were once commonly dispensed as such in powder papers, this practice is virtually unknown in pharmacy practice today. With the possible exception of the anesthetic gases, all drugs in legitimate commerce are now presented to the patient as drug products. It is now well recognized that the therapeutic efficacy and the therapeutic index [ratio of LD_{50} (lethal dose in 50% of the subjects) to ED_{50} (effective dose in 50% of the subjects)] of a drug product is not totally defined by the chemical constitution of the drug and its inherent pharmacokinetic profile. The actual performance of many drugs in clinical practice is now known to be greatly affected by the method of presentation of the drug to the patient. Factors affecting the presentation include:

- The portal of drug entry
- The physical form of the drug product
- The design and formulation of the product
- The method of manufacture of the drug product
- Various physicochemical properties of the drug and excipients
- Physicochemical properties of the drug product
- Control and maintenance of the location of the drug product at the absorption site(s)
- Control of the release rate of the drug from the drug product

In the late 1940s and early 1950s, sustained-release products appeared as a major new class of pharmaceutical product, in which product design was intended to modify and improve drug performance, by increasing the duration of drug action and reducing the required frequency of dosing. In the mid- to late 1960s, the term "controlled drug delivery" came into being to describe new concepts of dosage form design, which also usually involved controlling and retarding drug dissolution from the dosage form, but with additional or alternative objectives to sustained drug action. These new objectives included improving safety, enhancing bioavailability, improving drug efficiency and effectiveness, enhancing reliability of performance, reducing side effects, facilitating patient use and compliance, or other beneficial effects. In the 1970s, yet another term and concept of drug product design and administration appeared: the therapeutic system. The objective of the therapeutic system is to optimize drug therapy by design of a product that incorporates an advanced engineering systems control approach. Three types of therapeutic systems have been proposed, the first of which is already in use: (a) the "passive preprogrammed" therapeutic system—one containing a controlling "logic element," such as a membrane or series of plastic laminates, which preprograms at the time of fabrication or assembly a predetermined delivery pattern (usually constant zero-order release) that is ideally independent of all in vivo, physical, chemical, and biological processes; (b) the "active, externally programmed or controlled" therapeutic system—wherein the logic element is capable of receiving and converting a signal (such as an electromagnetic signal) sent from a source external to the body to control and properly modulate drug release from the device within the body; and (c) the "active, self-programmed" therapeutic system—containing a sensing element that responds to the biological environment (such as blood glucose concentration in diabetes) to modulate drug delivery in response to that information. Before the sustained-release concepts of the 1940s and 1950s, which also included depot forms of parenteral products, no significant new oral drug-delivery concepts had occurred in the preceding 75 years (since the enteric-coating concept).

B. The Concept of the Optimized Drug Product

The optimized drug product may be viewed as a drug-delivery system for the one or more drugs that it contains. The goal of this drug-delivery system is to release the drug(s) to produce the maximum simultaneous safety, effectiveness, and reliability, as depicted in Fig. 9. Various physicochemical product properties that influence the quality features of safety, effectiveness, and reliability are shown in Table 6. Some physicochemical properties can affect two or all three quality features of Fig. 9. For example, consider chemical stability. As a drug decomposes, if the

Copyright © 2002 Marcel Dekker, Inc.

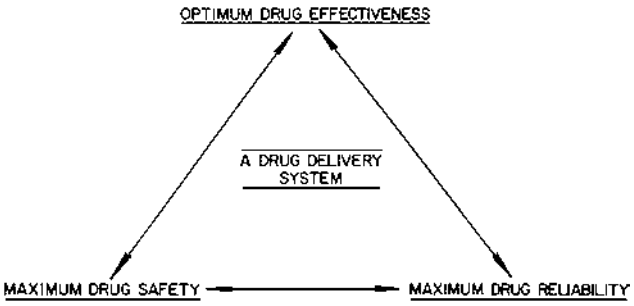


Fig. 9 Features of the optimized drug product.

decomposition product(s) are inactive, this is equivalent to a reduction of the drug dose remaining in the product— in other words, a reduction in product reliability and eventually effectiveness. If the decomposition products are toxic or irritating to the body, product safety is also reduced as the product degrades.

In examining Table 6, the manner in which each physicochemical, physiological, or therapeutic property affects the various quality features will generally be apparent. You will think of and read about many other physicochemical properties that influence the quality features of drugs and drug products as you read various chapters of this book. It should also be noted that the three basic quality features of Fig. 9 are connected by double-headed arrows. Thus, as the pharmaceutical formulator modifies the design of a drug product or its method of manufacture to improve one quality feature or one physicochemical property related primarily to one quality feature, the other properties or quality features may be, and usually are, altered. As an example, it may be our goal to increase the hardness of a tablet by formulation (adding more binder) and/or processing (compressing the tablets harder) to improve tablet gloss and appearance or to reduce tablet friability (powdering and chipping in the bottle). This is a

Table 6 Factors Affecting Drug Product Safety, Effectiveness, and Reliability

Safety	Effectiveness	Reliability
Acute safety quantification Therapeutic index = LD_{50}/ED_{50}	Clinical effectiveness	Chemical stability
	Generic effectiveness	Physical stability
Long-range safety considerations Onset of side effects Accumulation	Blood levels Urinary elimination Pharmacological response(s)	Microbiological stability Unit-dose precision
Nature of side effects Severity Reversibility	Bioavailability	Patient acceptance Convenience Pharmaceutical elegance
Frequency of side effects		Bioavailability High percentage
Untoward and other reactions Idiosyncratic responses Anaphylaxis Tolerance Addiction		Uniformity Stability
Drug interactions Number of drugs involved Probability of interaction in therapy Severity of the interactions Frequency of interactions		
Stability considerations Chemical stability Physical stability Microbiological stability Bioavailability stability		

Copyright © 2002 Marcel Dekker, Inc.

worthy objective, but it may also reduce the rate and extent of drug dissolution from the tablet. This, in turn, could reduce the reliability of drug absorption and drug performance from patient to patient, influence transit rate and drug dissolution along the gastrointestinal tract within a given patient, or even reduce effectiveness if the dissolution rate now limits or reduces bioavailability. In the example just cited, maximizing tablet hardness and appearance is a "competing objective" to maximizing drug dissolution and bioavailability.

In Fig. 9 we see the definition of the optimized drug product as the drug-delivery system that balances all these factors against each other to produce the maximum possible effectiveness as the primary objective, while producing the best possible simultaneous safety and reliability as secondary objectives, with mathematical certainty. An alternative optimization approach would be to produce the maximum possible (optimized) product safety as the primary objective, while producing the best effectiveness and reliability as the secondary objectives. Yet a third approach would be to optimize safety and effectiveness as equally weighted primary objectives, while maximizing reliability as the secondary objective. Chapter 18 is devoted to the topic of optimization and treats the manner in which experiments may be designed to establish the necessary factors and relationships between factors (independent and controllable processing, formulation, and other variables) as these influence one another and the product quality features (dependent or response variables). Optimization methods then treat this database to design and manufacture the best possible product from an overall standpoint, considering quality features, which may be competing (i.e., as you improve one feature, another degrades), taking into account primary versus secondary features and numerous possible trade-off decisions. Although it is true that the vast majority of drug products on the market today are reasonably safe and effective, it is also true that relatively few products have been designed as optimized systems. Indeed, until about 25 years ago, formal optimization methods were unknown in the pharmaceutical and most other industries. The significance of drug products not being optimum systems varies with drug product class. For drugs and drug product classes with a high therapeutic index (ratio of LD_{50} to ED_{50}) and minimal dose-related side effects, maximizing safety is of less concern, and if the drug is well absorbed, a good, stable, pure, and potent drug product that is reasonably reliable may be nearly optimum. For drugs that have less of a safety margin, it may be argued that the conventional, rapidly releasing, effective, stable, and typical reliable product that is currently marketed is not optimum.

Figure 10 illustrates the types of blood level concentration profiles that are produced with different doses of a rapidly releasing product versus a controlled-release product. This figure is representative of the oral route of administration but may be extrapolated to other routes of administration, with the blood-level time frame simply being shifted to reflect changes in absorption (and possibly distribution) patterns. For the rapidly releasing product, whether given in a single 100-unit dose (curve A) or three divided doses of 33 units (curve B), the inherent ability of the individual to absorb the drug determines the rate of absorption and the peak blood level obtained. (At the peak, the rate of absorption and elimination are equal.) The conventional rapidly releasing drug product is not controlling the blood level versus time profiles; such a product is simply an uncontrolled "dump system," dumping the drug in the stomach for rapid dissolution and uncontrolled absorption. The body's inherent ability to absorb the drug under the patient's physiological state at a particular time point and the drug's pharmacokinetics dictate the shape of the blood level versus time profile at any particular dosing level or dosing frequency. For the optimized controlled-release form (curve C in Fig. 10), the drug product, which is now a drug-delivery system, is controlling the rate of release of drug in solution for absorption, and the release rate has been optimized to match drug inactivation and elimination, so that nearly constant levels may be maintained while the drug is in an absorption region in the gut. It has been clearly demonstrated that controlled drug delivery can also substantially increase the therapeutic index and safety margins of certain drugs, while retaining full therapeutic effectiveness. This is because many controlled-release systems produce more rounded blood level versus time profiles, without sharp peaks (compare curves A and C in Fig. 10), so that much larger doses of the controlled-release product are required to reach toxic levels. In one study [21], the effectiveness and duration of action of an antihistamine was followed in animals using a histamine vapor challenge test. The drug as a rapidly soluble dispersion or solution had a duration of action of 3.8 hours, whereas the controlled-release form provided protection to histamine vapor of 8.7 hours. Many antihistamines are depressant drugs and are dangerous on overdosage or in combination with other depressant drugs. In the same study, the drug in conventional form was lethal to 85% of the rats dosed within 30

Copyright © 2002 Marcel Dekker, Inc.

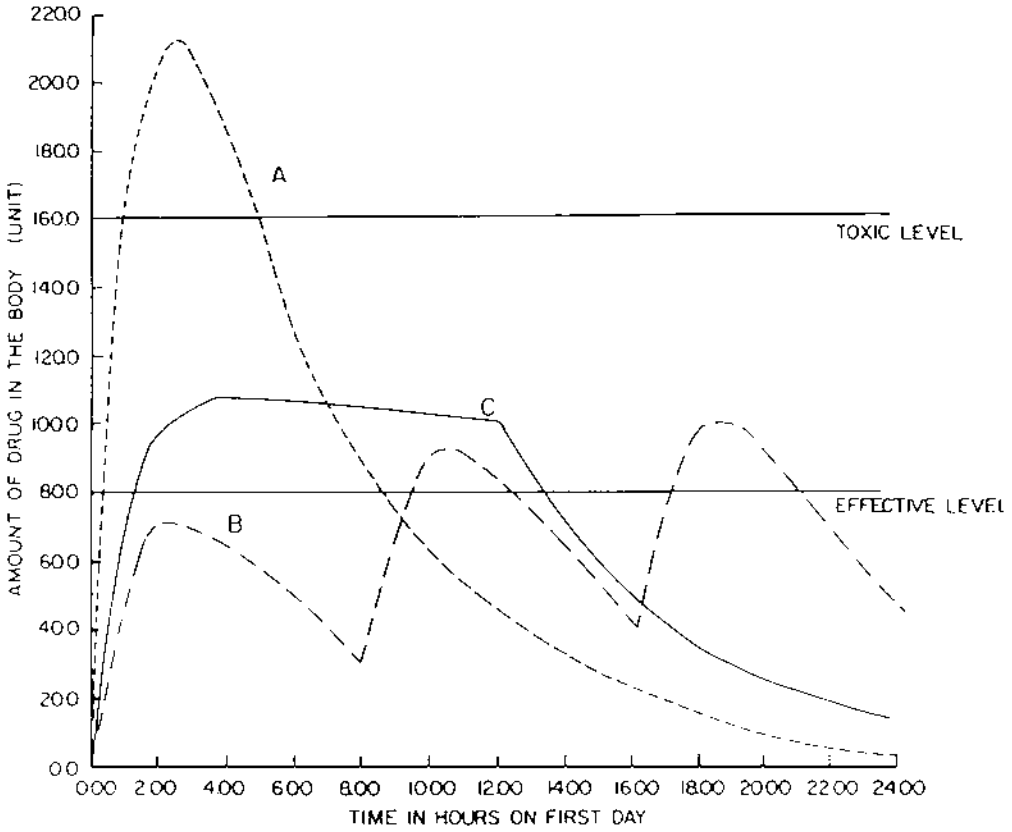


Fig. 10 Blood level versus time profile simulations following: (A) a single dose representing 100 units of a drug from a rapidly releasing dosage; (B) Three divided ddoses of 33 units each from the same rapidly releasing product; and (C) a single 100 unit dose from an optimized controlled-release dosage form. A hypothetical effective level (80 units) and toxic level (160 units) are depicted. The dosing units are typically in mg and the blood level concentration units in μg or ng.

minutes at a dose of 200 mg/kg. The same dose of drug when administered as the active controlled-release form killed none of the rats dosed (LD_0 at 24 h) [21]. Similar studies have demonstrated the ability to improve the safety of a barbiturate drug. In one such study, conventional phenobarbital had an LD_{50} in rats at a dose of 200 mg/kg. Twice the dose from an active controlled-release form had only an LD_{40} [22]. The cost of developing, testing, and gaining approval through the FDA of depressant or hazardous drugs as new products, optimized from a safety as well as effectiveness standpoint, is apparently currently too high to warrant much activity in this area. Nevertheless, we predict that such products will be the rule, rather than the exception, at some point in the future. Not only would such products save lives as a result of the consequences of purposeful or accidental overdosing of potent, depressant, and low therapeutic index drugs, but they might also reduce the severity of some drug interaction effects (especially when alcohol is one of the drugs). By eliminating the spikes in the blood profiles of immediate release dosage forms, sustained release dosage forms have the ability to reduce the frequency and severity of the side effects of some drugs. A number of cardiovascular and other drugs are designed as sustained release products for this reason. Other controlled-delivery forms have the potential of reducing some drug abuse problems, since the formulations and slow drug dissolution from the forms makes it more difficult to produce extemporaneous illegal and hazardous injectable solution forms.

Drug-delivery systems have been designed that are held in the stomach or are bioadhesive at other absorption sites for prolonged and controlled time periods (6–10 h) while releasing drug at controlled rates. Such concepts are bringing us closer to being able to achieve the ideal attributes of drug-delivery systems (Table 7), including control of such factors as variable

Copyright © 2002 Marcel Dekker, Inc.

Table 7 Ideal Attributes of a Drug-Delivery System

1. Capable of controlled delivery rates to accommodate the pharmacokinetics of various drugs (flexible programming)
2. Capable of precise control of a constant delivery rate (precise programming)
3. Not highly sensitive to physiological variables, such as:
   - Gastric motility and emptying, pH, fluid volume, and contents of the gut
   - Presence/absence or concentration of enzymes
   - State of fasting and type of food present
   - Physical position and activity of subject
   - Individual variability
   - Disease state
4. Predicated on physicochemical principles (not pharmaceutical art)
5. Capable of a high order of drug dispersion (the ultimate is molecular in scale)
6. Drug stability is maintained or enhanced
7. The controlling mechanism adds little mass to the dosage form
8. Applicable to a wide range and variety of drugs

and uncontrolled gastric emptying and transit along the gut, which currently cause the oral route of administration to be the least reliable route of drug administration, even though it is the most popular method of achieving systemic drug effects. Future drug-delivery systems that have controlled, prolonged retention in the stomach (or accessible body cavities) with simultaneously controlled delivery rates will no doubt improve drug effectiveness for some agents by enhancing absorption, reliability, and efficiency. They also offer new possibilities of truly optimizing drug safety by prolonging the time over which the drug can be recovered or neutralized, not only on overdosing, but in acute drug interaction episodes. Bioadhesive systems that can retain drugs on mucosal surfaces for prolonged periods, while promoting drug absorption and delivery, will play a growing role in developing future delivery systems and drug products for some proteins, peptides, and other biotechnology-generated drugs. Such systems will also render the oral mucosal, nasal, vaginal, and rectal routes more useful and reliable for drug delivery. During their working lifetimes pharmacy students of today are certain to see many new drug-delivery concepts reach the marketplace that permit true optimization of drug action and the attainment of the ideal attributes of drug delivery.

REFERENCES

1. HIV and Africa's future. Science 288: 2149–2178, 23 June 2000.
2. XIII International AIDS Conference, Durban, South Africa, reported in Newsweek, July 24, 2000, and other public press.
3. Emerging diseases. Science 289: 518–519, 28 July 2000.
4. R. Forbes. Life and Death in Shakespere's London. Am Sci 58: 511–520 1970.
5. Statistical Abstract. U.S. Dept. of Commerce, Washington, DC, 1999.
6. Report of the National Institute on Aging. Older Americans 2000: Key Indicators of Well Being, Washington, DC, Aug. 10, 2000.
7. World Health Report 2000. World Health Organization, Geneva, Switzerland.
8. F. Deardorff. Changes and Trends in the R_x Market. IMS Health, 1999.
9. K. Dychtwald. Age Power, How the 21st Century Will be Ruled by the New Old. Putnam, New York, 1999.
10. W. Model and A. Lansing. Drugs. Time-Life Books, New York, 1969.
11. A. Hechtlinger. The Great Patent Medicine Era. Galahad Books, New York, 1970.
12. G. Carson. One for a Man, Two for a Horse. Bramhall House, New York, 1971.
13. W. Screiber and F. K. Mathys. Infectious Diseases in the History of Medicine. Kreis and Co., Basel, 1987.
14. P. Boussel, H. Bonnemain, and F. Bove. History of Pharmacy and the Pharmaceutical Industry. Asklepious Press, Paris, 1983.
15. R. Carlisle. A Century of Caring—The Upjohn Story. Benjamin Co., Elmsford, NY, 1987.
16. W. D. Pratt. The Abbott Almanac. Benjamin Co., Elmsford, NY, 1987.
17. L. Galambos et al. Values and Visions. A Merck Century. Merck and Co., 1992.
18. C. F. Williams. A Century of Service and Beyond, 1898 NARD-NCPA 1998. National Community Pharmacy Association, Alexandria, VA 1998.
19. J. Mobley. Prescription for Success, The Chain Drug Story. The Lowell Press, Kansas City, MO, 1990.

Copyright © 2002 Marcel Dekker, Inc.

20. F. Mullan. Plagues and Politics, A Story of the United States Public Health Service. Basic Books Inc., New York, 1989.
21. H. Goodman and G. Banker. Molecular-scale drug entrapment as a precise method of controlled drug release I: Entrapment of cationic drugs by Polymeric flocculation, J. Pharm. Sci. 59: 1131–1137, 1970.
22. J. Boylan and G. Banker. Molecular-scale drug entrapment as a precise method of controlled drug release II: Entrapment of anionic drugs by polymeric gelation. J. Pharm. Sci. 62: 1177–1184, 1973.

Copyright © 2002 Marcel Dekker, Inc.

Chapter 2

Principles of Drug Absorption

Michael Mayersohn

University of Arizona, Tucson, Arizona

I. INTRODUCTION

Drugs are most often introduced into the body by the oral route of administration. In fact, the vast majority of drug dosage forms are designed for oral ingestion, primarily for ease of administration. It should be recognized, however, that this route may result in inefficient and erratic drug therapy. Whenever a drug is ingested orally (or by any nonvascular route), one would like to have rapid and complete absorption into the bloodstream for the following reasons:

1. Assuming that there is some relationship between drug concentration in the body and the magnitude of the therapeutic response (which is often the case), the greater the concentration achieved, the greater the magnitude of response.
2. In addition to desiring therapeutic concentrations, one would like to obtain these concentrations rapidly. The more rapidly the drug is absorbed, in general, the sooner the pharmacological response is achieved.
3. In general, one finds that the more rapid and complete the absorption, the more uniform and reproducible the pharmacological response becomes.
4. The more rapidly the drug is absorbed, the less chance there is of drug degradation or interactions with other materials present in the gastrointestinal tract.

In a broad sense, one can divide the primary factors that influence oral drug absorption and thus govern the efficacy of drug therapy into the following categories: physicochemical variables, physiological variables, and dosage form variables. For the most part, these variables will determine the clinical response to any drug administered by an extravascular route. Although often the total response to a drug given orally is a complex function of the aforementioned variables interacting together, the present discussion is limited primarily to the first two categories involving physicochemical and physiological factors. Dosage form variables influencing the response to a drug and the effect of route of administration are discussed in Chapters 4 and 5.

The vast majority of drugs in current use and those under development are relatively simple organic molecules obtained from either natural sources or by synthetic methods. It is important to note, however, the virtual revolution in development of new therapeutic entities—those based upon the incredible advances being made in the application of molecular biology and biotechnology. These new drugs, especially peptides and proteins, are not the traditional small organic molecules stressed in this chapter. Indeed, those compounds have unique physicochemical properties, which are quite different from those of small organic molecules, and they offer remarkable challenges for drug delivery. As a result, new and more complex physical delivery systems are being designed in conjunction with an examination of other, less traditional routes of administration (e.g., nasal, pulmonary, transdermal, etc.). Because of issues

Copyright © 2002 Marcel Dekker, Inc.

of instability in the gastrointestinal tract and poor intrinsic membrane permeability, it appears unlikely that these new biotechnology-derived drugs will employ the oral route for administration. Numerous strategies, however, are being explored and there is evidence that some measure of gastrointestinal absorption can be achieved for certain peptides [1]. One recently reported approach that shows promise involves conjugating the poorly absorbed compound to a so-called molecular transporter. The latter are oligomers of arginine that undergo active cellular uptake [2,3].

II. ANATOMICAL AND PHYSIOLOGICAL CONSIDERATIONS OF THE GASTROINTESTINAL TRACT

The gastrointestinal tract (GIT) is a highly specialized region of the body whose primary functions involve the processes of secretion, digestion, and absorption. Since all nutrients needed by the body, with the exception of oxygen, must first be ingested orally, processed by the GIT, and then made available for absorption into the bloodstream, the GIT represents a significant barrier and interface with the environment. The primary defense mechanisms employed by the gut to rid it of noxious or irritating materials are vomiting and diarrhea. In fact, emesis is often a first approach to the treatment of oral poisoning. Diarrhea conditions, initiated by either a pathological state or a physiological mechanism, will result in the flushing away of toxins or bacteria or will represent the response to a stressful condition. Indeed, the GIT is often the first site of the body's response to stress, a fact readily appreciated by students taking a final exam! The nearly instinctive gut response to stress may be particularly pertinent to patients needing oral drug therapy. Since stress is a fact of our daily lives, and since any illness requiring drug therapy may to some degree be considered stressful, the implications of the body's response to stress and the resulting influence on drug absorption from the gut may be extremely important.

Figure 1 illustrates the gross functional regions of the GIT. The liver, gallbladder, and pancreas secrete materials vital to the digestive and certain absorptive functions of the gut. The lengths of various regions of the GIT are presented in Table 1. The small intestine, comprising the duodenum, jejunum, and ileum, represents greater than 60% of the length of the GIT, which is consistent with its primary digestive and absorptive functions. In addition to daily food and fluid

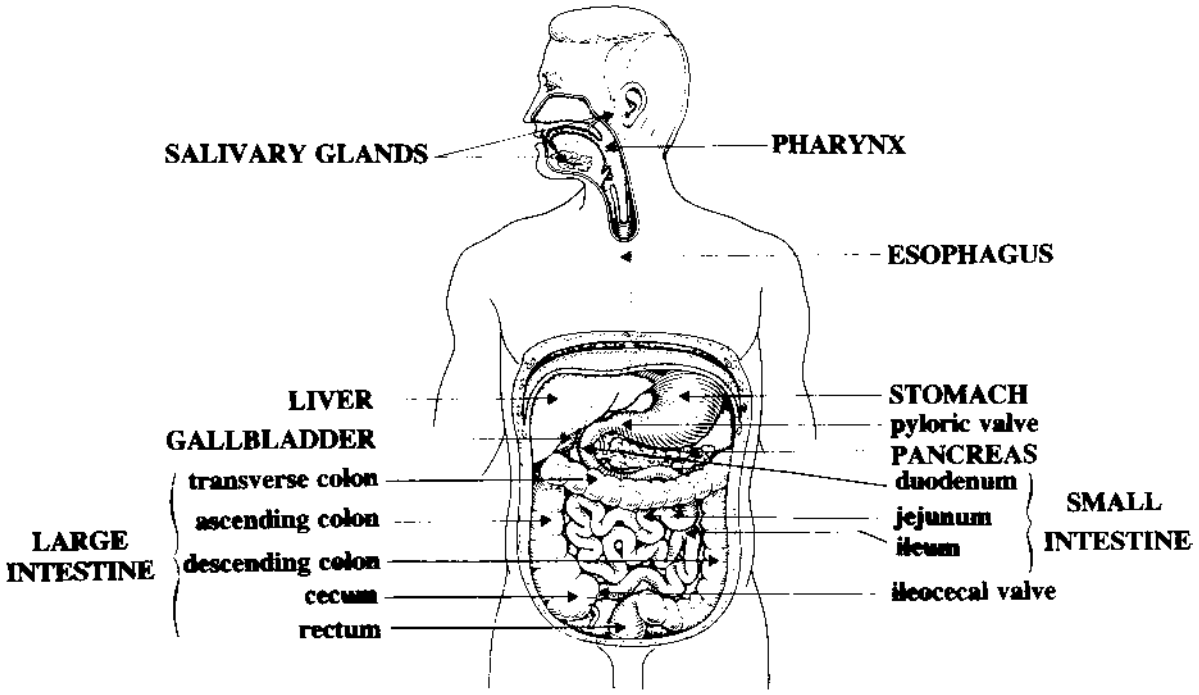


Fig. 1 Diagrammatic sketch of the gastrointestinal tract (and subdivisions of the small and large intestines) along with associated organs. (Modified from Ref. 4.)

Copyright © 2002 Marcel Dekker, Inc.

Table 1 Approximate Lengths of Various Regions of the Human Gastrointestinal Tract

Region	Length (m)
Duodenum	0.3
Jejunum	2.4
Ileum	3.6
Large Intestine	0.9–1.5

intake (about 1–2 L), the GIT and associated organs secrete about 8 L of fluid per day. Of this total, only 100–200 mL of stool water is lost per day, indicating efficient absorption of water throughout the tract.

A. Stomach

After oral ingestion, materials are presented to the stomach whose primary functions are storage, mixing, and reducing all components to a slurry with the aid of gastric secretions and then emptying these contents in a controlled manner into the upper small intestine (duodenum). All of these functions are accomplished by complex neural, muscular, and hormonal processes. Anatomically, the stomach has classically been divided into three parts: fundus, body, and antrum (or pyloric part), as illustrated in Fig. 2. Although there are no sharp distinctions among these regions, the proximal stomach, made up of the fundus and body, serves as a reservoir for ingested material, and the distal region (antrum) is the major site of mixing motions and acts as a pump to accomplish gastric emptying. The fundus and body regions of the stomach have relatively little tone in their muscular wall, and, as a result, these regions can distend outward to accommodate a meal of up to one liter.

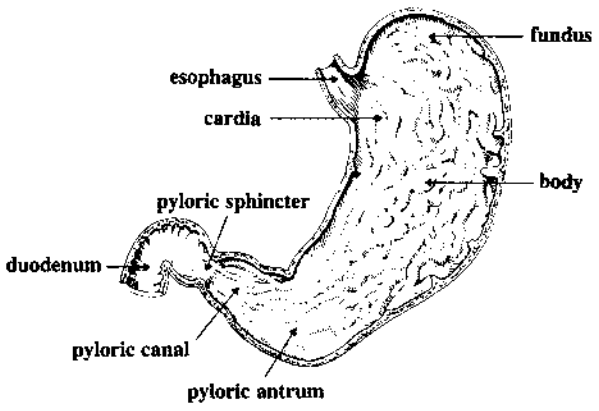


Fig. 2 Diagrammatic sketch of the stomach and anatomical regions. (Modified from Ref. 5.)

A common anatomical feature of the entire GIT is its four concentric layers. Beginning with the luminal (i.e., inner or absorbing) surface these are the mucosa, submucosa, muscularis mucosa, and serosa. The three outer layers are similar throughout most of the tract; however, the mucosa has distinctive structural and functional characteristics. The mucosal surface of the stomach is lined by an epithelial layer of columnar cells, the surface mucous cells. Along this surface are many tubular invaginations, referred to as gastric pits, at the bottom of which are found specialized secretory cells. These secretory cells form part of an extensive network of gastric glands, which produce and secrete about 2 L of gastric fluid daily. The epithelial cells of the gastric mucosa represent one of the most rapidly proliferating epithelial tissues, being shed by the normal stomach at the rate of about a half-million cells per minute. As a result, the surface epithelial layer is renewed every 1 to 3 days. Covering the epithelial cell surface is a layer of mucus 1.0–1.5 mm thick. This material, made up primarily of mucopolysaccharides, provides a protective lubricating coat for the cell lining.

The next region, the muscularis mucosa, consists of an inner circular and an outer longitudinal layer of smooth muscle. This area is responsible for the muscular contractions of the stomach wall needed to accommodate a meal by stretching and for the mixing and propulsive movements of gastric contents. An area known as the lamina propria lies below the muscularis mucosa and contains a variety of tissue types, including connective and smooth muscles, nerve fibers, and the blood and lymph vessels. It is the blood flow to this region and to the muscularis mucosa that delivers nutrients to the gastric mucosa. The major vessels providing a vascular supply to the GIT are the celiac and the inferior and superior mesenteric arteries. Venous return from the GIT is through the splenic and the inferior and superior mesenteric veins. The outermost region of the stomach wall provides structural support for the organ.

B. Small Intestine

The small intestine has the shape of a convoluted tube and represents the major length of the GIT. The small intestine, comprising the duodenum, jejunum, and ileum, has a unique surface structure, making it ideally suited for its primary role of digestion and absorption. The most important structural aspect of the small intestine is the means by which it greatly increases its

Copyright © 2002 Marcel Dekker, Inc.

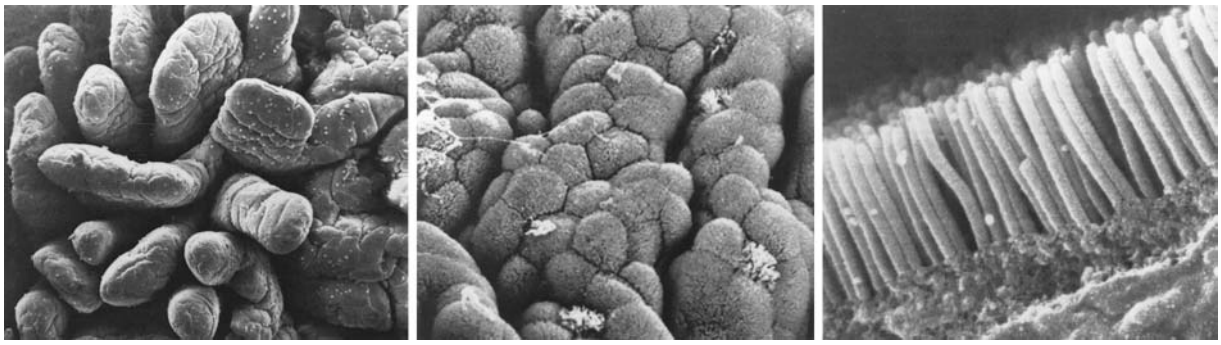


Fig. 3 (A) Photomicrograph of the human duodenal surface illustrating the projection of villi into the lumen (magnification $\times 75$). The goblet cells appear as white dots on the villus surface. (B) Photomicrograph of a single human duodenal villus illustrating surface coverage by microvilli and the presence of goblet cells (white areas) (magnification $\times 2400$). (C) Photomicrograph illustrating the microvilli of the small intestine of the dog (magnification $\times 33{,}000$). (From Ref. 6.)

effective luminal surface area. The initial increase in surface area, compared to the area of a smooth cylinder, is due to the projection within the lumen of folds of mucosa, referred to as the folds of Kerckring. Lining the entire epithelial surface are finger-like projections, the villi, extending into the lumen. These villi range in length from 0.5 to 1.5 mm, and it has been estimated that there are about 10–40 villi per square millimeter of mucosal surface. Projecting from the villi surface are fine structures, the microvilli (average length, 1 mm), which represent the final large increase in the surface area of the small intestine. There are approximately 600 microvilli protruding from each absorptive cell lining the villi. Relative to the surface of a smooth cylinder, the folds, villi, and microvilli increase the effective surface area by factors of 3, 30, and 600, respectively. The resulting area represents a surface about equal to two thirds of a regulation tennis court! These structural features are clearly indicated in the photomicrographs shown in Fig. 3. A diagrammatic sketch of the villus is shown in Fig. 4.

The mucosa of the small intestine can be divided into three distinct layers. The muscularis mucosa, the deepest layer, consists of a thin sheet of smooth muscle

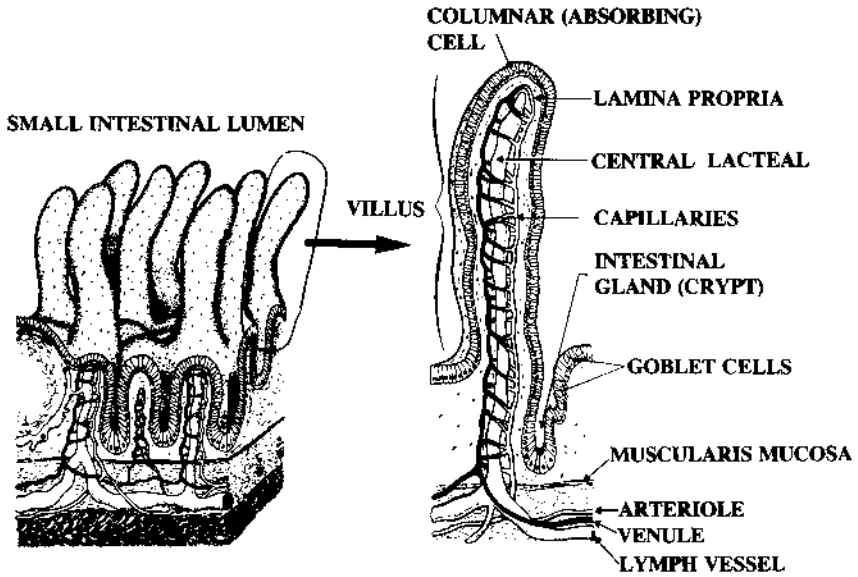


Fig. 4 Diagrammatic sketch of the small intestine illustrating the projection of the villi into the lumen (left) and the anatomic features of a single villus (right). (Modified from Ref. 7.)

Copyright © 2002 Marcel Dekker, Inc.

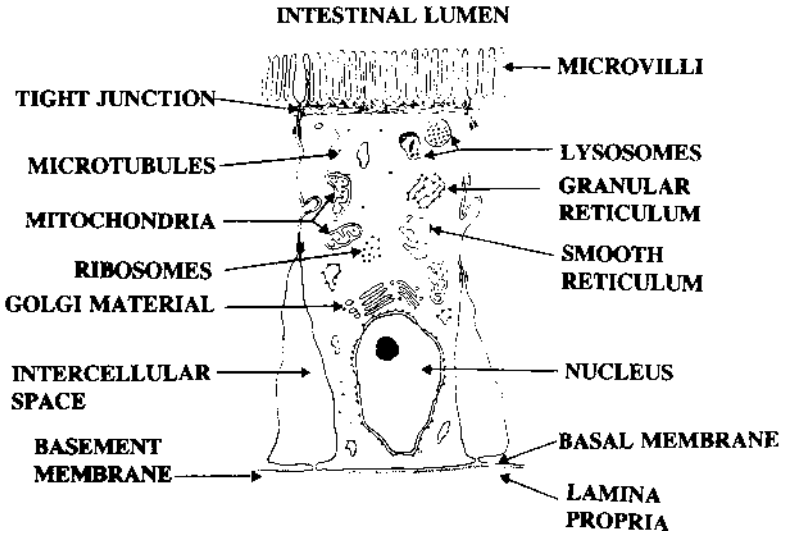


Fig. 5 Diagrammatic sketch of the intestinal absorptive cell. (Modified from Ref. 8.)

3–10 cells thick and separates the mucosa from the submucosa. The lamina propria, the section between the muscularis mucosa and the intestinal epithelia, represents the subepithelial connective tissue space and, together with the surface epithelium, forms the villi structure. The lamina propria contains a variety of cell types, including blood and lymph vessels and nerve fibers. Molecules to be absorbed must penetrate into this region to gain access to the bloodstream.

The third mucosal layer is that lining the entire length of the small intestine and which represents a continuous sheet of epithelial cells. These epithelial cells (or enterocytes) are columnar in shape, and the luminal cell membrane, upon which the microvilli reside, is called the apical cell membrane. Opposite this membrane is the basal (or basolateral) plasma membrane, which is separated from the lamina propria by a basement membrane. A sketch of this cell is shown in Fig. 5. The primary function of the villi is absorption.

The microvilli region has also been referred to as the striated or "brush" border. It is in this region where the process of absorption is initiated. In close contact with the microvilli is a coating of fine filaments composed of weakly acidic sulfated mucopolysaccharides. It has been suggested that this region may serve as a relatively impermeable barrier to substances within the gut such as bacteria and other foreign materials. In addition to increasing the effective luminal surface area, the microvilli region appears to be an area of important biochemical activity.

The surface epithelial cells of the small intestine are renewed rapidly and regularly. It takes about two days for the cells of the duodenum to be renewed completely. As a result of its rapid renewal rate, the intestinal epithelium is susceptible to various factors that may influence proliferation. Exposure of the intestine to ionizing radiation and cytotoxic drugs (such as folic acid antagonists and colchicine) reduces the cell renewal rate.

C. Large Intestine

The large intestine, often referred to as the colon, has two primary functions: the absorption of water and electrolytes and the storage and elimination of fecal material. The large intestine, which has a greater diameter than the small intestine ($\sim$ 6 cm), is connected to the latter at the ileocecal junction. The wall of the ileum at this point has a thickened muscular coat called the ileocecal sphincter, which forms the ileocecal valve, whose principal function is to prevent backflow of fecal material from the colon into the small intestine. From a functional point of view the large intestine may be divided into two parts. The proximal half, concerned primarily with absorption, includes the cecum, ascending colon, and portions of the transverse colon. The distal half, concerned with storage and mass movement of fecal matter, includes part of the transverse and descending colon, the rectum, and anal regions, terminating at the internal anal sphincter (see Fig. 1).

In humans, the large intestine usually receives about 500 mL of fluid-like food material (chyme) per day. As

Copyright © 2002 Marcel Dekker, Inc.

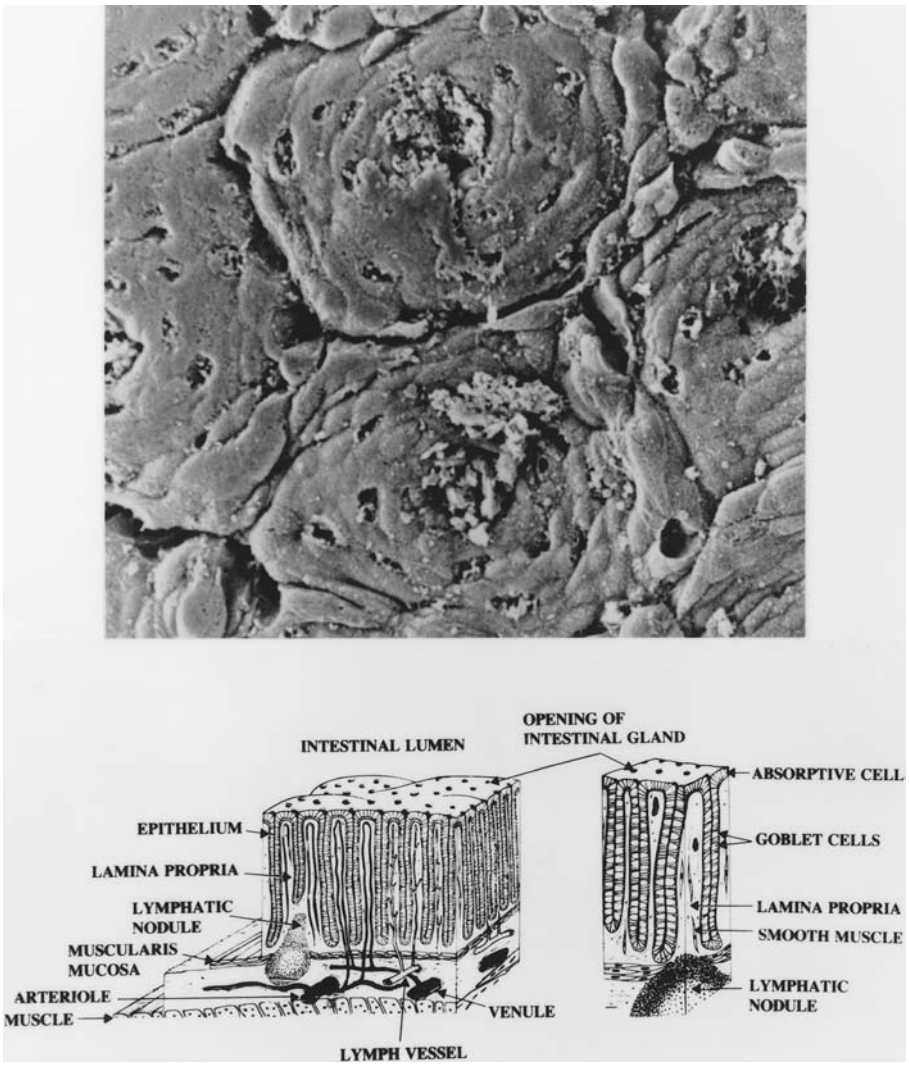


Fig. 6 (A) Scanning electron micrograph of the luminal surface of the large intestine (transverse colon; magnification ×60). (From Ref. 9.) (B) Schematic diagram showing a longitudinal cross section of the large intestine. (C) Enlargement of cross section shown in B. (A and B modified from Ref. 10.)

this material moves distally through the large intestine, water is absorbed, producing a viscous and finally a solid mass of matter. Due to efficient water absorption, of the 500 mL normally reaching the large intestine, approximately 80 mL are eliminated from the gut as fecal material.

Structurally, the large intestine is similar to the small intestine, although the luminal surface epithelium of the former lacks villi. The muscularis mucosa, as in the small intestine, consists of inner circular and outer longitudinal layers. Figure 6 illustrates a photomicrograph and diagrammatic sketches of this region.

D. Pathways of Drug Absorption

Once a drug molecule is in solution, it has the potential to be absorbed. Whether or not it is in a form available for absorption depends on the physicochemical characteristics of the drug (i.e., its inherent absorbability) and the characteristics of its immediate environment (e.g., pH, the presence of interacting materials, and the local properties of the absorbing membrane). Assuming that there are no interfering substances present to impede absorption, the drug molecule must come in contact with the absorbing membrane. To accomplish this, the drug molecule must diffuse from the gastro-

Copyright © 2002 Marcel Dekker, Inc.

intestinal fluids to the membrane surface. The most appropriate definition of drug absorption is the penetration of the drug across the intestinal "membrane" and the appearance of the unchanged form in the blood draining the GIT. The latter blood flow will drain into the portal circulation on the way to the liver. A clear distinction must be made between *absorbed* drug and *bioavailable* drug. The former was defined above; the latter refers to the appearance of unaltered drug in the systemic circulation (i.e., beyond the liver). There are two important points to this definition. First, it is often assumed that drug disappearance from the GI fluids represents absorption. This is true only if disappearance from the gut represents appearance in the bloodstream. This may not be the case, for example, if the drug degrades in GI fluids or if it is metabolized within the intestinal cells. Second, the term intestinal "membrane" is rather misleading, since this membrane is not a unicellular structure but a number of unicellular membranes parallel to one another. In fact, relative to the molecular size of most drug molecules, the compound must diffuse a considerable distance. Thus, for a drug molecule to reach the blood, it must penetrate the mucous layer and brush border covering the GI lumen, the apical cell surface, the fluids within this cell, the basal membrane, the basement membrane, the tissue region of the lamina propria, the external capillary membrane, the cytoplasma of the capillary cell, and finally, the inner capillary membrane. Therefore, when the expression intestinal "membrane" is used, we are discussing a barrier to absorption consisting of several distinct unicellular membranes and fluid regions bounded by these membranes. Throughout this chapter the term "intestinal membrane" will be used in that sense.

For a drug molecule to be absorbed from the GIT and gain access to the portal circulation (on its way to the liver), it must effectively penetrate all the regions of the intestine just cited. There are primarily three factors governing this absorption process once a drug is in solution: the physicochemical characteristics of the molecule, the properties and components of the GI fluids, and the nature of the absorbing membrane. Although penetration of the intestinal membrane is obviously the first part of absorption, we discuss the factors controlling penetration extensively in the following section. At this point, assume that the drug molecule has penetrated most of the barriers in the intestine and has reached the lamina propria region. Once in this region the drug may either diffuse through the blood capillary membrane and be carried away in the bloodstream or penetrate the central lacteal and reach the lymph. These functional units of the villi are illustrated in Fig. 4. Most drugs, if not all, reach the systemic circulation via the bloodstream of the capillary network in the villi. The primary reason for this route being dominant over lymphatic penetration is the fact that the villi are highly and rapidly perfused by the bloodstream. Blood flow to the GIT in humans is approximately 500–1000 times greater than lymph flow. Thus, although the lymphatic system is a potential route for drug absorption from the intestine, under normal circumstances it will account for only a small fraction of the total amount absorbed. The major exception to this rule will be drugs (and environmental toxicants, such as insecticides, etc.) that have extremely large oil/water partition coefficients (greater than about 10^5 or log partition of 5). By increasing lymph flow or, alternatively, reducing blood flow, drug absorption via the lymphatic system may become more important. The capillary and lymphatic vessels are rather permeable to most low molecular weight and lipid-soluble compounds. The capillary membrane, however, represents a more substantial barrier than the central lacteal to the penetration of very large molecules or combinations of molecules as a result of frequent separations of cells along the lacteal surface. The lymphatic route of movement is important, for example, for the absorption of triglycerides or emulsified fats in the form of chylomicrons, which are rather large (about 0.5 μm in diameter).

III. PHYSICOCHEMICAL FACTORS GOVERNING DRUG ABSORPTION

A. Oil/Water Partition Coefficient and Chemical Structure

As a result of extensive experimentation it has been found that the primary physicochemical properties of a drug influencing its passive absorption into and across biological membranes are its oil/water partition coefficient ($K_{o/w}$), extent of ionization in biological fluids determined by its pK_a value and pH of the fluid in which it is dissolved, and its molecular weight or volume. The fact that these variables govern drug absorption is a direct reflection of the nature of biological membranes. The cell surface of biological membranes (including those lining the entire GIT) is lipid in nature; as a result, one may view penetration into the intestinal cells as a competition for drug molecules between the aqueous environment, on one hand, and the lipid-like materials of the membrane, on the other. To a large extent, then, the principles of solution

Copyright © 2002 Marcel Dekker, Inc.

chemistry and the molecular attractive forces to which the drug molecules are exposed will govern movement from an aqueous phase to the lipid-like phase of the membrane.

At the turn of this century, Overton examined the osmotic behavior of frog sartorius muscle soaked in a buffer solution containing various dissolved organic compounds. He reasoned that if the solute entered the tissue, the weight of the muscle would remain essentially unchanged, whereas loss of weight would indicate an osmotic withdrawal of fluid and hence impermeability to the solute. He noted that, in general, the tissue was most readily penetrated by lipid-soluble compounds and poorly penetrated by lipid-insoluble substances. Overton was one of the first investigators to illustrate that compounds penetrate cells in the same relative order as their oil/water partition coefficients, suggesting the lipid nature of cell membranes. Using animal or plant cells, other workers provided data in support of Overton's observations. The only exception to this general rule was the observation that very small molecules penetrate cell membranes faster than would be expected based on their $K_{o/w}$ values. To explain the rapid penetration of these small molecules (e.g., urea, methanol, formamide), it was suggested that cell membranes, although lipid in nature, were not continuous but interrupted by small water-filled channels or "pores"; such membranes are best described as being lipid-sieve membranes. As a result, one could imagine lipid-soluble molecules readily penetrating the lipid regions of the membrane while small water-soluble molecules pass through the aqueous "pores". Fordtran et al. [11] estimated the effective pore radius to be 7–8.5 and 3–3.8 Å in human jejunum and ileum, respectively. There may be a continuous distribution of pore sizes—a smaller fraction of larger ones and a greater fraction of smaller pores.

Our knowledge of biological membrane ultrastructure has increased considerably over the years as a result of rapid advances in instrumentation. Although there is still controversy over the most correct biological membrane model, the concept of membrane structure presented by Davson and Danielli of a lipid bilayer is perhaps the one best accepted [12,13]. The most current version of that basic model, illustrated in Fig. 7, is referred to as the "fluid mosaic" model of membrane structure. This model is consistent with what we have learned about the existence of specific ion channels and receptors within and along surface membranes.

Table 2 summarizes some literature data supporting the general dependence of the rate of absorption on

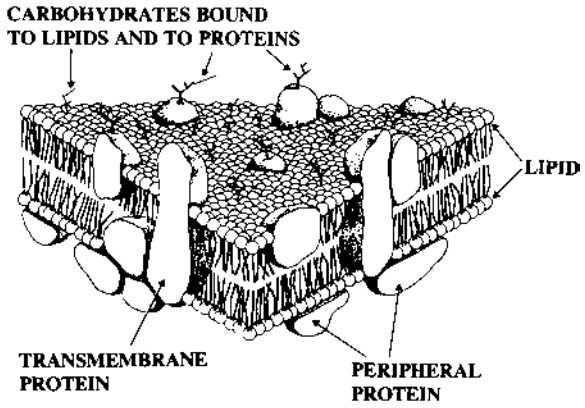


Fig. 7 Diagrammatic representation of the fluid mosaic model of the cell membrane. The basic structure of the membrane is that of a lipid bilayer in which the lipid portion (long tails) points inward and the polar portion (round "head") points outward. The membrane is penenetrated by transmembrane (or integral) proteins. Attached to the surface of the membrane are peripheral proteins (inner surface) and carbohydrates that bind to lipid and protein molecules (outer surface). (Modified from Ref. 14.)

$K_{o/w}$ as measured in the rat intestine [15,16]. As with other examples that are available, as $K_{o/w}$ increases, the rate of absorption increases. One very extensive study [17–19] has examined in depth the physicochemical factors governing nonelectrolyte permeability for several hundred compounds. This study employed an in vitro preparation of rabbit gallbladder, an organ whose mucosal surface is lined by epithelial cells. The

Table 2 Influence of Oil/Water Partition Coefficient ($K_{o/w}$) on Absorption from the Rat Intestine

Compound	$K_{o/w}$	Percentage absorbed
	Olive Oil/Water	
Valeramide	0.023	85
Lactamide	0.00058	67
Malonamide	0.00008	27
	Chloroform/Water	
Hexethal	>100	44
Secobarbital	50.7	40
Pentobarbital	28.0	30
Cyclobarbital	13.9	24
Butethal	11.7	24
Allybarbituric acid	10.5	23
Phenobarbital	4.8	20
Aprobarbital	4.9	17
Barbital	0.7	12

Copyright © 2002 Marcel Dekker, Inc.

method used to assess solute permeability is based upon measurement of differences in electrical potential (streaming potentials) across the membrane. The more permeable the compound, the smaller the osmotic pressure it exerts and the smaller the osmotic fluid flow it produces in the opposite direction; this results in a small potential difference. If the compound is impermeable, it produces a large osmotic pressure and osmotic fluid flow, resulting in a large potential difference. Experimentally, one exposes the mucosal membrane surface to a buffer solution containing a reference compound to which the membrane is completely impermeable and measures the resulting potential difference. This is followed by exposing the same membrane to a solution of a test compound and again measuring the resulting potential difference. The ratio of the potential difference of the test compound to that of the reference compound is referred to as the reflection coefficient (σ). The reflection coefficient is a measure of the permeability of the test compound relative to a reference solute with the particular membrane being used. The less permeable the test compound, the closer the reflection coefficient approaches one; the more permeable the test compound, the closer the coefficient approaches zero.

Using this method, Wright and Diamond [17–19] were able to reach a number of important conclusions concerning patterns of nonelectrolyte permeability. In general, membrane permeability of a solute increases with $K_{o/w}$, supporting previous findings mentioned earlier. The two classes of exceptions to this pattern are highly branched compounds, which penetrate the membrane more slowly than would be expected based on their $K_{o/w}$, and smaller polar molecules, which penetrate the membrane more readily than would be expected based on their $K_{o/w}$. The latter observation has been noted by other workers, and, as noted earlier, it has resulted in the development of the lipid-sieve membrane concept whereby one envisions aqueous "pores" in the membrane surface. The authors postulate that these small, polar, relatively lipid-insoluble compounds penetrate the membrane by following a route lined by the polar groupings of membrane constituents (i.e., localized polar regions). This concept is an attractive structural explanation of what have been referred to as "pores." The accessibility of this route would be limited primarily by the molecular size of the compound as a result of steric hindrance. In fact, it is the first one or two members of a homologous series of compounds that are readily permeable, but beyond these members it is primarily $K_{o/w}$ that dictates permeability. Table 3 illustrates this effect for several members of various homologous series. Recall that the smaller the σ, the more permeable the compound. In each instance, permeability decreases after the first member, reaches a minimum, and then increases again.

Table 3 Influence of Chain Length on Membrane Permeability within Several Homologous Series[a]

Compound	Reflection coefficient, σ
Urea	0.29 ↑
Methyl urea	0.54 |
Ethyl urea	0.92 |
Propyl urea	0.93 -
Butyl urea	0.70 ↓
Malononitrile	0.09 ↑
Succinonitrile	0.30 -
Glutaronitrile	0.21 ↓
Methylformamide	0.28 ↑
Methylacetamide	0.51 -
Methylproprionamide	0.22 ↓

[a] The reflection coefficient σ is defined in the text. The direction of the arrows indicates an increase in permeability from the least permeable member of the series.

The other anomalous behavior was the smaller-than-expected permeability of highly branched compounds. This deviation has been explained on the basis that membrane lipids are subject to a more highly constrained orientation (probably a parallel configuration of hydrocarbon chains of fatty acids) than are those in a bulk lipid solvent. As a result, branched compounds must disrupt this local lipid structure of the membrane and will encounter greater steric hindrance than will a straight-chain molecule. This effect with branched compounds is not adequately reflected in simple aqueous-lipid partitioning studies (i.e., in the $K_{o/w}$ value).

With the exception of rather small polar molecules, the majority of compounds, including drugs, appear to penetrate biological membranes via a lipid route. As a result, the membrane permeability of most compounds is dependent on $K_{o/w}$. The physicochemical interpretation of this general relationship is based on the atomic and molecular forces to which the solute molecules are exposed in the aqueous and lipid phases. Thus, the ability of a compound to partition from an aqueous to a lipid phase of a membrane involves the balance between solute-water and solute-membrane intermolecular forces. If the attractive forces of the solute-water interaction are greater than those of the solute-membrane interaction, membrane permeability will be relatively poor and vice versa. In examining the permeability of a homologous series of compounds

Copyright © 2002 Marcel Dekker, Inc.

and, therefore, the influence of substitution or chain length on permeability, one must recognize the influence of the substituted group on the intermolecular forces in aqueous and membrane phases (e.g., dipole-dipole, dipole-induced dipole, or van der Waals forces). The membrane permeabilities of the nonelectrolytes studied appear to be largely determined by the number and strength of hydrogen bonds the solute can form with water. Thus, nonelectrolyte permeation is largely a question of physical organic chemistry in aqueous solution. Table 4 summarizes some of the interesting findings of Diamond and Wright with respect to the influence of substituent groups on membrane permeation. These data have been interpreted based on the solutes' ability to form hydrogen bonds with water.

Within a homologous series of compounds the first few small members are readily permeable, due to the polar route of membrane penetration. Permeability decreases for the next several members (i.e., σ increases) and then increases as the carbon chain length increases. The regular influence of chain length on permeability is a result not of increased solubility in the lipid phase of the membrane but of the unique interaction of hydrocarbon chains with water. The nonpolar hydrocarbon molecules are surrounded by a local region of water that has a more highly ordered structure than bulk water. This "iceberg" structure of water results in increased $K_{o/w}$ and membrane permeability as the carbon chain length is increased due to the compound being "pushed out" of the aqueous phase by the resulting gain in entropy.

Based upon a review of the physical chemical properties of marketed drugs, Lipinski and coworkers have proposed an empirical "rule of 5" (20). This rule may help pharmaceutical scientists in reaching an early decision about the potential candidacy for further development of a new chemical entity. The rule states that a chemical candidate is likely to display poor absorption or poor membrane permeability if

1. There are more than 5 hydrogen bond donors
2. There are more than 10 hydrogen bond acceptors
3. The molecular weight is greater than 500
4. Log $K_{o/w}$ is greater than 5
5. The above rules only apply to compounds that undergo passive membrane transport

There have been several, albeit limited, attempts to develop quantitative, structure-activity relationships in drug absorption (e.g., Ref. 21). Such relationships could prove extremely useful in the early stages of drug design in order to produce optimum absorption characteristics. Another very practical approach involves "data mining," whereby large data bases are examined to characterize the properties of those compounds exhibiting good absorption.

B. pK_a and pH

Most drug molecules are either weak acids or bases that will be ionized to an extent determined by the compound's pK_a and the pH of the biological fluid in which it is dissolved. The importance of ionization in drug absorption is based on the observation that the nonionized form of the drug has a greater $K_{o/w}$ than the ionized form and, since $K_{o/w}$ is a prime determinant of membrane penetration, ionization would be expected to influence absorption. The observation that pH influences the membrane penetration of ionizable drugs is not a recent finding. At the turn of the previous century Overton was able to relate pH to the rate of penetration of various alkaloids into cells, and he noted the resulting influence in toxicity. Other investigators have made similar observations with respect to the influence of pH on the penetration of alkaloids through the conjunctival and mammalian skin [22,23]. The rate of penetration of these weak bases is enhanced by alkalinization due to a greater fraction of the nonionized species being present. Travell [24] examined the influence of pH on the absorption of several alkaloids from the stomach and intestine of the cat. After ligation of the proximal and distal ends of the stomach of an anesthetized cat, a 5.0 mg/kg solution of strychnine at pH 8.5 produced death within 24 minutes; however, the same dose at pH 1.2 produced no toxic response. Identical results were found with nicotine, atropine, and cocaine. The same trend was also seen when the drug solution was instilled into ligated intestinal segments and after oral administration (via stomach tube) to ambulatory animals. These results indicate that alkaloids, which are weak bases, will be more rapidly absorbed in the nonionized form (i.e., at high fluid pH) compared to the ionized form (low pH). This fundamental observation has sometimes been overlooked in oral acute drug toxicity studies.

In 1940 Jacobs [25] made use of the Henderson-Hasselbalch equation to relate pH and pK_a to membrane transport of ionizable compounds. Extensive experimentation by a group of investigators in the early 1950s [16,26–30] quantitated many of the aforementioned observations concerning the influence of

Copyright © 2002 Marcel Dekker, Inc.

pH and pK_a on drug absorption from the GIT. These studies have resulted in the so-called "pH partition hypothesis." In essence, this hypothesis states that ionizable compounds penetrate biological membranes primarily in the nonionized form (i.e., nonionic diffusion). As a result, acidic drugs should best be absorbed from acidic solutions where $pH < pK_a$, while basic compounds would best be absorbed from alkaline solutions where $pH > pK_a$. The data in Table 5 illustrate this principle [31].

The investigators noted some inconsistencies in their data, however, as some compounds (e.g., salicylic acid) that were essentially completely ionized in the buffer solution were nevertheless rapidly absorbed. To explain these exceptions it was suggested that there was a "virtual membrane pH" (~5.3), different from the bulk pH of the buffer solution, which was the actual pH determining the fraction of drug nonionized and hence dictating the absorption pattern. Although there may indeed be an effective pH at the immediate surface of the intestinal membrane different from the pH of solutions bathing the lumen, there is overwhelming experimental evidence indicating that many drugs in the ionic form may be well absorbed. Over the years there has been an unqualified acceptance of the pH partition hypothesis, and, as a result, many texts and considerable literature on drug absorption indicate that acidic drugs are best absorbed from the acidic gastric fluids of the stomach and basic drugs best absorbed from the relatively more alkaline intestinal fluids. If all other conditions were the same, the nonionized form of the solution would be more rapidly absorbed than the ionized form. However, conditions along the GIT are not uniform, and hence most drugs, whether ionized or nonionized (i.e., regardless of pH), are best absorbed from the small intestine as a result of the large absorbing surface area of this region. A good example to illustrate this point is presented in Table 6. There are three important comparisons that should be made in examining these data:

1. By comparing gastric absorption at pH 3 and pH 6 where surface area and factors other than pH are constant, one sees that the general principle is supported; acid drugs are more rapidly absorbed from acidic solution, whereas basic drugs are more rapidly absorbed from relatively alkaline solution.
2. At the same pH (i.e., pH 6) acidic and basic drugs are more rapidly absorbed from the intestine compared to the stomach, by virtue of the larger intestinal surface area.
3. Acidic drugs are more rapidly absorbed from the intestine (pH 6), although there is substantial ionization, compared to the rate of gastric absorption, even at a pH where the drug is in a far more acidic solution (pH 3). Again, this is primarily a result of surface area differences.

Interestingly, in an analysis of the original data used in developing the pH partition hypothesis, Benet [32] has shown that these data support the findings in point 3 above. The pH partition hypothesis provides a useful guide in predicting general trends in drug absorption, and it remains an extremely useful concept. There are numerous examples illustrating the general relationship among pH, pK_a, and drug absorption developed in that hypothesis. The primary limitation of this concept is the assumption that only nonionized drug is absorbed, when in fact the ionized species of some compounds can be absorbed, albeit at a slower rate. There is also the presence of unstirred water layers at the epithelial membrane surface, which can alter the rate of drug diffusion. Furthermore, the hypothesis is based on data obtained from drug in solution. In a practical sense, there are other considerations that may also govern the pattern of drug absorption and these include dissolution rate from solid dosage forms, the large intestinal surface area, and the relative residence times of the drug in different parts of the GIT. These factors are discussed below. In general then, drug absorption in humans takes place primarily from the small intestine regardless of whether the drug is a weak acid or base.

C. Mechanisms of Drug Absorption

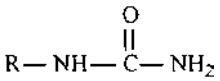
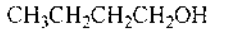
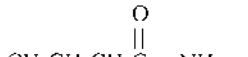
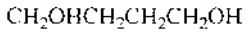
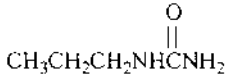
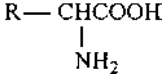
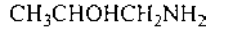
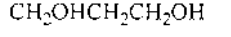
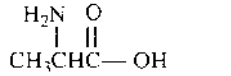
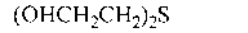
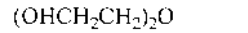
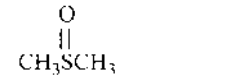
A thorough discussion of the mechanisms of absorption is provided in Chapter 4. Water-soluble vitamins (B_2, B_{12}, and C) and other nutrients (e.g., monosaccharides, amino acids) are absorbed by specialized mechanisms. With the exception of a number of antimetabolites used in cancer chemotherapy, L-dopa, and certain antibiotics (e.g., aminopenicillins, aminocephalosporins), virtually all drugs are absorbed in humans by a passive diffusion mechanism. Passive diffusion indicates that the transfer of a compound from an aqueous phase through a membrane may be described by physicochemical laws and by the properties of the membrane. The membrane itself is passive in that it does not partake in the transfer process but acts as a simple barrier to diffusion. The driving force for diffusion across the membrane is the concentration gradient (more correctly, the activity gradient) of the compound across that membrane. This mechanism of

Copyright © 2002 Marcel Dekker, Inc.

Table 4 Influence of Chemical Substitution on the Membrane Permeability of Several Series of Nonelectrolytes

Substituent group	Influence on membrane permeability	Compound	Example	σ^a
Oxygen and nitrogen functional groups				
Alcoholic hydroxyl group (-OH)	(a) At any given chain length, permeability decreases as the number of -OH groups increases	*n*-Propanol	$CH_3CH_2CH_2OH$	0.02 ↑
		1,2-Propanediol	$CH_3CHOHCH_2OH$	0.84
		Glycerol	$CH_2OHCHOHCH_2OH$	0.95
	(b) Intramolecular H-bonds formed between adjacent -OH groups result in greater permeability compared to the same compound with nonadjacent -OH groups due to decreased H-bond formation with water	2,3-Butanediol	$CH_3CHOHHCHOHCH_3$	0.74 ↑
		1,3-Butanediol	$CH_3CHOHCH_2CH_2OH$	0.77
		1,4-Butanediol	$CH_2OHCH_2CH_2CH_2OH$	0.86
Ether group (-O-)	Has less of an influence than an -OH group in decreasing permeability	*n*-Propanol	$CH_3CH_2CH_2OH$	0.02 ↑
		Ethyleneglycol-methyl ether	$CH_3–O–CH_2CH_2OH$	0.15
		1,2-Propanediol	$CH_3CHOHCH_2OH$	0.84
Carbonyl group Ketone (–C=O) Aldehyde (–HC=O)	Has less of an influence than an -OH gourp in decreasing permeability; difficulty in measuring permeability of these compounds per se as many are unstable in solution forming diols and enolic tautomers	Acetone	$CH_3\overset{\overset{O}{\|}}{C}CH_3$	0.01 ↑
		2-Propanol	$CH_3CHOHCH_3$	0.10
		2-5-Hexanedione	$CH_3\overset{\overset{O}{\|}}{C}CH_2CH_2\overset{\overset{O}{\|}}{C}CH_3$	0.00 ↑
		2,5-Hexanediol	$CH_3CHOHCH_2CH_2CHOHCH_3$	0.59
Ester group $(-\overset{\overset{O}{\|}}{C}-O-)$	Has less of an influence than an -OH group in decreasing permeability	1,2-Propanediol-1-acetate	$CH_3\overset{\overset{O}{\|}}{C}-O-CH_2CHOHCH_3$	0.31 ↑
		1,5-Pentanediol	$CH_2OH(CH_2)_3CH_2OH$	0.71
Amide group $-\overset{\overset{O}{\|}}{C}-NH_2$	Causes a greater decrease in permeability than any of the above groups	*n*-Propanol	$CH_3CH_2CH_2OH$	0.02 ↑
		Acetone	$CH_3\overset{\overset{O}{\|}}{C}CH_3$	0.08
		Ethyleneglycol-methyl ether	$CH_3-O-CH_2CH_2OH$	0.15
		Proprionamide	$CH_3CH_2\overset{\overset{O}{\|}}{C}NH_2$	0.66

Copyright © 2002 Marcel Dekker, Inc.

Urea derivatives 	These compounds have lower permeability than amides with the same numer of carbons and are about as impermeable as the corresponding dihydroxyl alcohols	*n*-Butanol		0.01 ↑
		n-Butryamide		0.42
		1,4-Butanediol		0.86
		n-Propyl urea		0.89 ↑
α-Amino acids 	These compounds have the lowest $K_{o/w}$ values of all orgainc molecules and are essentially impermeable due to large dipole-dipole interactions with water	Proprionamide		0.66 ↑
		1-Amino-2-propanol		0.89
		1,3-Propanediol		0.92
		Alanine		1.06 ↑
Sulfur functional groups				
Sulfur replacement of oxygen	(a) Sulfur compounds have greater $K_{o/w}$ values and permeate membranes more readily than the corresponding oxygen compound; this is a result of poor H-bond formation between sulfur and water compared to the oxygen analog	1-Thioglycerol		0.69 ↑
		Glycerol		0.95
		Thiodiglycol		0.71
		Diethylene glycol		0.92 ↑
	(b) Sulfoxides ($R_2S{=}O$) are less permeable than the corresponding ketone ($R_2C{=}O$), due to stronger H-bond formation with water	Acetone		0.01 ↑
		Dimethyl sulfoxide		0.92 ↑

[a] The reflection coefficient σ is defined in the text. The direction of the arrows indicates an increase in permeability.
Source: Refs. 17–19.

Copyright © 2002 Marcel Dekker, Inc.

Table 5 Influence of pH on Drug Absorption from the Small Intestine of the Rat[a]

Drug	pK_a	Percentage absorbed pH 4	pH 5	pH 7	pH 8
Acids					
5-Nitrosalicylic acid	2.3	40	27	<2	<2
Salicylic acid	3.0	64	35	30	10
Acetylsalicylic acid	3.5	41	27	—	—
Benzoic acid	4.2	62	36	35	5
Bases					
Aniline	4.6	40	48	58	61
Aminopyrine	5.0	21	35	48	52
p-Toluidine	5.3	30	42	65	64
Quinine	8.4	9	11	41	54

[a] Drug buffer solutions were perfused through the in situ rat intestine for 30 min and percentage of drug absorbed was determined from four subsequent 10 min samples of the buffer solution.

Table 6 Influence of pH on Drug Absorption from the Stomach and Intestine of the Rat[a]

Drug	Apparent first-order absorption rate constant (min^{-1}) pH 3	Stomach pH 6	Intestine pH 6
Acids			
Salicylic acid	0.015	0.0053	0.085
Barbital	0.0029	0.0026	0.037
Sulfaethidole	0.004	0.0023	0.022
Bases			
Prochlorperazine	<0.002	0.0062	0.030
Haloperidol	0.0028	0.0041	0.028
Aminopyrine	<0.002	0.0046	0.022

[a] Drug buffer solutions were placed into the GIT of an in situ rat preparation. The apparent first-order absorption rate constants are based upon drug disappearance from the buffer solution.

membrane penetration may be described mathematically by Fick's first law of diffusion, which has been simplified by Riggs [33] and discussed by Benet [32].

$$\left(\frac{dQ_b}{dt}\right)_{g\to b} = D_m \cdot A_m \cdot R_{m/aq}\left(\frac{C_g - C_b}{\Delta X_m}\right) \quad (1)$$

The derivative on the left side of the equation represents the rate of appearance of drug in the blood (amount/time) when the drug diffuses from the gut fluids (g) to the blood (b). The expression reads, the rate of change of the quantity (Q) entering the blood stream. The other symbols have the following meanings (and units): D_m, the diffusion coefficient of the drug through the membrane (area/time); A_m, the surface area of the absorbing membrane available for drug diffusion (area); $R_{m/aq}$, the partition coefficient of the drug between the membrane and aqueous gut fluids (unitless); $C_g - C_b$, the concentration gradient across the membrane, representing the difference in the effective drug concentration (i.e., activity) in the gut fluids (C_g) at the site of absorption and the drug concentration in the blood (C_b) at the site of absorption (amount/volume); and ΔX_m, the thickness of the membrane (length). This equation nicely explains several of the observations discussed previously. Thus, rate of drug absorption is directly dependent on the membrane surface area available for diffusion, indicating that one would expect more rapid absorption from the small intestine compared to the stomach. Furthermore, the greater the membrane aqueous fluid partition coefficient ($R_{m/aq}$), the more rapid the rate of absorption, supporting the previous discussion indicating the dependence of absorption rate on $K_{o/w}$. We know that pH will produce a net effect on absorption rate by altering several of the parameters in Eq. (1). As the pH for a given drug will determine the fraction nonionized, the value of $R_{m/aq}$ will change with pH, generally increasing as the fraction nonionized increases. Depending on the relative ability of the membrane to permit the diffusion of the nonionized and ionized forms, C_g will be altered appropriately. Finally, the value of D_m may be different for the ionized and nonionized forms of the compound. For a given drug and membrane and under specified conditions, Eq. (1) is made up of a number of constants that may be incorporated into a large constant (P), referred to as the permeability coefficient:

$$\left(\frac{dQ_b}{dt}\right)_{g\to b} = P \cdot (C_g - C_b) \quad (2)$$

where P incorporates D_m, A_m, $R_{m/aq}$, and ΔX_m and has units of volume/time, which is analogous to a flow or clearance term. Since the volume into which the drug may distribute from the blood is large compared to the gut fluid volume and since the rapid circulation of blood through the GIT continually moves absorbed drug away from the site of absorption, $C_g \gg C_b$. This is often referred to as a "sink condition," indicating a relatively small drug concentration in the bloodstream at the absorption site. As a result, Eq. (2) may be simplified:

$$\left(\frac{dQ_b}{dt}\right)_{g\to b} \cong P \cdot C_g \quad (3)$$

Copyright © 2002 Marcel Dekker, Inc.

Equation (3) is in the form of a differential equation describing a first-order kinetic process, and, as a result, drug absorption generally adheres to first-order kinetics. The rate of absorption should increase directly with an increase in drug concentration in the GI fluids.

Figure 8 illustrates the linear dependence of absorption rate on concentration for several compounds placed into the in situ rat intestine. The slopes of these lines represent the constant (P) for absorption in Eq. (3). Alternatively, one may express these data as the percentage absorbed per unit of time as a function of concentration or amount. Several examples illustrating such an analysis are listed in Table 7. As can be seen, the percentage absorbed in any given period is independent of concentration, indicating that these compounds are absorbed by a passive diffusion or first-order kinetic process over the concentration ranges studied. Similar studies by other investigators employing an in situ rat intestine preparation indicate that several other drugs (those listed in Table 6) are absorbed in a first-order kinetic fashion.

It is far more difficult to establish the mechanism(s) of drug absorption in humans. Most investigators analyze drug absorption data in humans (from blood or urine data) by assuming first-order absorption kinetics. For the most part this assumption seems quite

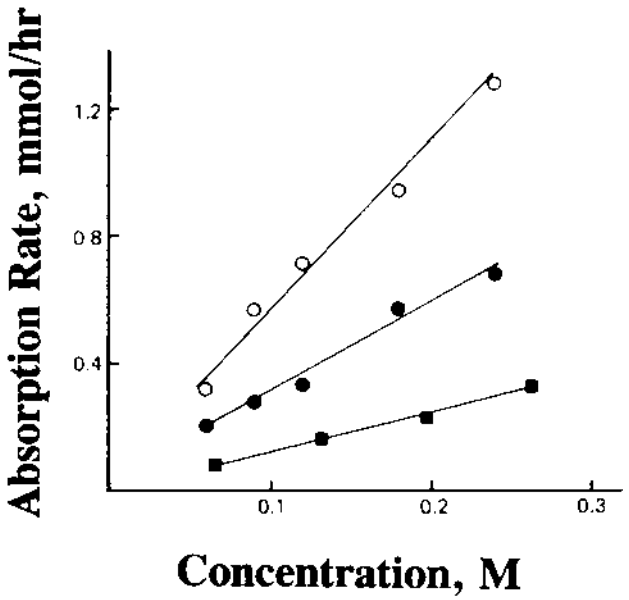


Fig. 8 Influence of concentration on the rate of absorption from the in situ rat intestine. The linear dependence of absorption rate on concentration suggests an apparent first-order absorption process over the range studied. Absorption rates have been calculated from the data in Ref. 15 and the straight lines are from linear regression of the data. Key: (○) erythritol; (●) urea; (■) malonamide.

Table 7 Influence of Concentration on the Absorption of Various Solutes from the In Situ Rat Intestine

Compound	Concentration (mM)	Percentage absorbed	Compound	Concentration (mM)	Percentage absorbed
Urea[a]	60	20.9	Salicylic acid[b]	1	12
	90	19.0		2	12
	120	17.0		10	13
	180	20.0			
	240	17.8	Aniline[b]	1	44
				10	43
Erythritol[a]	60	54.1			
	90	65.0	Benzoic acid[b]	1	12
	120	62.2		2	12
	180	54.4		10	13
	240	55.5			
			Quinine[b]	1	20
Malonamide[a]	66	16.9		10	20
	132	16.8		0.1	58
	198	16.5	Aniline[c]		
	264	18.4		1	54
				10	59
				20	54

[a] Based on data in Ref. 15.
[b] Based on data in Ref. 16.
[c] Based on data in Ref. 29.

Copyright © 2002 Marcel Dekker, Inc.

valid, and the results of such analyses are consistent with that assumption. As discussed in Chapter 3, one method used to assess the mechanism of drug absorption in humans is based on a pharmacokinetic treatment of blood or urine data and the preparation of log percentage unnabsorbed versus time plots. If a straight-line relationship is found, this is indicative of an apparent first-order absorption process, where the slope of that line represents the apparent first-order absorption rate constant. Some cautions must be taken in the application of this method. Although the overall absorption process in humans for many drugs appears consistent with the characteristics of a first-order kinetic process, there are some questions as to which of the sequential steps in the absorption process is rate-limiting. As discussed in a thorough review of mass transport phenomena [34], the oil/water partition coefficient of a solute ($K_{o/w}$) will govern its movement across a lipid-like membrane as long as the membrane is the predominant barrier to diffusion. However, for such membranes, when the $K_{o/w}$ becomes very large, the barrier controlling diffusion may no longer be the membrane but rather an aqueous diffusion layer surrounding the membrane. Thus, for some molecules, depending on their physicochemical characteristics, the rate-limiting step in membrane transport will be movement through or out of the membrane, while for other compounds the rate-limiting step will be diffusion through an aqueous layer. Our incomplete understanding of drug transport across biological membranes is not that surprising given the complexity of the system and the experimental requirements needed to make unequivocal statements about this process on a molecular level.

The analysis of absorption data in humans has moved away from the more traditional modeling and data fitting techniques [35]. Absorption processes are now more often characterized by a mean absorption (or input) time (i.e., the average amount of time that the drug molecules spend at the absorption site) or by a process called deconvolution. The former analysis results in a single value (similar to absorption half-life) and the latter results in a profile of the absorption process as a function of time (e.g., absorption rate or cumulative amount absorbed vs. time). These approaches offer additional ways of interpreting the absorption process.

Most drugs appear to be absorbed in humans by passive diffusion (linear or first-order kinetics). The predominant pathway taken by most drugs is through the epithelial cell, the *transcellular* route. It is this route that requires the compound to have a reasonable $K_{o/w}$

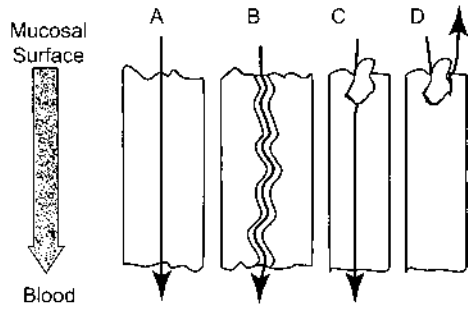


Fig. 9 Schematic representation depicting the movement of molecules from the absorbing (mucosal or apical) surface of the GIT to the basolateral membrane and from there to blood. (A) transcellular movement through the epithelial cell. (B) Paracellular transport via movement between epithelial cells. (C) Specialized carrier-mediated transport into the epithelial cell. (D) Carrier-mediated efflux transport of drug out of the epithelial cell. (Copyright © 2000 Saguaro Technical Press, Inc., used with permission.)

(>1 and $<10^5$). This route is indicated in Fig. 9A as the arrow moving through the cell from the mucosal (or apical) absorbing surface to the basolateral membrane. In contrast, small polar molecules ($K_{o/w} < 1$) may have access to a convoluted route that exists between adjacent epithelial cells but which has a tight junction along the absorbing surface. This route is referred to as being *paracellular* (Fig. 9B). The molecular size cut-off for this route is about 500 Da [36].

Since many essential nutrients (e.g., monosaccharides, amino acids, and vitamins) are water-soluble, they have low oil/water partition coefficients, which would suggest poor absorption from the GIT. However, to ensure adequate uptake of these materials from food, the intestine has developed specialized absorption mechanisms that depend on membrane participation and require the compound to have a specific chemical structure. Since these processes are discussed in Chapter 4, we will not dwell on them here. This carrier transport mechanism is illustrated in Fig. 9C. Absorption by a specialized carrier mechanism (from the rat intestine) has been shown to exist for several agents used in cancer chemotherapy (5-fluorouracil and 5-bromouracil) [37,38], which may be considered "false" nutrients in that their chemical structures are very similar to essential nutrients for which the intestine has a specialized transport mechanism. It would be instructive to examine some studies concerned with riboflavin and ascorbic acid absorption in humans, as these illustrate how one may treat urine data to explore the mechanism of absorption. If a compound is

Copyright © 2002 Marcel Dekker, Inc.

absorbed by a passive mechanism, a plot of amount absorbed (or amount recovered in the urine) versus dose ingested will provide a straight-line relationship. In contrast, a plot of percentage of dose absorbed (or percentage of dose recovered in the urine) versus dose ingested will provide a line of slope zero (i.e., a constant fraction of the dose is absorbed at all doses). If the absorption process requires membrane involvement, the absorption process may be saturated as the oral dose increases, making the process less efficient at larger doses (i.e., there are more drug molecules than sites on the transporter). As a result, a plot of amount absorbed versus dose ingested will be linear at low doses, curvilinear at larger doses, and approach an asymptotic value at even larger doses. One sees this type of relationship for riboflavin and ascorbic acid in Figs. 10A and C, suggesting nonpassive absorption mechanisms in humans [39,40]. This nonlinear relationship is reminiscent of Michaelis-Menten saturable enzyme kinetics, from which one may estimate the kinetic parameters (K_m and V_{max}) associated with the absorption of these vitamins. Figures 10B and D illustrate an alternative plot—percentage absorbed versus dose ingested. For a nonpassive absorption process, the percentage dose absorbed will decrease as the dose increases as a result of saturation of the

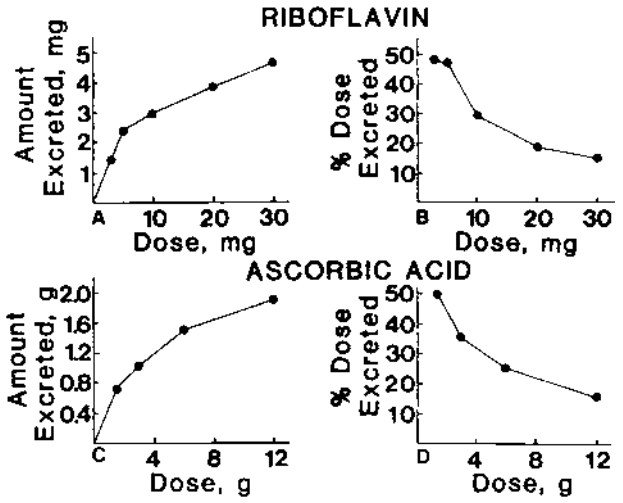


Fig. 10 Urinary excretion of riboflavin (A, B) and ascorbic acid (C, D) in humans as a function of oral dose. Graphs A and C illustrate the nonlinear dependence of absorption on dose, which is suggestive of a saturable specialized absorption process. Graphs B and D represent an alternative graph of the same data and illustrate the reduced absorption efficiency as the dose increases. (Graphs A and C based on data in Ref. 39 and graphs B and D based on data in Ref. 40.)

transport mechanism and of there being a reduction in absorption efficiency. It has been suggested [40] that one means of overcoming the decrease in absorption efficiency is to administer small divided doses rather than large single doses, as illustrated later for ascorbic acid.

Several investigators suggest that L-dopa (L-dihydroxyphenylalanine) absorption may be impaired if the drug is ingested with meals containing proteins [41]. Amino acids formed from the digestion of protein meals, which are absorbed by a specialized mechanism, may competitively inhibit L-dopa absorption if the drug is also transported by the same mechanism. There is evidence (in animals) indicating a specialized absorption mechanism for phenylalanine and L-dopa, and there are data illustrating L-dopa inhibition of phenylalanine and tyrosine absorption in humans [42,43]. L-Dopa appears to be absorbed by the same specialized transport mechanism responsible for the absorption of other amino acids [44]. In a later section several of the complicating factors in L-dopa absorption that influence therapy are discussed.

In addition to some anticancer agents being absorbed by a specialized process in humans (e.g., methotrexate) [45], there is recent evidence to suggest that a similar mechanism exists for the absorption of aminopenicillins (e.g., amoxicillin) [46] and aminocephalosporins (e.g., cefixime) [47]. Absorption of these compounds appears to be linked to cellular amino acid or peptide transporters. Other compounds that have the requisite structural properties may also benefit from those transporting systems (e.g., gabapentin) [48]. This behavior may represent an important observation for the new generation of drugs being developed through the application of biotechnology (e.g., peptides), assuming such compounds are sufficiently stable in the GIT. Calcium channel blockers, such as nifedipine, have been shown to increase the absorption of amoxicillin and cefixime [46,47]. This may result from the role of calcium in the transport process, the inhibition of which (i.e., calcium channel blockers) enhances absorption.

In direct contrast with specialized transport of a drug into the epithelial cell is a process referred to as "efflux transport"; the facilitated movement of a drug out of the cell. This phenomenon explains the resistance of cancerous cells to some chemotherapeutic agents [49]. It is now apparent that this mechanism exists in epithelial cells of numerous organ systems, including the GIT, liver, lung, kidney and brain [50]. There is a cell surface glycoprotein (P-glycoprotein; P-gp), a multidrug-resistance protein (MDR), that is responsible for the efflux mechanism. This protein

Copyright © 2002 Marcel Dekker, Inc.

belongs to a large family of glycosylated membrane proteins referred to as ATP-binding cassette (ABC) transporters. As depicted in Fig. 9D, the cell surface glycoprotein will attach to the drug molecule and escort it out of the cell and back into the gut lumen, hence the term "efflux transporter." The presence of this protein in the GIT has significant implications for drug absorption and bioavailability (e.g., Refs. 51,52). Substrates for this transporter have a diverse range of structures and include compounds such as [51]: anticancer agents (e.g., anthracyclines, taxol, and vinblastine); cardioactive drugs (e.g., digoxin, phenytoin, quinidine, and verapamil); immunosuppressants (e.g., cyclosporine and tacrolimus); erythromycin, quinine, etc. The P-gp protein may be inhibited or induced and the efflux transport process may be saturated. The former is the basis for many potential drug-drug and nutrient-drug interactions, and the latter limits the efficacy of the transporter.

It is difficult to unequivocally implicate P-gp as the mechanism of an interaction resulting in altered absorption. The reason for this is that the absorbing intestinal cells are rich in metabolizing enzymes, especially those responsible for phase I oxidative metabolism, the cytochrome P450 family (CYP450). Within that family the most significant isozyme is CYP3A4, since it is less selective of substrates (a "promiscuous" enzyme), and, as a result, it accounts for about 50% of all drug-drug and nutrient-drug interactions. P-gp and CYP3A4 are present in the same locations, and they both appear to share the same substrates. As a consequence, it is almost impossible to ascribe alterations in absorption to be the exclusive result of either the efflux transporter or the enzyme. This is the case, for example, for cyclosporine, whose absorption is enhanced in the presence of several drugs (e.g., ketoconazole), which may be due to inhibition of P-gp efflux or inhibition of CYP3A4 metabolic activity or both [52,53]. A further complication is the need to consider the systemic effect of altered P-gp activity on drug clearance, an effect that is independent of the absorption process.

The significance of P-gp, however, in affecting absorption and bioavailability of P-gp substrate drugs can be seen in studies in knockout mice that do not have intestinal P-gp. The gene responsible for producing that protein has been knocked out of the genetic repertoire. Those animals evidenced a sixfold increase in plasma concentrations (and AUC, area under the plasma concentration-time curve) of the anticancer drug paclitaxel (Taxol) compared to the control animals [54]. Another line of evidence is the recent report of an interaction between the β-adrenergic blocking agent talinolol and digoxin [55]. Talinolol co-administration resulted in a significant increase in digoxin plasma concentrations (and AUC), and, since talinolol does not appear to be a substrate for CYP3A3, the effect on absorption may be attributed to inhibition of P-gp, which normally modulates digoxin absorption.

IV. PHYSIOLOGICAL FACTORS GOVERNING DRUG ABSORPTION

A. Components and Properties of Gastrointestinal Fluids

The characteristics of aqueous GI fluids to which a drug product is exposed will exert an important influence on what happens to that dosage form in the tract and on the pattern of drug absorption. To appreciate clearly how physiological factors influence drug absorption, one must consider the influence of these variables on the dosage form per se, that is, how these variables influence drug dissolution in the aqueous GI fluids, and finally what influence these variables exert on absorption once the drug is in solution.

One important property of Gl fluids is pH, which varies considerably along the length of the tract. The gastric fluids are highly acidic, usually ranging from pH 1 to 3.5. There appears to be a diurnal cycle of gastric acidity, the fluids being more acidic at night and fluctuating during the day, primarily in response to food ingestion. Gastric fluid pH generally increases when food is ingested and then slowly decreases over the next several hours, fluctuating from pH 1 to about 5 [56]. There is considerable intersubject variation, however, in GI fluid pH, depending on the general health of the subject, the presence of local disease conditions along the tract, types of food ingested, and drug therapy. Upper GI fluid pH appears to be independent of gender.

An abrupt change in pH is encountered when moving from the stomach to the small intestine. Pancreatic secretions (200–800 mL/day) have a high concentration of bicarbonate, which neutralizes gastric fluid entering the duodenum and thus helps regulate the pH of fluids in the upper intestinal region. Neutralization of acidic gastric fluids in the duodenum is important to avoid damage to the intestinal epithelium, prevent inactivation of pancreatic enzymes, and prevent precipitation of bile acids, which are poorly soluble at acid pH. The pH of intestinal fluids gradually increases when moving in the distal direction, ranging from approximately 5.7 in the pylorus to 7.7

Copyright © 2002 Marcel Dekker, Inc.

in the proximal jejunum. The fluids in the large intestine are generally considered to have a pH of between 7 and 8.

Gastrointestinal fluid pH may influence drug absorption in a variety of ways. Since most drugs are weak acids or bases, and since the aqueous solubility of such compounds is influenced by pH, the rate of dissolution from a dosage form, particularly tablets and capsules, is dependent on pH. This is a result of the direct dependence of dissolution rate on solubility, as discussed in Chapter 6. Acidic drugs dissolve most readily in alkaline media and, therefore, will have a greater rate of dissolution in intestinal fluids compared to gastric fluids. Basic drugs will dissolve most readily in acidic solutions, and, thus, the dissolution rate will be greater in gastric fluids compared to intestinal fluids. Since dissolution is a prerequisite step to absorption and is often the slowest process, especially for poorly water-soluble drugs, pH will exert a major influence on the overall absorption process. Furthermore, since the major site of drug absorption is the small intestine, it would seem that poorly soluble basic drugs (e.g., dipyridamole, ketaconazole, and diazepam) must first dissolve in the acidic gastric fluids in order to be well absorbed from the intestine, as the dissolution rate in intestinal fluids will be low. In addition, the disintegration of some dosage forms, depending on their formulation, will be influenced by pH if they contain certain components (e.g., binding agents or disintegrants) whose solubility is pH sensitive. Several studies (e.g., Ref. 57) have indicated that if the specific products being examined were not first exposed to an acidic solution, the dosage form would not disintegrate and thus dissolution could not proceed.

A complication here, however, is noted with those drugs that exhibit a limited chemical stability in either acidic or alkaline fluids. Since the rate and extent of degradation is directly dependent on the concentration of drug in solution, an attempt is often made to retard dissolution in the fluid where degradation is seen. There are preparations of various salts or esters of drugs (e.g., erythromycin) that do not dissolve in gastric fluid and thus are not degraded there but which dissolve in intestinal fluid prior to absorption. A wide variety of chemical derivatives are used for such purposes.

As mentioned previously, pH will also influence the absorption of an ionizable drug once it is in solution, as outlined in the pH partition hypothesis. Most drugs, however, are best absorbed from the small intestine regardless of pK_a and pH. In some instances, especially lower down the GIT, there is the possibility of insoluble hydroxide formation of a drug or insoluble film formation with components of a dosage form, which reduces the extent of absorption of, for example, aluminum aspirin (in chewable tablets) [58,59], and iron [60]. The coadministration of acidic or alkaline fluids with certain drugs may exert an effect on the overall drug absorption process for any of the foregoing reasons.

Moreover, in addition to pH considerations, the Gl fluids contain various materials that have been shown to influence absorption, particularly bile salts, enzymes, and mucin. Bile salts, which are highly surface active, may enhance the rate and/or extent of absorption of poorly water-soluble drugs by increasing the rate of dissolution in the GI fluids. This effect has been noted in in vitro experiments and has been seen as well with other natural surface-active agents (e.g., lysolecithin). Increased absorption of the poorly water-soluble drug griseofulvin after a fatty meal [61,62] may reflect the fact that bile is secreted into the gut in response to the presence of fats and the bile salts which are secreted increase the dissolution rate and absorption of the drug. The contrast agent, iopanoic acid, used in visualizing the gallbladder dissolves more rapidly and is better absorbed from the dog intestine in the presence of bile salts. Studies in rats have indicated enhanced intestinal drug absorption from bile salt solutions; however, the implication of these findings for humans are uncertain. Bile salts may also reduce drug absorption (e.g., neomycin and kanamycin) through the formation of water-insoluble, nonabsorbable complexes.

Since intestinal fluids contain large concentrations of various enzymes needed for digestion of food, it is reasonable to expect certain of these enzymes to act on a number of drugs. Pancreatic enzymes hydrolyze chloramphenicol palmitate. Pancreatin and trypsin are able to deacetylate *N*-acetylated drugs, and mucosal esterases appear to attack various esters of penicillin. Oral cocaine ingestion is generally ineffective in producing a pharmacological response because of efficient hydrolysis by esterase enzymes in the gut. This is not true at very large doses, however, such as those resulting from the rupture of bags containing the drug which are ingested to avoid detection at international borders.

Mucin, a viscous mucopolysaccharide that lines and protects the intestinal epithelium, has been thought to bind certain drugs nonspecifically (e.g., quarternary ammonium compounds) and thereby prevent or reduce absorption. This behavior may partially account for the erratic and incomplete absorption of such charged

Copyright © 2002 Marcel Dekker, Inc.

compounds. Mucin may also represent a barrier to drug diffusion prior to reaching the intestinal membrane.

B. Gastric Emptying

Physiologists have for many years been interested in factors that influence gastric emptying and the regulatory mechanisms controlling this process. Our interest in gastric emptying is based on the fact that, since most drugs are best absorbed from the small intestine, any factor that delays movement of drug from the stomach to the small intestine will influence the rate (and possibly the extent) of absorption and therefore the time needed to achieve maximal plasma concentrations and pharmacological response. As a result, and in addition to rate of dissolution or inherent absorbability, gastric emptying may represent a limiting factor in drug absorption. Only in those rare instances where a drug is absorbed by a specialized process in the intestine will the amount of drug leaving the stomach exceed the capacity of the gut to absorb it.

Gastric emptying is determined with a variety of techniques using liquid or solid meals or other markers. Gastric emptying is quantitated by one of several mesurements including emptying time, emptying half-time ($t_{50\%}$), and emptying rate. Emptying time is the time needed for the stomach to empty the total initial stomach contents. Emptying half-time is the time it takes for the stomach to empty one half of its initial contents. Emptying rate is a measure of the speed of emptying. Note that the last two measures are inversely related (i.e., the greater the rate, the smaller the value for emptying half-time).

Gastric emptying and factors that affect that process need to be understood because of the implications for drug absorption and with regard to optimal dosage form design [63]. Gastric-emptying patterns are distinctly different depending upon the absence or presence of food. In the absence of food, the empty stomach and the intestinal tract undergo a sequence of repetitious events referred to as the interdigestive migrating motor (or myoelectric) complex (MMC) [64]. This complex results in the generation of contractions beginning with the proximal stomach and ending with the ileum. The first of four stages is one of minimal activity that lasts for about 1 hour. Stage 2, which lasts 30–45 minutes, is characterized by irregular contractions that gradually increase in strength leading to the next phase. The third phase, while only lasting 5–15 minutes, consists of intense peristaltic waves, which results in the emptying of all remaining gastric contents into the pylorus. The latter phase is sometimes referred to as the ''housekeeper'' wave. The fourth stage represents a transition of decreasing activity leading to the beginning of the next cycle (i.e., phase 1). The entire cycle lasts about 2 hours. Thus, a solid dosage form ingested on an empty stomach will remain in the stomach for a period of time dependent upon time of dosing relative to the occurrence of the housekeeper. The gastric residence time of a solid dosage form will vary from perhaps 5–15 minutes (if ingested at the beginning of the housekeeper) to about 2 hours or longer (if ingested at the end of the housekeeper wave). It would not be surprising, however, for gastric residence time to be substantially longer. This variability in gastric residence time may explain some of the intersubject variation in rate of absorption, and it raises some question concerning the term ''ingested on an empty stomach.'' While it is quite common in clinical research studies for a panel of subjects to ingest a solid test dosage form following an overnight fast and, therefore, on an empty stomach, it is unlikely that all subjects will be in the same phase of the migrating motor complex. It is the latter point rather than an empty stomach per se, that will determine when emptying occurs and, consequently, when drug absorption is initiated. The above considerations will not apply to liquid dosage forms, however, which are generally able to empty during all phases of the migrating motor complex.

Various techniques have been used to visualize the gastric emptying of dosage forms. Radiopaque tablets were found to undergo relatively mild agitation in the stomach—a point that needs to be considered in the design and interpretation of disintegration and dissolution tests. While single large solid dosage forms (e.g., tablets and capsules) rely upon the housekeeper wave for entry into the small intestine, some controversy remains about the influence of particle (or pellet) size (diameter and volume), shape, and density on gastric emptying. There has been a great deal of recent interest in this issue, which has been investigated primarily with use of gamma scintigraphy (a gamma-emitting material is ingested and externally monitored with a gamma camera). These studies are generally performed with the use of nondisintegrating pellets so that movement throughout the tract may be estimated. Particles as large as 5–7 mm may leave the stomach. It is likely that there is a range of particle sizes that will empty from the stomach, rather than there being an abrupt cut-off value. The range of values among individuals will be affected by the size of the pylorus diameter and the relative force of propulsive contrac-

Copyright © 2002 Marcel Dekker, Inc.

tions generated by the stomach. The interest in this issue stems from the desire to develop sustained-release dosage forms that would have sufficient residence time in the GIT to provide constant drug release over an extended time. Experimental dosage forms that have been investigated include floating tablets, bioadhesives (that attach to the gastric mucosa), dense pellets, and large dimension forms.

Eating interrupts the interdigestive migrating motor complex. Gastric emptying in the presence of solid or liquid food is controlled by a complex variety of mechanical, hormonal, and neural mechanisms. Receptors lining the stomach, duodenum, and jejunum that assist in controlling gastric emptying include mechanical receptors in the stomach which respond to distension; acid receptors in the stomach and duodenum; osmotic receptors in the duodenum, which respond to electrolytes, carbohydrates and amino acids; fat receptors in the jejunum; and L-tryptophan receptors. Neural control appears to be through the inhibitory vagal system (the exact neurotransmitter is not known but may be dopamine and enkephalin). Hormones involved in controlling emptying include cholecystokinin and gastrin, among others.

As food enters the stomach the fundus and body regions relax to accommodate the meal. Upon reaching the stomach, food tends to form layers that are stratified in the order in which the food was swallowed, and this material is mixed with gastric secretions in the antrum. Nonviscous fluid moves into the antrum, passing around any solid mass. Gastric emptying will begin once a considerable portion of the gastric contents become liquid enough to pass the pylorus. Peristaltic waves begin in the fundus region, travel to the prepyloric area, and become more intense in the pylorus. The antrum and pyloric sphincter contract and the proximal duodenum relaxes. A moment later the antrum relaxes and the duodenum regains its tone. The pyloric sphincter will remain contracted momentarily to prevent regurgitation, and the contents in the duodenum are then propelled forward. Emptying is accomplished by the antral and pyloric waves, and the rate of emptying is regulated by factors controlling the strength of antral contraction. Gastric emptying is influenced primarily by meal volume, the presence of acids, certain nutrients, and osmotic pressure. Distension of the stomach is the only natural stimulus known to increase the emptying rate. Fat in any form in the presence of bile and pancreatic juice produces the greatest inhibition of gastric emptying. This strong inhibitory influence of fats permits time for their digestion, as they are the slowest of all foods to be digested. Meals containing substantial amounts of fat can delay gastric emptying for 3–6 hours or more. These various factors appear to alter gastric emptying by interacting with the receptors noted above.

Other than meal volume per se all of the other factors noted above result in a slowing of gastric emptying (e.g., nutrients, osmotic pressure, and acidity). It is important to recognize that a host of other factors are known to influence emptying rate. Thus, a variety of drugs can alter absorption of other drugs via their effect on emptying. For example, anticholinergics and narcotic analgesics reduce gastric-emptying rate, while metoclopramide increases that rate. A reduced rate of drug absorption is expected in the former instance and an increased rate in the latter. The following factors should be recognized: body position (reduced rate lying on left side), viscosity (rate decreases with increased viscosity), and emotional state (reduced rate during depression, increased rate during stress). As an illustration, one recent report indicates that absorption rate (and potentially completeness of absorption) may be altered when comparing posture—lying on the left or right side [65]. Acetaminophen and nifedipine absorption rates were faster when the subjects were lying on the right compared to the left side, suggesting more rapid gastric emptying. In the case of nifedipine, the extent of absorption was greater when the subjects were lying on the right side, which may be due to transient saturation of a presystemic metabolic process (this is discussed in a later section). Miscellaneous factors whose exact effect on emptying may vary, include gut disease, exercise, obesity, gastric surgery, and bulimia.

Many investigators have suggested that gastric emptying takes place by an exponential (i.e., first-order kinetic) process. As a result, plots of log volume remaining in the stomach versus time will provide a straight-line relationship. The slope of this line will represent a rate constant associated with emptying. This relationship is not strictly linear, however, especially at early and later times, but the approximation is useful in that one can express a half-time for emptying ($t_{50\%}$). Hopkins [66] has suggested a linear relationship between the square root of the volume remaining in the stomach versus time. There may be a physical basis for this relationship, since the radius of a cylinder varies with the square root of the volume and the circumferential tension is proportional to the radius. Methods for analyzing gastric emptying data have been reviewed [67].

Gastric-emptying rate is influenced by a large number of factors, as noted above. Many of these

Copyright © 2002 Marcel Dekker, Inc.

factors account for the large variation in emptying among different individuals and variation within an individual on different occasions. Undoubtedly, much of this variation in emptying is reflected in variable drug absorption. Although gastric emptying probably has little major influence on drug absorption from solution, emptying of solid dosage forms does exert an important influence on drug dissolution and absorption. A prime example is enteric-coated tablets, which are designed to prevent drug release in the stomach. Any delay in the gastric emptying of these forms will delay dissolution, absorption, and the onset time for producing a response. Since these dosage forms must empty as discrete units, the drug is either in the stomach or the intestine. The performance of this dosage form can be seriously hampered if it is taken with or after a meal, as emptying is considerably delayed. Furthermore, if the drug is to be taken in a multiple-dosing fashion, there is a possibility that the first dose will not leave the stomach until the next dose is taken, resulting in twice the desired dose getting into the intestine at one time. Blythe et al. [68] administered several enteric-coated aspirin tablets containing $BaSO_4$ and radiologically examined emptying of these tablets. The tablets emptied in these subjects anywhere from 0.5 to 7 hours after ingestion. Tablets will empty more rapidly when given prior to a meal compared to administration after a meal. One potential way of improving the emptying and release pattern of enteric-coated products is to use capsules containing enteric-coated microgranules. The median time for 50% and 90% emptying of such a dosage form has been shown to be 1 and 3–3.5 hours, respectively [69].

Several recent publications have reviewed the effects of food on drug absorption in humans [70–72]. The effect of food on the gastrointestinal absorption of drugs is complex and multidimensional. We are only now beginning to understand this complexity. The physical presence of food in the GIT may play a significant role in affecting the efficient absorption of a drug from an oral dosage form. The ultimate effect of food on the rate and/or extent of gastrointestinal absorption is a function of numerous interacting variables. While some general rules may be postulated, the effect of food on a given drug and its dosage form will require, in general, individual investigation. The U.S. Food and Drug Administration (FDA) has recognized this complexity and requires that all dosage forms that do not immediately release drug (e.g., controlled release formulations) undergo a food-effects study in humans, for which a "Guidance" has been written (these are available on the FDA webpage—fda.gov). The precedence for this requirement is the observation that a sustained release formulation may "dose-dump" its entire contents of drug in the presence of food, and, since the dose is several times the usual single dose, this may lead to toxicity [73,74].

The extent to which food will alter absorption depends upon factors such as physical chemical characteristics of the drug (e.g., aqeuous solubility, oil/water partition coefficient, and stability in gut fluids), the dose of drug, the characteristics of the dosage form, time of drug administration relative to food ingestion, amount of food, and type of food. The schematic in Fig. 11 summarizes the variables that food is able to affect, which in turn are associated with the sequential steps of drug release, dissolution, absorption, systemic availability, and elimination. Notice the two heavy arrows for the dissolution and absorption steps. One of those two processes is generally associated with the rate-limiting step in the overall absorption of a drug. Therefore, when food affects drug absorption, it most often does so by affecting the factors influencing dissolution or transport. Most of the factors noted in Fig. 11 are self-explanatory, but many will be discussed with examples in the following sections.

It is useful to consider the schematic in Fig. 11 and the most important rate-limiting steps in drug absorption in conjunction with what has become known as the Biopharmaceutical Classification System (BCS). The latter system attempts to classify drugs in terms of their aqueous solubility and membrane permeability and, from that classification, predict the most likely behavior of the drug with regard to absorption following oral administration. FDA is proposing to use this approach in establishing regulatory standards for new drug entities and for generic versions of marketed drugs. The BCS can be easily understood from the simple two-by-two matrix illustrated in Fig. 12. Drug behavior with respect to aqueous solubility and membrane permeability may be described by one of four possible conditions noted as I to IV. The effects of food on the absorption behavior of compounds classified according to the BCS has been discussed thoroughly elsewhere [71], and they will be highlighted here.

Class I compounds have both good solubility and permeability and generally offer no problems with regard to having a good absorption profile (e.g., acetaminophen, disopyramide, ketoprofen, metoprolol, nonsteroidal anti-inflammatory agents, valproic acid, verapamil). In general, one would not expect the presence of food to influence the absorption of this class

Copyright © 2002 Marcel Dekker, Inc.

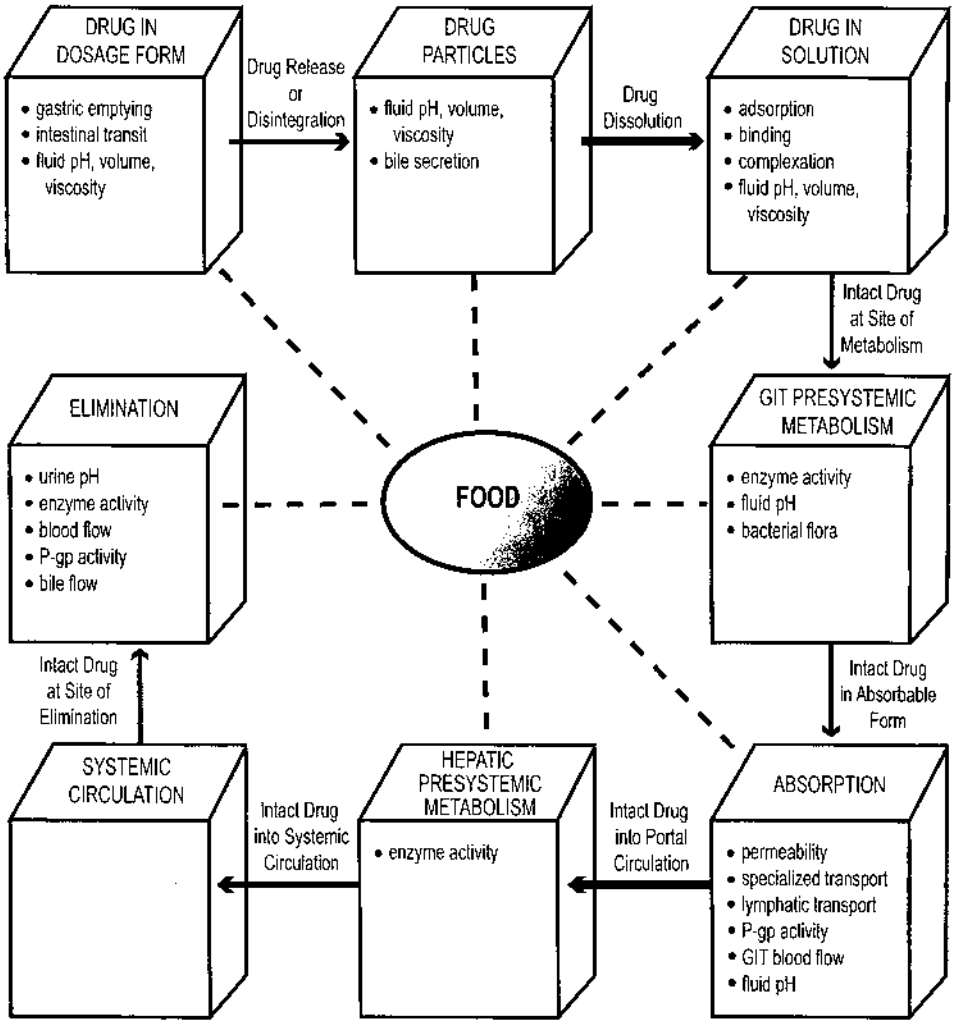


Fig. 11 The influence of food on factors that affect the sequential steps leading to GIT drug absorption, systemic availability, and elimination. (Copyright © 2000 Saguaro Technical Press, Inc., used with permission.)

of drug. The onset or rate of absorption may be delayed by the presence of food as a consequence of reduced gastric-emptying rate, but the completeness of absorption should not be compromised.

Drugs in Class II have low aqueous solubility (but high membrane permeability), and any factor affecting dissolution rate would be expected to have an impact on the absorption of such compounds. Factors that are noted in Fig. 11, such as fluid pH, volume and viscosity, and bile secretion (especially in response to fatty foods), might be expected to play a role in dissolution rate and thereby affect absorption. Compounds that fall into this class include carbamazepine, cyclosporin, digoxin, griseofulvin, and spironolactone. Food would be expected to exert a potentially significant affect on the absorption of the drugs in this class. The absorption of many compounds in this class is improved in the presence of a fatty meal. Improved absorption occurs because of the secretion of bile in response to fat and the surfactant and solubilizing properties that bile salts exert on poorly water-soluble compounds (griseofulvin is a classic example). The presence of food, especially fatty foods, will delay gastric emptying and intestinal transit, which in turn allow for more time for drug dissolution and absorption.

Drugs in Class III have good aqueous solubility but poor membrane permeability (e.g., bidisomide, bisphosphonates, captopril, and furosemide). Food and food components would only be expected to influence absorption of drugs in this class if they affected some aspect

Copyright © 2002 Marcel Dekker, Inc.

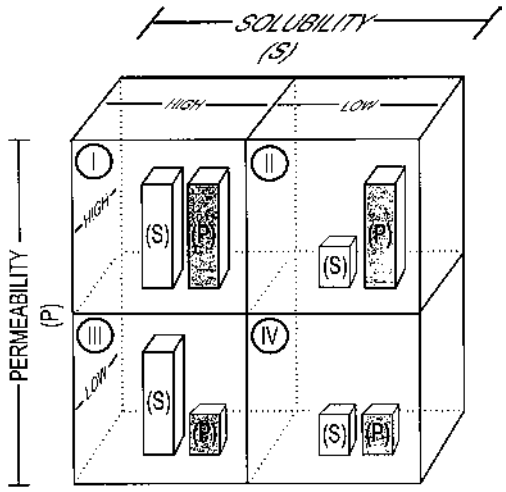


Fig. 12 Illustration of the Biopharmaceutical Classification System (BCS), which classifies drug absorption potential on the basis of aqueous solubility or membrane permeability. (Copyright © 2000 Saguaro Technical Press, Inc., used with permission.)

of membrane permeability or function or if the permeable form of the molecule was altered (e.g., through ionization). Many of these compounds tend to be absorbed in relatively well defined and limited areas of the upper small intestine. As a consequence, the physical presence of food may affect the time during which the drug resides at sites of maximal absorption. Food and the viscous milieu that it creates in gut fluids may also create a significant physical barrier to the diffusion of the drug to sites of absorption along the GIT membrane. Several compounds that fall into this class illustrate reduced absorption in the presence of food.

Class IV drugs have low aqueous solubility and poor membrane permeability and as such are often considered as poor drug candidates for oral administration. Other routes of administration may need to be considered. For example, neomycin falls into this category, and its oral use is to achieve sterilization of the gut. There is too little information about these compounds and the effect of food to offer general observations.

A drug should always be ingested with a cup of water ($\sim$ 8 oz) to insure easy transit down the esophagus and to provide fluid for disintegration and dissolution. Whether or not the drug should be taken on an empty stomach (e.g., enteric-coated tablets) or with food will depend upon the specific drug as noted above.

Drugs that should be taken with food include those compounds that are irritating to the tract (e.g., phenylbutazone or nitrofurantoin), those compounds absorbed high in the tract by a specialized mechanism (e.g., riboflavin and ascorbic acid), and those compounds for which the presence of certain food constituents are known to enhance absorption (e.g., griseofluvin). For those compounds that irritate the tract, perhaps the best recommendation is to ingest the drug with or after a light meal that does not contain fatty foods or constituents known to interact with the drug. Nitrofurantoin absorption is improved in the presence of food [75,76]. Riboflavin and ascorbic acid, which are absorbed by a specialized process high in the small intestine, are best absorbed when gastric emptying is delayed by the presence of food [39,77]. As the residence time of the vitamins in the upper portion of the intestine is prolonged, contact with absorption sites is increased and absorption becomes more efficient (i.e., saturation of transporters is avoided). The influence of food on the absorption of those vitamins is illustrated in Fig. 13 along with improved efficacy of ascorbic acid absorption achieved by administering divided doses.

The absorption of griseofulvin, which is a very poorly water-soluble drug, is enhanced when it is coadministered with a fatty meal as discussed previously [61,62]. The importance of gastric emptying can probably be most readily appreciated by those investigators who have examined drug absorption in patients after a partial or total gastrectomy. Muehlberger [78] notes that,

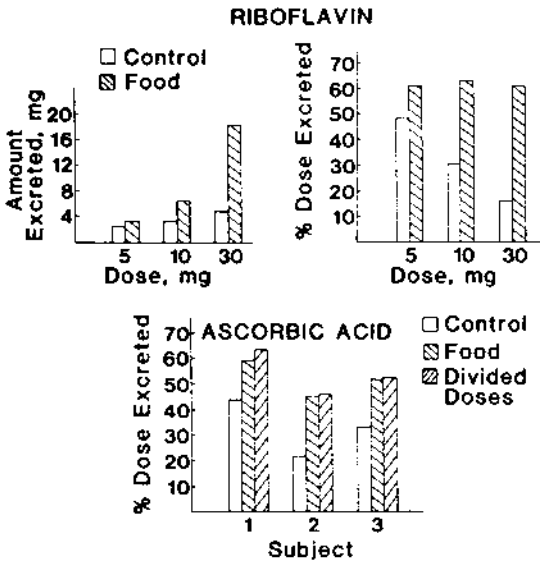


Fig. 13 (Top) Influence of food on the absorption of different doses of riboflavin. (Bottom) Influence of food and divided doses on ascorbic acid absorption in three subjects. (Based on data from Refs. 39 and 77.)

Copyright © 2002 Marcel Dekker, Inc.

following a near-total gastrectomy, patients often complain of a "sensitivity" to alcohol. This is probably best explained by ethanol moving rapidly from the poorly absorbing surface of the stomach to the small intestine, where absorption will be rapid. Gastric emptying has been shown to be important in oral L-dopa therapy, and it has been noted [79] that patients with a partial gastrectomy or gastrojejunostomy exhibit a prompt response with less than average doses of the drug. This observation is consistent with rapid absorption from the small intestine in such patients and is essentially equivalent to introduction of the drug into the duodenum. Similar conclusions have been reached for aspirin and warfarin absorption [80,81].

For many drugs, as has been shown for acetaminophen, there will be a direct relationship between gastric-emptying rate and maximal plasma concentration and an inverse relationship between gastric-emptying rate and the time required to attain maximal plasma concentrations. Those relationships are illustrated in Figs. 14A and B. Also shown is the influence of a narcotic (heroin) on the gastric emptying and absorption of acetaminophen (Figs. 14C and D). In attempting to predict such relationships, however, it is essential that one consider the physicochemical characteristics of the drug. While increased gastric-emptying rate will probably increase the rate (and possibly the extent) of absorption for drugs best absorbed from the small intestine from rapidly dissolving dosage forms, the converse may be true in other circumstances. For example, if the dosage form must first be exposed to the acidic gastric fluids to initiate disintegration or dissolution, rapid emptying may reduce the rate and extent of absorption. Similarly, if the drug dissolves slowly from the dosage form, a shortened residence time in the gut may reduce the extent of dissolution and absorption. One needs a good deal of fundamental understanding of the chemistry of the drug, its dosage form, and the absorption mechanism before being able to anticipate or rationalize the influence of these various factors on the efficacy of absorption.

A final point that should be mentioned here, although it has received relatively little attention, is that of esophageal transit. Delay in movement down the esophagus will delay absorption and, in addition, for certain drugs, may also cause local mucosal damage. Capsule disintegration has been observed to occur in the esophagus within 3–5 minutes. Esophageal transit is delayed when solid dosage forms are swallowed with little fluid or when the subject is supine [84,85]. Antipyrine absorption from capsules [86] has been shown to be delayed when esophageal transit was prolonged. To avoid this delay the dosage form should be swallowed with water or other fluids, and the subject should be in a standing or sitting position.

C. Intestinal Transit

Once a dosage form empties from the stomach and enters the small intestine, it will be exposed to an

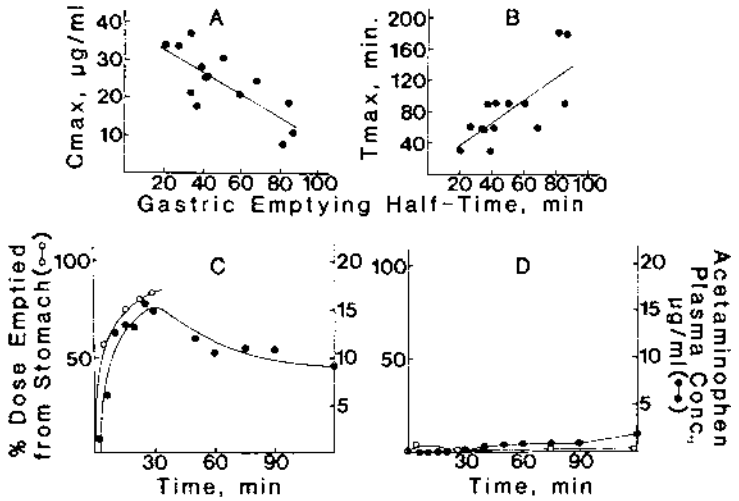


Fig. 14 (A, B) Maximum acetaminophen plasma concentration (C_{max}) and time to achieve that concentration (T_{max}) as a function of gastric emptying half-time. (From Ref. 82.) (C) Percent of an acetaminophen dose emptied from the stomach (○) and acetaminophen plasma concentrations (●) as a function of time in one subject. (D) The same plot and for the same subject as in (C) after a 10 mg intramuscular dose of heroin. (C and D from Ref. 83.)

Copyright © 2002 Marcel Dekker, Inc.

environment totally different from that in the stomach, as discussed previously. Since the small intestine is the primary site of drug absorption, the longer the residence time in this region, the greater the potential for complete absorption, assuming that the drug is stable in the intestinal fluids and will not form water-insoluble derivatives.

There are primarily two types of intestinal movements: propulsive and mixing. Propulsive movements, generally synonymous with peristalsis, will determine intestinal transit rate and therefore the residence time of a drug in the intestine. This time of residence is important since it will dictate the amount of time the dosage form has in which to release the drug, permit dissolution, and allow for absorption. Obviously, the greater the intestinal motility, the shorter the residence time and the less time there is for those processes to proceed. Intestinal motility will be most important for those dosage forms that release drug slowly (e.g., sustained-release products) or require time to initiate release (e.g., enteric-coated products), as well as those drugs that dissolve slowly or where absorption is maximal only in certain regions of the intestine. Peristaltic waves propel intestinal contents down the tract at about 1–2 cm/s. Peristaltic activity is increased after a meal as a result of the gastroenteric reflex initiated by distension of the stomach and results in increased motility and secretion.

Mixing movements of the small intestine are a result of contractions dividing a given region of the intestine into segments producing an appearance similar to a chain of sausages. These contractions result in the mixing of the intestinal contents with secretions several times a minute. These movements bring the gut contents into optimal contact with the surface epithelium and thereby provide a larger effective area for absorption. In addition, the muscularis mucosa produces folds in the surface epithelium, resulting in an increased surface area and rate of absorption. The villi contract during this process, which results in a ''milking'' action so that lymph flows from the central lacteal into the lymphatic system.

These mixing motions will tend to improve drug absorption for two reasons. Any factor that increases rate of dissolution will increase the rate (and possibly the extent) of absorption, especially for poorly water-soluble drugs (BCS Classes II and IV). Since rate of dissolution depends on agitation intensity, mixing movements will tend to increase dissolution rate and thereby influence absorption. As rate of absorption depends directly on membrane surface area, and since mixing increases the contact area between drug and membrane, these motions will tend to increase rate of absorption.

Metoclopramide will increase the rate of gastric emptying, which will often, but not always, increase the rate of drug absorption. However, metoclopramide will also increase the rate of intestinal transit and thus reduce the residence time in the intestine. These two effects may have an opposing influence on absorption. The net effect on absorption depends on the characteristics of the drug and its dosage form as well as the mechanism of absorption. Metoclopramide or similar-acting drugs will probably have little if any effect on absorption of a drug given orally in solution, unless the compound (e.g., riboflavin) is absorbed by a specialized process high in the small intestine, in which case there is likely to be a reduction in the amount absorbed. Metoclopramide will probably increase the rate of absorption of a drug from a solid dosage form because of its effect on gastric emptying if the drug is rapidly released and readily dissolved. On the other hand, if the drug dissolves slowly from the dosage form, the extent of absorption may be reduced as a result of shortened residence time in the intestine, even though gastric-emptying rate is increased. Similar reasoning may be applied to the influence on drug absorption of various anticholinergics (e.g., atropine and propantheline) and narcotic analgesics that reduce gastric emptying and intestinal transit rates. While there will be a reduction in gastric-emptying rate and thus a delay in absorption, these compounds will increase intestinal transit time and possibly increase the extent of absorption, particularly for slowly dissolving drugs or dosage forms that release drug slowly.

Transit through the small intestine appears to be quite different in a variety of ways from movement through the stomach. Once emptied from the stomach, material (such as pellets and tablets) will move along the small intestine and reach the ileocecal valve in about 3 hours. While this value may range from about 1 to 6 hours, intestinal residence time appears to be relatively consistent among normal subjects [87]. Values similar to this have been found for food and water movement along the small intestine. Transit appears to be less dependent upon the physical nature of the material (liquid vs. solid and size of solids) compared to the response of the stomach. Furthermore, food appears not to influence intestinal transit as it does gastric emptying.

Three to four hours in the small intestine is a relatively short time for a dosage form to completely dissolve or release drug and then be absorbed. This time

Copyright © 2002 Marcel Dekker, Inc.

would be even more critical to the performance of poorly water-soluble drugs, slowly dissolving coated dosage forms (enteric or polymer coated), and sustained-release forms. Assuming minimal absorption from the colon (discussed below), gastric residence time may prove a critical issue to the performance of certain drugs and drug dosage forms (especially those in the latter categories) as a result of the relatively short intestinal residence time.

There is less information available concerning the factors that may influence intestinal transit time compared to what we know about gastric residence time. Although based upon small populations, there appear to be no gender-related differences in intestinal transit time [88], while vegetarians appear to have longer intestinal transit times compared to nonvegetarians [89]. The latter point may have implications for drug therapy in the third world where the diet is primarily vegetarian. Other factors that result in an increased transit time include reduced digestive fluid secretion, reduced thyroxin secretion, and pregnancy [90–92].

The distal portion of the GIT, the colon (see Fig. 1), has as its primary function water and electrolyte absorption and the storage of fecal matter prior to its being expelled. The proximal half of the colon is concerned with absorption and the distal half with storage. Although there are mixing and propulsive movements in the colon, they tend to be rather sluggish. Large circular constrictions occur in the colon that are similar to the segmenting contractions seen in the small intestine. The longitudinal muscles lining the colon also contract, producing a bulging similar in appearance to sacs and referred to as haustrations. These movements increase the surface area of the colon and result in efficient water absorption.

Contents within the colon are propelled down the tract not by peristaltic waves but by a "mass movement," which occurs only several times a day, being most abundant the first hour after breakfast as a result of a duodenocolonic reflex. The greatest proportion of time moving down the GIT is spent by a meal moving through the colon. In the presence of a diarrheal condition, fluid absorption is incomplete, which results in a watery stool.

Colonic residence time is considerably longer than in other parts of the GIT, and it is also more variable. The transit time can be as short as several hours to as long as 50–60 hours. Transit along the colon is characterized by abrupt movement and long periods of stasis. In one study of 49 healthy subjects the average colonic residence time was 35 hours with the following times associated with different regions: 11 hours in the right (ascending) colon, 11 hours in the left (descending) colon, and 12 hours in the rectosigmoid colon [93]. The latter values do not appear to be influenced by particle size (i.e., pellets vs. tablet), but these times are highly variable and are shortened in response to ingestion of a laxative (average time for a 5 mm tablet in the ascending colon of 8.7 vs. 13.7 hours) [94,95]. Furthermore, the ingestion of food, which is known to increase colonic activity, does not appear to have a dramatic effect on the movement of dosage forms from the ileum into the colon or on the movement within the colon [96]. Any differences in colonic transit times as a function of age and gender are not clear at this time due to conflicting reports and investigation in small populations of subjects.

The colonic mucosal pH varies along the length of the colon: right colon, pH 7.1; transverse colon, pH 7.4; left colon, pH 7.5; sigmoid colon, pH, 7.4; rectum, pH 7.2. These values were determined in a group of 21 subjects (mean age 54 years), and they are somewhat higher than previous estimates (pH $\sim$ 6.7 in the right colon) [97]. Those values contrast with the proximal small intestine with a pH of about 6.6 and the terminal ileum with a pH of about 7.4. This near-neutral pH in conjunction with low enzymatic activity has made the colon an interesting potential site for drug absorption. Indeed, there is active interest in delivery of drug dosage forms to the colon for site-specific absorption, especially for peptides (e.g., Refs. 63,98). Characteristics of the colon that are thought to provide a good environment for drug absorption include mild pH, little enzyme activity, and long residence time. The disadvantages of the colon, however, include several considerations that substantially limit this area for providing good absorption: small surface area, relatively viscous fluid-like environment (which varies along the length of the colon), and the large colonies of bacteria. The latter factors would limit dissolution and contact with the absorbing surface membrane and could result in presystemic drug metabolism.

The intention of colon-specific drug delivery is to prevent the drug from being released from the dosage form (by coating or other release-controlling mechanism) until it reaches the distal end of the large intestine (i.e., the ileocecal valve). Drug release needs to be delayed for about 5 hours, but clearly this delay time will vary from patient to patient and will depend upon a host of factors that may affect gastric emptying and intestinal transit (e.g., food, drugs, etc.). The dosage form should then release drug over the next 10–15 hours while in the colon. The results of studies that have examined colonic absorption are not that

Copyright © 2002 Marcel Dekker, Inc.

encouraging, although they do indicate that absorption does occur but to a variable extent (depending on the drug). The hormone calcitonin provided an absolute bioavailability from the colon of <1%; however, no comparison to oral dosing was made [99]. The relative bioavailability of ranitidine solution from the cecum was about 15% of that following gastric or jejunal administration [100]. Benazepril relative bioavailability following a colonic infusion was about 23% that of an oral solution [101]. Figure 15 illustrates the plasma concentration–time profiles for those two drugs. The long-lasting analog of vasopressin, dDAVP, a nonapeptide, had a relative bioavailability of about 17 and 21% compared to duodenal and gastric (and jejunal) solution administration, respectively. Rectal administration provided absorption comparable to that from the colon [102]. Sumatriptan solution was absorbed from the cecum to an extent of about 23% compared to an oral (and jejunal) dose [103]. In all cases rate of absorption is substantially slower than from the upper regions of the GIT. Furthermore, in some instances the metabolite to parent drug concentration ratios change depending upon the site of administration, which may reflect a number of causes (e.g., different extent of presystemic metabolism, differences in metabolite absorption, etc.). The latter needs to be a consideration for those compounds whose metabolites are either pharmacologically active or toxic.

A relevant consideration of absorption from the colonic area is with regard to rectal drug administration. While this is not a frequently used route, it is employed to some extent, especially in infants, children, and those unable to swallow medication. Absorption from the rectum is generally considered to be relatively poor, at least in comparison to absorption from regions of the upper GIT. The reasons for this are essentially those outlined above for the colon: small absorbing surface area, little fluid content, and poor mixing movements. There are, in addition, two other considerations. First, the presence of fecal material may provide a site for adsorption, which can effectively compete for absorption. Second, the extent of absorption will be dependent upon the retention time of the dosage form in the rectum. This may be a critical issue for infants, who often have irregular bowel movements. The readers are referred to a review of this topic [104].

D. Blood Flow

The entire GIT is highly vascularized and therefore well perfused by the bloodstream. The splanchnic

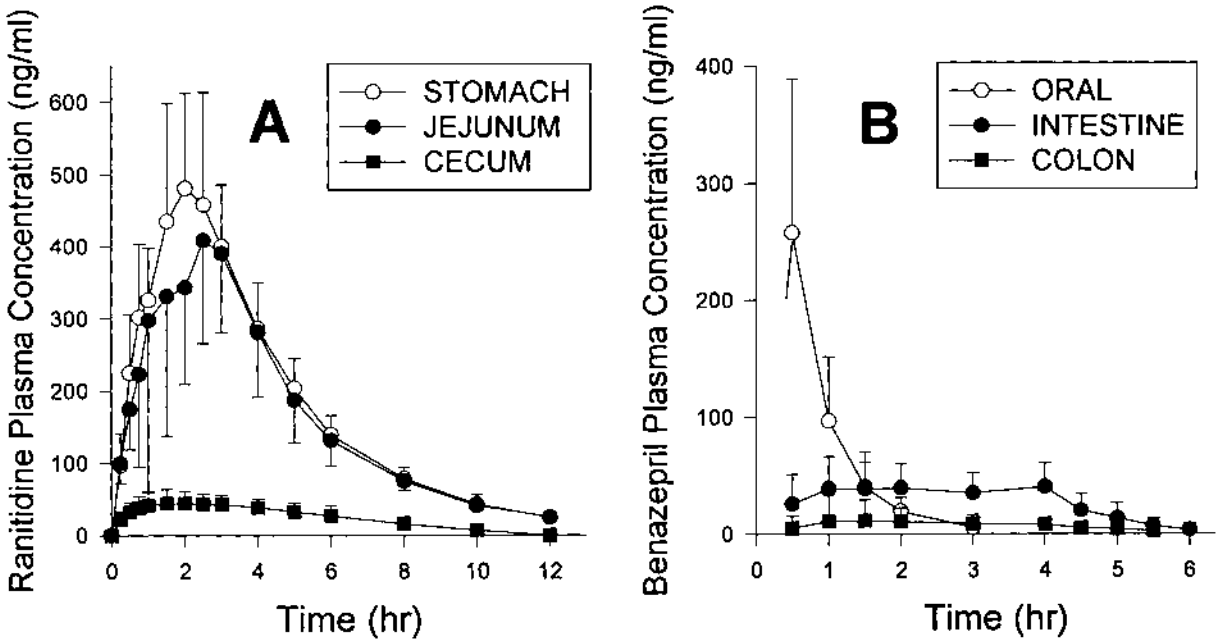


Fig. 15 (A) Ranitidine plasma concentrations as a function of time following administration of a solution into the stomach, jejunum, or cecum. Each value is the mean of eight subjects (the cross-hatched vertical bars are standard deviations). (Based on data in Ref. 100.) (B) Benazepril plasma concentrations as a function of time following a solution dose taken orally or administered as a 4-hour intestinal or 4-hour colonic infusion. Each value is the mean of 7–13 subjects. (Based on data in Ref. 101.)

Copyright © 2002 Marcel Dekker, Inc.

circulation, which perfuses the GIT, receives about 28% of cardiac output, and this flow drains into the portal vein and then goes to the liver prior to reaching the systemic circulation. An absorbed drug will first go to the liver, which is the primary site of drug metabolism in the body; the drug may be metabolized extensively prior to systemic distribution. This has been referred to as the hepatic "first-pass" effect or presystemic hepatic elimination, and it has important implications in bioavailability and drug therapy.

The fact that the GIT is so well perfused by the bloodstream permits efficient delivery of absorbed materials to the body. As a result of this rapid blood perfusion, the blood at the site of absorption represents a virtual "sink" for absorbed material. Under normal conditions, then, there is never a buildup in drug concentration in the blood at the site of absorption. Therefore, the concentration gradient will favor further unidirectional transfer of drug from the gut to the blood. Usually, then, blood flow is not an important consideration in drug absorption. Generally, the properties of the dosage form (especially dissolution rate) or the compound's inherent absorbability will be the limiting factors in absorption.

There are circumstances, however, where blood flow to the GIT may influence drug absorption. Those compounds absorbed by active or specialized mechanisms require membrane participation in transport, which in turn depends on the expenditure of metabolic energy by intestinal cells. If blood flow and therefore oxygen delivery is reduced, there may be a reduction in absorption of those compounds. This has been shown to be the case in rats for the active absorption of phenylalanine [105].

The rate-limiting step in the absorption of those compounds that readily penetrate the intestinal membrane (i.e., have a large permeability coefficient) may be the rate at which blood perfuses the intestine. However, absorption will be independent of blood flow for those compounds that are poorly permeable. Extensive studies have illustrated this concept in rats [106,107]. The absorption rate of tritiated water, which is rapidly absorbed from the intestine, is dependent on intestinal blood flow, but a poorly absorbed compound, such as ribitol, penetrates the intestine at a rate independent of blood flow. In between these two extremes are a variety of intermediate compounds whose absorption rate is dependent on blood flow at low flow rates but independent of blood flow at higher flow rates. By altering blood flow to the intestine of the dog, as blood flow decreased the rate of sulfaethidole absorption also decreased [108]. These relationships are illustrated in Fig. 16.

An interesting clinical example of the influence of blood flow on drug absorption is that provided by Rowland et al. [109]. After oral ingestion of aspirin, one subject fainted while blood samples were being taken. Absorption ceased at the time of fainting but continued when the subject recovered. Interestingly, there was no reduction in the total amount of aspirin absorbed compared to another occasion when the subject did not faint. Another investigator observed a

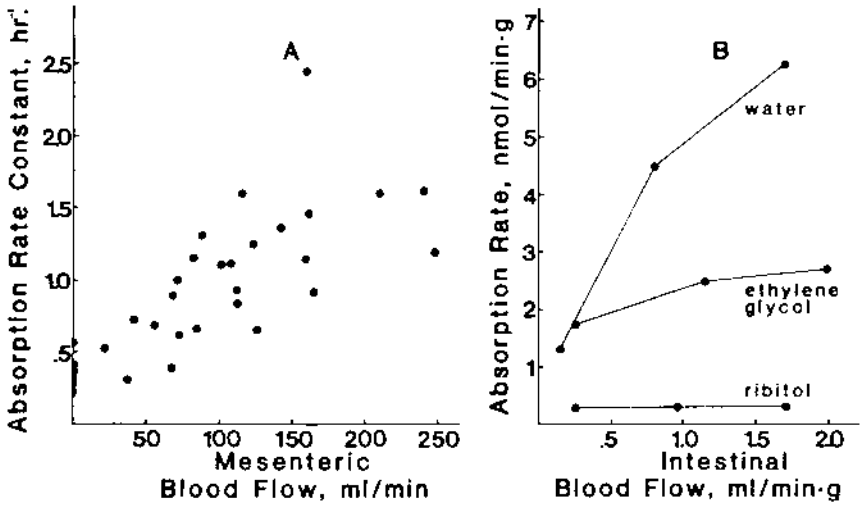


Fig. 16 (A) Absorption rate constant of sulfaethidole in dogs as a function of mesenteric blood flow. (Based on data from Ref. 108.) (B) Absorption rate of several compounds in rats as a function of intestinal blood flow. (Based on data from Ref. 107.)

Copyright © 2002 Marcel Dekker, Inc.

3-hour delay in the absorption of sulfamethoxy-pyridazine in a patient who fainted [110]. The most reasonable explanation of these observations is the fact that in a fainting episode blood is preferentially shunted away from the extremities and other body organs, including the GIT, thereby reducing blood perfusion of the tract and resulting in a decreased rate of absorption. It is possible that generalized hypotensive conditions may be associated with altered drug absorption. In that regard, consideration needs to be given to the influence of congestive heart failure and other disease conditions that will alter gut blood flow as well as the presence of other drugs that may alter flow. For example, it has been suggested that digoxin absorption is impaired in congestive heart failure but improves after compensation [111]. The influence of such conditions on absorption has been reviewed elsewhere [112], but there is relatively little information available.

Blood flow to the GIT increases shortly after a meal and may last for several hours. Digestive processes in general seem to enhance blood flow to the tract. For the reasons discussed previously, however, coadministration of a drug with a meal would normally not be expected to improve drug absorption. Strenuous physical exercise appears to reduce blood flow to the tract and may reduce absorption rate.

V. COMPLICATING FACTORS IN DRUG ABSORPTION

A. Drug-Food and Drug-Drug Interactions

There are a variety of factors that may affect the rate or extent of absorption. Interactions in absorption are mediated by physical chemical, physiological, or biochemical factors. Physical chemical considerations include the characteristics of the dosage form and altered solubility, dissolution, and chemical stability within the GIT. Physiological factors include residence time in the tract (i.e., gastric emptying and intestinal transit rates) and blood flow to and the characteristics of the absorbing membrane. Biochemical factors would include enzymatic activity in the GIT. Drug-food and drug-drug interactions will alter absorption by one or more of the foregoing mechanisms.

As discussed previously, drug absorption may be impaired or improved when food is present in the GIT. Food may reduce the rate or extent of absorption by virtue of reduced gastric-emptying rate, which is particularly important for compounds unstable in gastric fluids and for dosage forms designed to release drug slowly. Food provides a rather viscous environment that will reduce the rate of drug dissolution and drug diffusion to the absorbing membrane. Drugs may also bind to food particles or react with gastrointestinal fluids secreted in response to the presence of food. An interesting example of the influence of food and gastric emptying on the fate of a drug in the body is that of *para*-aminobenzoic acid. This compound is metabolized in the body (acetylation) by a saturable process. The rate of presentation (i.e., the rate of absorption) of the drug to its site of metabolism will influence the extent of acetylation. The more rapid the absorption, the less the extent of acetylation since the capacity to metabolize is exceeded by the rate of presentation to the enzyme system. When absorption rate is reduced by slowing gastric emptying (e.g., with fat or glucose), a greater fraction of the dose is metabolized [113]. Similar observations have been made with respect to salicylamide absorption rate and metabolism [114].

There are problems as well in the absorption of certain drugs in the presence of specific food components. L-Dopa absorption may be inhibited in the presence of certain amino acids formed from the digestion of proteins [43]. The absorption of tetracycline is reduced by calcium salts present in dairy foods and by several other cations, including magnesium and aluminum [115–117], which are often present in antacid preparations. In addition, iron and zinc have been shown to reduce tetracycline absorption [118]. Figure 17 illustrates several of these interactions. These cations react with tetracycline to form a water-insoluble and nonabsorbable complex. Obviously, these offending materials should not be co-administered with tetracycline antibiotics.

The tetracycline example cited above is one type of physical-chemical interaction that may alter absorption. The relative influence of complexation on drug absorption will depend on the water-solubility of the drug, the water-solubility of the complex, and the magnitude of the interaction (i.e., the complexation stability constant). If the drug itself is poorly water-soluble, the absorption pattern will be governed by rate of dissolution. Often such compounds are incompletely and erratically absorbed. As a result, complexation will probably exert more of an influence on the absorption of such a compound than on one that is normally well absorbed, although this will depend on the nature of the complex. If the complex is water-insoluble, as in the case of tetracycline interactions with various metal cations, the fraction complexed will be unavailable for absorption. Although most complexation interactions are reversible, the greater the stability constant of the

Copyright © 2002 Marcel Dekker, Inc.

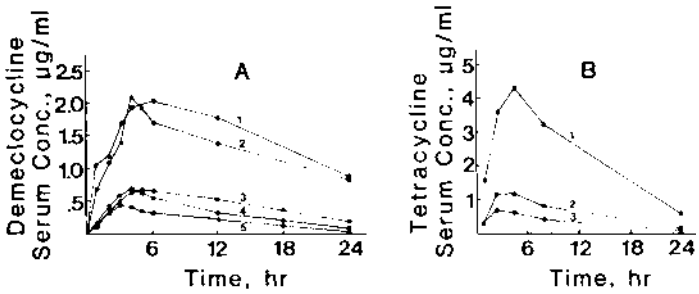


Fig. 17 (A) Demeclocycline serum concentrations as a function of time in four to six subjects after oral ingestion of demeclocycline in the absence or presence of dairy meals. Key: 1, meal (no dairy products); 2, water; 3, 110 g cottage cheese; 4, 240 mL buttermilk; 5, 240 mL whole milk. (Based on data from Ref. 117.) (B) Tetracycline serum concentrations as a function of time in six subjects after oral ingestion of tetracycline in the absence or presence of iron salts (equivalent to 40 mg elemental iron). Key: 1, control; 2, ferrous gluconate; 3, ferrous sulfate. (Based on data from Ref. 118.)

complex, the greater the relative influence on absorption. Generally, however, because the interaction is reversible, complexation is more likely to influence the rate than the extent of absorption.

Drug complexation is sometimes used in preparing pharmaceutical dosage forms to improve stability or solubility or to prolong drug release. There are several examples, however, where a drug complex results in reduced absorption. Amphetamine interacts with sodium carboxymethyl cellulose to form a poorly water-soluble derivative, and a decrease in absorption is seen [119]. Phenobarbital absorption is reduced as a result of interaction with polyethylene glycol 4000 [120]. These large macromolecules have the potential to bind many drugs.

Surface-active agents, because they are able to form micelles above the critical micelle concentration, may bind drugs either by inclusion within the micelle (solubilization) or by attachment to its surface. Below the critical micelle concentration, surfactant monomers have a membrane-disrupting effect, which can enhance drug penetration across a membrane. The latter influence has been seen in drug absorption studies in animals. The influence of surface-active agents on drug absorption will depend on the surfactant concentration and the physicochemical characteristics of the drug. If the drug is capable of partitioning from the aqueous to the micellar phase, and if the micelle is not absorbed, which is usually the case, there may be a reduction in rate of absorption. Micellar concentrations of sodium lauryl sulfate or polysorbate 80 (Tween 80) increase the rectal absorption rate of potassium iodide in the rat but reduce the absorption rate of iodoform and triiodophenol [121,122]. Since potassium iodide is not solubilized by the micelle, the enhanced rate of absorption is attributed to the influence of the surfactant on the mucosal membrane. The other compounds, which partition into the micelle, exhibit a reduced rate of absorption since there is a decrease in their effective concentration. Similar observations, using pharmacological response data in goldfish, have been made for several barbiturates in the presence of varying surfactant concentrations.

In addition to the aforementioned effects of surfactants, one must consider their influence on drug dissolution from pharmaceutical dosage forms. If the drug is poorly water-soluble, enhanced dissolution rate in the presence of a surface-active agent, even if part of the drug is solubilized, will result in increased drug absorption. The absorption rate of sulfisoxazole suspensions given rectally to rats increased with increasing polysorbate 80 concentration. At surfactant concentrations in excess of that needed to solubilize the drug completely, there was a reduced rate of absorption; however, the rate was greater than that from the control suspension (i.e., without surfactant) [123].

Another important type of physical chemical interaction that may alter absorption is that of drug binding or adsorption onto the surface of another material. As with complexation and micellarization, adsorption will reduce the effective concentration gradient between gut fluids and the bloodstream, which is the driving force for passive absorption. While adsorption frequently reduces the rate of absorption, the interaction is often readily reversible and will not affect the extent of absorption. A major exception is adsorption onto charcoal, which in many cases appears to be irreversible, at least during the time of residence within the GIT. As a result, charcoal often reduces the extent of drug absorption. Indeed, this fact

Copyright © 2002 Marcel Dekker, Inc.

along with the innocuous nature of charcoal is what make it an ideal antidote for oral drug overdose. The effectiveness of that form of therapy will depend on the amount of charcoal administered and the time delay between overdose and charcoal dosing. Another interesting aspect of charcoal dosing is its influence on shortening the elimination half-life of several drugs. This is a particularly attractive noninvasive means of enhancing drug elimination from the body (phenobarbital is a good example).

In addition to charcoal, adsorption is often seen with pharmaceutical preparations that contain large quantities of relatively water-insoluble components. A good example is antidiarrheal products and perhaps antacids. The importance of the strength of binding as it influences absorption has been illustrated by Sorby [124], who showed that both attapulgite and charcoal reduce the rate of drug absorption but only charcoal reduced the extent of absorption. Lincomycin is an example of a drug whose absorption in impaired by an antidiarrheal preparation [125]. Another type of compound that has been shown to alter drug absorption due to binding are the anion-exchange resins cholestyramine and colestipol. The foregoing physical chemical interactions, which may alter drug therapy, may be minimized by not coadministering the interacting compounds at the same time but separating their ingestion by several hours.

We noted previously, when discussing mechanisms of drug absorption, that certain calcium channel blockers have been shown to enhance the gastrointestinal absorption of several aminopenicillin and aminocephalosporin derivatives [46,47]. This type of interaction, while perhaps not of practical clinical importance, is very intriguing in terms of promoting absorption efficiency, and it illustrates the useful (rather than deleterious) aspect of a drug-drug interaction. Another interesting and recently observed drug-food (or drug-nutrient) interaction is that between grapefruit juice and certain drugs, especially those with a high hepatic clearance that undergo substantial presystemic (first-pass) metabolism (discussed later). The discovery of this interaction is an excellent example of serendipity and an illustration of Pasteur's famous statement, "In the field of experimentation, chance favors only the prepared mind." In their original study [126], the experimenters were attempting to examine the influence of ethanol on the pharmacokinetics of co-administered drugs. The ethanol was mixed (serendipitously) with grapefruit juice to mask the taste of the ethanol. A substantial interaction was noted with felodipine—an approximate three fold increase in bioavailability. Since that initial observation, numerous studies have indicated that many other drugs with diverse chemical structures having a high metabolic clearance participate in this interaction with grapefruit juice (not ethanol!). Many studies have been devoted to determining the mechanism(s) of this interaction, and the readers are referred to a recent symposium and book chapter that cover this topic [127,128]. While we have a better understanding of the nature of the interaction, it now appears that there may be several processes going on.

The drugs that undergo this interaction are substrates for one of the major drug-metabolizing isozymes in the body, cytochrome P450 3A4 (CYP3A4). This enzyme has been referred to as being "promiscuous" (in a biochemical sense), in that it will accommodate and metabolize a very wide range of drug molecules. Since this isozyme is responsible for the metabolism of many drugs, it is also involved in many drug-drug and, in this case, nutrient-drug interactions. This enzyme and many other isozymes are present in the mucosal surface cells (enterocytes) of the GIT. Therefore, the gut surface represents a potential site for drug metabolism. Some component(s) of grapefruit juice is able to inhibit the CYP3A4 enzyme, resulting in less than usual presystemic gastrointestinal metabolism of the drug. This, in turn, results in greater oral bioavailability of the drug. While it was initially thought that bioflavanoids present in the grapefruit juice were the critical chemical components, it has since been determined that furanocoumarin derivatives are the most likely enzyme inhibitors [129]. Enzyme inhibition could also be expressed in liver enzyme activity (also resulting in increased bioavailability); however, liver enzymes appear to play a less significant role in this interaction than those in the GIT. The latter, however, may be more or less true depending upon the drug whose interaction is under investigation.

A complicating factor, however, in this interaction is that many compounds that are substrates for CYP3A4 are also substrates for P-glycoprotein (P-gp), the efflux transporter protein that resides in the membrane surface of the GIT enterocyte. This transporter was noted in a previous section that discussed the transport process. There is evidence to suggest that the drug interaction with grapefruit juice also involves inhibition of this efflux transporter [127]. Both CYP3A4 metabolic inhibition and P-gp efflux transport inhibition will result in increased drug bioavailability. It is possible that these two mechanisms are acting in tandem in an attempt to limit foreign compounds (such as drugs) from entering the systemic circulation.

Copyright © 2002 Marcel Dekker, Inc.

Table 8 Drugs Known to Interact with Grapefruit Juice

Anti-infective: artemether, saguinavir
Antilipemic: atorrastatin, lovastatin, simvastatin
Cardiovascular: carvedilol, cilostazole, felodipine, nicardipine, nifedipine, nimodipine, nisoldipine, nitrendipine, verapamil
CNS: buspirone, carbamazepine, diazepam, midazolam, trizolam
Other:
Antihistamine: astemizole, terfenadine
Estrogen: ethinyl estradiol
GI: cisapride
Immunosuppressant: cyclosporine, tacrolimus

Source: Ref. 130.

A recent paper [130] has also indicated that this interaction occurs in elderly patients. This is especially important since most interaction studies are typically performed in young subjects, yet the elderly are more likely to be taking the drugs that undergo this interaction. Table 8 lists drugs in different therapeutic categories that are known to participate in this grapefruit juice interaction [130]. It should also be recognized that these compounds will also interact with any other drug (e.g., ketoconazole) capable of inhibiting CYP3A4 and/or P-gp.

The above illustration should be a clear caution that components of food may interact with drugs, resulting in substantial positive or negative therapeutic effects. As will be noted later, this principle also applies to so-called dietary supplements, including botanicals, used for the treatment of numerous medical conditions.

B. Metabolism

Drug molecules may be chemically or metabolically altered at various sites along the GIT including within gut fluids, within the gut wall, and by microorganisms present in the low end of the tract. These sites are noted in Fig. 18. Several examples of enzymatic

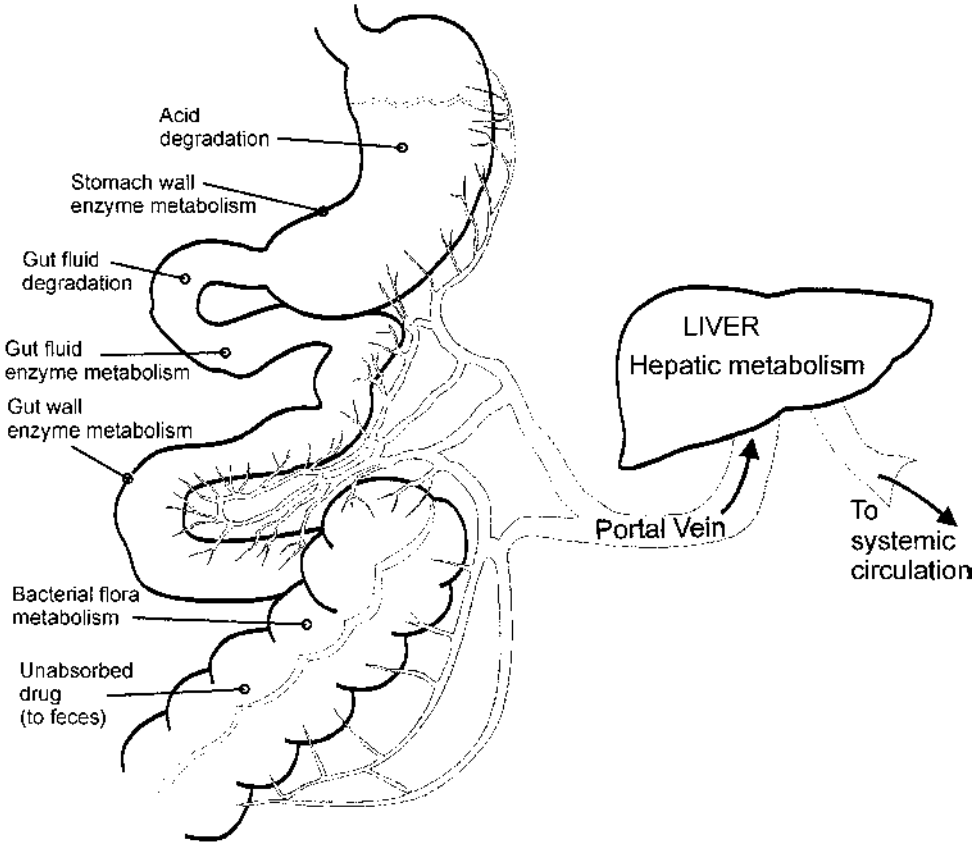


Fig. 18 Diagrammatic sketch indicating sites along the GIT where drug may be chemically altered or enzymatically metabolized. (Copyright © 2000 Saguaro Technical Press, Inc., used with permission.)

Copyright © 2002 Marcel Dekker, Inc.

alteration of certain drugs in gut fluids have been noted previously. Gut fluids contain appreciable quantities of a variety of enzymes, which are needed to accomplish digestion of food. An additional consideration in that regard is that of acid- or base-mediated drug breakdown. Numerous drugs are unstable in acidic media (e.g., erythromycin, penicillin) and will therefore degrade and provide lower effective doses depending on the pH of the gastric fluid, solubility of the drug, and the residence time of the dosage form in the stomach. Chemical modification of the drug by, for example, salt or ester formation may provide a more stable derivative whose absorption will be influenced to a smaller degree by the factors noted above. An interesting example of a prodrug that must first be acid-hydrolyzed to produce the active chemical form is clorazepate. Upon hydrolysis in the gut fluids the active form, *N*-desmethyldiazepam, is produced. In this instance, unlike the examples cited above, acid hydrolysis is a prerequisite for absorption of the pharmacologically active form. As a result, pH of gastric fluids and gastric-emptying time and variables that influence those factors are expected to affect the absorption profile of clorazapate. Greater concentrations of *N*-desmethyldiazepam are achieved at the lower gastric pH, which is consistent with the more rapid acid hydrolysis at acidic pH [131].

As noted in the preceding section, mucosal cells lining the gut wall represent a significant potential site for drug metabolism. The metabolic activity of this region has been studied by a variety of techniques, ranging from subcellular fractions to tissue homogenates to methods involving the whole living animal. Metabolic reactions include both phase I and II processes. It appears that the entire small intestine, especially the jejunum and ileum, have the greatest enzymatic activity, although most regions of the GIT can partake in metabolism. It is not a simple matter, especially in the whole animal, to distinguish among the several sites of metabolism responsible for presystemic elimination. The latter refers to all processes of chemical or metabolic alteration prior to the drug reaching the systemic circulation and which take place primarily in the gut and liver. It is this presystemic elimination that contributes to differences in drug effects as a function of route of administration and which may seriously compromise the clinical efficiency of certain drugs given orally. Many drugs have been shown to undergo metabolism in the gut wall, including those listed in Table 8 and, among others, aspirin, acetaminophen, salicylamide, *p*-aminobenzoic acid, morphine, ethanol, pentazocine, isoproterenol, L-dopa, lidocaine, and certain steroids. L-Dopa appears to be metabolized by decarboxylase enzymes present in the gastric mucosa, which, as discussed previously, suggests the importance of rapid gastric emptying to achieve maximal absorption of the unchanged compound [132].

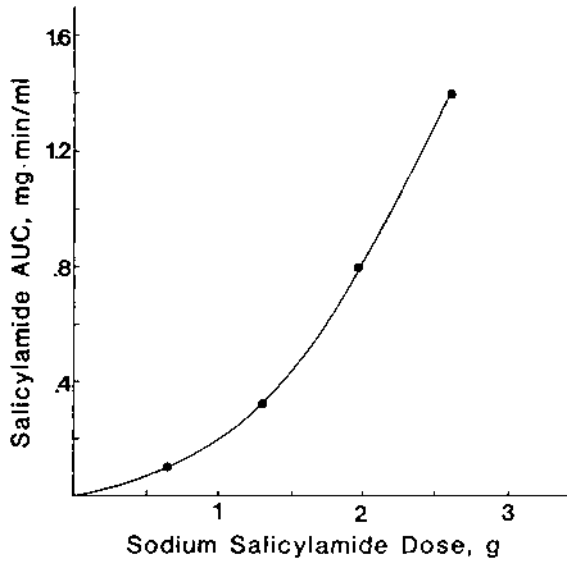


Fig. 19 Area under the salicylamide plasma concentration-time curve (AUC) as a function of the oral dose of sodium salicylamide. Each point is the average of five subjects. (Based on data from Ref. 133.)

Salicylamide and *p*-aminobenzoic acid are interesting examples because they illustrate another aspect of gut metabolism, that of saturation. In addition, factors affecting absorption rate will influence the fraction of the dose that reaches the systemic circulation in the form of intact drug. Figure 19 illustrates the relationship between the area under the salicylamide plasma concentration–time curve (AUC) as a function of the oral dose of sodium salicylamide [133]. Normally, that relationship is expected to be linear and the line should go through the origin. The curvilinearity, especially at low doses, suggests some form of presystemic elimination, which becomes saturated above a certain dose. The drug is metabolized in the gut wall to sulfate and glucuronide conjugates, although metabolism in the liver also occurs on the first pass. Extrapolation of the straight-line segment to the x-axis in Fig. 19 gives an intercept that has been referred to as the "break through" dose, which approximates the dose needed to saturate the enzyme system (about 1–1.5 g for salicylamide). Doses less than that value produce only small plasma concentrations of the unchanged drug.

Copyright © 2002 Marcel Dekker, Inc.

Another interesting aspect of this phenomenon is that the rate of drug presentation to the enzyme system will influence the fraction of the dose reaching the systemic circulation unchanged and will alter the metabolic pattern. The latter factors, therefore, will be influenced by the dosage form characteristics and the rate of gastric emptying. The more rapid the rate of absorption, the more likely the enzyme system will become saturated. The latter will result in greater plasma concentrations of unchanged drug and a metabolic pattern with a lower percentage of the drug recovered as the saturable metabolite. At any given dose, the more rapidly soluble sodium salicylamide produces greater plasma concentrations of unchanged drug compared to salicylamide. The more rapidly soluble dosage forms of saticylamide (solution and suspension) produce smaller fractions of the dose in the form of the saturable metabolite (sulfate conjugate) compared to a slowly dissolving tablet [134]. Observations similar to the latter have been made for *p*-aminobenzoic acid. In that instance, however, it was shown that a delay in gastric emptying produced a greater percentage of the dose in the form of the saturable metabolite (acetyl derivative), which is consistent with a reduced rate of absorption [135].

An interesting example of gut metabolism and gender-dependent differences is that of ethanol. Females appear to have greater blood ethanol concentrations following an oral dose compared to males given the same dose. The latter is true even if the data are corrected for weight and lean body mass differences. Ethanol is metabolized by alcohol dehydrogenase present in the gastric mucosa, and it appears that this enzyme is present in smaller quantities in females. The lower enzyme concentration results in a greater fraction of the dose not being metabolized compared to males and a subsequent higher blood ethanol concentration. Estimates of absorption suggest that females absorb about 91% of the dose as ethanol compared to about 61% in males [136]. Evidence suggests that alcohol deydrogenase activity is lower in young women (less than about 50 years of age), elderly males, and alcoholics [137]. One review has examined first-pass metabolism in oral absorption and factors that affect that process [138].

The gastrointestinal microflora provide another potential site for drug metabolism within the GIT, and it has received some attention. In normal subjects the stomach and proximal small intestine contain small numbers of microorganisms. Concentrations of these organisms increase toward the distal end of the intestine. A wide variety of aerobic and anaerobic organisms are present in the gut. The microflora, derived primarily from the environment, tend to adhere to the luminal surface of the intestine. Within an individual, the microflora tend to remain rather stable over long periods of time. The primary factors governing the numbers and kinds of microorganisms present in the tract include the activity of gastric and bile secretions, which tend to limit the growth of these organisms in the stomach and upper part of the GIT, and the propulsive motility of the intestine, which is responsible for continually cleansing the tract, thereby limiting the proliferation of microorganisms. Gastric atrophy permits increased numbers of microorganisms to pass into the small intestine, and reduced intestinal motility results in overgrowth.

Studies conducted mostly in animals indicate a wide range of primarily phase I metabolic pathways. Various drugs that are glucuronidated in the body are secreted into the intestine via the bile, and these are subject to cleavage by bacterial glucuronidase enzymes. The cleavage product may then be in a form available for absorption. Various drug conjugates may be similarly deconjugated by other bacterial enzymes (e.g., the glycine conjugate of isonicotinic acid). Although some drugs may be rendered inactive, bacterial metabolism of other drugs may give rise to more active or toxic products. The formation of the toxic compound cyclohexylamine from cyclamate is an example [139].

Salicylazosulfapyridine (sulfasalazine), which is used in treating ulcerative colitis, provides an interesting example of a drug whose metabolites represent the active pharmacological species. The parent drug is metabolized to 5-aminosalicylate and sulfapyridine. In conventional rats, both metabolites and their conjugates appear in urine and feces. In germ-free rats, however, the metabolites are not excreted. This suggests that the intestinal flora play a role in reducing the parent compound and formation of the two metabolites. If this is the case, factors influencing the population and types of intestinal microorganisms may in turn influence the absorption and effectiveness of the drug. For example, concomitant antibiotic therapy, by reducing the population of microorganisms, may prevent the parent drug from being metabolized.

The impact of presystemic elimination may be clearly understood by considering the following relationships among the several steps involved in making the drug available to the systemic circulation. The significance of this relationship is its multiplicative nature, since most of the processes are sequential. These relationships assume linear or first-order kinetics (i.e., there is no nonlinearity or saturation effects).

$$F_{\text{systemic}} = f_{\text{released}} \times f_{\text{absorbed}} \times f_{\text{hepatic}} \tag{4}$$

Copyright © 2002 Marcel Dekker, Inc.

The fraction of the orally administered dose that is bioavailable to the systemic circulation ($F_{systemic}$) is dependent upon the fraction of the dose that is released from the dosage form ($f_{released}$), multiplied by the fraction that is absorbed into the portal circulation on its way to the liver ($f_{absorbed}$; this is the fraction that escapes gut metabolism), multiplied by the fraction of the dose that escapes the hepatic first-pass effect ($f_{hepatic}$). Since this is a multiplicative process if, for example, $f_{released}$ is 1.0 and $f_{absorbed}$ is 0.5 and $f_{hepatic}$ is 0.5, then the overall $F_{systemic}$ is 0.25 (i.e., 25% of the dose reaches the systemic circulation). The absorption process can be broken down further into its component parts

$$F_{systemic} = f_{released}(f_{gut\ fluid} \times f_{gut\ wall} \times f_{gut\ flora}) f_{hepatic} \quad (5)$$

where $f_{gut\ fluid}$, $f_{gut\ wall}$ and $f_{gut\ flora}$ are the fractions of the dose of drug that escape metabolism in the gut fluid, gut wall, and gut flora, respectively. In the above example, an $f_{absorbed}$ of 0.5 represents the product of the three processes noted that affect absorption. It is often common to express the fraction surviving a specific process as one minus the extraction ratio (ER) for that process. Thus, the above relationship may be written as,

$$F_{systemic} = f_{released}(1 - ER_{gut\ fluid}) \times (1 - ER_{gut\ wall}) \times (1 - ER_{gut\ flora}) \times (1 - ER_{hepatic}) \quad (6)$$

A change in the ER value for any of the above processes will have an effect on the systemically available dose of the orally administered drug. For example, in the previous section it was noted that grapefruit juice inhibits CYP3A4 metabolism of many drugs. The factor most affected by this interaction is the intrinsic metabolic clearance of the drug by enzymes present in the gut wall. The latter would be expressed by a decrease in the value of $ER_{gut\ wall}$, and a consequent increase in $F_{systemic}$.

In addition to food- or nutrient-based interactions in the metabolism of drugs, it has become quite clear in recent years that so-called dietary supplements including botanicals have the potential to participate in such interactions. The latter observation has special relevance because of the extensive use of such products worldwide ($\sim$ \$12 billion per year in the United States alone), their easy commercial availability (no prescription required), and their common use with prescribed drugs. Furthermore, many people consider such ''natural'' products to be safe and free of any bad effects (it should be pretty easy to recall many poisons that occur in nature and are, therefore, ''natural''). Several recent examples of botanical-drug interactions should serve as a warning that we have just begun to see the tip of the iceberg. This should not be at all surprising when one considers the large number and variety of chemicals (phytochemicals) present in botanicals that are ingested for medicinal value.

St. John's wort, which has become very popular for the treatment of depression, has been shown to induce intestinal P-gp and intestinal and hepatic CYP3A4 in humans [140]. That mechanism explains the significant reduction in cyclosporin and anti-aids (e.g., indinavir) plasma concentrations. It is likely that similar effects will be noted with the compounds listed in Table 8 (although the effects noted in Table 8 are in the opposite direction of those seen in the presence of St. John's wort).

C. Disease States

Gastrointestinal disorders and disease states have the potential to influence drug absorption. Although this important area has not been explored thoroughly, numerous studies have addressed this issue. Unfortunately, many of these studies have not been correctly designed, and this has resulted in conflicting reports and our inability to reach generally valid conclusions. The majority of these studies are conducted by administration of an oral dose and measurement of the area under the plasma concentration–time curve (AUC). The latter parameter is frequently used in assessing bioavailability. The resulting AUC is compared to that from a control group of different subjects or within the same subject during the time the disorder is present and compared to the value prior to or after the disorder is resolved. The problem here is that a value for AUC depends as much on the body's ability to clear or eliminate the drug as it does on absorption (i.e., AUC = dose absorbed/clearance). Differences in the former parameter are likely to be present between subjects as well as within a subject from time to time (especially in the presence of a disease). Therefore, values of AUCs after oral dosing may lead to incorrect conclusions. To use such a value properly, one must be certain that drug clearance is not different between or within a subject. In the ideal situation an intravenous dose would be given to establish the correctness of that assumption. This is an approach, unfortunately, that is not generally used. The influence of gastrointestinal disease on drug absorption has been reviewed [141].

Elevated gastric pH is seen in subjects with achlorhydria as a result of reduced acid secretion. The

Copyright © 2002 Marcel Dekker, Inc.

absorption of tetracycline, which is most soluble at acidic pH, appears to be unaffected by achlorhydria and after surgery where the acid-secreting portion of the stomach was removed [142,143]. The absorption of clorazepate would be expected to be reduced in achlorhydria (for the reasons discussed above), but as yet the data are not conclusive. The clinical significance of altered gut pH with regard to drug absorption is not clearly established. Alterations in drug absorption due to changes in gut pH will most likely be mediated by its influence on dissolution rate.

Changes in gastric emptying are expected to influence the rate and possibly the extent of absorption for the reasons discussed previously. Emptying may be severely hampered and absorption altered soon after gastric surgery as a result of pyloric stenosis and in the presence of various disease states. Riboflavin absorption is increased in hypothyroidism and reduced in hyperthyroidism, conditions that alter gastric emptying and intestinal transit rates [144].

Diarrheal conditions may decrease drug absorption as a result of reduced intestinal residence time. The absorption of several drugs was decreased in response to lactose- and saline-induced diarrhea [145]. Digoxin absorption from tablets was impaired in one subject who developed chronic diarrhea as a result of x-ray treatment [146]. Abdominal radiation or the underlying disease has been shown to reduce digoxin and clorazepate absorption [147]. A dosage form that provides rapid drug dissolution (e.g., solution) may partially resolve this problem.

Various malabsorption syndromes are known to influence the absorption of certain nutrients. Although not thoroughly investigated, such syndromes may exert an influence on the efficacy of drug absorption. Heizer et al. [148] noted reduced absorption of digoxin in patients with sprue, malabsorption syndrome, and pancreatic insufficiency. The dosage form of digoxin, especially dissolution rate from tablets, will partially determine the influence of malabsorption states on absorption, the problem being compounded by poorly dissolving tablets. Phenoxymethyl penicillin absorption is reduced in patients with steatorrhea [149], and ampicillin and nalidixic acid absorption appears to be impaired in children with shigellosis [150].

There are a variety of other disease states whose influence on drug absorption has been reported, including cystic fibrosis, villous atrophy, celiac disease, diverticulosis, and Crohn's disease. The results of these studies are frequently divergent, and therefore general statements cannot be made. A thorough discussion of these findings is beyond the scope of this chapter, and the interested reader is referred to a recent review [141].

As most drugs are best absorbed from the small intestine, any surgical procedure that removes a substantial portion of the small intestine is likely to influence absorption; however, and as discussed previously, the characteristics of the dosage form may affect the findings. Although the procedure has fallen out of favor, intestinal bypass surgery has been used in treatment of the morbidly obese. A number of studies have been conducted to examine absorption prior to and after surgery. As noted in the introduction to this section, care must be exercised in study design and evaluation of data, as large weight loss may alter drug elimination from the body compared to the presurgery condition. Further complications include the time that the study is conducted relative to the time of surgery and the length and sections of the intestine removed.

One excellent study [151] employed intravenous and oral dosing at each of several times postsurgery (1–2 weeks, 6 and 12 months). This design permits valid conclusions about the absorption process. There was a significant reduction after surgery in ampicillin absorption but no change in propylthiouracil absorption.

In those instances where a patient's response to a drug is less than expected, and there is reason to believe that this is a result of impaired absorption due to any of the pathological conditions or disease states cited earlier, a first attempt in seeking to improve drug therapy is to optimize absorption from the GIT. To do this, a practical approach might well be to administer the drug in a form readily available for absorption. In most cases, if such a form is marketed or easily prepared, administration of a drug in solution will represent the best way to achieve maximal absorption, as this will eliminate the time for drug dissolution in the gut needed by solid oral dosage forms. When absorption cannot be sufficiently improved by use of a drug solution, alternative routes of administration must be considered (e.g., intramuscular).

D. Age

The majority of the information discussed to this point and most of the literature concerned with drug absorption involve studies performed in young, healthy (usually male) adults. In contrast, there is considerably less information concerning absorption in subjects at either end of the age spectrum (i.e., pediatric and geriatric populations). For a variety of reasons, one would expect the absorption process in the latter

Copyright © 2002 Marcel Dekker, Inc.

groups to be different from that in young adults; unfortunately, at this time there is little information to present valid general statements.

The pediatric population (neonates, infants, and children) presents a particularly difficult group in which to conduct clinical experimentation because of ethical considerations. A further complication is the rapid development of organ function, which is likely to influence results even over a relatively short experimental period (e.g., 2–4 weeks), especially in neonates and infants. An additional consideration in the latter groups is whether the neonate is premature or full-term. Most often, plasma concentration–time data are obtained after an oral dose for the purpose of estimating elimination half-life or to provide a basis for the development of a multiple dosing regimen. Such data provide very limited information about the rate or extent of absorption. Indeed, most reviews of drug disposition in the pediatric population indicate the lack of rigorous information about absorption. There have been recent initiatives to encourage the study of drugs in the pediatric population, and this will undoubtedly result in useful new information.

Gastric fluid is less acidic in the newborn than in adults since acid secretion is related to the development of the gastric mucosa. This condition appears to last for some time, as pH values similar to the adult are not reached until after about 2 years. The higher gastric fluid pH along with smaller gut fluid volume may influence dissolution rate and the stability of acid-unstable drugs. The gastric-emptying rate appears to be slow, approaching adult values after about 6 months. An interesting example in support of that suggestion is a study that examined riboflavin absorption in a 5-day-old and a 10-month-old infant [152]. The maximum urinary excretion rate was considerably greater in the infant, while excretion rate in the neonate was constant and prolonged. These data suggest more rapid absorption in the infant, while absorption in the neonate proceeds for a longer time. For the reasons discussed previously, these data suggest slower emptying and/or intestinal transit rates in the neonate (recall that riboflavin is absorbed by a specialized process high in the small intestine). Intestinal transit tends to be irregular and may be modified by the type of food ingested and the feeding pattern.

Intestinal surface area and total blood flow to the GIT are smaller than in adults and may influence the efficiency of absorption. With regard to the use of rectal suppositories, one must keep in mind that the completeness of absorption will be a function of retention time in the rectum. Since bowel movements in the neonate and infant are likely to be irregular, the retention time may limit the efficiency of absorption by that route. In light of the little information available about absorption in the young, it would seem reasonable to attempt to optimize absorption by using solution or suspensions rather than solid dosage forms.

Only in recent years has there been any substantial progress made in better understanding drug disposition in the elderly. Active research programs in gerontology have begun to provide more information about rational drug dosing in the elderly. There are a number of important and unique characteristics of the elderly that make a compelling argument for the need for such information (e.g., they ingest more drugs per capita, their percentage of the population is increasing, they suffer from more disease and physical impairments, etc.). As noted for the pediatric population, a variety of complex issues are associated with the conduct of research in the elderly. Careful consideration must be given to experimental design and data analysis. Some considerations include the appropriate definition of age, cross-sectional versus longitudinal study design, and health status of the subject [153,154].

There have been numerous statements in the literature to the effect that GI absorption in the elderly is impaired and less efficient than in young adults. Although there have been few data to support the suggestion, one basis for that statement has been the results obtained from the application of the so-called xylose tolerance test, which is often used in assessing malabsorption. This conclusion of impaired absorption in the elderly presents a good example of the need for careful study design and appropriate pharmacokinetic analysis of data. Figure 20 illustrates the results of several studies that have examined xylose absorption [155]. Most studies indicate an inverse relationship between urinary xylose recovery and age after an oral dose (lines B and C). It is this observation that has suggested reduced absorption with age. However, the same inverse relationship is found after an intravenous dose (line A), an observation that is not explained by impaired absorption but, rather, by reduced renal clearance of xylose. Line D shows the ratio of urinary recovery (oral to intravenous), which suggests that absorption is not altered with age. A more complete study in which each subject (age range, 32–85 years) received both an oral and an intravenous dose indicates no relationship between xylose bioavailability and age [156].

Copyright © 2002 Marcel Dekker, Inc.

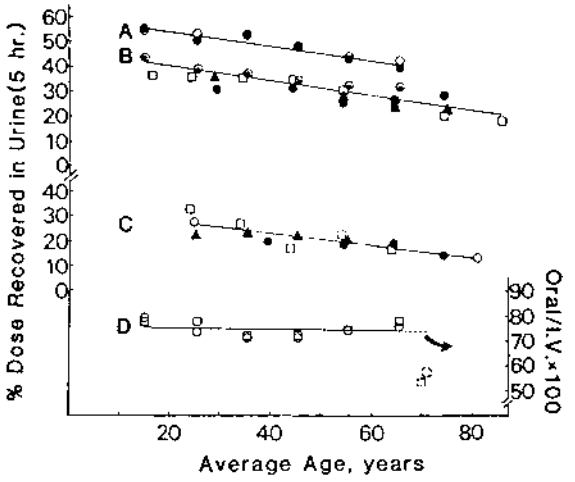


Fig. 20 Percentage xylose dose recovered in urine as a function of age after a 5 g intravenous dose (A), 5 g oral dose (B), and 25 g oral dose (C). Line D is the ratio of urinary recoveries (oral to intravenous) after 5 g doses (y-axis on right). Symbols represent data obtained from different studies. (From Ref. 155.)

There are substantial changes in a variety of physiological functions in the elderly, which may influence drug absorption [153], including a greater incidence of achlorhydria, altered gastric emptying, reduced gut blood flow, and smaller intestinal surface area. One recent example indicates that gastric pH may be an important determinant of drug absorption in the elderly [157]. Dipyridamole is a poorly water-soluble weak base whose dissolution would be optimal in an acidic environment. Elevated gastric pH due to achlorhydria (a condition more prevalent in the elderly than the young) results in impaired absorption of dipyridamole. The ingestion of glutamic acid by achlorhydric subjects improves absorption. Other factors that may influence absorption include a greater incidence of GI disease, altered nutritional intake and eating habits, and ingestion of drugs, which may affect the absorption of other drugs. Although data are still somewhat limited, the general impression is that the rate of absorption is frequently reduced, while there is little if any change in the extent of absorption [153]. This is a tentative statement that needs to be qualified for the specific drug and for the health status of the subject. For example, the absorption of drugs that undergo hepatic first-pass metabolism (e.g., propranolol) may be improved in the elderly as a consequence of reduced hepatic clearance with age.

REFERENCES

1. JD Verhoff, HE Bodde, AG deBoer, JA Bouwstra, HE Junginger, FWHM Merkus DD Breimer. Transport of peptide and protein drugs across biological membranes. Eur J Drug Metab Pharmacokin 15:83–93, 1990.
2. PA Wender, DJ Mitchell, K Pattabiraman, ET Pelkey, L Steinman, JB Rothbard. The design, synthesis, and evaluation of molecules that enable or enhance cellular uptake: Peptoid molecular transporters. Proc Natl Acad Sci USA 97:13003–13008, 2000.
3. JB Rothbard, S Garlington, Q Lin, T Kirschberg, E Kreider, PL McGrane, PA Wender, PA Khavari. Conjugation of arginine digomers to cyclosporin: A facilitates topical delivery and inhibition of inflammation. Nat Med 6:1253–1257, 2000.
4. JW Hole. Human Anatomy and Physiology. 2nd ed. Dubuque, IA: Wm. C. Brown Company, 1981, p. 414.
5. H Leonhardt. Color Atlas/Text of Human Anatomy. Vol. 2. 4th ed. New York: Thieme Medical Publishers, 1993, p. 213.
6. T Fujita, K Tanaka, J Tokunaga. SEM Atlas of Cells and Tissues. New York: Igaku-Shoin, 1981, pp. 122, 123, 129.
7. JW Hole. In: SEM Atlas of Cells and Tissues. New York: Igaku-Shoin, 1981, p. 439.
8. JS Trier. Morphology of the epithelium of the small intestine. In: CF Code, ed. Handbook of Physiology. Vol. III, Intestinal Absorption. Washington, DC: American Physiological Society,1968, p. 1133.
9. T Fujita, K Tanaka, J Tokunaga. In: CF Code, ed. Handbook of Physiology. Washington DC: American Physiological Society, 1968, p. 135.
10. GJ Tortora, NP Anagnostakos. Principles of Anatomy and Physiology. 6th ed. New York: Harper and Row, 1990, p. 770.
11. JS Fordtran, FC Rector, Jr., MF Ewton, N Soter, J Kinney. Permeability characteristics of the human small intestine. J Clin Invest 44:1935–1944, 1965.
12. JF DanielIi, H Davson. A contribution to the theory of permeability of thin films. J Cell Comp Physiol 5:495–508, 1935.
13. H Davson, JF DanielIi. The Permeability of Natural Membranes. 2nd ed. New York: Cambridge University Press, 1952.
14. LC Junqueira, J Carneiro, RO Kelley. Basic Histology. 6th ed. East Norwalk, CT: Appleton and Lange, 1989, p. 32.
15. R Hober, J Hober. Experiments on the absorption of organic solutes in the small intestine of rats. J Cell Comp Physiol 10:401–422, 1937.
16. LS Schanker. Absorption of drugs from the rat colon. J Pharmacol Exp Ther 126:283–290, 1959.
17. EM Wright, JM Diamond. Patterns of non-electrolyte permeability. Proc Royal Soc B 172:203–225, 1969.

Copyright © 2002 Marcel Dekker, Inc.

18. EM Wright, JM Diamond. Patterns of non-electrolyte permeability. Proc Royal Soc B 172:227–271, 1969.
19. JM Diamond, EM Wright. Molecular forces governing nonelectrolyte permeation through cell membranes. Proc Royal Soc B 172:273–316, 1969.
20. CA Lipinski, F Lombardo, BW Dominy, PF Feeney. Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. Adv Drug Deliv Rev 23:3–25, 1997.
21. Y Yoshimura, N Kakeya. Structure-gastrointestinal absorption relationship of penicillins. Int J Pharm 17:47–57, 1983.
22. T Sollmann. The comparative efficiency of local anesthetics. J Am Med Assoc 70:216–219, 1918.
23. JM Faulkner. Nicotine poisoning by absorption through the skin. J Am Med Assoc 100:1664–1665, 1933.
24. J Travell. The influence of the hydrogen ion concentration on the absorption of alkaloids from the stomach. J Pharmacol Exp Ther 69:21–33, 1940.
25. MH Jacobs. Some aspects of cell permeability to weak electrolytes. Cold Spring Harbor Symp Quant Biol 8:30–399, 1940.
26. PA Shore, BB Brodie, CAM Hogben. The gastric secretion of drugs: a pH partition hypothesis. J Pharmacol Exp Ther 119:361–369, 1957.
27. LS Schanker, PA Shore, BB Brodie, CAM Hogben. Absorption of drugs from the stomach. I. The rat. J Pharmacol Exp Ther 120:528–539, 1957.
28. CAM Hogben, LA Schanker, DJ Tocco, BB Brodie. Absorption of drugs from the stomach. II. The human. J Pharmacol Exp Ther 120:540–545, 1957.
29. LS Schanker, DJ Tocco, BB Brodie, CAM Hogben. Absorption of drugs from the rat small intestine. J Pharmacol Exp Ther 123:81–87, 1958.
30. CAM Hogben, DJ Tocco, BB Brodie, LS Schanker. On the mechanism of intestinal absorption of drugs. J Pharmacol Exp Ther 125:275–282, 1959.
31. JT. Doluisio, NF Billups, LW Dittert, ET Sugita, JV Swintosky. Drug absorption. I. An in situ rat gut technique yielding realistic absorption rates. J Pharm Sci 58:1196–1200, 1969.
32. LZ Benet. Biopharmaceutics as a basis for the design of drug products. In: EJ Ariens, ed. Drug Design, Vol. 4. New York: Academic Press, 1973, pp. 26–28.
33. DS Riggs. The Mathematical Approach to Physiological Problems. Baltimore: Williams & Wilkins, 1963, pp. 181–185.
34. GL Flynn, SK Yalkowsky, TJ Roseman. Mass transport phenomena and models: theoretical concepts. J Pharm Sci 63:479– 510, 1974.
35. M Mayersohn. Drug absorption. J Clin Pharmacol 27:634–638, 1987.
36. Y-L He, S Murby, G Warhurst, L Gifford, D Walker, J Ayrton, R Eastmond, M Rowland. Species differences in size discrimination in the paracellular pathway reflected by oral bioavailability of poly (ethylene glycol) and D-peptides. J Pharm Sci 87: 626–633, 1998.
37. LS Schanker, JJ Jeffrey. Active transport of foreign pyrimidines across the intestinal epithelium. Nature 190:727–728, 1961.
38. DF Evered, HG Randall. The absorption of amino acid derivatives of nitrogen mustard from rat intestine in vitro. Biochem Pharmacol 11:371–376, 1962.
39. G Levy, WJ Jusko. Factors affecting the absorption of riboflavin in man. J Pharm Sci 55:285–289, 1966.
40. M Mayersohn. Ascorbic acid absorption in man-pharmacokinetic implications. Eur J Pharmacol 19:140–142, 1972.
41. NG Gillespie, I Mena, GC Cotzias, MA Bell. Diets affecting treatment of parkinsonism with levodopa. J Am Diet Assoc 62:525–528, 1973.
42. DN Wade, PT Mearrick, JL Morris. Active transport of L-Dopa in the intestine. Nature 242:463–465, 1973.
43. MH van Woert. Phenylalanine and tyrosine metabolism in Parkinson's disease treated with levodopa. Clin Pharmacol Ther 12:368–375, 1971.
44. H Lennernas, D Nilsson, S-M Aquilonius, D Ahrenstedt, L Knutson, LK Paalzow. The effect of L-leucine on the absorption of levodopa, studied by regional jejunal perfusion in man. Br J Clin Pharmacol 35:343–250, 1993.
45. J Zimmerman. Methotrexate transport in the human intestine. Biochem Pharmacol 43:2377–2383, 1992.
46. J-F Westphal, J-H Trouvin, A Deslandes, C Carbon. Nifedipine enhances amoxicillin absorption kinetics and bioavailability in human. J Pharmacol Exp Ther 255:312–317, 1990.
47. C Duverne, A Bouten, A Deslandes, J-F Westphal, J-H Trouvin, R Farinotti, C Carbon. Modification of cifixime bioavailability by nifedipine in humans: involvement of the dipeptide carrier system. Antimicrob Agents Chemother 36:2462–2467, 1992.
48. BH Stewart, AR Kugler, PR Thompson, NN Bockbrader. A saturable transport mechanism in the intestinal absorption of gabapentin is the underlying cause of the lack of proportionality between increasing dose and drug levels in plasma. Pharmaceut Res 10:276–281, 1993.
49. II Gottesman, I Pastan. Biochemistry of multidrug resistance mediated by the multidrug transporter. Ann Rev Biochem 62:385–427, 1993.
50. Y Tanigawara. Role of p-glycoprotein in drug disposition. Ther Drug Monitor 22:137–140, 2000.
51. JV Asperen, OV Tellingen, JH Beijnen. The pharmacological role of *p*-glycoprotein in the intestinal epithelium. Pharmacol Res 37:429–435, 1998.

Copyright © 2002 Marcel Dekker, Inc.

52. V Wacher, JA Silverman, Y Zhang, LZ Benet. Role of *p*-glycoprotein and cytochrome P4503A in limiting oral absorption of peptides and peptidomimetics. J Pharm Sci 87:1322–1330, 1998.
53. J Wacher, L Salphati, LZ Benet. Active secretion and enterocyte drug metabolism barriers to drug absorption. Adv Drug Del Rev 20:99–112, 1996.
54. A Sparreboom, J van Asperen, U Mayer, AH Schinkel, JW Smit, DKF Meijer, P Borst, WJ Nooijen, JH Beijnen, O van Tillingen. Limited oral bioavailability of taxol and active epithelial secretion of paclitaxel (Taxol) caused by *p*-glycoprotein in the intestine. Proc Nat Acad Sci USA 94:2031–2035, 1997.
55. K Westphal, A Weinbrenner, T Giessmann, M Stuhr, G Franke, M Zschiesche, R Oertel, B Terhaag, HK Kroemer, W Siegmund. Oral bioavailability of digoxin is enhanced by talinolol: evidence for involvement of intestinal p-glycoprotein. Clin Pharmacol Ther 68:6–12, 2000.
56. JB Dressman, GL Amidon, C Reppas, VP Shah. Dissolution testing as a prognostic tool for oral drug absorption: immediate release dosage forms. Pharm Res 15:11–22, 1998.
57. TR Bates, JM Young, CM Wu, HA Rosenberg. pH-dependent dissolution rate of nitrofurantoin from commercial suspensions, tablets and capsules. J Pharm Sci 63:643–645, 1974.
58. G Levy, BA Sahli. Comparison of the gastrointestinal absorption of aluminum acetylsalicylate and acetylsalicylic acid in man. J Pharm Sci 51:58–62, 1962.
59. G Levy, JA Procknal. Unusual dissolution behavior due to film formation. J Pharm Sci 51:294, 1962.
60. CE Blezek, JL Lach, JK Guillory. Some dissolution aspects of ferrous sulfate tablets. Am J Hosp Pharm 27:533–539, 1970.
61. RG Crounse. Human pharmacology of griseofulvin. The effect of fat intake on gastrointestinal absorption. J Invest Dermatol 37:529–533, 1961.
62. M Kraml, J Dubuc, D Beall. Gastrointestinal absorption of griseofulvin. I. Effect of particle size, addition of surfactants and corn oil on the level of griseofulvin in the serum of rats. Can J Biochem Physiol 40:1449–1451, 1962.
63. AJ Moes. Gastroretentive dosage forms. Crit Rev Therap Drug Carrier Sys 10:143–195, 1993.
64. H Minami, RW McCallum. The physiology and pathophysiology of gastric emptying in humans. Gastroenterol 86:1592–1610, 1984.
65. AG Renwick, CH Ahsan, VF Challenor, R Daniels, BS MacKlin, DG Waller, CF George. The influence of posture on the pharmacokinetics of orally administered nifedipine. Br J Clin Pharmacol 34:332–336, 1992.
66. A Hopkins. The pattern of gastric emptying: a new view of old results J Physiol 182:144–149, 1966.
67. JD Elashoff, TJ Reedy, JH Meyer. Analysis of gastric emptying data. Gastroenterology 83:1306–1312, 1982.
68. RH Blythe, GM Grass, DR MacDonnell. The formulation and evaluation of enteric coated aspirin tablets. Am J Pharm 131:206–216, 1959.
69. M Alpsten, C Bogentoft, G Ekenved, L Solveil. Gastric emptying and absorption of acetylsalicylic acid administered as enteric-coated micro-granules. Eur J Clin Pharmacol 22:57–61, 1982.
70. P Welling. Effects of food on drug absorption. Ann Rev Nutr 16:383–415, 1996.
71. D Fleisher, C Li, Y Zhou, L-H Pao, A Karim. Drug, meal and formulation interactions influencing drug absorption after oral administration. Clin Pharmacokin 36:233–254, 1999.
72. WN Charman, CJH Porter, S Mithani, JB Dressman. Physicochemical and physiological mechanisms for the effects of food on drug absorption: the role of lipids and pH. J Pharm Sci 86:269–282, 1997.
73. L Hendeles, M Weinberger, G Milavetz, M Hill III, L Vaughan. Food-induced 'dose-dumping' from a once-a-day theophylline product as a cause of theophylline toxicity. Chest 87:758–765, 1985.
74. MN Gai, A Isla, MT Andonaegui, AM Thielemann, C Seitz. Evaluation of the effect of 3 different diets on the bioavailability of 2 sustained release theophylline matrix tablets. Int J Clin Pharmacol Ther 35:565–571, 1997.
75. TR Bates, JA Sequeira, AV Tembo. Effect of food on nitrofurantoin absorption. Clin Pharmacol Ther 16:63–68, 1974.
76. HA Rosenberg, TR Bates. The influence of food on nitrofurantoin bioavailability. Clin Pharmacol Ther 20:227–232, 1976.
77. S Yung, M Mayersohn, JB Robinson. Ascorbic acid absorption in man: influence of divided dose and food. Life Sci 28:2505–2511, 1981.
78. CW Meuhlberger. The physiological action of alcohol. J Am Med Assoc 167:1842–1845, 1958.
79. J Fermaglich, S O'Doherty. Effect of gastric motility on levodopa. Dis Nerv Syst 33:624–625, 1972.
80. M Siurala, O Mustala, J Jussila. Absorption of acetylsalicylic acid by a normal and strophic gastric mucosa. Scand J Gastroenterol 4:269–273, 1969.
81. M Kekki, K Pyorola, O Justala, H Salmi, J Jussila, M Siurala. Multicompartment analysis of the absorption kinetics of warfarin from the stomach and small intestine. Int J Clin Pharmacol 5:209–214, 1971.
82. RC Heading, J Nimmo, LF Prescott, P Tothill. The dependence of paracetamol absorption on the rate of gastric emptying. Br J Pharmacol 47:415–421, 1973.
83. LF Prescott, WS Nimmo, RC Heading. Drug absorption interactions. In: DG Grahame-Smith, ed. Drug Interactions. Baltimore: Macmillan, 1977, p. 45.
84. KS Channer, J Virjee. Effect of posture and drink volume on the swallowing of capsules. Br Med J 285:1702, 1982.

Copyright © 2002 Marcel Dekker, Inc.

85. H Hey, F Jorgenson, K Sorensen, H Hasselbalch, T Wamberg. Oesophageal transit of six commonly used tablets and capsules. Br Med J 285:171–179, 1982.
86. KS Channer, CJC Roberts. Effect of delayed esophageal transit on acetaminophen absorption. Clin Pharmacol Ther 37:72–76, 1985.
87. SS Davis, JG Hardy, JW Fara. Transit of pharmaceutical dosage forms through the small intestine. Gut 27:886–892, 1986.
88. JL Madsen. Effects of gender, age, and body mass index on gastrointestinal transit times. Digest Dis Sci 37:1548–1553, 1992.
89. JMC Price, SS Davis, IR Wilding. The effect of fibre on gastrointestinal transit times in vegetarians and omnivores. Int J Pharmaceut 76:123–141, 1991.
90. F Pirk. Changes in the motility of the small intestine in digestive disorders. Gut 8:486–490, 1967.
91. AC Guyton. Basic Human Physiology: Normal Functions and Mechanisms of Disease. Philadelphia: WB Saunders, 1971, p. 428.
92. E Parry, R Shields, AC Turnbull. Transit time in the small intestine in pregnancy. J Obstet Gynaecol 77:900–901, 1970.
93. AM Metcalf, SF Phillips, AR Zinsmeister, RL MacCarty, RW Beart, BG Wolff. Simplified assessment of segmental colonic transit. Gastroenterology 92:40–47, 1987.
94. PJ Watts, L Barrow, KP Steed, CG Wilson, RC Spiller, CD Melia, MC Davis. The transit rate of different sized model dosage forms through the human colon and the effects of a lactulose-induced catharsis. Int J Pharmaceut 87:215–221, 1992.
95. DA Akin, SS Davis, RA Sparrow, IR Wilding. Colonic transit of different sized tablets in healthy subjects. J Controlled Release 23:147–156, 1993.
96. JMC Price, SS Davis, RA Sparrow, IR Wilding. The effect of meal composition on the gastrocolonic response: implications for drug delivery to the colon. Pharmaceut Res 10:722–726, 1993.
97. CJ McDougall, R Wong, P Scudera, M Lesser, JJ De Cosse. Colonic mucosal pH in humans. Digest Dis Sci 38:542–545, 1993.
98. DR Friend. Colon-specific drug delivery. Adv Drug Del Rev 7:149–199, 1991.
99. KH Antonin, V Saano, P Bleck, J Hastewell, R Fox, P Low, M MacKay. Colonic absorption of human calcitonin in man. Clin Sci 83:627–631, 1992.
100. MF Williams, GF Dukes, W Heizer, Y-H Han, DJ Hermann, T Lampkin, LJ Hak. Influence of gastrointestinal site of drug delivery on the absorption characteristics of ranitidine. Pharmaceut Res 9:1190–1194, 1992.
101. KKH Chan, A Buch, RD Glazer, VA John, WH Barr. Site-differential gastrointestinal absorption of benazepril hydrochloride in health volunteers. Pharmaceut Res 11:432–437, 1994.
102. L d'Agay-Abensour, A Fjellestad-Paulsen, P Hoglund, Y Ngo, O Paulsen, JC Rambaud. Absolute bioavailability of an aqueous solution of 1-deamino-8-D-arginine vasopressin from different regions of the gastrointestinal tract in man. Eur J Clin Pharmacol 44:473–476, 1993.
103. PE Warner, KLR Brouwer, EK Hussey, GE Dukes, WD Heizer, KH Donn, IM Davis, JR Powell. Sumatriptan absorption from different regions of the human gastrointestinal tract. Pharmaceut Res 12:138–143, 1995.
104. AG De Boer, F Moolenaar, LGJ de Leede, DD Breimer. Rectal drug administration: clinical pharmacokinetic considerations. Clin Pharmacokinet 7:285–311, 1982.
105. D Winne. The influence of blood flow on the absorption of L- and D-phenylalanine from the jejunum of the rat. Arch Pharmacol 277:113–138, 1973.
106. D Winne. Formal kinetics of water and solute absorption with regard to intestinal blood flow. J Theoret Biol 2:1–18, 1970.
107. D Winne, J Remischovsky. Intestinal blood flow and absorption of nondissociable substances. J Pharm Pharmacol 22:640–641, 1970.
108. WG Crouthamel, L Diamond, LW Dittert, JT Doluisio. Drug absorption.VII. Influence of mesenteric blood flow on intestinal drug absorption in dogs. J Pharm Sci 64:664–671, 1975.
109. M Rowland, S Riegelman, PA Harris, SD Sholkoff. Absorption kinetics of aspirin in man following oral administration of an aqueous solution. J Pharm Sci 61:379–385, 1972.
110. E Kruger-Thiemer. Pharmacokinetics and dose-concentration relationships. In: EJ Ariens, ed. Physico-Chemical Aspects of Drug Action. Elmsford, NY: Pergamon Press, 1968, pp. 63–113.
111. GC Oliver, R Tazman, R Frederickson. Influence of congestive heart failure on digoxin level. In: O Storstein, ed. Symposium on Digitalis. Oslo: Gyldendal Norsk Forlag, 1973, pp. 336–347.
112. NL Benowitz. Effects of cardiac disease on pharmacokinetics: pathophysiologic considerations. In: LZ Benet, N Massoud, JG Gainbertoglio, eds. Pharmacokinetics Basis for Drug Treatment. New York: Raven Press, 1984, pp. 89–103.
113. MM Drucker, SJ Blondheim, L Wislicki. Factors affecting acetylation in vivo of *para*-aminobenzoic acid by human subjects. Clin Sci 27:133–141, 1964.
114. G Levy, T. Matsuzawa. Pharmacokinetics of salicylamide elimination in man. J Pharmacol Exp Ther 156:285–293, 1967.
115. KE Price, Z Zolli, Jr, JC Atkinson, HG Luther. Antibiotic inhibitors. I. The effect of certain milk constituents. Antibiot Chemother 7:672–688, 1957.
116. KE Price, Z Zolli, Jr, JC Atkinson, HG Luther. Antibiotic inhibitors. II. Studies on the inhibitory action

Copyright © 2002 Marcel Dekker, Inc.

of selected divalent cations for oxytetracyctine. Antibiot Chemother 7:689–701, 1957.
117. J Scheiner, WA Altemeier. Experimental study of factors inhibiting absorption and effective therapeutic levels of declomycin. Surg Gynecol Obstet 114:9–14, 1962.
118. PJ Neuvonen, H Turakka. Inhibitory effect of various iron salts on the absorption of tetracycline in man. Eur J Clin Pharmacol 7:357–360, 1974.
119. JG Wagner. Biopharmaceutics: absorption aspects. J Pharm Sci 50:359–387, 1961.
120. P Singh, JK Guillory, TD Sokoloski, LZ Benet, VN Bhatia. Effect of inert tablet ingredients on drug absorption. I. Effect of polyethylene glycol 4000 on the intestinal absorption of four barbiturates. J Pharm Sci 55:63–68, 1966.
121. S Riegelman, WJ Crowell. The kinetics of rectal absorption. II. The absorption of anions. J Am Pharm Assoc Sci Ed 47:123–127, 1958.
122. S Riegelman, WJ Crowell. The kinetics of rectal absorption. III. The absorption of undissociated molecules. J Am Pharm Assoc Sci Ed 47:127–133, 1958.
123. K Kakemi, T Arita, S Muranishi. Absorption and excretion of drugs. XXVII. Effect of nonionic surface-active agents on rectal absorption of sulfonamides. Chem Pllarm Bull 13:976–985, 1965.
124. DL Sorby. Effect of adsorbents on drug absorption. I. Modification of promazine absorption by activated attapulgite and activated charcoal. J Pharm Sci 54:677–683, 1965.
125. JG Wagner. Design and data analysis of biopharmaceutical studies in man. Canad J Pharm Sci 1:55–68, 1966.
126. DG Bailey, JD Spence, C Munoz, JMO Arnold. Interaction of citrus juices with felodipine and nifedipine. Lancet 337:268–269, 1991.
127. SD Hall, KE Thummel, PB Watkins, KS Lown, LZ Benet, MF Paine, RR Mayo, DK Turgeon, DG Bailey, RJ Fontana, SA Wrighton. Molecular and physical mecahnisms of first-pass extraction. Drug Metab Dispos 27:161–166, 1999.
128. DG Bailey, JMO Arnold, JD Spence. Inhibitors in the diet: grapefruit juice-drug interactions. In: RH Levy, KE Thummel, WE Trager, PD Hansten, M Eichelbaum, eds. Metabolic Drug Interactions: New York: Lippincott Williams & Wilkins, 2000, pp. 661–667.
129. P Schmeidlin-Ren, DJ Edwards, ME Fitzsimmons, K He, KS Lown, PM Woster, A Rahman, KE Thummel, JM Fisher, PF Hollenberg, PB Watkins. Mechanisms of enhanced oral availability of CYP3A4 substrates by grapefruit constituents: decreased enterocyte CYP3A4 concentration and mechanism-based inactivation by furanocoumarins. Drug Metab Dispos 25:1228–1233, 1997.
130. GK Dresser, DG Bailey, SG Carruthers. Grapefruit juice-felodipine interaction in the elderly. Clin Pharmacol Ther 68:28–34, 2000.
131. CW Abruzzo, T Macasieb, R Weinfeld, JA Rider, SA Kaplan. Changes in the oral absorption characteristics in man of dipotassium clorasepate at normal and elevated gastric pH. J Pharmacokinet Biopharm 5:377–390, 1977.
132. L Rivera-Calimlim, CA Dujovne, JP Morgan, L Lasagna, JR Bianchine. Absorption and metabolism of L-dopa by the human stomach. Eur J Clin Invest 1:313–320, 1971.
133. L Fleckenstein, GR Mundy, RA Horovitz, JM Mazzullo. Sodium salicylamide: relative bioavailability and sujective effects. Clin.Pharmacol Ther 19:451–458, 1976.
134. G Levy, T Matsuzawa. Pharmacokinetics of salicylamide in man. J Pharmacol Ther 156:285–293, 1967.
135. MM Drucker, SJ Blondheim, L Wislicki, Factors affecting acetylation in vivo of *para*-aminobenzoic acid by human subjects. Clin Sci 27:133–141, 1964.
136. M Frezza, C Di Padova, G Pozzato, M Terpin, E Baraona, CS Lieber. High blood alcohol levels in women: the rold of decreased gastric alcohol dehydrogenase and first-pass metabolism. N Engl J Med 322:95–99, 1990.
137. HK Seitz, G Egerer, UA Simanowski, R Waldherr, R Eckey, DP Agarwal, HW Goedde, J-P von Wartourg. Human gastric alcohol dehydrogenase activity: effect of age, sex and alcoholism. Gut 34:1433–1437, 1993.
138. YK Tam. Individual variation in first-pass metabolism. Clin Pharmacokin 25:300–332, 1993.
139. RL Smith. The role of the gut flora in the conversion of inactive compounds to active metabolites. In: A Symposium on Mechanisms of Toxicity. WN Aldridge, ed. New York: Macmillan, 1971, pp. 228–244.
140. D Durr, B Stieger, GA Kullak-Ublick, KM Rentsch, HC Steinert, PJ Meier, K Fattinger. St. John's wort induces intestinal *p*-glycoprotein/MDR1 and intestinal and hepatic CYP3A45. Clin Pharmacol Ther 68:598–604, 2000.
141. WA Ritschel, DD Denson. Influence of disease on bioavailability. In: PG Welling, FLS Tse, SV Dighe, eds. Pharmaceutical Bioequivalence. New York: Marcel Dekker, 1991, pp. 67–115.
142. PA Kramer, DJ Chapron, J Benson, SA Merik. Tetracycline absorption in elderly patients with achlorhydria. Clin Pharmacol Ther 23:467–472, 1978.
143. HR Ochs, DJ Greenblatt, HJ Dengler. Absorption of oral tetracycline in patients with Billroth-II gastrectomy. J Pharmacokinet Biopharm 6:295–303, 1978.
144. G Levy, MH MacGillivray, JA Procknal. Riboflavin absorption in children with thyroid disorders. Pediatrics 50:896–900, 1972.

Copyright © 2002 Marcel Dekker, Inc.

145. LF Prescott. Gastrointestinal absorption of drugs. Med Clin North Am 58:907–916, 1974.
146. WJ Jusko, DR Conti, A Molson, P Kuritzky, J Giller, R Schultz. Digoxin absorption from tablets and elixir: the effect of radiation-induced malabsorption. J Am Med Assoc 230:1554–1555, 1974.
147. GH Sokol, DJ Greenblatt, BL Lloyd, A Gecrgotas, MD Allen, JS Harmatz, TW Smith, RI Shader. Effect of abdominal radiation therapy on drug absorption in humans. J Clin Pharmacol 18:388–396, 1978.
148. WD Heizer, TW Smith, SE Goldfinger. Absorption of digoxin in patients with malabsorption syndromes. N Engl J Med 285:257–259, 1971.
149. AE Davis, RC Pirola. Absorption of phenoxymethyl penicillin in patients with steatorrhea. Aust Ann Med 17:63–65, 1968.
150. JD Nelson, S Shelton, HT Kusmiesz, KC Haltalin. Absorption of ampicillin and nalidixic acid in infants and children with acute shigellosis. Clin Pharmacol Ther 13:879–886, 1972.
151. JP Kampmann, H Klein, B Lumholtz, JEM Hansen. Ampicillin and propylthiouracil pharmacokinetics in intestinal bypass patients followed up to a year after operation. Clin Pharmacokinet 9:168–176, 1984.
152. WJ Jusko, N Khanna, G Levy, L Stern, SJ Yaffe. Riboflavin absorption and excretion in the neonate. Pediatrics 45:945–949, 1970.
153. M Mayersohn. Special pharmacokinetic considerations in the elderly. In: WE Evans, JJ Schentag, WJ Jusko, eds. Applied Pharmacokinetics: Principles of Therapeutic Drug Monitoring. 3rd ed. Vancouver, WA: Applied Therapeutics, 1992, pp. 9-1–9-43.
154. M Mayersohn. Pharmacokinetics in the elderly. Environ Health Perspect 102 (suppl 11):119–124, 1994.
155. M Mayersohn, The "xylose test" to assess gastrointestinal absorption in the elderly: a pharmacokinetic evaluation of the literature. J Gerontol 37:300–305, 1982.
156. SL Johnson, M Mayersohn, KA Conrad. Gastrointestinal absorption as a function of age: xylose absorption in healthy adult subjects. Clin Pharmacol Ther 38:331–335, 1985.
157. TL Russell, RR Berardi, JL Barnett, TL O'Sullivan, JG Wagner, JB Dressman. pH-related changes in the absorption of dipyridamole in the elderly. Pharmaceut Res 11:136–143, 1994.

Copyright © 2002 Marcel Dekker, Inc.

Chapter 3

Pharmacokinetics

David W. A. Bourne

University of Oklahoma, Oklahoma City, Oklahoma

I. INTRODUCTION

Drug therapy is a dynamic process. When a drug product is administered, absorption usually proceeds over a finite time interval, and distribution, metabolism, and excretion (ADME) of the drug and its metabolites proceed continuously at various rates. The relative rates of these "ADME processes" determine the time course of the drug in the body, most importantly at the receptor sites that are responsible for the pharmacological action of the drug.

The usual aim of drug therapy is to achieve and maintain effective concentrations of drug at the receptor site. However, the body is constantly trying to eliminate the drug, and, therefore, it is necessary to balance absorption against elimination so as to maintain the desired concentration. Often the receptor sites are tucked away in a specific organ or tissue of the body, such as the central nervous system, and it is necessary to depend upon the blood supply to distribute the drug from the site of administration, such as the gastrointestinal tract, to the site of action.

Since the body may be viewed as a very complex system of compartments, at first it might appear to be hopeless to try to describe the time course of the drug at the receptor sites in any mathematically rigorous way. The picture is further complicated by the fact that, for many drugs, the locations of the receptor sites are unknown. Fortunately, body compartments are connected by the blood system, and distribution of drugs among the compartments usually occurs much more rapidly than absorption or elimination of the drug. The net result is that the body behaves as a single homogeneous compartment with respect to many drugs, and the concentration of the drug in the blood directly reflects or is proportional to the concentration of the drug in all organs and tissues. Thus, it may never be possible to isolate a receptor site and determine the concentration of drug around it, but the concentration at the receptor site usually can be controlled if the blood concentration can be controlled.

The objective of pharmacokinetics is to describe the time course of drug concentrations in blood in mathematical terms so that (a) the performance of pharmaceutical dosage forms can be evaluated in terms of the rate and amount of drug they deliver to the blood, and (b) the dosage regimen of a drug can be adjusted to produce and maintain therapeutically effective blood concentrations with little or no toxicity. The primary objective of this chapter will be to describe the graph paper and calculator-level mathematical tools needed to accomplish these aims when the body behaves as a single homogeneous compartment and when all pharmacokinetic processes obey first-order kinetics.

On some occasions, the body does not behave as a single homogeneous compartment, and multicompartment pharmacokinetics are required to describe the time course of drug concentrations. In other instances certain pharmacokinetic processes may not obey first-order kinetics and saturable or nonlinear models may be required. Additionally, advanced pharmacokinetic analyses require the use of various computer programs, such as those listed on the website http://www.boomer.org/pkin/soft.html.

Copyright © 2002 Marcel Dekker, Inc.

Readers interested in such advanced topics are referred to a number of texts that describe these more complex pharmacokinetic models in detail [1–5] and to the website http://www.boomer.org/pkin/.

II. PRINCIPLES OF FIRST-ORDER KINETICS

A. Definition and Characteristics of First-Order Processes

The science of kinetics deals with the mathematical description of the rate of the appearance or disappearance of a substance. One of the most common types of rate processes observed in nature is the first-order process in which the rate is dependent upon the concentration or amount of only one component. An example of such a process is radioactive decay in which the rate of decay (i.e., the number of radioactive decompositions per minute) is directly proportional to the amount of undecayed substance remaining. This may be written mathematically as follows:

$$\text{Rate of radioactivity decay} \propto [\text{undecayed substance}] \tag{1}$$

or

$$\text{Rate of radioactive decay} = k\,(\text{undecayed substance}) \tag{2}$$

where k is a proportionality constant called the first-order rate constant.

Chemical reactions usually occur through collision of at least two molecules, very often in a solution, and the rate of the chemical reaction is proportional to the concentrations of all reacting molecules. For example, the rate of hydrolysis of an ester in an alkaline buffered solution depends upon the concentration of both the ester and hydroxide ion:

$$\text{Ester} + \text{OH}^- \rightarrow \text{Acid}^- + \text{Alcohol} \tag{3}$$

The rate of hydrolysis may be expressed as follows:

$$\text{Rate of hydrolysis} \propto [\text{Ester}] \bullet [\text{OH}^-] \tag{4}$$

or

$$\text{Rate of hydrolysis} = k[\text{Ester}] \bullet [\text{OH}^-] \tag{5}$$

where k is the proportionality constant called the second-order rate constant.

But in a buffered system, $[\text{OH}^-]$ is constant. Therefore, at given pH, the rate of hydrolysis is dependent only on the concentration of the ester and may be written:

$$\text{Rate of hydrolysis}_{(\text{pH})} = k^* \bullet [\text{Ester}] \tag{6}$$

where k^* is the pseudo-first-order rate constant at the pH in question. (The pseudo-first-order rate constant, k^*, is the product of the second-order rate constant and the hydroxide ion concentration: $k^* = k \bullet [\text{OH}^-]$).

Fortunately, most ADME processes behave as pseudo-first-order processes—not because they are so simple, but because everything except the drug concentration is constant. For example, the elimination of a drug from the body may be written as follows:

$$\begin{bmatrix}\text{Drug}\\ \text{in}\\ \text{body}\end{bmatrix} + \begin{bmatrix}\text{Enzymes;}\\ \text{membranes;}\\ \text{pH; protein}\\ \text{binding; etc.}\end{bmatrix} \longrightarrow \begin{matrix}\text{Metabolized or}\\ \text{excreted drug}\end{matrix} \tag{7}$$

If everything except the concentration of drug in the body is constant, the elimination of the drug will be a pseudo-first-order process. This may seem to be a drastic oversimplification, but most in vivo drug processes, in fact, behave as pseudo-first-order processes.

B. Differential Rate Expressions

In the previous discussion of radioactive decay, it was noted that the rate of decay is directly proportional to the amount of undecayed substance remaining. In a solution of a radioactive substance, a similar relationship would hold for the concentration of undecayed substance remaining. If a solution of a radioactive substance were allowed to decay and a plot were constructed of the concentration remaining versus time, the plot would be a curve such as that shown in Fig. 1.

In this system, the rate of decay might be expressed as a change in concentration per unit time, $\Delta C/\Delta t$, which corresponds to the slope of the line. But the line in Fig. 1 is curved, which means that the rate is constantly changing and therefore cannot be expressed in terms of a finite time interval. By resorting to differential calculus, it is possible to express the rate of decay in terms of an infinitesimally small change in concentration (dC) over an infinitesimally small time interval (dt). The resulting function, dC/dt, is the slope of the line, and it is this function that is proportional to concentration in a first-order process. Thus,

Copyright © 2002 Marcel Dekker, Inc.

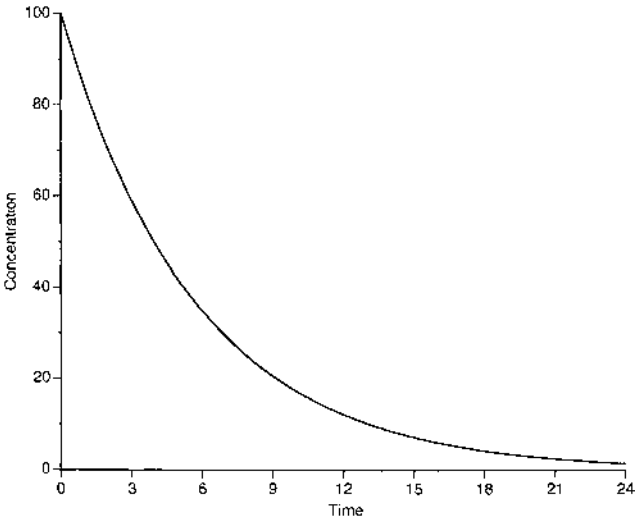


Fig. 1 Plot of concentration remaining versus time for a first-order process (e.g., radioactive decay).

$$\text{Rate} = \frac{dC}{dt} = -kC \tag{8}$$

The negative sign is introduced because the concentration is falling as time progresses.

Equation (8) is the differential rate expression for a first-order reaction. The value of the rate constant, k, could be calculated by determining the slope of the concentration versus time curve at any point and dividing by the concentration at that point. However, the slope of a curved line is difficult to measure accurately, and k can be determined much more easily using integrated rate expressions.

C. Integrated Rate Expressions and Working Equations

Equation (8) can be rearranged and integrated as follows:

$$\frac{dC}{C} = -k\,dt \tag{9}$$

$$\int \frac{dC}{C} = -k \int dt$$

$$\ln C = -kt + \text{constant} \tag{10}$$

where ln C is the natural logarithm (base e) of the concentration.

The constant in Eq. (10) can be evaluated at zero time when $kt = 0$ and $C = C_0$, the initial concentration. Thus,

$$\ln C_0 = \text{constant}$$

and

$$\log C = -kt + \ln C_0 \tag{11}$$

Equation (11) is the integrated rate expression for a first-order process and can serve as a working equation for solving problems. It is also in the form of the equation of a straight line:

$$y = mx + b$$

Therefore, if ln C is plotted against t, as shown in Fig. 2, the plot will be a straight line with an intercept (at $t = 0$) of ln C_0, and the slope of the line (m) will be $-k$. Such plots are commonly used to determine the order of a reaction; that is, if a plot of ln C versus time is a straight line, the reaction is assumed to be a first-order or pseudo-first-order process.

The slope of the line and the corresponding value of k for a plot such as that shown in Fig. 2 may be calculated using the following equation:

$$\text{slope}(m) = \frac{\ln C_1 - \ln C_2}{t_1 - t_2} = -k \tag{12}$$

Example. A solution of ethyl acetate in pH 10.0 buffer (25°C) 1 hour after preparation was found to contain 3 mg/mL. Two hours after preparation, the solution contained 2 mg/mL. Calculate the

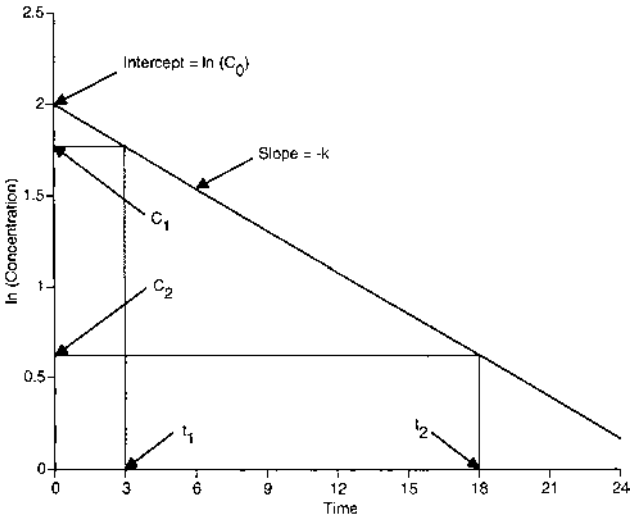


Fig. 2 Plot of ln (Concentration remaining) versus time for a first-order process.

Copyright © 2002 Marcel Dekker, Inc.

pseudo-first-order rate constant for hydrolysis of ethyl acetate at pH 10.0 (25°C).

$$\text{slope}(m) = \frac{\ln 3 - \ln 2}{(1-2)\,\text{h}} = -k$$
$$= \frac{1.099 - 0.693}{-1}$$
$$= -0.406\,\text{h}^{-1} = -k$$

$$k = 0.406\,\text{h}^{-1}$$

Note that since ln C is dimensionless, the rate constant, k, has the dimensions of reciprocal time (i.e., day^{-1}, h^{-1}, min^{-1}, s^{-1}, etc.).

Another useful working equation can be obtained by rearranging Eq. (11) as follows:

$$\begin{aligned} &\ln C - \ln C_0 = -kt \\ &\ln C_0 - \ln C = kt \\ &\ln \frac{C_0}{C} = kt \end{aligned} \tag{13}$$

Equation (13) shows that since k is a constant for a given process, the value of t is determined solely by the ratio C_0/C. For example, when C_0/C is equal to 2, the value t will be the same, no matter what was the value of the initial concentration (C_0).

Example. For the ethyl acetate hydrolysis above ($k = 0.406\,\text{h}^{-1}$), if $C_0 = 3\,\text{mg/mL}$, when would $C = 1.5\,\text{mg/mL}$?

$$\ln \frac{C_0}{C} = kt$$

$$\ln \frac{3}{1.5} = 0.406 \times t$$

$$\ln 2 = 0.406 \times t$$

$$0.693 = 0.406 \times t$$

$$t = 1.71\,\text{h}$$

If $C_0 = 1.5\,\text{mg/mL}$, when would $C = 0.75\,\text{mg/mL}$?

$$\ln \frac{C_0}{C} = kt$$

$$\ln \frac{1.5}{0.75} = 0.406 \times t$$

$$\log 2 = 0.406 \times t$$

$$t = 1.71\,\text{h}$$

The time required for the concentration to fall to $C_0/2$ is called the half-life, and the foregoing example shows that the half-life for a first-order or pseudo-first-order process is a constant throughout the process; it also demonstrates that a first-order process theoretically never reaches completion, since even the lowest concentration would only fall to half its value in one half-life.

Table 1 Approach to Completeness with Increasing Half-Lives

Number of half-lives elapsed	Initial concentration remaining (%)	"Completeness" of process (%)
0	100.0	0.0
1	50.0	50.0
2	25.0	75.0
3	12.5	87.5
4	6.25	93.75
5	3.13	96.87
6	1.56	98.44
7	0.78	99.22

For most practical purposes, a first-order process may be deemed complete if it is 95% or more complete. Table 1 shows that five half-lives must elapse to reach this point. Thus the elimination of a drug from the body may be considered to be complete after five half-lives have elapsed (i.e., 97% completion). This principle becomes important, for example, in crossover bioavailability studies in which the subjects must be rested for sufficient time between each drug administration to ensure that "washout" is complete.

The half-life of a first-order process is very important. Since it is often desirable to convert a half-life to a rate constant, and vice versa, a simple relationship between the two is very useful. The relationship may be derived as follows:

$$\ln \frac{C_0}{C} = kt$$

When $C_0/C = 2$ and $t = t_{1/2}$. Thus,

$$\ln 2 = kt_{1/2}$$

$$0.693 = kt_{1/2}$$

$$k = \frac{0.693}{t_{1/2}} \tag{14}$$

$$t_{1/2} = \frac{0.693}{k} \tag{15}$$

D. Examples of Calculations

Equations (13), (14), and (15) can be used to solve three types of problems involving first-order processes.

Copyright © 2002 Marcel Dekker, Inc.

These types of problems are illustrated in the following examples:

Type 1

Given the rate constant or half-life and the initial concentration, calculate the concentration at some time in the future.

Example. A penicillin solution containing 500 units/mL has a half-life of 10 days. What will the concentration be in 7 days?

$$k = \frac{0.693}{t_{1/2}} = \frac{0.693}{10\,\text{day}} = 0.069\,\text{day}^{-1}$$

$$\ln\frac{C_0}{C} = kt$$

$$\ln\frac{500\ \text{units/mL}}{C} = 0.069\,\text{day}^{-1} \times 7\,\text{days} = 0.483$$

$$\frac{500}{C} = \text{anti ln}(0.483) = e^{0.483} = 1.62$$

$$C = 308\,\text{units/mL}$$

Type 2

Given the half-life or rate constant and the initial concentration, calculate the time required to reach a specified lower concentration.

Example. A penicillin solution has a half-life of 21 days. How long will it take for the potency to drop to 90% of the initial potency?

$$k = \frac{0.693}{21\,\text{days}} = 0.033\,\text{day}^{-1}$$

$$\ln\frac{C_0}{C} = kt$$

$$\ln\frac{100\%}{90\%} = 0.033 \times t$$

$$t = 3.2\,\text{days}$$

Type 3

Given an initial concentration and the concentration after a specified elapsed time, calculate the rate constant or half-life.

Example. A penicillin solution has an initial potency of 125 mg/5 mL. After 1 month in a refrigerator, the potency is found to be 100 mg/5 mL. What is the half-life of the penicillin solution under these storage conditions?

$$\ln\frac{C_0}{C} = kt$$

$$\ln\frac{125\,\text{mg/5 mL}}{100\,\text{mg/5 mL}} = k \times 30\,\text{day}$$

$$k = 0.0074\,\text{day}^{-1}$$

$$t_{1/2} = \frac{0.693}{0.0074\,\text{day}^{-1}} = 94\,\text{days}$$

For each type of problem the following assumptions are made: (a) the process follows first-order kinetics, at least over the time interval and concentration range involved in the calculations, and (b) all time and concentration values are accurate.

The latter assumption is particularly critical in solving problems such as type 3, where a rate constant is being calculated. It would be unwise to rely on only two time points to calculate such an important value. Normally, duplicate or triplicate assays would be performed at six or more time points throughout as much of the reaction as possible. The resulting mean assay values and standard deviation values would be plotted on semilogarithmic graph paper and a straight line carefully fitted to the data points. The half-life could then the determined using Eq. (14).

Example. A solution of ethyl acetate in pH 9.5 buffer (25°C) was assayed in triplicate several times over a 20-hour period. The data obtained are presented in Table 2. The results were plotted on semilogarthmic graph paper as shown in Fig. 3. Calculate the psuedo-first-order rate constant for the hydrolysis of ethyl acetate at pH 9.5 (25°C).

By fitting a straight line ("best-fit" line) through the data points in Fig. 3 (this can be done by eye using a transparent straight edge) and extrapolating to $t = 0$,

Table 2 Assay of Ethyl Acetate

Time (h)	Concentration (mg/mL) ± SD
2	1.83 ± 0.15
4	1.01 ± 0.09
6	0.58 ± 0.07
8	0.33 ± 0.06
10	0.18 ± 0.04
12	0.10 ± 0.02
16	0.031 ± 0.006
20	0.012 ± 0.002

Copyright © 2002 Marcel Dekker, Inc.

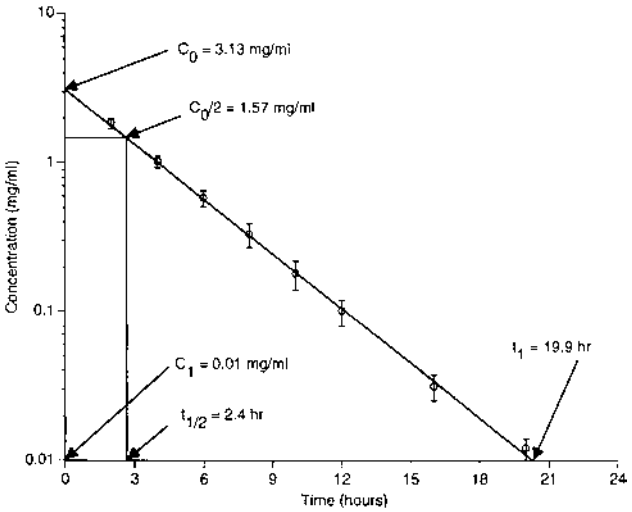


Fig. 3 Semilogarithmic plot of concentration versus time for the hydrolysis of ethyl acetate. (Data shown in Table 2. One standard deviation is indicated by error bars.)

the intercept C_0 is found to be 3.13 mg/mL. The half-life is the time at which the concentration equals 1.57 mg/mL, and this is found by interpolation to be 2.4 hours. The value of k can be given by

$$k = \frac{0.693}{t_{1/2}} = \frac{0.693}{2.4\,\text{h}} = 0.289\,\text{h}^{-1}$$

A value for k can also be determined directly from points on the line using Eq. (12). In Fig. 3, values for C_0, C_1, t_0, and t_1 can read from the "best-fit" line as 3.13 mg/mL, 0.01 mg/mL, 0 hour, and 19.9 hours. Thus:

$$k = -\text{slope} = \frac{\ln C_0 - \ln C_1}{t_1 - t_0}$$

$$k = \frac{\ln 3.13 - \ln 0.01}{19.9 - 0.0} = 0.289\,\text{h}^{-1}$$

Semilogarithmic graph paper is readily available from many graph paper manufactures. It consists of a logarithmic scale on the y-axis and a cartesian scale on the x-axis (see Fig. 3). On the logarithmic scale, the spatial distribution of lines is such that the position of each line is proportional to the logarithm of the value represented by the mark. For example, plotting a concentration of 1.83 mg/mL on semilog paper is equivalent to looking up the logarithm of 1.83 and plotting it on a Cartesian scale. This type of graph paper is extremely useful for kinetic calculations because raw concentration data can be plotted directly without converting to logarithms, and concentration values can be extrapolated and interpolated from the plot without converting logs to numbers.

For example, to determine the half-life in the preceding example, the C_0 value and the time at which $C = C_0/2$ were both read directly from the graph. If Fig. 3 had been a plot of ln C (on a Cartesian scale) versus time, it would have been necessary to read ln C_0 from the graph, convert it to C_0, divide by 2, convert back to ln $(C_0/2)$, then read the half-life off the graph. If the rate constant is determined for this example using Eq. (12), the slope must be calculated. To calculate the slope of the line it is necessary first to read two concentrations from the graph and then take the logarithm of each concentration as described in Eq. (12).

III. FIRST-ORDER PHARMACOKINETICS: DRUG ELIMINATION FOLLOWING RAPID INTRAVENOUS INJECTION

It was mentioned previously that drug elimination from the body most often displays the characteristics of a first-order process. Thus, if a drug is administered by rapid intravenous (IV) injection, after mixing with the body fluids its rate of elimination from the body is proportional to the amount remaining in the body.

Normally, the plasma concentration is used as a measure of the amount of drug in the body, and a plot of plasma concentration versus time has the same characteristics as the plot in Fig. 1. A semilogarithmic

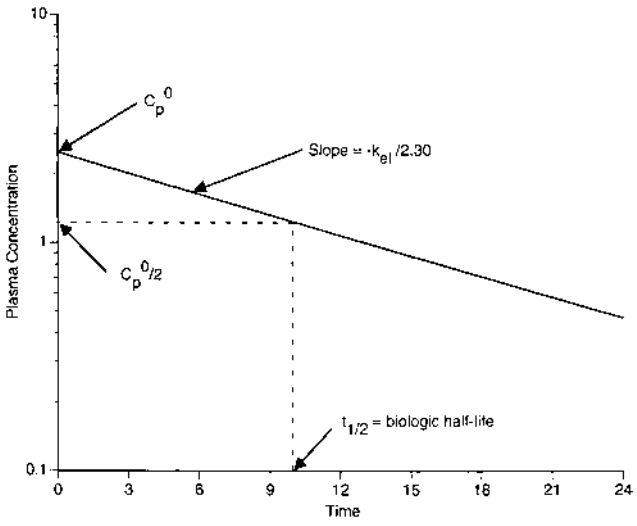


Fig. 4 Semilogarithmic plot of plasma concentration versus time for a drug administered by rapid intravenous injection.

Copyright © 2002 Marcel Dekker, Inc.

plot of plasma concentration versus time is a straight line and allows the calculation of k_{el} from the slope, where k_{el} is the overall elimination rate constant. The intercept at $t = 0$ is C_p^0, the hypothetical plasma concentration after the drug is completely mixed with body fluids but before any elimination has occurred.

A typical semi-log plasma concentration versus time plot is shown in Fig. 4. This figure shows that pharmacokinetic data can also be expressed in terms of a half-life, called the biological half-life, which bears the same relationship to k_{el} as that shown in Eqs. (14) and (15).

Since all the kinetic characteristics of the disappearance of a drug from plasma are the same as those for the pseudo-first-order disappearance of a substance from a solution by hydrolysis, the same working equations [Eqs. (11) and (13)] and the same approach to solving problems can be used.

Example. A 250 mg dose of tetracycline was administered to a patient by rapid IV injection. The initial plasma concentration (C_p^0) was 2.50 μg/mL. After 4 hours the plasma concentration was 1.89 μg/mL. What is the biological half-life ($t_{1/2}$) of tetracycline in this patient?

$$\ln \frac{C_p^0}{C_p} = k_{el} t$$

$$\ln \frac{2.50}{1.89} = k_{el} \times 4$$

$$k_{el} = \frac{0.280}{4} = 0.0699\ \text{h}^{-1}$$

$$t_{1/2} = \frac{0.693}{0.0699\ \text{h}^{-1}} = 9.91\ \text{h}$$

Note that this approach involves the following assumptions: (a) the drug was eliminated by a psuedo-first-order process, and (b) the drug was rapidly distributed so that an "initial plasma concentration" could be measured before any drug began to leave the body. The latter assumption implies that the body behaves as a single homogeneous compartment throughout which the drug distributes instantaneously following IV injection. In pharmacokinetic terms, this is referred to as the one-compartment model. Although most drugs do not, in fact, distribute instantaneously, they do distribute very rapidly, and the one-compartment model can be used for may clinically important pharmacokinetic calculations.

An important parameter of the one-compartment model is the apparent volume of the body compartment, because it directly determines the relationship between the plasma concentration and the amount of drug in the body. This volume is called the apparent volume of distribution, V_d, and it may be calculated using the relationship:

$$\text{Volume} = \frac{\text{amount}}{\text{concentration}}$$

The easiest way to calculate V_d is to use C_p^0, the plasma concentration when distribution is complete (assumed to be instantaneous for a one-compartment model) and the entire dose is still in the body. Thus,

$$V_d = \frac{\text{dose}}{C_p^0} \tag{16}$$

Example. Calculate V_d for the patient in the previous example;

$$V_d = \frac{250\ \text{mg}}{2.50\ \mu\text{g/mL}}$$

$$V_d = 100\ \text{L}$$

Note: Since 1 μg/mL = 1 mg/L, dividing the dose in milligrams by the plasma concentration in micrograms per milliliter will give V_d in liters.

The apparent volume of distribution of a drug very rarely corresponds to any physiological volume, and even in cases where it does, it must never be construed as showing that the drug enters or does not enter various body spaces. For example, the 100 L volume calculated in the foregoing example is much greater than either plasma volume (about 3 L) or whole blood volume (about 6 L) in a standard (70 kg) man; it is even greater than the extracellular fluid volume (19 L) and total body water (42 L) in the same average man. Based on the calculated value of V, it cannot be said that tetracycline is restricted to the plasma or that it enters or does not enter red blood cells, or that it enters or does not enter any or all extracellular fluids.

A discussion of all the reasons for this phenomenon is beyond the scope of this chapter, but a simple example will illustrate the concept. Highly lipid-soluble drugs, such as pentobarbital, are preferentially distributed into adipose tissue. The result is that plasma concentrations are extremely low after distribution is complete. When the apparent volumes of distribution are calculated, they are frequently found to exceed total body volume, occasionally by a factor of 2 or more. This would be impossible if the concentration in the entire body compartment were equal to the plasma concentration. Thus, V_d is an empirically fabricated number relating the

Copyright © 2002 Marcel Dekker, Inc.

concentration of drug in plasma (or blood) with the amount of drug in the body. For drugs such as pentobarbital the ratio of the concentration in adipose tissue to the concentration in plasma is much greater than unity, resulting in a large value for V_d.

IV. PHARMACOKINETIC ANALYSIS OF URINE DATA

Occasionally, it is inconvenient or impossible to assay the drug in plasma, but it may be possible to follow the appearance of the drug in urine. If the drug is not metabolized to any appreciable degree, the pharmacokinetic model may be written as shown in Scheme 1.

$$D_B \xrightarrow{k_{el}} D_U$$

Scheme 1

A plot of cumulative amount of drug appearing in urine (D_U) versus time will be the mirror image of a plot of amount of drug remaining in the body (D_B) versus time. This is illustrated in Fig. 5, which shows that the total amount of drug recovered in urine throughout the entire study (D_U^∞) is equal to the dose (D_B^0) and, at any time, the sum of drug in the body (D_B) plus drug in urine (D_U) equals the dose (D_B^0).

A kinetic equation describing urine data can be developed as follows. If

$$\frac{dD_B}{dt} = -k_{el}D_B$$

then,

$$\frac{dD_U}{dt} = +k_{el}D_B$$

But if

$$D_B + D_U = D_U^\infty$$
$$= \text{amount recovered in urine}$$

Then,

$$D_B = D_U^\infty - D_U$$

Therefore,

$$\frac{dD_U}{dt} = k_{el}(D_U^\infty - D_U)$$

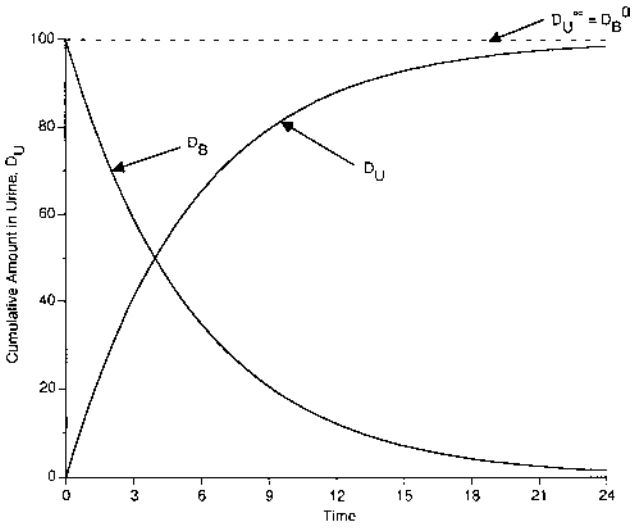


Fig. 5 Plot of cumulative amount of drug in urine, D_U (solid line), and the amount of drug in body, D_B (dashed line), versus time according to Scheme 1.

or

$$\frac{dD_U}{D_U^\infty - D_U} = k_{el}\,dt$$

Integration gives

$$\int \frac{dD_U}{D_U^\infty - D_U} = k_{el}\int dt$$

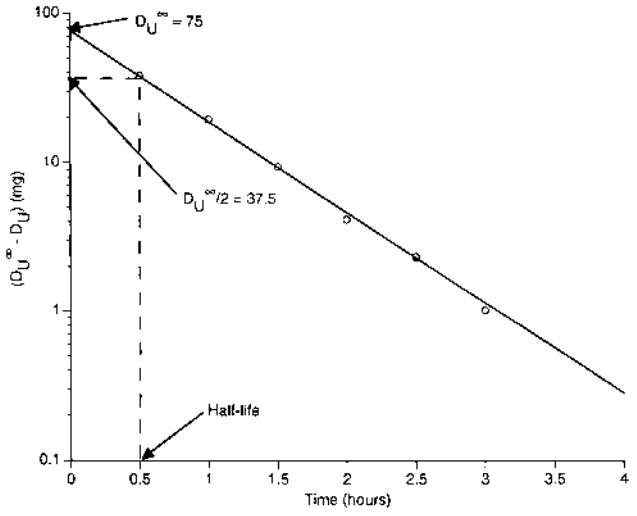


Fig. 6 Semilogarithmic plot of amount of drug remaining to be excreted (ARE) into urine, $D_U^\infty - D_U$, versus time.

Copyright © 2002 Marcel Dekker, Inc.

Table 3 Drug Excreted into the Urine Versus Time

Time interval (h)	Amount excreted (mg)	Cumulative amount excreted, D_U^a (mg)	$D_U^\infty - D_U$ (mg)
0.0–0.5	37.5	37.5	37.8
0.5–1.0	18.5	56.0	19.3
1.0–1.5	10.0	66.0	9.3
1.5–2.0	5.2	71.2	4.1
2.0–2.5	1.8	73.0	2.3
2.5–3.0	1.3	74.3	1.0
3.0–6.0	1.0	75.3	0.0
6.0–12.0	0.0	75.3	0.0

[a] $D_U^\infty = 75.3$ mg.

$$-\ln(D_U^\infty - D_U) + \ln(D_U^\infty - D_U^0) = k_{el}t$$

Since $D_U^0 = 0$ (there is no drug in urine when $t = 0$),

$$\ln(D_U^\infty - D_U) - \ln D_U^\infty = -k_{el}t$$

$$\ln(D_U^\infty - D_U) = -k_{el}t + \ln D_U^\infty \tag{17}$$

Equation (17) is in the form of the equation for a straight line ($y = mx + b$), where t is one variable (x), $-k_{el}$ is determined from the slope (m), $\ln D_U^\infty$ is the constant (b), and $\ln (D_U^\infty - D_U)$ the other variable (y). Thus a plot of $\ln (D_U^\infty - D_U)$ versus time is a straight line with a slope providing a value of k_{el} and an intercept of $\ln D_U^\infty$. Since D_U^∞ is the total amount excreted and D_U is the amount excreted up to time t, $D_U^\infty - D_U$ is the amount remaining to be excreted (ARE). A typical ARE plot is shown in Fig. 6.

Example. The plot in Fig. 6 was constructed using the data shown in Table 3. Note that the concentration of the drug in each urine specimen is not the information analyzed. The total amount excreted over each time interval and throughout the entire study must be determined. As a result, the experimental details of a urinary excretion study must be very carefully chosen, and strict adherence to the protocol is required. Loss of a single urine specimen, or even an unknown part of a urine specimen, makes construction of an ARE plot impossible.

V. CLEARANCE RATE AS AN EXPRESSION OF DRUG-ELIMINATION RATE

A clearance rate is defined as the volume of blood or plasma completely cleared of drug per unit time. It is a useful way to describe drug elimination because it is related to blood or plasma perfusion of various organs of elimination, and it can be directly related to the physiological function of these organs. For example, the renal clearance rate (RCR) of a drug can be calculated using the following equation:

$$\text{RCR} = \frac{\text{amount excreted in urine per unit time}}{\text{plasma concentration}} \tag{18}$$

Example. In the example plotted in Fig. 6, the amount of drug excreted over the 0- to 0.5-hour interval was 37.5 mg. If the plasma concentration at 0.25 hour (the middle of the interval) was 10 μg/mL, what was the renal clearance rate? From Eq. (18),

$$\begin{aligned}\text{RCR} &= \frac{37.5\,\text{mg}/0.5\,\text{h}}{10\,\mu\,\text{g/mL}} \\ &= 7.5\,\text{L/h} \\ &= 125\,\text{mL/min}\end{aligned}$$

The glomerular filtration rate (GFR) in normal males is estimated to be 125 mL/min, and the results of the example calculation suggest that the drug is cleared by GFR. If the RCR had been less than 125 mL/min, tubular reabsorption of the drug would have been suspected. If it had been greater than 125 mL/min, tubular secretion would have been involved in the drug elimination.

Drugs can be cleared from the body by metabolism as well as renal excretion, and when this occurs it is not possible to measure directly the amount cleared by metabolism. However, the total clearance rate (TCR), or total body clearance, of the drug can be calculated from its pharmacokinetic parameters using the following equation:

$$\text{TCR} = k_{el}V_d \tag{19}$$

Example. The biological half-life of procaine in a patient was 35 minutes, and its volume of distribution was estimated to be 58 L. Calculate the TCR of procaine.

$$k_{el} = \frac{0.693}{35\ \text{min}} = 0.0198\ \text{min}^{-1}$$

$$\begin{aligned}\text{TCR} &= k_{el}V_d \\ &= 0.0198\,\text{min}^{-1} \times 58\,\text{L} \\ &= 1.15\,\text{L/min}\end{aligned}$$

When a drug is eliminated by both metabolism and urinary excretion, it is possible to calculate the metabolic clearance rate (MCR) by the difference between TCR and RCR:

Copyright © 2002 Marcel Dekker, Inc.

$$\text{MCR} = \text{TCR} - \text{RCR} \tag{20}$$

The RCR can be determined from urine and plasma data using Eq. (18), and the TCR can be determined from the pharmacokinetic parameters using Eq. (19). Alternately, the RCR can be calculated by multiplying the TCR by the fraction of the dose excreted unchanged into urine, f_e:

$$\text{RCR} = f_e\text{TCR} \tag{21}$$

If it is assumed that the fraction of the dose not appearing as unchanged drug in urine has been metabolized, the MCR can be calculated as follows:

$$\text{MCR} = (1 - f_e)\text{TCR} \tag{22}$$

Example. Sulfadiazine in a normal volunteer had a biological half-life of 16 hours and a volume of distribution of 20 L. Sixty percent of the dose was recovered as unchanged drug in urine. Calculate TCR, RCR, and MCR for sulfadiazine in this person.

$$k_{el} = \frac{0.693}{16\,\text{h}} = 0.0433\,\text{h}^{-1}$$

$$\begin{aligned}\text{TCR} &= k_{el}V_d \\ &= 0.0433\,\text{h}^{-1} \times 20\,\text{L} \\ &= 0.866\,\text{L/h} \\ &= 14.4\,\text{mL/min}\end{aligned}$$

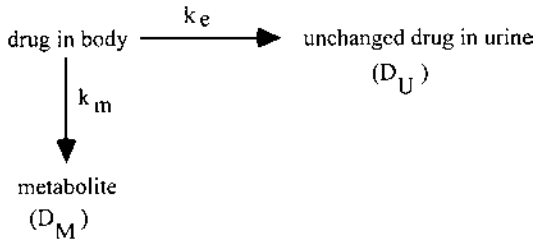
$$\begin{aligned}\text{RCR} &= f_e\text{TCR} \\ &= 0.6 \times 14.4\,\text{mL/min} \\ &= 8.64\,\text{mL/min}\end{aligned}$$

$$\begin{aligned}\text{MCR} &= (1 - f_e)\text{TCR} \\ &= (1 - 0.6) \times 14.4\,\text{mL/min} \\ &= 5.76\,\text{mL/min}\end{aligned}$$

It should be emphasized that the assumption that any drug not appearing as unchanged drug in urine has been metabolized may introduce a great amount of error into the values of the clearance rates estimated using Eqs. (21) and (22). By this assumption unchanged drug eliminated in the feces would be included with metabolized drug, as would any orally administered drug that was unabsorbed.

VI. PHARMACOKINETICS OF DRUG ELIMINATED BY SIMULTANEOUS METABOLISM AND EXCRETION

Although some drugs are excreted unchanged in urine, most are partially eliminated by metabolism. Usually



Scheme 2

both the urinary excretion of unchanged drug and the metabolism are first-order processes, with the rate of excretion and metabolism dependent upon the amount of unchanged drug in the body. This results in a "branch" in the kinetic chain representing exit of drug in the body as depicted in the accompanying pharmacokinetic model (Scheme 2).

In this scheme, the rate of loss of drug from the body is determined by both k_e and k_m, and this can be written in differential form as follows:

$$\begin{aligned}\frac{dD_B}{dt} &= -k_eD_B - k_mD_B \\ &= -(k_e + k_m)D_B\end{aligned} \tag{23}$$

Thus, the overall elimination rate constant (k_{el}) here is the sum of the urinary excretion rate constant (k_e) and the metabolism rate constant (k_m):

$$k_{el} = k_e + k_m \tag{24}$$

For drugs that are both metabolized and excreted unchanged, semilogarithmic plots of plasma concentrations versus time will provide values of k_{el}.

Urine data are required to determine the individual values of k_e and k_m. The required equations are derived next.

Derivation. From Scheme 2, the differential equation describing overall rate of disappearance of drug from the body may be written:

$$\frac{dD_B}{dt} = -k_{el}D_B$$

and the following integrated equation can be written [see also Eq. (10)]:

$$\ln D_B = \ln D_B^0 - k_{el}t$$

Taking antilogs yields

$$D_B = D_B^0 \exp(-k_{el}t) \tag{25}$$

It should be noted that Eq. (25) is another form of an integrated rate equation. This form makes use of

Copyright © 2002 Marcel Dekker, Inc.

an exp $(-x)$ term and may be referred to as an exponential rate expression. These expressions are useful for visualizing the characteristics of a first-order process. For example, when $t = 0$, $\exp(-k_{el}t) = 1$, and $D_B = D_B^0$. When $t = t_{1/2}$, $\exp(-k_{el}t_{1/2}) = 0.5$, and $D_B = 0.5 \times D_B^0$. When $t = \infty$, $\exp(-k_{el}t) = 0$, and $D_B = 0$. Thus, the value of exp $(-k_{el}t)$ varies from 1 to 0 as time varies from 0 to ∞. At any time between 0 and ∞, the fraction of the dose remaining in the body is equal to exp $(-k_{el}t)$.

Exponential rate expressions are also useful in deriving kinetic equations because they can be substituted into differential equations, which can then be integrated. For example, from Scheme 2 the differential equation describing the rate of appearance of unchanged drug in urine may be written:

$$\frac{dD_U}{dt} = +k_e D_B \tag{25a}$$

substituting Eq. (25) into Eq. (25a) gives:

$$\frac{dD_U}{dt} = +k_e[D_B^0 \exp(-k_{el}t)]$$

$$dD_U = +k_e[D_B^0 \exp(-k_{el}t)]dt$$

Integration yields:

$$D_U = -\frac{k_e}{k_{el}} D_B^0 \exp(-k_{el}t) + \text{constant}$$

at $t = 0$, $D_U = 0$, and $\exp(-k_{el}t) = 1$; therefore, the constant equals $(k_e/k_{el})D_B^0$, and

$$D_U = \frac{k_e}{k_{el}} D_B^0[1 - \exp(-k_{el}t)] \tag{26}$$

At $t = \infty$ after elimination is complete, the total amount of drug excreted unchanged in urine (D_U^∞) can be calculated using Eq. (26) as follows:

$$D_U^\infty = \frac{k_e}{k_{el}} D_B^0(1 - 0)$$

$$\frac{D_U^\infty}{D_B^0} = \frac{k_e}{k_{el}} = f_e \tag{27}$$

Equation (27) shows that the fraction of the dose appearing as unchanged drug in urine (f_e) is equal to the fraction of k_{el} attributable to k_e. [An equation analogous to Eq. (27) for D_M^∞ and k_m could be derived in much the same way.]

Substituting Eq. (27) into Eq. (26) and rearranging gives:

$$D_U^\infty - D_U = D_U^\infty \exp(-k_{el}t)$$

Taking logarithms yeilds

$$\ln(D_U^\infty - D_U) = -k_{el}t + \ln D_U^\infty \tag{28}$$

Equation (28) is identical to Eq. (17), for the case in which all eliminated drug was excreted unchanged in urine. $(D_U^\infty - D_U)$ is the amount remaining to be excreted (ARE), and Eq. (28) shows that an ARE plot of unchanged drug in urine versus time will be a straight line with a slope providing a value of k_{el} even when the drug is partially eliminated by metabolism (see Figs. 6 and 7). Using Eq. (27) and the total amount of unchanged drug excreted in urine (D_U^∞), it is possible to calculate k_e. Also, k_m can be calculated from Eq. (24). Thus, all the rate constants in Scheme 2 can be calculated solely on the basis of urinary excretion of unchanged drug.

Example. Five hundred milligrams of a drug was administered IV to a normal healthy volunteer, and various amounts of unchanged drug were recovered from the urine over the 24-hour postdrug period (Table 4). Calculate k_{el}, k_e, and k_m for this drug. A plot of $(D_U^\infty - D_U)$ on a log scale versus time is shown in Fig. 7. A half-life of 1.2 hours can be estimated from the line in Fig. 7.

$$k_{el} = \frac{0.693}{1.2\,\text{h}} = 0.578\,\text{h}^{-1}$$

From Eq. (27):

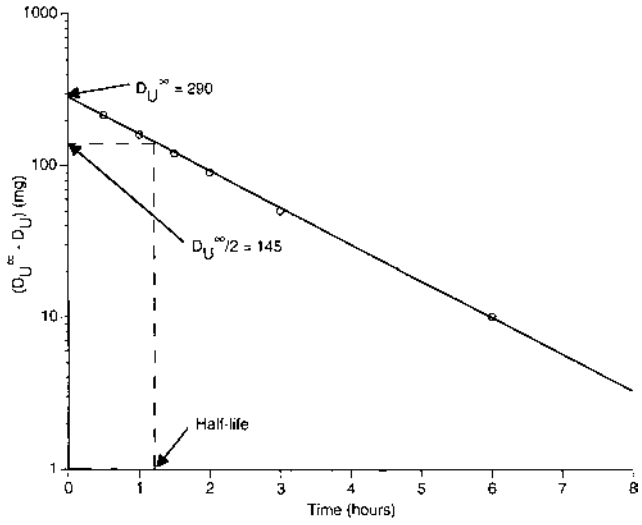


Fig. 7 Semilogarithmic plot of amount of unchanged drug remaining to be excreted into urine, $D_U^\infty - D_U$, versus time according to Scheme 2.

Copyright © 2002 Marcel Dekker, Inc.

Table 4 Drug Recovered in Urine

Time interval (h)	Amount excreted unchanged (mg)	Cumulative amount excreted unchanged, D_U^a (mg)	$D_U^\infty - D_U$ (mg)
0.0–0.5	75	75	215
0.5–1.0	55	130	160
1.0–1.5	40	170	120
1.5–2.0	30	200	90
2.0–3.0	40	240	50
3.0–6.0	40	280	10
6.0–12.0	10	290	0
12.0–24.0	0	290	0

[a] $D_U^\infty = 290\,\text{mg}$.

$$\frac{D_U^\infty}{D_B^0} = \frac{k_e}{k_{el}}$$

$$\frac{290\,\text{mg}}{500\,\text{mg}} = \frac{k_e}{0.578\,\text{h}^{-1}}$$

$$k_e = 0.335\,\text{h}^{-1}$$

From Eq. (24),

$$k_{el} = k_e + k_m$$

$$k_m = 0.578 - 0.335 = 0.243\,\text{h}^{-1}$$

It is important to reemphasize the following assumptions inherent in this type of calculation:

1. It must be assumed that urine collections were accurately timed and that complete urine specimens were obtained at each collection time. It is also assumed that the assay procedure is accurate and reproducible.
2. It is assumed that all processes of elimination obey first-order kinetics.
3. It is assumed that any drug not appearing unchanged in urine has been metabolized. Furthermore, if the drug is not administered by IV injection, it must also be assumed that the dose is completely absorbed. (The IV route was chosen for the preceding example specifically to avoid the need to introduce this assumption.)

A. Significance of k_e and k_m in Patients with Kidney or Liver Disease

In the foregoing example, the drug was administered to a healthy subject who had normal kidney and liver function. The estimated biological half-life in this person was 1.2 hours. If the same drug were administered to a person with no kidney function but with a normal liver, it would be impossible for this individual to excrete unchanged drug. They would, however, be able to metabolize the drug at the same rate as a normal individual. The net result would be that the overall k_{el} would be reduced to the value of k_m, and the biological half-life would increase to

$$\frac{0.693}{k_m} = \frac{0.693}{0.243\,\text{h}^{-1}} = 2.85\,\text{h}$$

Thus, the biological half-life of a drug can increase dramatically when the organs of elimination are diseased or nonfunctional; it may increase to varying degrees if these organs are partially impaired.

At the present time no simple relationship exists between clinical measurements of liver function and the value of k_m. Fortunately, kidney function can be measured quantitatively using standard clinical tests, and it is directly related to k_e for a number of drugs. Great success has been achieved in using kidney clearance measurements to predict the biological half-lives of a number of drugs. This is best illustrated with a drug that is eliminated exclusively by urinary excretion.

Example. Kanamycin is a member of the aminoglycoside class of antibiotics, all of which are eliminated exclusively by glomerular filtration. Creatinine is a natural body substance that is cleared almost exclusively by glomerular filtration, and creatinine clearance rate is frequently used as a diagnostic tool to determine glomerular filtration rate. The relationship

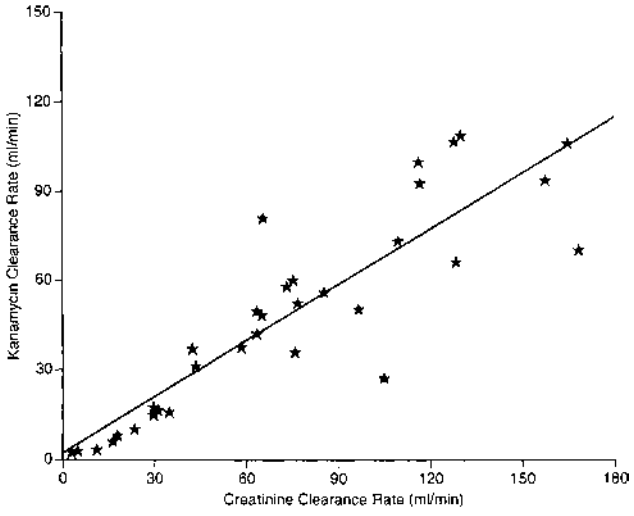


Fig. 8 Plot of kanamycin clearance rate versus creatinine clearance rate.

Copyright © 2002 Marcel Dekker, Inc.

between creatinine clearance rate and kanamycin clearance rate is shown in Fig. 8. Creatinine clearance rate can be determined as a standard clinical procedure, and the corresponding kanamycin clearance rate can be determined by interpolation on the plot in Fig. 8. Since clearance rate $= k_{el}V_d$ [see Eq. (19)], the kanamycin clearance rate can be converted to kanamycin elimination rate constant by dividing by the V_d value for kanamycin, estimated to be about 27% of the patient's body weight.

Although determination of creatinine clearance rate is a standard clinical procedure, it is difficult to carry out mainly because accurate collection of total urine output over a 24-hour period is required. It can never be certain that this requirement has been met. Since creatinine is produced continuously in muscle and is cleared by the kidney, renal failure is characterized by elevated serum creatinine levels. The degree of elevation is directly related to the degree of renal failure—if it is assumed that the production of creatinine in the muscle mass is constant and that renal function is stable. When these assumptions are valid, there is a direct relationship between serum creatinine level and kanamycin half-life, as shown in Fig. 9. The equation of the line in Fig. 9 is

Kanamycin half-life (h)
$= 3 \times$ serum creatinine concentration (mg/100 mL)

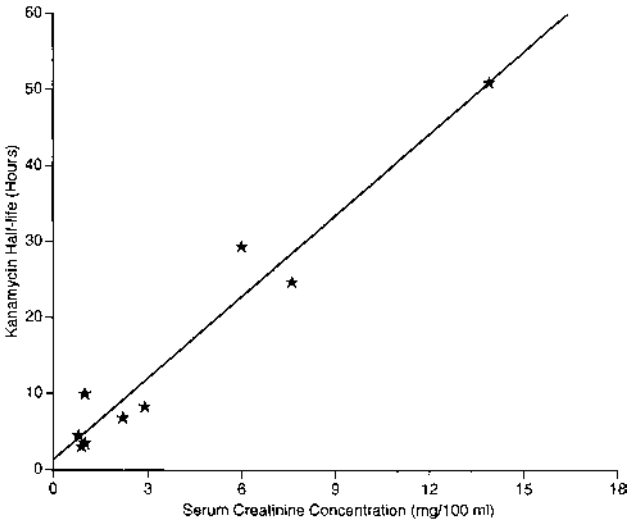


Fig. 9 Plot of kanamycin elimination half-life versus serum creatinine concentration in patients with varying degrees of (stable) renal failure.

Thus, kanamycin half-lives (h) can be predicted in patients with varying degrees of (stable) renal failure by multiplying the serum creatinine level (in mg/100 mL) by 3.

VII. KINETICS OF DRUG ABSORPTION

For all commonly used routes of administration except intravenous, the drug must dissolve in body fluids and diffuse through one or more membranes to enter the plasma. Thus, all routes except intravenous are classed as extravascular routes, and absorption is defined as appearance of the drug in plasma.

The most common extravascular route is oral. When a solution or a rapidly dissolving solid dosage form is given orally, the absorption process often obeys first-order kinetics. In these cases, absorption can be characterized by evaluating the absorption rate constant, k_a, using plasma concentration versus time data.

A. The Method of "Residuals" ("Feathering" the Curve)

When absorption is first-order, the kinetic model may be written as shown in Scheme 3:

$$D_G \xrightarrow{k_a} D_B \xrightarrow{k_{el}} D_E$$

Scheme 3

where D_G = drug at the absorption site (gut)
D_B = drug in the body
D_E = eliminated drug
k_a = first-order absorption rate constant
k_{el} = overall elimination rate constant

The differential equations describing the rates of change of the three components of Scheme 3 are:

$$\frac{dD_G}{dt} = -k_a D_G \tag{29}$$

$$\frac{dD_B}{dt} = k_a D_G - k_{el} D_B \tag{30}$$

$$\frac{dD_E}{dt} = +k_{el} D_B \tag{31}$$

To determine k_a from plasma concentration versus time data, it is necessary to integrate Eq. (30). This is

Copyright © 2002 Marcel Dekker, Inc.

best achieved through exponential expressions. First, integration of Eq. (29) gives

$$D_G = D_G^0 \exp(-k_a t) \tag{32}$$

where D_G^0 is the initial amount of drug presented to the absorbing region of the gut. (D_G^0 = dose, if absorption is complete.)

Substituting Eq. (32) into Eq. (30) gives

$$\frac{dD_B}{dt} = +k_a D_G^0 \exp(-k_a t) - k_{el} D_B \tag{33}$$

Integration of Eq. (33) may be accomplished with Laplace transforms.* The result is

$$D_B = \frac{D_G^0 k_a}{k_a - k_{el}} [\exp(-k_{el} t) - \exp(-k_a t)] \tag{34}$$

Thus, amount of drug in the body following administration of an extravascular dose is a constant $[(D_G^0 k_a)/(k_a - k_{el})]$ multiplied by the difference between two exponential terms—one representing elimination $[\exp(-k_{el} t)]$ and the other representing absorption $[\exp(-k_a t)]$.

Dividing both sides of Eq. (34) by V_d yields an equation for plasma concentration versus time:

$$C_p = \frac{D_G^0 k_a}{V_d(k_a - k_{el})} [\exp(-k_{el} t - \exp(-k_a t)] \tag{35}$$

Equation (35) describes the line in Fig. 10, which is a semilog plot of C_p versus time for an orally administered drug absorbed by a first-order process. The plot begins as a rising curve and becomes a straight line with a negative slope after 6 hours. This behavior is the result of the biexponential nature of Eq. (35). Up to 6 hours, both the absorption process $[\exp(-k_a t)]$ and the elimination process $[\exp(-k_{el} t)]$ influence the plasma concentration. After 6 hours, only the elimination process influences the plasma concentration.

This separation of the processes of absorption and elimination is the result of the difference in the values of k_a and k_{el}. If k_a is much larger than k_{el} (a good rule is that it must be at least five times larger), the second exponential term in Eq. (35) will approach zero much more rapidly than the first exponential term. And at large values of t, Eq. (35) will reduce to

$$C_p = \frac{D_G^0 k_a}{V_d(k_a - k_{el})} [\exp(-k_{el} t)] \tag{36}$$

* Full details of this integration may be found in M. Mayersohn, M. Gibaldi, Am J Pharm Ed 34: 608, 1970, Eq. (27).

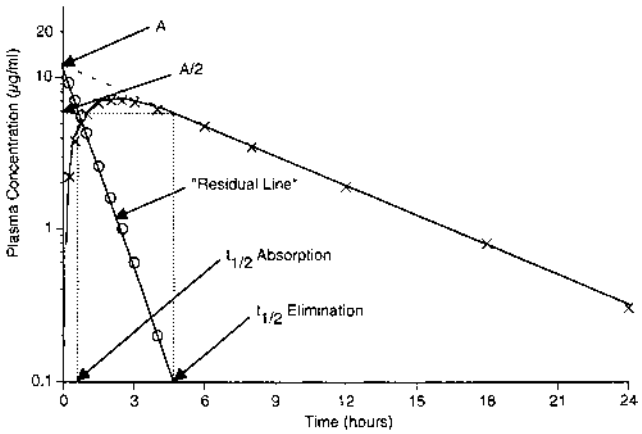


Fig. 10 Semilogarithmic plot of observed plasma concentrations (crosses) and "residuals" (circles) versus time for an orally administered drug absorbed by a first-order process.

or

$$C_p = A \cdot \exp(-k_{el} t)$$

where A is a constant term.

Converting to logarithms, we obtain

$$\ln C_p = -k_{el} t + \ln A \tag{37}$$

Thus after 6 hours the semilog plot of C_p versus time shown in Fig. 10 becomes a straight line and k_{el} can be determined from the slope. Therefore, the overall elimination rate constant for a drug may be accurately determined from the "tail" of a semilog plot of plasma concentration versus time following extravascular administration if k_a is at least five times larger than k_{el}.

The value of k_a can also be determined from plots like Fig. 10 using the following logic: In Fig. 10 the curved line up to 6 hours is given by

$$C_{p1} = A \cdot \exp(-k_{el} t) - A \cdot \exp(-k_a t)$$

The straight line after 6 hours and the extrapolated (dashed) line before 6 hours is given by

$$C_{p2} = A \cdot \exp(-k_{el} t)$$

The difference ("residual") between the curved line and the extrapolated (dashed) line up to 6 hours is given by

$$\begin{aligned} \text{Residual} &= C_{p2} - C_{p1} \\ &= A \cdot \exp(-k_a t) \end{aligned}$$

Converting to logs:

$$\ln(\text{residual}) = -k_a t + \ln A \tag{38}$$

Copyright © 2002 Marcel Dekker, Inc.

Table 5 Plasma Concentrations and "Residuals" Versus Time

Time (h)	Observed C_p (μg/mL)	Extrapolated C_p (μg/mL)	Residuals (μg/mL)
0.0	0.0	11.8	11.8
0.25	2.2	11.4	9.2
0.5	3.8	10.9	7.1
0.75	5.0	10.6	5.6
1.0	5.8	10.1	4.3
1.5	6.8	9.4	2.6
2.0	7.1	8.7	1.6
2.5	7.1	8.1	1.0
3.0	6.9	7.5	0.6
4.0	6.2	6.4	0.2
6.0	4.8	4.8	—
8.0	3.5	3.5	—
12.0	1.9	1.9	—
18.0	0.8	0.8	—
24.0	0.3	0.3	—

As shown in Fig. 10, a semilog plot of residuals versus time is a straight line with a slope of $-k_a$.

It should be noted that the intercepts (A) for both the extrapolated (dashed) line [Eq. (37)] and the residuals line [Eq. (38)] are the same and are equal to the constant in Eq. (35):

$$A = \frac{D_G^0 k_a}{V_d(k_a - k_{el})} \tag{39}$$

A is a function of the two rate constants (k_a and k_{el}), the apparent volume of distribution (V_d), and the amount of drug absorbed (D_G^0). After k_a and k_{el} have been evaluated and A has been determined by extrapolation, a value for V_d can be calculated if it is assumed that D_G^0 is equal to the dose administered, i.e., absorption is 100% complete.

Example. Fig. 10 is a plot of the data shown in Table 5. The extrapolated value of A is 11.8 μg/mL.

The $t_{1/2}$ (elimination) is the time at which the elimination line crosses $A/2 = 4.5$ hour:

$$k_{el} = \frac{0.693}{t_{1/2}(\text{elimination})} = 0.154 \text{ h}^{-1}$$

The $t_{1/2}$ (absorption) is the time at which the residuals line crosses $A/2 = 0.7$ hour:

$$k_a = \frac{0.693}{t_{1/2}(\text{absorption})} = 0.990 \text{ h}^{-1}$$

Assuming that the 100-mg dose of drug was completely absorbed, the V_d can be calculated from Eq. (39):

$$A = 11.8 \text{ μg/mL} = \frac{100 \text{ mg} \times 0.990 \text{ h}^{-1}}{V_d(0.990 - 0.154) \text{ h}^{-1}}$$

$$V_d = 10.0 \text{ L}$$

This method of calculation is often referred to as the "method of residuals" or "feathering the curve." It is important to remember that the following assumptions were made:

1. It is assumed that k_a is at least five times larger than k_{el}; if not, neither constant can be determined accurately.
2. It is assumed that the absorption and elimination processes are both strictly first order; if not, the residuals line and, perhaps, the elimination line will not be straight.
3. It is assumed that absorption is complete; if not, the estimate of V_d will be erroneously high.

B. The Wagner-Nelson Method*

A major shortcoming of the method of residuals for determining the absorption rate constant from plasma concentration versus time data following administration of oral solid dosage forms is the necessity to assume that the absorption process obeys first-order kinetics. Although this assumption is often valid for solutions and rapidly dissolving dosage forms for which the absorption process itself is rate determining, if release of drug from the dosage form is rate determining, the kinetics are often zero-order, mixed zero- and first-order, or even more complex processes.

The Wagner-Nelson method of calculation does not require a model assumption concerning the absorption process. It does require the assumption that (a) the body behaves as a single homogeneous compartment and (b) drug elimination obeys first-order kinetics. The working equations for this calculation are developed next.

Derivation. For any extravascular drug administration, the mass balance equation can be written as amount absorbed (A) equals amount in body (W) plus amount eliminated (E), or

$$A = W + E$$

Taking the derivative with respect to time yields

$$\frac{dA}{dt} = \frac{dW}{dt} + \frac{dE}{dt}$$

* See J.G. Wagner, E. Nelson, J Pharm Sci 53:1392, 1964.

Copyright © 2002 Marcel Dekker, Inc.

But

$$W = V_d C_p$$

or

$$\frac{dW}{dt} = V_d \frac{dC_p}{dt}$$

and

$$\frac{dE}{dt} = k_{el} W$$
$$= k_{el} V_d C_p$$

Therefore,

$$\frac{dA}{dt} = V_d \frac{dC_p}{dt} + k_{el} V_d C_p$$

$$dA = V_d dC_p + k_{el} V_d C_p \, dt$$

Integrating from $t = 0$ to $t = t$

$$\int_0^t dA = V_d \int_0^t dC_p + k_{el} V_d \int_0^t C_p \, dt$$

$$A_t = V_d C_p^t + k_{el} V_d \int_0^t C_p \, dt$$

Rearranging, we have

$$\frac{A_t}{V_d} = C_p^t + k_{el} \int_0^t C_p \, dt \tag{40}$$

where A_t/V_d is the amount of drug absorbed up to time t divided by the volume of distribution, C_p^t is plasma (serum or blood) concentration at time t, and $\int_0^t C_p \, dt$ is the area under the plasma (serum or blood) concentration versus time curve up to time t (see Sec. VIII.A). An equation similar to Eq. (40) can be derived by integration from $t = 0$ to $t = \infty$. Since $C_p = 0$ at $t = \infty$, the equation becomes

$$\frac{A_{\max}}{V_d} = k_{el} \int_0^\infty C_p \, dt \tag{41}$$

where $A_{\max}$ is the total amount of drug absorbed from the dosage form divided by the volume of distribution and $\int_0^\infty C_p \, dt$ is the area under the entire plasma (serum or blood) concentration versus time curve (see Sec. VIII.A).

Equation (41) is useful for comparing the bioavailabilities of two dosage forms of the same drug administered to the same group of subjects. If it is assumed that k_{el} and V_d are the same for both administrations, it can be seen that the relative availabilities of the dosage forms is given by the ratio of the areas under the plasma concentration versus time curves:

$$\frac{A_{\max_1}}{A_{\max_2}} = \frac{\int_0^\infty (C_p \, dt)_1}{\int_0^\infty (C_p \, dt)_2} \tag{42}$$

Other methods of comparing bioavailabilities will be discussed in a later section.

A great deal can be learned about the absorption process by applying Eqs. (40) and (41) to plasma concentration versus time data. Since there is no model assumption with regard to the absorption process, the calculated values of A_t/V_d can often be manipulated to determine the kinetic mechanism that controls absorption. This is best illustrated by an example.

Example. A tablet containing 100 mg of a drug was administered to a healthy volunteer and the plasma concentration (C_p) versus time data shown in Table 6 were obtained. Figure 11 shows a semi-log plot of these C_p versus time data. The half-life for elimination of the drug can be estimated from the straight line "tail" of the plot to be 4.7 hours. The overall elimination rate constant is then

$$k_{el} = \frac{0.693}{4.7\,\text{h}} = 0.147\,\text{h}^{-1}$$

Table 6 illustrates the steps involved in carrying out the Wagner-Nelson calculation. The third column ($\int_0^t C_p \, dt$) shows the area under the C_p versus time curve calculated sequentially from $t = 0$ to each of the time points using the trapezoidal rule (see Sec. VIII.A). The fourth column ($k_{el} \int_0^t C_p \, dt$) shows each of the preceding areas multiplied by k_{el} (as estimated from the "tail")

Table 6 Data Illustrating the Wagner-Nelson Calculation

Time (h)	C_p (μg/ml)	$\int_0^t C_p \, dt$	$k_{\text{el}} \int_0^t C_p \, dt$	$\frac{A_t}{V_d}$	$\frac{A_{\max}}{V_d} - \frac{A_t}{V_d}$
0.25	0.6	0.1	0.0	0.6	9.4
0.50	1.2	0.3	0.1	1.3	8.7
0.75	1.8	0.7	0.1	1.9	8.1
1.0	2.3	1.2	0.2	2.5	7.5
1.5	3.4	2.6	0.4	3.8	6.2
2.0	4.3	4.5	0.7	5.0	5.0
3.0	6.0	9.7	1.5	7.5	2.5
6.0	5.6	27.1	4.1	9.7	0.3
12.0	2.3	50.8	7.6	9.9	0.1
18.0	0.9	60.4	9.1	10.0	—
24.0	0.4	64.3	9.6	10.0	—

[a] $A_{\max}/V_d = 10.0$.

Copyright © 2002 Marcel Dekker, Inc.

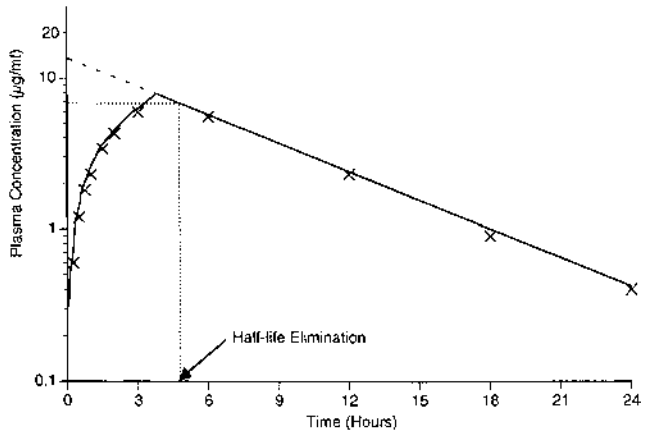


Fig. 11 Semilogarithmic plot of observed plasma concentrations (crosses) versus time for an orally administered drug absorbed by a zero-order process. (Data shown in Table 6.)

constituting the second term of the Wagner-Nelson equation [see Eq. (40)]. The fifth column (A_t/V_d) shows the sums of the values indicated in the second and fourth columns according to Eq. (40). A_{max}/V_d is the maximum value in the fifth column (i.e., 10.0), and the sixth column shows the residual between A_{max}/V_d and each sequential value of A_t/V_d in the fifth column.

If the absorption process obeyed first-order kinetics, a semilog plot of the residuals in the sixth column would be a straight line and k_a can be determined from the slope. However, the regular Cartesian plot of the residuals shown in Fig. 12 is a straight line showing the absorption process obeys zero-order kinetics; that is, the process proceeds at a constant rate (25 mg/h), stopping abruptly when the dose has been completely absorbed.

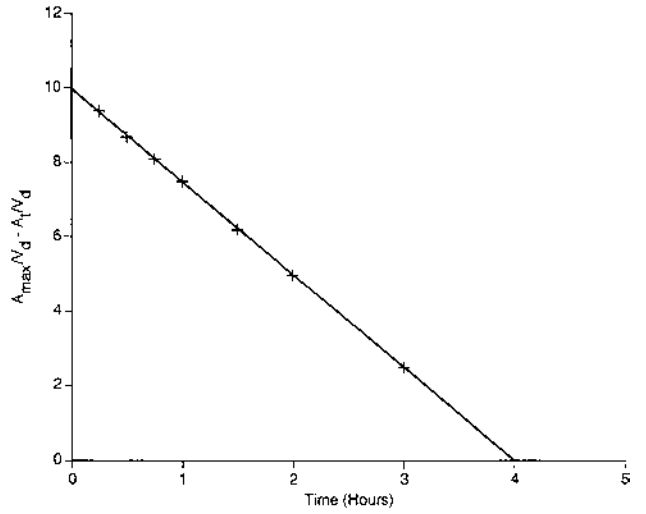


Fig. 12 Plot of the residuals (crosses) between A_{max}/V_d and A_t/V_d versus time (last column of Table 6).

This example illustrates the usefulness of the Wagner-Nelson calculation for studying the mechanism of release of drugs from dosage forms in vivo. Whereas the absorption process itself usually obeys first-order kinetics, dissolution of capsules, tablets, and especially sustained-release dosage forms often must be described by more complex kinetic mechanisms. Although pure zero-order absorption, such as that just illustrated, is almost never observed in practice, many sustained-release dosage forms are designed to produce as close to zero-order release as possible since constant absorption produces constant plasma levels.

C. The Method of "Inspection"

Often it is unnecessary to calculate an exact value for an absorption rate constant. For example, when several oral tablets containing the same drug substance are all found to be completely absorbed, it may be sufficient to merely determine if the absorption rates are similar to conclude that the products would be therapeutically equivalent. In another instance, it would be possible to choose between an elixir and a sustained-release tablet for a specific therapeutic need without assigning accurate numbers to the absorption rate constant for the two dosage forms.

In these instances, the "*time of the peak*" in the plasma concentration versus time curve provides a convenient measure of the absorption rate. For example, if three tablets of the same drug are found to be completely absorbed and all give plasma peaks at 1 hour, it can be safely concluded that all three tablets are absorbed at essentially the same rate. (In fact, if all tablets are completely absorbed and all peak at the same time, it would be expected that all three plasma concentration versus time curves would be identical, within experimental error.)

The time of the peak can also be used to roughly estimate the absorption rate constant. If it is assumed that k_a is at least $5 \times k_{el}$, then it can be assumed that absorption is at least 95% complete at the peak time; that is, the peak time represents approximately five absorption half-lives (see Table 1). The absorption half-life can then be calculated by dividing the time of the peak by 5, and the absorption rate constant can be calculated by dividing the absorption half-life into 0.693.

Copyright © 2002 Marcel Dekker, Inc.

Example. Inspection of Fig. 10 gives a peak time of about 2.5 hours. The absorption half-life can be estimated to be 0.5 h and the absorption rate constant, to be $1.4\,h^{-1}$.

VIII. BIOAVAILABILITY (EXTENT OF ABSORPTION)

If a drug is administered by an extravascular route and acts systemically, its potency will be directly related to the amount of drug the dosage form delivers to the blood. Also, if the pharmacological effects of the drug are related directly and instantaneously to its plasma concentration, the rate of absorption will be important because the rate will influence the height of the plasma concentration peak and the time at which the peak occurs. Thus, the *bioavailability of a drug product is defined in terms of the amount of active drug delivered to the blood and the rate at which it is delivered.*

Whenever a drug is administered by an extravascular route, there is a danger that part of the dose may not reach the blood (i.e., absorption may not be complete). When the intravenous route is used, the drug is placed directly in the blood; therefore an IV injection is, by definition, 100% absorbed. The absolute bioavailability of an extravascular dosage form is defined relative to an IV injection. If IV data are not available, the relative bioavailability may be defined relative to a standard dosage form. For example, the bioavailability of a tablet may be defined relative to an oral solution of the drug.

In Sec. VII we dealt with methods of determining the rate (and mechanism) of absorption. In this section we will deal with methods of determining the extent of absorption. In every example, the calculation will involve a comparison between two studies carried out in the same group of volunteers on different occasions. Usually it will be necessary to assume that the volunteers behaved identically on both occasions, especially with regard to their pharmacokinetic parameters.

A. Area Under the Plasma Concentration Versus Time Curve

In the development of equations for the Wagner-Nelson method of calculation, the following equation was derived [see Eq. (42)]:

$$\frac{A_{\max_1}}{A_{\max_2}} = \frac{\int_0^\infty (C_p\,dt)_1}{\int_0^\infty (C_p\,dt)_2}$$

This equation shows that the amounts of drug absorbed from two drug products (i.e., the relative bioavailability of product 1 compared with product 2) can be calculated as the ratio of the areas under the plasma concentration versus time curves (AUCs), assuming k_{el} and V_d were the same in both studies. This assumption is probably valid when the studies are run with the same group of volunteers and within a few weeks of one another.

If dosage form 2 [Eq. (42)] is an intravenous dosage form, the *absolute bioavailability* of the extravascular dosage form (dosage form 1) is given by:

$$\text{Absolute bioavailability (extravascular dosage form)} = \frac{\text{AUC}_{\text{extravascular}}}{\text{AUC}_{\text{IV}}} \tag{43}$$

The AUC for a plasma concentration versus time curve can be determined using the trapezoidal rule. For this calculation, the curve is divided into vertical segments, as shown in Fig. 13. The top line of each segment is assumed to be straight rather than slightly curved, and the area of the segment is calculated as though it were a trapezoid; for example, the area of segment 10 is

$$\text{area}_{10} = \frac{C_{p9} + C_{p10}}{2} \times (t_{10} - t_9) \tag{44}$$

The total AUC is then obtained by summing the areas of the individual segments. [Equation (44) can be programmed into a microcomputer that will calculate the areas and sum them as rapidly as the C_p values can be entered.]

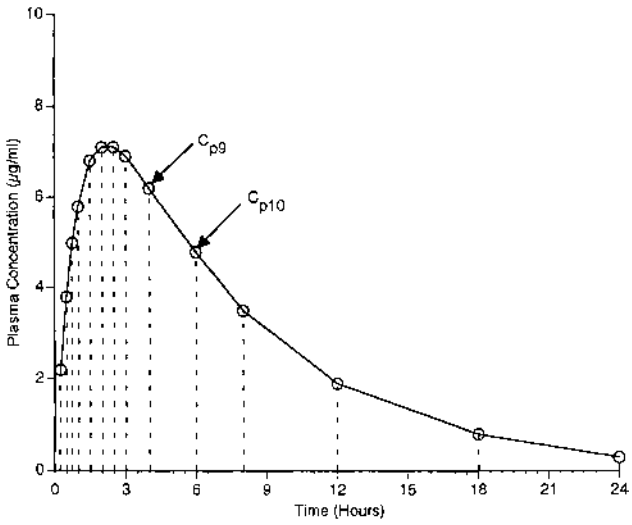


Fig. 13 Plot of plasma concentrations (circles) versus time with the curve divided into vertical segments. (Data shown in Table 7.)

Copyright © 2002 Marcel Dekker, Inc.

It should be readily apparent that the trapezoidal rule does not measure AUC exactly. However, it is accurate enough for most bioavailability calculations, and the segments are chosen on the basis of the time intervals at which plasma was collected.

Example. The AUC for Fig. 10 can be calculated from the data given in Table 7.

Assuming that the AUC for a 100 mg IV dose given to the same group of volunteers was 86.7 h.μg/mL, the absolute bioavailability of the extravascular dosage form is

Table 7 Calculation of Area Under the Plasma Concentration Versus Time Curve (AUC) Using the Trapezoidal Rule

Time (h)	C_p (μg/mL)	Area of segment (h·μg/mL)	Cumulative area up to time t (h·μg/mL)
0.0	0.0		
		0.275	0.275
0.25	2.2		
		0.75	1.025
0.5	3.8		
		1.1	2.125
0.75	5.0		
		1.35	3.475
1.0	5.8		
		3.15	6.625
1.5	6.8		
		3.475	10.10
2.0	7.1		
		3.55	13.65
2.5	7.1		
		3.50	17.15
3.0	6.9		
		6.55	23.70
4.0	6.2		
		11.0	34.70
6.0	4.8		
		8.3	43.0
8.0	3.5		
		10.8	53.8
12.0	1.9		
		8.1	61.9
18.0	0.8		
		3.3	65.2
24.0	0.3		
		2.0	67.2[a]
∞	—		

[a] $AUC_\infty = 67.2$ h.μg/m.

$$\text{Absolute bioavailability} = \frac{AUC_\infty}{AUC_{IV}} \times 100 = \frac{67.2}{86.7} \times 100 = 77.5\%$$

It is not necessary to apply the trapezoidal rule to the entire plasma concentration versus time curve in order to calculate the total AUC. After the semilog plot becomes a straight line, the remaining area out to $t =$ can be calculated using the following equation:

$$AUC_{(t \text{ to } \infty)} = \frac{C_p^t}{k_{el}} \tag{45}$$

Once a semilog plasma concentration versus time plot begins to follow simple first-order elimination kinetics, the remaining AUC can be calculated in one step using Eq. (45).

Example. In the previous problem, the AUC from 24 hours to infinity is given by

$$AUC_{(24\,h \text{ to } \infty)} = \frac{0.3\,\mu g/mL}{0.15\,h^{-1}} = 2.0\,h \cdot \mu g/mL$$

It follows that if the entire semilog plot were straight, as would be the case for a one-compartment drug following IV administration, the total AUC would be given by

$$AUC_{IV} = \frac{C_p^0}{k_{el}} \tag{45a}$$

Example. For IV administration in the foregoing problem, the AUC was calculated as follows:

$$AUC_{IV} = \frac{13.0\,\mu g/mL}{0.15\,h^{-1}} = 86.7\,h.\mu g/mL$$

B. Cumulative Urinary Excretion

In the development of equations for calculating urine data when the drug is partially metabolized and partially excreted unchanged in urine, the following equation was derived [see Eq. (27)]:

$$\frac{D_U^\infty}{D_B^0} = f_e$$

where D_U^∞ is the amount of drug recovered from urine, D_B^0 is the amount of drug absorbed, and f_e is the fraction of the absorbed amount recovered as unchanged drug in urine. Equation (27) may be rearranged and written for two dosage forms as follows:

$$D_{U_1}^\infty = D_{B_1}^0 \times f_{e_1}$$

Copyright © 2002 Marcel Dekker, Inc.

and

$$D_{U_2}^{\infty} = D_{B_2}^{0} \times f_{e_2}$$

Dividing the first equation by the second gives:

$$\frac{D_{U_1}^{\infty}}{D_{U_2}^{\infty}} = \frac{D_{B_2}^{0} \times f_{e_2}}{D_{B_2}^{0} \times f_{e_2}}$$

Assuming that $f_{e_1} = f_{e_2}$, we have

$$\frac{D_{U_1}^{\infty}}{D_{U_2}^{\infty}} = \frac{D_{B_2}^{0}}{D_{B_2}^{0}} = \text{relative bioavailability} \tag{46}$$

Similarly,

$$\frac{D_{U(\text{extravascular})}^{\infty}}{D_{U(\text{IV})}^{\infty}} = \text{absolute bioavailability} \tag{47}$$

Thus, if it is assumed that the same fraction of absorbed drug always reaches the urine unchanged, the bioavailability can be calculated as the ratio of total amounts of unchanged drug recovered in urine.

Example. When potassium penicillin G was administered IV to a group of volunteers, 80% of the 500 mg dose was recovered unchanged in urine. When the same drug was administered orally to the same volunteers, 280 mg was recovered unchanged in urine. What is the absolute bioavailability of potassium penicillin G following oral administration? From Eq. (47),

$$\text{Absolute bioavailability} = \frac{280}{400} \times 100 = 70\%$$

For this calculation, it is unnecessary to assume that V_d and/or k_{el} are the same for the two studies. It is only necessary that f_e be the same in both studies. This is usually a valid assumption unless the drug undergoes a significant amount of "first-pass" metabolism in the gut wall or liver following oral administration or a significant amount of decomposition at an intra muscular (IM) injection site. When this occurs, the availability of the extravascular dosage form may appear to be low, but the fault will not lie with the formulation. The bioavailability will be a true reflection of the therapeutic efficacy of the drug product, and reformulation may not increase bioavailability.

C. The Method of "Inspection"

Bioavailability studies are frequently carried out for the sole purpose of comparing one drug product with another with the full expectation that the two products will have identical bioavailabilities; that is, their rates and extents of absorption will be identical. Such studies are called *bioequivalence studies* and are often employed when a manufacturer wishes to market a "generic equivalent" of a product already on the market. To take advantage of the safety and efficacy data the product's originator has filed with the FDA, the second manufacturer must show that their product gives an *identical plasma concentration versus time curve*.

In these cases it is not necessary to determine the absolute bioavailability or the absorption rate constant for the product under study. It is only necessary to prove that the plasma concentration versus time curve is not significantly different from the reference product's curve. This is done by comparing the means and standard deviations of the plasma concentrations for the two products at each sampling time using an appropriate statistical test.

A discussion of the statistical methods used in analyzing the data from bioequivalence studies is beyond the scope of this chapter. For a discussion of these considerations, the reader is referred to a description by Westlake [6].

IX. MULTIPLE DOSING REGIMENS (REPETITIVE DOSING)

Drugs are infrequently used in single doses to produce an acute effect, the way aspirin is used to relieve a headache. More often, drugs are administered in successive doses to produce a repeated or prolonged effect, the way aspirin is used to relieve the pain and inflammation of arthritis. A properly designed multiple dosing regimen will maintain therapeutically effective plasma concentrations of the drug while avoiding toxic concentrations. Such regimens are easily designed if the pharmacokinetic parameters of the drug are known.

When drugs are administered on a multiple dosing regimen, each dose (after the first) is administered before the preceding doses are completely eliminated. This results in a phenomenon known as *accumulation*, during which the amount of drug in the body (represented by plasma concentration) builds up as successive doses are administered. The phenomenon of accumulation for a drug administered IV is shown in Fig. 14.

Figure 14 shows that the plasma concentrations do not continue to build forever but reach a plateau where the same maximum ($C_{\max}$) and minimum ($C_{\min}$) con-

Copyright © 2002 Marcel Dekker, Inc.

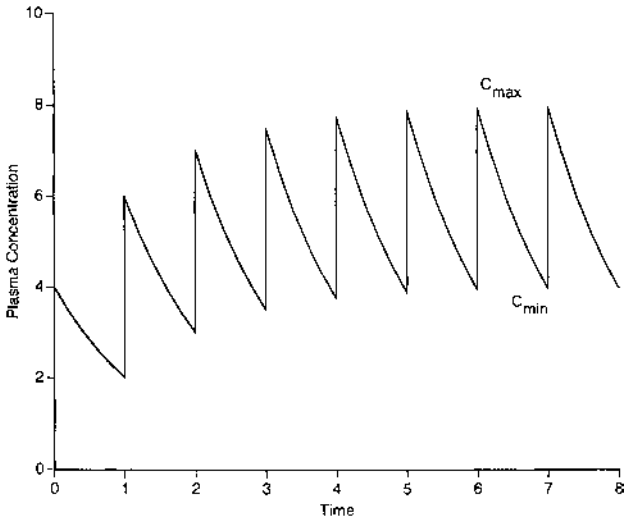


Fig. 14 Plot of plasma concentration versus time showing accumulation following multiple intravenous injections.

centrations are reproduced over and over. The objectives of designing a dosing regimen are to keep $C_{\min}$ above the *minimum effective concentration* (MEC) and to keep $C_{\max}$ below the *minimum toxic concentration* (MTC).

A. Repetitive Intravenous Dosing

The plasma concentrations in Fig. 14 can be calculated as follows: from Eq. (25) the plasma concentration at the end of the first dosing interval, T, is given by:

$$C_{p_1}^T = C_{p_1}^0 \exp(-k_{el}T) \tag{48}$$

Immediately after the second dose is given, the plasma concentration will be,

$$C_{p_2}^0 = C_{p_1}^T + C_{p_1}^0 = C_{p_1}^0 \exp(-k_{el}T) + C_{p_1}^0 \tag{49}$$

and so on.

It is now helpful to define the parameter R as the fraction of the initial plasma concentration that remains at the end of any dosing interval; R is given by the following equation:

$$R = \exp(-k_{el}T) \tag{50}$$

As was pointed out in Sec. VI, when $T = t_{1/2}$, $R = 0.5$. The plot in Fig. 14 was constructed using these conditions; therefore, the plasma concentration at the end of each dosing interval is half the concentration at the beginning of the dosing interval.

Equations (48) and (49) can be simplified to

$$C_{p_1}^T = C_{p_1}^0 R$$

for the plasma concentration at the end of the first dosing interval, and

$$C_{p_2}^0 = C_{p_1}^0 R + C_{p_1}^0$$

for the plasma concentration at the beginning of the second dosing interval.

The series can be carried further for more doses:

$$C_{p_2}^T = (C_{p_1}^0 R + C_{p_1}^0)R$$

$$C_{p_3}^0 = (C_{p_1}^0 R + C_{p_1}^0)R + C_{p_1}^0$$

$$C_{p_3}^T = [(C_{p_1}^0 R + C_{p_1}^0)R + C_{p_1}^0]R$$

etc.

The plasma concentrations at the beginning and end of the nth dosing interval are given by the following power series:

$$\text{Beginning} = C_{p_1}^0 + C_{p_1}^0 R + C_{p_1}^0 R^2 + \cdots + C_{p_1}^0 R^{n-1} \tag{51}$$

$$\text{End} = C_{p_1}^0 R + C_{p_1}^0 R^2 + C_{p_1}^0 R^3 + \cdots + C_{p_1}^0 R^n \tag{52}$$

Since R is always smaller than 1, R^n becomes smaller as n increases. For example, if $R = 0.5$, $R^{10} = 0.001$. Therefore, the high power terms in Eqs. (51) and (52) become negligible as n increases, and additional doses do not change the value of $C_{p_n}^0$ or $C_{p_n}^T$ significantly. This explains why the plasma concentrations reach a plateau instead of continuing to rise as more doses are given.

Hence, $C_{\max}$ and $C_{\min}$ (see Fig. 14) are defined as the plasma concentrations at the beginning and end, respectively, of the nth dosing interval after the plateau has been reached (i.e., $n = \infty$). When $n = \infty$, Eqs. (51) and (52) become

$$C_{\max} = \frac{C_{p_1}^0}{1 - R} \tag{53}$$

$$C_{\min} = C_{\max} R = \frac{C_{p_1}^0 R}{1 - R} \tag{54}$$

Thus, the maximum and minimum plasma concentrations on the plateau of a repetitive IV dosing regimen can be calculated if the dosing interval (T), the overall elimination rate constant (k_{el}), and the initial plasma concentration (C_p^0) are known.

Copyright © 2002 Marcel Dekker, Inc.

Example. A drug has a biologic half-life of 4 hours. Following an IV injection of 100 mg, C_p^0 is found to be 10 μg/mL. Calculate C_{max} and C_{min} if the 100-mg IV dose is repeated every 6 hour until a plasma concentration plateau is reached.

$$k_{el} = \frac{0.693}{4\text{ h}} = 0.173\text{ h}^{-1}$$

$$\begin{aligned} R &= \exp(-k_{el}T) \\ &= \exp(-0.173 \times 6) \\ &= \exp(-0.451) \\ &= 0.354 \end{aligned}$$

$$C_{max} = \frac{10\text{ μg/mL}}{1 - 0.354} = 15.5\text{ μg/mL}$$

$$C_{min} = 15.5\text{ μg/mL} \times 0.354 = 5.49\text{ μg/mL}$$

Example. As indicated earlier, when $T = t_{1/2}$, $R = 0.5$. As a result, on the plateau in Fig. 14,

$$C_{max} = \frac{C_p^0}{1 - 0.5} = 2 \times C_p^0$$

$$C_{min} = \frac{C_p^0 \times 0.5}{1 - 0.5} = C_p^0$$

Thus, when a dose is administered every half-life, C_{max} will be twice C_p^0 and C_{min} will be half C_{max} or equal to C_p^0.

The second example illustrates a very simple and often-used dosage regimen; i.e., administration of a maintenance dose every half-life. The calculations indicate that on this regimen, C_{min} will be $C_{max}/2$ and C_{max} will be $2 \times C_p^0$. Figure 14 indicates that approximately five half-lives will be required to reach the plasma concentration plateau. If the drug has a relatively long half-life, many hours, perhaps days, may be required for the plasma concentrations to reach the ideal range. If the patient's condition is serious, the physician may not want to wait for this to happen. It is under these circumstances that a loading dose is indicated. The loading dose immediately puts the plasma concentrations in the plateau range, and the maintenance dose maintains that condition.

For IV administration, the easiest way to determine the loading dose is in terms of C_p^0 and C_{max}. For example, if the desired C_{max} is 20 μg/mL and a dose of 100 mg gives a C_p^0 of 10 μg/mL, a loading dose of 200 mg should give a C_p^0 of 20 μg/mL, which is the desired C_{max}. If this loading dose is followed by maintenance doses of 100 mg every half-life, the plasma concentrations can be maintained at the plateau from the very beginning and throughout the entire dosing regimen. Thus, for the maintenance dose every half-life regimen, the ideal loading dose is twice the maintenance dose.

Example. Kanamycin is an aminoglycoside antibiotic that exerts a toxic effect on the hearing. If the plasma concentrations are allowed to remain above 35 μg/mL (MTC) for very long, permanent hearing loss may result. The minimum effective concentration (MEC) of kanamycin in plasma is estimated to be about 10 μg/mL for most organisms against which it is used. Thus, kanamycin is a classic example of a drug with a narrow therapeutic index for which a very precise dosing regimen is an absolute necessity. (In fact, this is true of all aminoglycosides.) When kanamycin is administered IV in a dose of 7.5 mg/kg to adults, it yields a C_p^0 of about 25 μg/mL and a half-life of about of 3 hours. What would be a good dosing regimen for kanamycin?

Since 25 μg/mL is well above the MEC but below the MTC, a loading dose of 7.5 mg/kg might be given initially. After one half-life (3 h), the plasma concentration should be 12.5 μg/mL. Since this is just above the MEC and corresponds to half the initial 25 μg/mL, a maintenance dose of 3.75 mg/kg could be administered. With repeated 3.75 mg/kg maintenance doses every 3 hours, C_{max} should be 25 μg/mL and C_{min} should be 12.5 μg/mL, which would allow some margin for error on either side.

Example.The kanamycin problem could be solved more aggressively as follows: let $C_{max} = 35$ μg/mL and $C_{min} = 10$ μg/mL. From Eqs. (53) and (54), the value of R on the plateau may be calculated as follows:

$$C_{p_1}^0 = C_{max}(1 - R) = (35\text{ μg/mL})(1 - R)$$

$$C_{p_1}^0 = \frac{C_{min}(1 - R)}{R} = \frac{(10\text{ μg/mL})(1 - R)}{R}$$

$$35\text{ μg/mL}(1 - R) = \frac{(10\text{ μg/mL})(1 - R)}{R}$$

$$\begin{aligned} R &= \frac{10}{35} = 0.286 \\ &= 10\ \exp(-k_{el}T) \qquad \left(k_{el} = \frac{0.693}{3\text{ h}} = 0.231\text{ h}^{-1}\right) \end{aligned}$$

$$0.268 = \exp(-0.231 \times T)$$

$$-1.25 = -0.231 \times T$$

$$T = 5.41\text{ h (dosing interval)}$$

Copyright © 2002 Marcel Dekker, Inc.

A loading dose that produces a $C_{p_1}^0$ of 35 μg/mL is desired, and this can be calculated as follows [from Eq. (16)]:

$$\frac{\text{dose}_1}{C_{p_1}^0} = \frac{\text{dose}_2}{C_{p_2}^0}$$

$$\frac{7.5\,\text{mg/kg}}{25\,\mu\text{g/mL}} = \frac{x\,\text{mg/kg}}{35\,\mu\text{g/mL}}$$

$$\text{Loading dose} = 10.5\,\text{mg/kg}$$

The amount of drug remaining in the body at the end of the first dosing interval can be calculated in a similar way from the known $C_{\min}$:

$$\frac{7.5\,\text{mg/kg}}{25\,\mu\text{g/mL}} = \frac{x\,\text{mg/kg}}{10\,\mu\text{g/mL}}$$

$$\text{Amount remaining} = 3\,\text{mg/kg}$$

The maintenance dose needed to replace the amount lost over the dosing interval is the difference between the loading dose and the amount remaining at the end of the interval:

$$\text{Maintenance dose} = (10.5 - 3)\,\text{mg/kg} = 7.5\,\text{mg/kg}$$

Thus, the regimen would be a loading dose of 10.5 mg/kg, followed by maintenance doses of 7.5 mg/kg every 5.41 hours. This regimen is not only impractical but, were it carried out, it would produce $C_{\max}$ and $C_{\min}$ concentrations too close to the limiting values to allow for any errors. A better approach would be to define clinicially relevant $C_{\max}$ and $C_{\min}$ values and use the approach in the previous section to develope a useful dosing regimen.

B. Repetitive Extravascular Dosing

Although the equations become considerably more complex than for the IV case, $C_{\max}$ and $C_{\min}$ can be calculated when the drug is administered by an extravascular route. The required equations may be developed as follows: The equation describing the plasma concentration versus time curve following one extravascular administration was discussed previously. Equation (35) may be written as follows:

$$C_p = \frac{FD}{V_d} \times \frac{k_a}{k_a - k_{el}} [\exp(-k_{el}t) - \exp(-k_a t)] \qquad (55)$$

where D is the dose administered and F is the fraction of the administered dose absorbed [$FD = D_G^0$ in Eq. (35)].

If n doses of the drug are administered at fixed time intervals (T), the plasma concentrations following the nth dose are given by

$$C_p = \frac{FD}{V_d} \times \frac{k_a}{k_a - k_{el}} \left[\frac{1 - \exp(-nk_{el}T)}{1 - \exp(-k_{el}T)} \exp(-k_{el}t') - \frac{1 - \exp(-nk_aT)}{1 - \exp(-k_aT)} \exp(-k_a t') \right] \qquad (56)$$

where t' is the time elapsed after the nth dose. When n is large (i.e., when the plasma concentrations reach a plateau), the terms $\exp(-nk_{el}T)$ and $\exp(-nk_aT)$ become negligibly small, and Eq. (56) simplifies to

$$C_p = \frac{FD}{V_d} \times \frac{k_a}{k_a - k_{el}} \times \left[\frac{\exp(-k_{el}t')}{1 - \exp(-k_{el}T)} - \frac{\exp(-k_a t')}{1 - \exp(-k_aT)} \right] \qquad (57)$$

Equation (57) can be used to calculate the $C_{\max}$ and $C_{\min}$ values on the plasma concentration plateau by substituting values for t' that correspond to the "peaks" and "valleys" in the C_p versus t curve. Thus, if $t' = t_{\max}$ (the time of the peak), Eq. (57) gives $C_{\max}$:

$$C_{\max} = \frac{FD}{V_d} \times \frac{k_a}{k_a - k_{el}} \times \left[\frac{\exp(-k_{el}t_{\max})}{1 - \exp(-k_{el}T)} - \frac{\exp(-k_a t_{\max})}{1 - \exp(-k_aT)} \right] \qquad (58)$$

If $t' = 0$ (the time at which another dose is to be given), Eq. (57) gives $C_{\min}$:

$$C_{\min} = \frac{FD}{V_d} \times \frac{k_a}{k_a - k_{el}} \times \left[\frac{1}{1 - \exp(-k_{el}T)} - \frac{1}{1 - \exp(-k_aT)} \right] \qquad (59)$$

Example. The results of a single IM dose of kanamycin show that the dose is completely absorbed (F = 1.0), $V_d = 20$ L, $k_{el} = 0.3\,\text{h}^{-1}$, and the time of the peak is about 1 hr ($k_a = 3.47\,\text{h}^{-1}$). If 800 mg doses of kanamycin are administered IM every 6 h, what will $C_{\max}$ and $C_{\min}$ be when the plasma concentration plateau is reached?

$$C_{\max} = \frac{1.0 \times 800\,\text{mg}}{20\,\text{L}} \times \frac{3.47\,\text{h}^{-1}}{(3.47 - 0.3)\,\text{h}^{-1}} \times \left[\frac{\exp(-0.3 \times 1)}{1 - \exp(-0.3 \times 6)} - \frac{\exp(-3.47 \times 1)}{1 - \exp(-3.47 \times 6)} \right] = 37.5\,\mu\text{g/mL}$$

Copyright © 2002 Marcel Dekker, Inc.

From Eq. (59),

$$C_{\min} = \frac{1.0 \times 800\,\text{mg}}{20\,\text{L}} \times \frac{3.47\,\text{h}^{-1}}{(3.47 - 0.3)\,\text{h}^{-1}} \times \left[\frac{1}{1 - \exp(-0.3 \times 6)^{-1}} \frac{1}{1 - \exp(-3.47 \times 6)}\right] = 8.67\,\mu\text{g/mL}$$

Note: The value of $\exp(-x)$ can be calculated directly with a scientific calculator or as the inverse or antilog(base e).

The foregoing example shows the calculation of the two most important features of a repetitive dosing regimen, the maximum and minimum plasma concentrations on the plateau. But if Eq. (56) had been used, it would have been possible to calculate the plasma concentration at any time throughout an entire dosing regimen. Although these calculations are complex and laborious when done by hand, relatively inexpensive programmable calculators and now personal computers can solve Eq. (56) in seconds. As a result, a plasma concentration versus time plot can be generated for an entire dosing regimen in a very short time. The next example illustrates such a calculation.

Example. The following personal computer spreadsheet will solve Eq. (56) with an added loading dose for any time during a repetitive dosing regimen (see Fig. 15).

Comments: This spreadsheet has been written to be as flexible as possible for use in clinical situations; therefore many parameters are entered in a form different from their form in Eq. (56). The following comments should clarify the relationships:

1. The spreadsheet accepts half-lives for the elimination and absorption processes and converts them into rate constants, but both half-lives must be in the same time units! (If the half-life for absorption is unknown, the time of the peak divided by 5 gives a reasonable estimate; see Sec. VII.C.)
2. The fraction of the dose absorbed must be estimated in a separate study or from past experience.
3. The volume of distribution is expressed as a fraction of the body weight, and the spreadsheet calculates V_d in liters using the subject weight in kilograms.
4. Since loading doses are often used clinically, the spreadsheet accepts both a loading dose and a

B14 =($B$7*$B$8*$E$4/($E$6*($E$4-$E$3)))*(EXP(-$E$3*($B$11*$B$10+A14))-EXP(-$E$4*($B$11*$B$10+A14)))+($B$7*$B$9*$E$4/($E$6*($E$4-$E$3)))*(((1-EXP(-$B$11*$E$3*$B$10))/(1-EXP(-$E$3*$B$10)))*EXP(-$E$3*A14)-((1-EXP(-$B$11*$E$4*$B$10))/(1-EXP(-$E$4*$B$10)))*EXP(-$E$4*A14))

Figure 15

Elimination Half-life	2.3	hr	kel	0.30	hr(-1)
Absorption Half-life	0.2	hr	ka	3.47	hr(-1)
Volume of Distribution	0.27	L/kg			
Body Weight	74	kg	Vd	19.98	L
Fraction Absorbed (F)	1				
Loading Dose	800	mg			
Maintenance Dose	800	mg			
Dosing Interval (T)	6	hr			
Number of Doses (n)	2				

Time Since Last Dose (t')	Cp (mg/L)
0	8.3684
0.25	29.9937
0.5	37.1647
0.75	38.3977
1	37.2635
1.25	35.2539

Fig. 15 Spreadsheet illustrating the calculation of plasma concentrations after multiple extravascular doses, calculated using Eq. (56).

Copyright © 2002 Marcel Dekker, Inc.

maintenance dose. If no loading dose is given, these two entries are the same value.

5. The time units for the dosing interval (T) and the step interval (int) must be the same as those for the half-lives of elimination and absorption.
6. Once the number of maintenance doses (n) is entered the spreadsheet calculates C_p for the nth dosing interval.
7. This spreadsheet can be used in the following ways:
   (a) The entire plasma concentration versus time profile for a repetitive dosing regimen can be calculated by starting with $n = 0$ and calculating C_p for the first dosing interval. Then $n = 1$ is entered, and the second interval C_p values calculated. The process can be repeated for as many dosing intervals as desired.
   (b) The plasma concentration versus time profile for a dosing interval on the plateau can be calculated by entering $n = 50$. The values of C_{min} and C_{max} can then be determined by noting the maximum and minimum values calculated. [Note: When $n = 50$ or some other large number, Eq. (56) becomes Eq. (57). Thus, calculations involving Eqs. (57), (58), and (59) can also be performed with this spreadsheet if a large number is entered for n.]
   (c) The entire plasma concentration versus time profile for a single dose administration can be calculated if the dose is entered as the loading dose (D_0), 50 (or any large number) is entered for T, and 0 is entered for d_0 and n. [Equation (56) becomes Eq. (55).]

Example. Calculate C_p at quarter-hourly intervals for the first 24 hours of the kanamycin dosing regimen described in the first example in this subsection. The entry data are as follows:

$$\begin{aligned}
\text{Half-life for elimination} &= 2.3\,\text{h}\\
\text{Half-life for absorption} &= 0.2\,\text{h}\\
\text{Fraction of body weight equal to volume of distribution} &= 0.27\\
\text{Subject weight} &= 74\,\text{kg}\\
\text{Fraction of the dose absorbed} &= 1.0\\
\text{Loading dose } (D_0) &= 800\,\text{mg}\\
\text{Maintenance dose } (d_0) &= 800\,\text{mg}\\
\text{Dosing interval } (T) &= 6\,\text{h}\\
\text{Time interval} &= 0.25\,\text{h}\\
\text{Number of maintenance doses } (n) &= 0 \text{ to } 4
\end{aligned}$$

The results of the calculation are shown in Fig. 16. Note that C_{max} and C_{min} are identical to the values calculated in the aforementioned example.

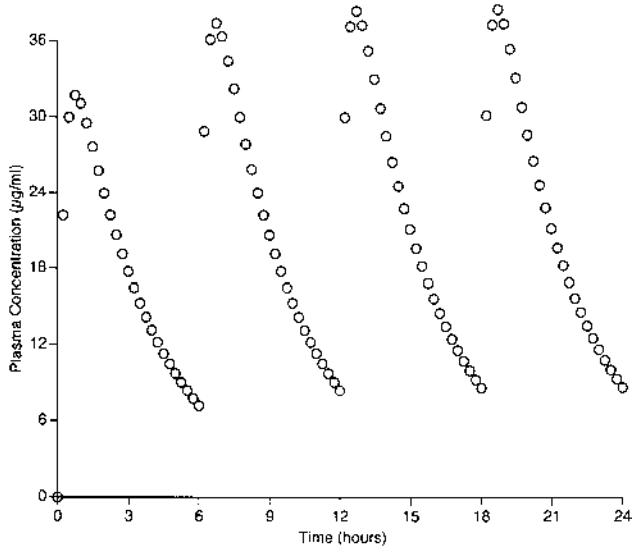


Fig. 16 Plot of kanamycin plasma concentrations (circles) versus time following multiple IM injections, calculated using Eq. (56).

REFERENCES

1. DWA Bourne. Mathematical Modeling of Pharmacokinetic Data. Lancaster, PA: Technomic Publishing Company, 1995.
2. RE Notari. Biopharmaceutics and Clinical Pharmacokinetics. 4th ed. New York, Marcel Dekker, 1987.
3. M Rowland, TN Tozer. Clinical Pharmacokinetics. 2nd ed. Philadelphia, PA: Lea & Febiger, 1989.
4. L. Shargel, ABC Yu. Applied Biopharmaceutics and Pharmacokinetics. 4th Ed, Stamford, CT: Appleton & Lange, 1999.
5. ME Winter. Basic Clinical Pharmacokinetics. 3rd ed. Vancouver, WA: Applied Therapeutics, 1994.
6. WJ Westlake. The design and analysis of comparative blood-level trails In: J Swarbrick, ed. Current Concepts in the Pharmaceutical Sciences: Dosage Form Design and Bioavailability. Philadelphia, PA: Lea & Febiger, 1973.

Copyright © 2002 Marcel Dekker, Inc.

Chapter 4

Factors Influencing Drug Absorption and Drug Availability

Betty-ann Hoener and Leslie Z. Benet

University of California, San Francisco, California

I. INTRODUCTION

In this chapter we examine dosage forms as drug-delivery systems. The scope of the examination will, however, be limited to those oral dosage forms in which the drug is not in solution when taken by the patient. Usually the absorption of drugs from aqueous solutions can be defined by the principles discussed in Chapter 2. However, occasionally drugs in solution may precipitate in the gastrointestinal (GI) fluids and effectively be considered as suspensions. Nevertheless, the dosage forms considered in this chapter are tablets, capsules, and suspensions. In addition, the discussion is limited to drugs absorbed by passive diffusion that thus appear to obey first-order linear absorption kinetics.

The successful transposition of a drug from an oral dosage form into the general circulation can be described as a four-step process: first, delivery of the drug to its absorption site; second, getting the drug into solution; third, movement of the dissolved drug through the membranes of the gastrointestinal tract (GIT); and, finally, movement of the drug away from the site of absorption into the general circulation. Each of these four steps is considered in turn. However, the order of the first two steps is not absolute; that is, the drug may dissolve either before or after reaching its absorption site. It is imperative, however, as discussed in Chapter 2, that the drug be in solution before it can be absorbed. The slowest of the four steps will determine the rate of availability of the drug from an oral dosage form. The rate and extent of availability of a drug from its dosage form will be influenced by many factors in all four of these steps. Those factors related to the physicochemical properties of the drug and the design and production of the dosage form will be called pharmaceutical variables. Those variables resulting from the anatomical, physiological, and pathological characteristics of the patient are called patient variables. The pharmacist, who is the one member of the health care team knowledgeable about both the patient and the dosage form, is uniquely qualified to evaluate the influence of both the pharmaceutical and patient variables on rational drug therapy.

II. GETTING THE DRUG TO ITS SITE OF ABSORPTION

When an oral dosage form is swallowed by a patient, it will travel through the gastrointestinal tract (Fig. 1). During this passage the dosage form will encounter great anatomical and physiological variations. These patient variables have been discussed in Chapter 2. Two of these variables are most important in effecting the delivery of the drug from its dosage form. The first variable, hydrogen ion concentration, exhibits a 10^7-fold difference between the mucosal fluids of the stomach and the intestine. The second variable, available surface area of the absorbing membranes, changes dramatically between different regions of the GIT (see Chapter 2).

Copyright © 2002 Marcel Dekker, Inc.

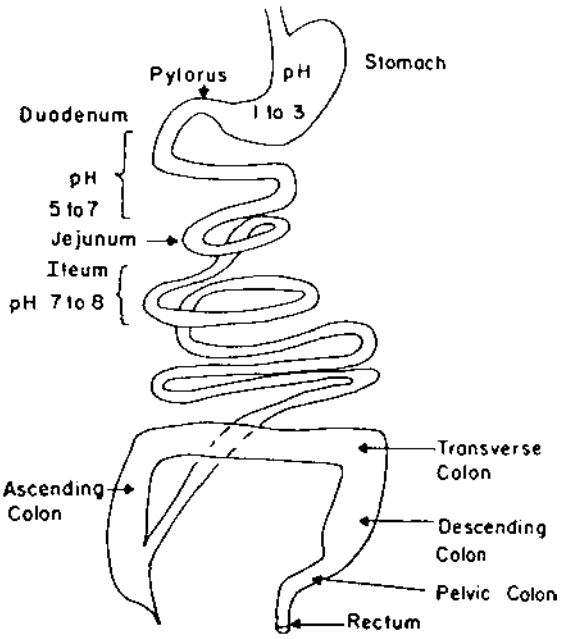


Fig. 1 Diagram of the gastrointestinal tract showing the variations of pH. (From Ref. 1.).

A. Gastric Emptying

Since various factors, principally increased membrane surface area and decreased thickness of the membrane, favor the small intestine versus the stomach as the primary site for drug absorption, the rate at which the drug gets to the small intestine can significantly affect its rate of absorption. Hence, gastric-emptying rate may well be the rate-determining step in the absorption of a drug. As has been discussed in Chapter 2, light physical activity will stimulate stomach emptying, but strenuous exercise will delay emptying [2,3]. If a patient is lying on the left side, the stomach contents will have to move uphill to get into the intestine [2,3]. The emotional state of the patient can either reduce or speed up the stomach-emptying rate [2,3]. In addition, numerous pathological conditions may alter gastric-emptying rate [4]. Other drugs may affect the stomach-emptying rate by influencing GI motility. It should be noted that any substance in the stomach delays emptying and, therefore, can delay the absorption of a drug. The stomach serves to protect the intestine from extreme conditions. Thus, if the stomach contents differ appreciably in pH, temperature, osmolarity, or viscosity from those conditions normally expected in the intestine, the stomach will delay emptying until those conditions approach normal [2,3]. Food, by altering the foregoing factors, can affect the rate of stomach emptying. Moreover, the volume, composition, and caloric content of a meal can alter the stomach-emptying rate [2,3].

Thus, the timing of meals relative to the timing of the oral dosing of a drug can influence the rate, and possibly the extent, of drug availability. It can be anticipated that taking a drug shortly before, after, or with a meal may delay the rate of drug availability as a function of decreased-emptying rate. However, the effect of food on the extent of availability cannot be so readily predicted. Figure 2 illustrates the mean serum concentration versus time curves obtained following administration of three 2.5-mg tablets of methotrexate to normal healthy volunteers [5]. The tablets were taken on an empty stomach or with a standard, high-fat content breakfast. Food significantly decreased the rate of availability of methotrexate from these tablets, as indicated by the shift to the right of the peak time. However, there was no significant change in the area under the curve and, therefore, no change in the extent of availability of methotrexate from these tablets.

In contrast with these results, Fig. 3 illustrates the decreased extent of availability of erythromycin when two 250-mg tablets of erythromycin were taken on a fasting stomach with 20 or 250 mL of water or with 250 mL of water immediately after high-fat, high-protein, and high-carbohydrate meals [6]. The rate of availability of this poorly absorbed water-soluble, acid-labile antibiotic was not affected by food (i.e., peak time was approximately the same for all dosings), indicating that stomach emptying is not the rate-determining step in the absorption of erythromycin from this drug-delivery system. The extent of availability was, however, markedly decreased. This decrease might be due to complexation between drug and food or, more likely, to degradation of the antibiotic when it is retained in the acid environment of the stomach for longer periods. (The effect of the volume of water is considered later in the chapter.)

A third possibility, an increase in the extent of availability of a drug taken with food, is illustrated in Fig. 4. Here the cumulative amount of unmetabolized nitrofurantoin excreted in the urine is plotted against time. Healthy male volunteers took nitrofurantoin as a capsule containing 100 mg of the drug in a *macro*crystalline form (circles) or as a tablet containing 100 mg of the drug in the *micro*crystalline form (squares). (The significance of the macro- and microcrystalline forms are discussed later in this chapter.) The dosage forms were taken with 240 mL of

Copyright © 2002 Marcel Dekker, Inc.

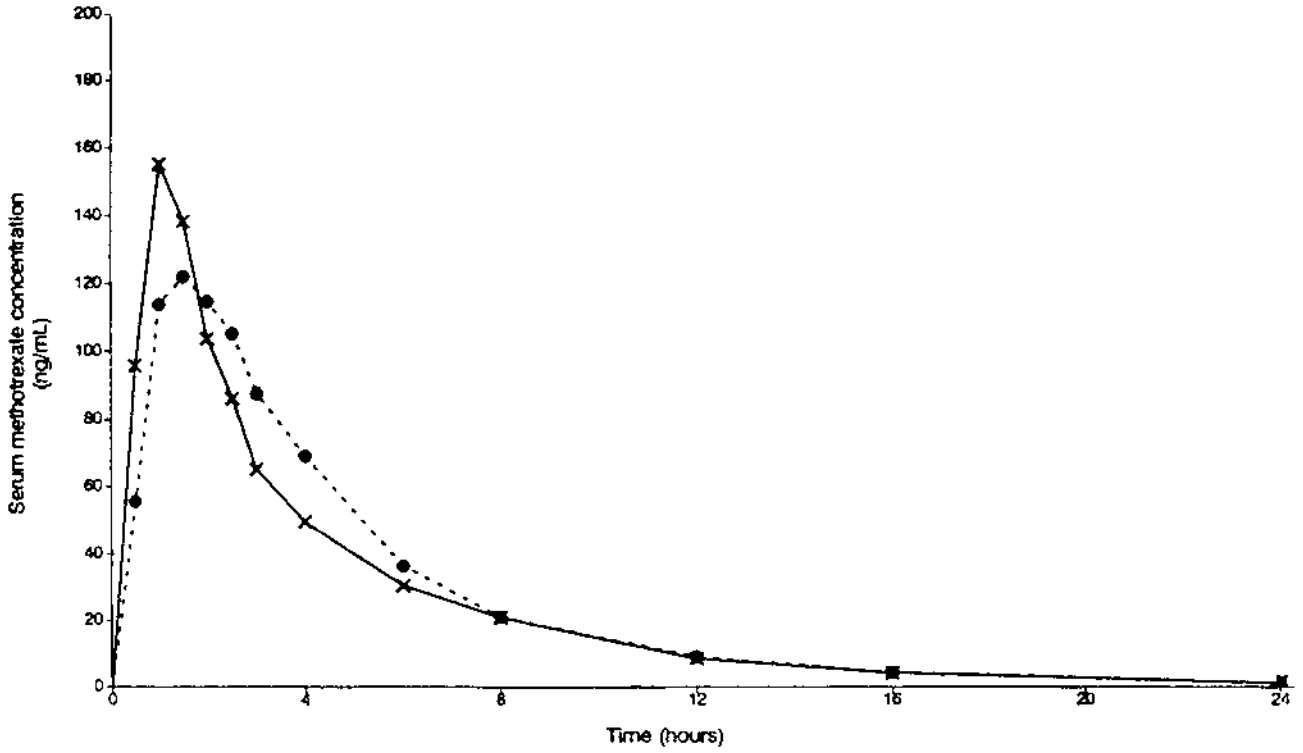


Fig. 2 Mean serum concentration of methotrexate in healthy volunteers who took three 2.5-mg tablets: ×, on a fasting stomach; ●, with breakfast. (From Ref. 5.).

water either on an empty stomach (open symbols) or immediately after a standard breakfast (solid symbols). It can be seen that food delays the absorption of nitrofurantoin from both the tablet and the capsule dosage forms, as indicated by the initial phase of each plot. However, food enhanced the extent of availability of the drug from both dosage forms, as indicated by the cumulative amount of drug excreted at 24 hours. It appears that delaying the rate of transit of nitrofurantoin through the GIT gives this poorly soluble drug more time to dissolve in the GI fluids. Thus, more nitrofurantoin gets into solution, and, therefore, more is absorbed.

In summary, food will delay stomach emptying and, as a result, may decrease the rate of availability of a drug from its oral dosage form. The extent of availability of that drug, however, may be increased, decreased, or unaffected by meals. Thus, it is important for the pharmacist to counsel patients on the importance of timing the taking of their medications relative to their mealtimes.

Large objects, such as nondisintegrating dosage forms, are handled by the stomach quite differently from liquids and particles smaller than about 2 mm in diameter [8]. These larger particles will remain in the stomach pouch until they are swept out by a series of contractions called a migrating motor complex (MMC). This MMC occurs about every 3 hours in the fasted stomach [9]. However, when a patient has eaten a meal, the MMC does not occur until after most of the food has left the stomach [8]. Thus, there could be a considerable delay in the absorption of drugs from solid dosage forms that do not disintegrate or release their contents by some other mechanism while in the stomach. This effect of size on the rate of delivery of the drug from its dosage form may or may not also affect the extent of availability.

B. Gastrointestinal Motility

Other patient variables may affect GI motility and, thereby, the extent or rate of availability of a drug from a delivery system. As illustrated in Chapter 2, the degree of physical activity, age, disease state, and emotional condition of a patient may increase or de-

Copyright © 2002 Marcel Dekker, Inc.

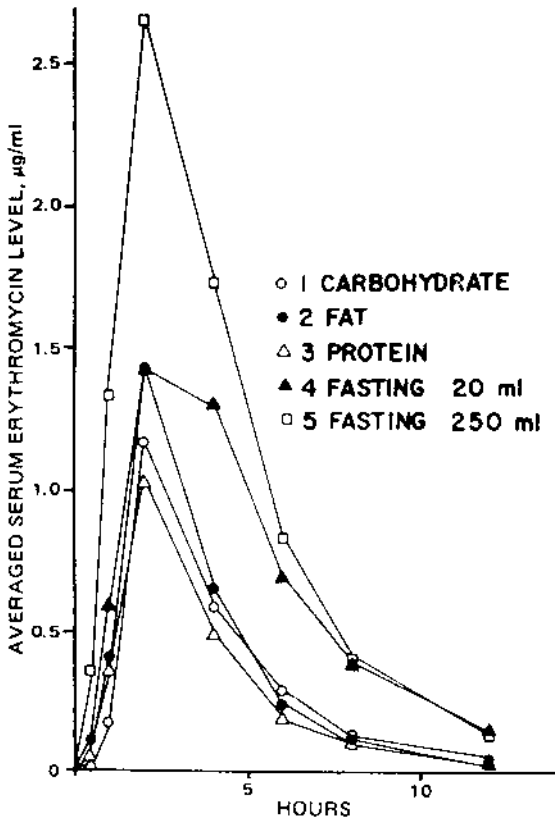


Fig. 3 Mean serum erythromycin levels in healthy volunteers given 500 mg of erythromycin stearate with 20 mL of water (▲), 250 ml of water (□), or 250 mL of water immediately after a high-carbohydrate meal (○), 250 mL of water immediately after a high-fat meal (●), and 250 mL of water immediately after a high-protein meal (△). (From Ref. 6.).

crease GI motility. Other drugs taken concurrently may affect GI motility and, thus, indirectly affect the rate and extent of availability of a particular drug. This concurrent therapy may increase or decrease GI motility, thereby increasing or decreasing the rate at which a drug reaches its absorption site. Such changes in GI motility may increase, decrease, or have no effect on the extent of availability of a drug from an oral dosage form.

Average serum digoxin levels for an elderly female patient taking 0.375 mg of digoxin daily in a tablet dosage form are shown in Fig. 5A. When the patient was also given three daily 10-mg oral doses of metoclopramide, a drug that increases GI motility, the serum digoxin levels dropped (see Fig. 5B), probably indicating a decrease in the extent of digoxin availability. It is possible that this decreased extent of availability occurred because there was insufficient time for the digoxin to be released from its dosage form, or to dissolve in the fluids of the stomach or small intestine, before it was moved completely through the GI tract.

An increase in GI motility will not, however, always decrease the extent of availability of a drug. Figure 6 illustrates the plasma level versus time curves obtained after oral administration of a 100-mg tablet of atenolol to six healthy male volunteers. When the atenolol was given 1 hour after a 25-mg dose of metoclopramide, a slight increase in the rate of availability was observed (i.e., shorter peak time), but the extent of availability of the atenolol was not significantly different from when atenolol was administered alone. Therefore it appears that, under these study conditions, stomach emptying is not the rate-determining step in absorption, nor does an increase in gastrointestinal motility significantly affect the amount of drug available to the general circulation.

Concurrent drug therapy may also decrease the motility of the GIT. Figure 5C shows again average digoxin serum levels in an elderly female patient receiving 0.375 mg of digoxin daily in tablet form. After receiving three daily doses of a 15-mg tablet of propantheline, a drug that slows GI motility, the patient's serum digoxin level increased significantly. Since the digoxin tablet was moving more slowly through its principal site of absorption, the small intestine, it is probable that there was more time for the tablet to disintegrate or for the digoxin to dissolve, thereby increasing the extent of availability.

Returning to Fig. 6, it can be seen that the oral administration of two 15-mg tablets of propantheline 1.5 hours before atenolol delayed the rate of availability of this β-blocker, while increasing its extent of availability [11]. This increased extent might be due to more complete dissolution of the drug, resulting from its increased time in the gastrointestinal tract.

Blood levels versus the time curves obtained after oral administration of two 0.5-g tablets of sulfamethoxazole to male patients are pictured in Fig. 7. (Each subject took the drug with 300 mL of lightly sugared tea.) When a 15-mg tablet of propantheline was taken 30 minutes before the sulfamethoxazole was given and a second propantheline tablet taken with it, both the rate and extent of availability of the sulfa drug from the tablet dosage form decreased (see dotted line in Fig. 7). Here, it is possible that slowing the passage of a drug through the GIT increased the opportunity for drug degradation to take place.

Copyright © 2002 Marcel Dekker, Inc.

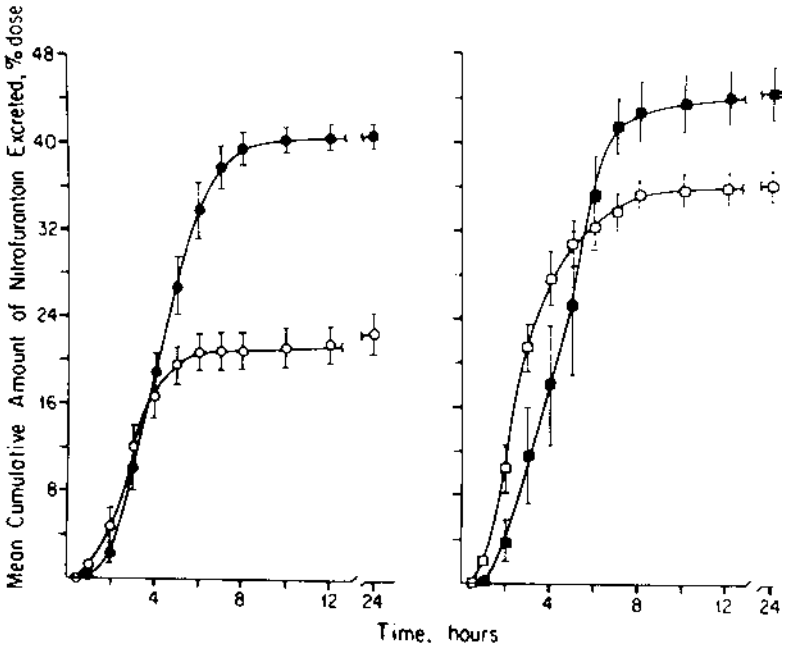


Fig. 4 Mean cumulative urinary excretion of nitrofurantoin after oral administration of a 100-mg macrocrystalline capsule of fasting (○) and nonfasting (●) subjects and a 100-mg microcrystalline tablet to fasting (■) and nonfasting (□) subjects. Vertical bars represent standard errors of the mean. (From Ref. 7.).

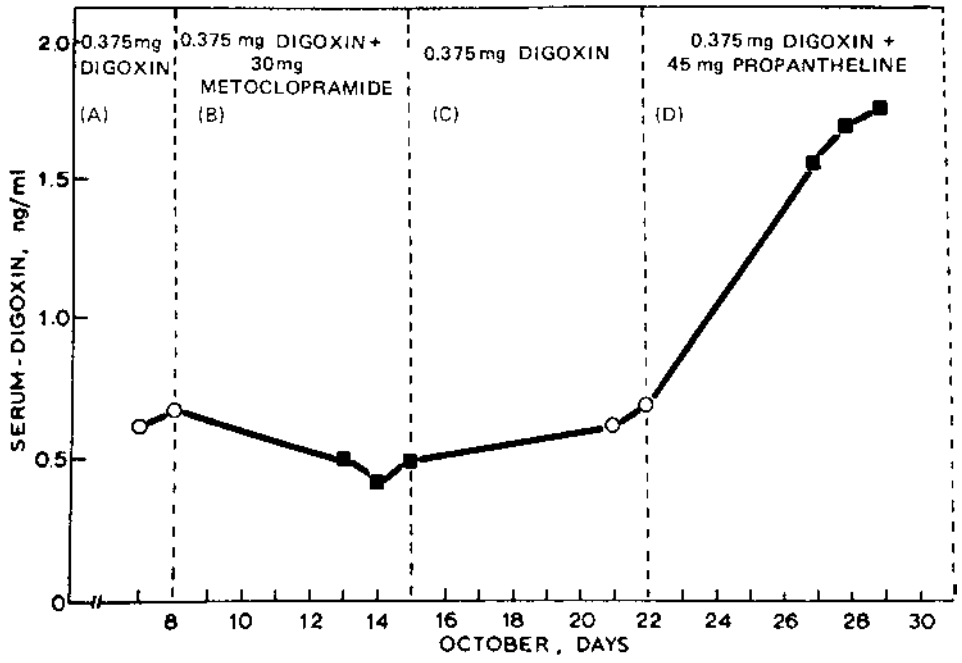


Fig. 5 Variations in serum digoxin concentration in an elderly female patient during treatment with digoxin and metoclopramide or propantheline. (From Ref. 10.).

Copyright © 2002 Marcel Dekker, Inc.

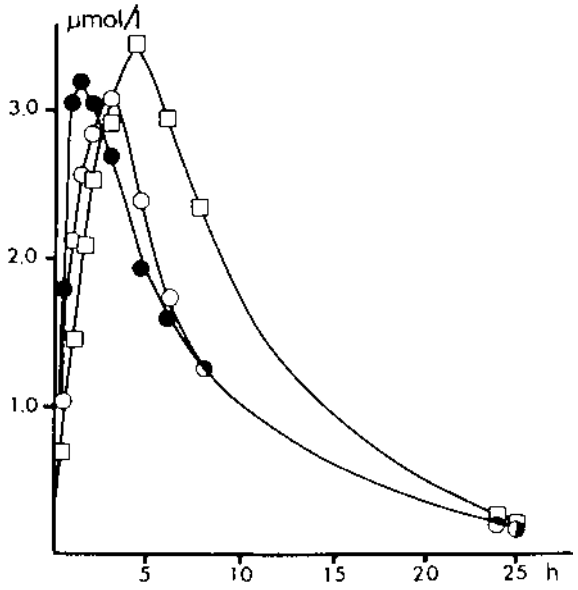


Fig. 6 Mean plasma levels of atenolol in six male volunteers after a 100-mg tablet alone (○), with 25 mg of metoclopramide (●), or with 30 mg of propantheline (□). (From Ref. 11.).

Thus, drugs that alter GI motility may increase or decrease the rate of availability of another drug. In addition, concurrent therapy may increase, decrease, or not affect the extent of availability of a drug. It should be noted that knowledge of the effect of a drug on GI motility usually can lead to an accurate prediction of its potential effect on the rate of availability of a second drug, if, and only if, this is the rate-determining step in absorption, but this knowledge usually does not allow the pharmacist to predict how changes in motility may affect the extent of availability for the second drug. To make such a prediction, additional information, such as degradation and binding mechanisms along the GI tract, must be known.

Since gastric emptying can significantly affect the availability of a drug, it might be suspected that patients who have undergone partial or total gastrectomies will exhibit abnormal rates or extents of availability. Figure 7 illustrates the increased rate of availability of sulfamethoxazole in a patient who had previously undergone a partial gastrectomy. Since the patient had most of his stomach removed, the drug could reach its principal site of absorption, the small intestine, without delay. Some pathological conditions are also accompanied by altered gastrointestinal motility, which may affect availability [4].

The pharmacist, as a drug therapist, should advise both the physician and the patient of the potential problems involved in either initiating or discontinuing concurrent therapy of drugs known to alter GI motility. The pharmacist can also advise the physician

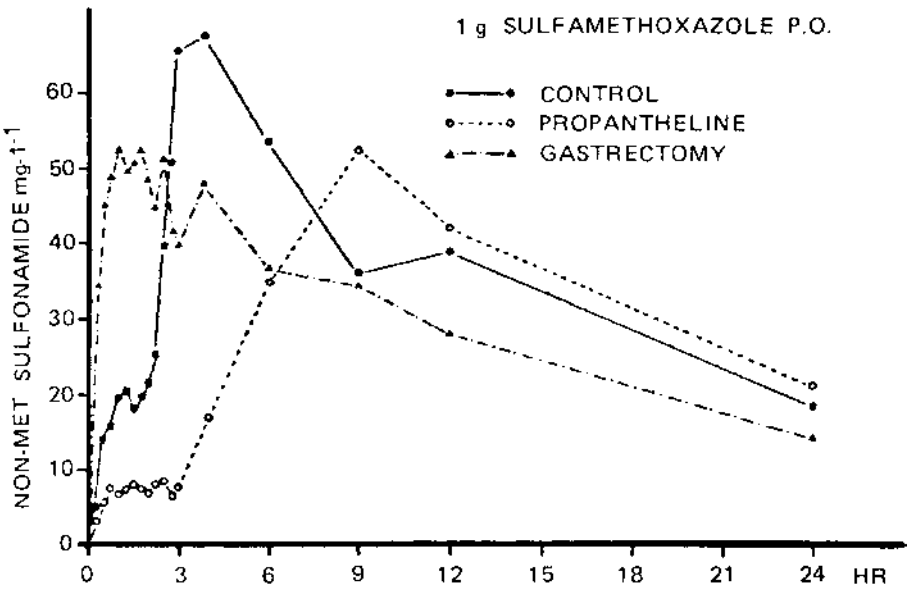


Fig. 7 Blood levels of nonmetabolized sulfamethoxazole following oral administration of 1.0 g, as two 0.5-g tablets, under three different conditions. (From Ref. 12.).

Copyright © 2002 Marcel Dekker, Inc.

on rational drug therapy in patients in whom the gastrointestinal tract has been altered surgically or by disease.

C. Degradation in the Gastrointestinal Tract

Absorption is not the only process that can occur along the GIT. A drug may be degraded or metabolized before it can be absorbed. Chemical degradation, especially pH-dependent reactions, can occur in the solubilizing fluids of the GIT. Drugs that structurally resemble nutrients, such as polypeptides, nucleotides, or fatty acids, may be especially susceptible to enzymatic degradation [2,3]. Digestion of these molecules would result as a consequence of the normal functions of the GIT. Moreover, the GIT is rich in microflora. These microorganisms could metabolize drug molecules before absorption. The variety of degradations attributable to bacteria has been reviewed [14]. In addition, both the fluids and the microflora of the GIT may be altered in disease states. As a result, drug therapy may need to be modified if the rate and extent of availability are affected by these changes [4].

It would seem that metabolism and degradation of a drug within the GIT would serve primarily to reduce its extent of availability. In some instances, however, these processes may enhance availability. For example, clindamycin palmitate is less soluble than clindamycin HCl. A suspension of the palmitate ester is a more stable, better-tasting dosage form than a solution of clindamycin HCl. The palmitate ester is rapidly hydrolyzed in the GIT to free the more soluble active parent antibiotic drug, which becomes rapidly available to the systemic circulation [13]. Molecules such as clindamycin palmitate are chemical derivatives of drugs. These derivatives, called prodrugs, are made to enhance the pharmaceutical properties of the parent molecule. These prodrugs may depend on degradation in the GIT to release the active parent molecule. Consequently, although metabolism or degradation in the GIT decreases the extent of availability for most drugs, with some prodrugs the degradative processes may be essential for complete bioavailability.

The pharmacist should certainly know which drug dosage forms have been designed as delivery systems for prodrugs and also understand the mechanism whereby the active form is made available within the body. Only then can the pharmacist counsel the physician about why a certain product may be inappropriate for such individuals as an achlorhydric patient, a patient with a gastrectomy, or even possibly a patient with a hepatoportal bypass, if liver metabolism is responsible for conversion of the prodrug into the active form.

III. GETTING THE DRUG INTO SOLUTION: FACTORS AFFECTING THE RATE OF DISSOLUTION

Once the dosage form reaches the absorption site, it must break down and release its therapeutic agent. Figure 8 depicts the disintegration and dissolution processes involved in the gastrointestinal absorption of a drug administered in a tablet dosage form. Figure 9 indicates more comprehensively the various solubility problems that may be encountered after the administration of a drug in an oral dosage form. Arrows drawn with heavy dark lines indicate primary pathways that most drugs administered in a particular dosage form undergo. Arrows drawn with dashed lines indicate that the drug is administered in this state in the dosage form. Arrows drawn as thin, continuous lines and labeled "precipitation" indicate situations in which a drug is already in solution, but then precipitates out as fine particles, usually owing to a change in pH of the aqueous environment that causes a change in drug solubility. Other thin arrows indicate secondary pathways that are usually inconsequential in achieving therapeutic efficacy. The dissolution process is primarily dependent on pharmaceutical variables, with the possible exception of a pH dependency, which may be a patient variable. The relative importance of the various processes in Fig. 9 may be explained in terms of the equation developed by Nernst and Brunner [16] whereby dissolution is described by a diffusion layer model:

$$\frac{dQ}{dt} = \frac{D}{h} S(C_s - C_g) \tag{1}$$

where

Q = amount of drug involved
t = time
D = diffusion coefficient of the drug in the solubilizing fluids of the GIT
S = effective surface area of the drug particles
h = thickness of a stationary layer of solvent around the drug particle
C_s = saturation solubility of the drug in the stationary layer, h
C_g = concentration of drug in the bulk fluids of the GIT

In deriving this equation, it was assumed that the drug dissolved uniformly from all surfaces of the

Copyright © 2002 Marcel Dekker, Inc.

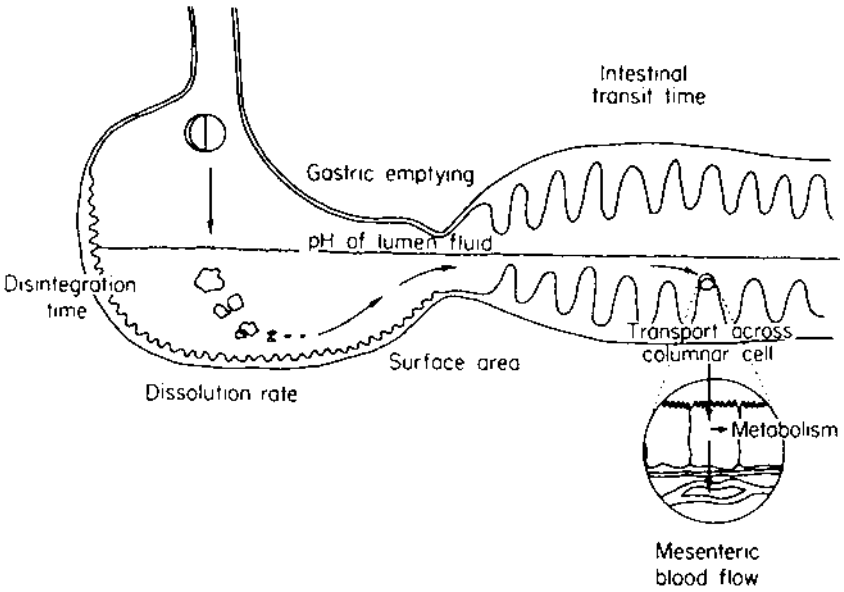


Fig. 8 Factors affecting the rate of absorption of drug from the gastrointestinal tract. (From Ref. 15.).

particles. The particles were assumed to be spherical and all of the same size. In addition, h was assumed to be constant, and both h and C_s were assumed to be independent of particle size. Figure 10 schematically illustrates the relationship of the terms in Eq. (1) to the dissolution process. It should be emphasized that this crude model does not completely describe the dissolution process. The model has, however, helped to explain many experimental results. The following subsections will consider the pharmaceutical implications of the terms given in Eq. (1).

A. Effective Surface Area of the Drug, *S*

Particle Size

The smaller the drug particles, the greater the surface area for a given amount of drug (e.g., 1 g). Thus Eq. (1) predicts that dissolution rate will increase as particle size decreases. Figure 11 confirms this expectation. The smaller granules of phenacetin dissolve more quickly than the larger granules, and there is a graded response; that is, for the five size ranges an increasing amount of drug dissolves over a specified time as the particle size decreases and surface area increases. However, in this example the hydrophobic drug phenacetin has been manipulated by a pharmaceutical manufacturing process called granulation (see Chapter 9 for further discussion) whereby the hydrophilic diluent gelatin has been incorporated into the particle. An opposite effect is seen when phenacetin particles themselves are dissolved in 0.1 *N* HCl (see the solid lines in the lower portion of Fig. 12). In this dissolution study of the hydrophobic drug, the dissolution rate increases with increasing particle size and decreasing particle surface area, in direct contradiction of Eq. (1). However, when a suface-active agent, Tween 80, is added to the 0.1 *N* HCl medium, the dissolution rate of phenacetin particles increases as the particle size decreases (see the dashed lines in the upper portion of Fig. 12). It is probable that decreasing the particle size of a hydrophobic drug actually decreases its *effective* surface area (i.e., the portion of the surface actually in contact with the dissolving fluids). In fact, the smaller phenacetin particles had more air adsorbed on their surfaces and actually floated on the dissolution medium. When a surface-active agent was added to the dissolution medium, the smaller particles were more readily wetted. Thus, their absolute surface area became their effective surface area. In Fig. 11 the granulation process has incorporated the hydrophobic phenacetin particles into the hydrophilic granules where the surface area of the granule approximated the effective surface area.

Figure 13 compares the dissolution rate of pure phenacetin particles with the granulated particles of the same size and with tablets prepared from these granules. These comparisons were made by using

Copyright © 2002 Marcel Dekker, Inc.

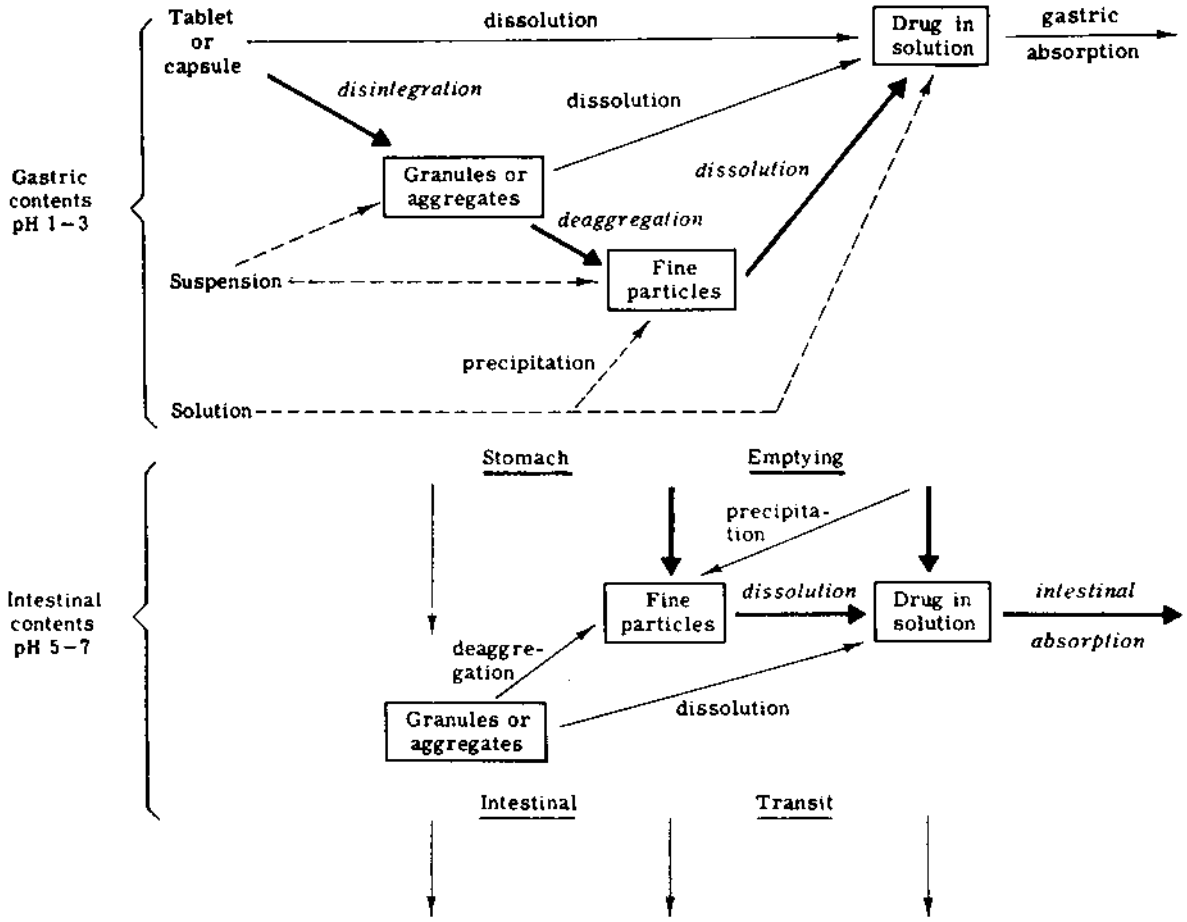


Fig. 9 Processes involved in getting a drug into solution in the GIT so that absorption may take place. Heavy arrows indicate primary pathways that the majority of drugs administered in a particular dosage form undergo. Dashed arrows indicate that the drug is administered in this state in the dosage form. Thin continuous arrows indicate secondary pathways, which are usually inconsequential in achieving therapeutic efficacy. (From Ref. 15.).

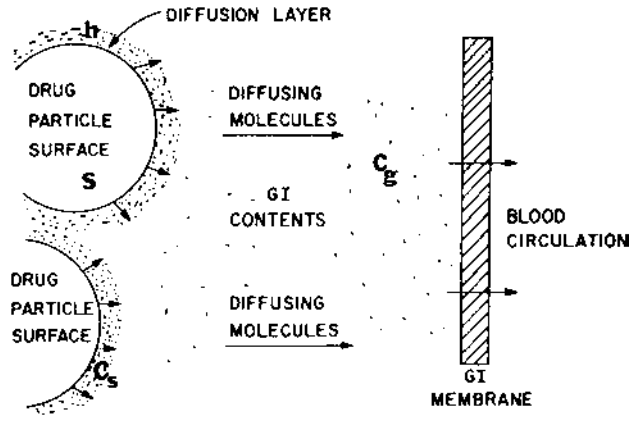


Fig. 10 Schematic diagram of the dissolution process. (From Ref. 1.).

dilute gastric juice as the dissolution medium. Gastric juice has a relatively low surface tension, 42.7 dyn/cm, compared with water, which has a surface tension of approximately 70 dyn/cm. The low surface tension of the gastric juice aids in the wetting of both the hydrophobic particles and the hydrophilic granules. When these hydrophilic granules were compressed into tablets, the dissolution rate of phenacetin decreased but remained greater than the rate of dissolution for the hydrophobic powder. It is possible that the large tablets do not rapidly disintegrate into the smaller granules, as indicated by the apparent lag time before dissolution begins. However, when the tablets do disintegrate, the hydrophilicity of the granules enables them to be more readily wetted than the powdered form.

Copyright © 2002 Marcel Dekker, Inc.

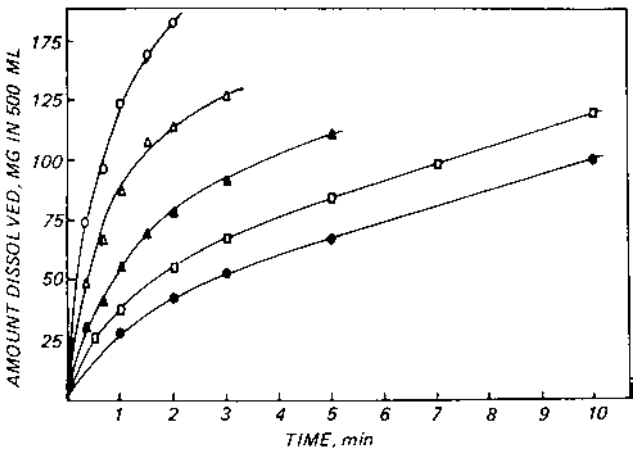


Fig. 11 Effect of particle size of phenacetin on dissolution of drug from granules containing starch and gelatin. ○, particle size 0.11–0.15 mm; △, particle size 0.15–0.21 mm; ▲, particle size 0.21–0.30 mm; □, particle size 0.30–0.50 mm; ●, particle size 0.50–0.71 mm. (From Ref. 17.).

In addition to these in vitro demonstrations of the importance of the effective surface area of drug particles on dissolution rate, many in vivo studies are available. Phenacetin plasma levels versus time are plotted for three different particle sizes of phenacetin in Fig. 14. Healthy adult volunteers received 1.5-g doses of phenacetin as an aqueous suspension on an empty stomach. The results show that both the rate and extent of availability of the phenacetin increase as particle size decreases. Since the large particles dissolve very slowly, the dosage form may pass through the GIT before dissolution is complete. When Tween 80 is added to the dosage form, the rate and extent of availability of the phenacetin increases even more, perhaps as a result of increasing the wettability of the particles, as shown in Fig. 12 for an in vitro study.

The effect of particle size reduction on the bioavailability of nitrofurantoin was shown in Fig. 4. The microcrystalline form (<10 μm) is more rapidly and completely absorbed from the tablet dosage form than is the macrocrystalline form (74–177 μm) from the capsule dosage form. This is not a completely satisfactory illustration of the effect of particle size on the rate and extent of availability, since other manufacturing variables have not been held constant. Nevertheless, it does suggest some correlation between particle size, dissolution rate, and rate of availability.

In summary, it is the effective surface area of a drug particle that determines its dissolution rate. The effective surface area may be increased by physically reducing the particle size, by adding hydrophilic diluents to the final dosage form, or by adding surface-active agents to the dissolution medium or to the dosage form.

Disintegration and Deaggregation

The rate of disintegration of the dosage form and the size of the resulting aggregates can be the rate-limiting

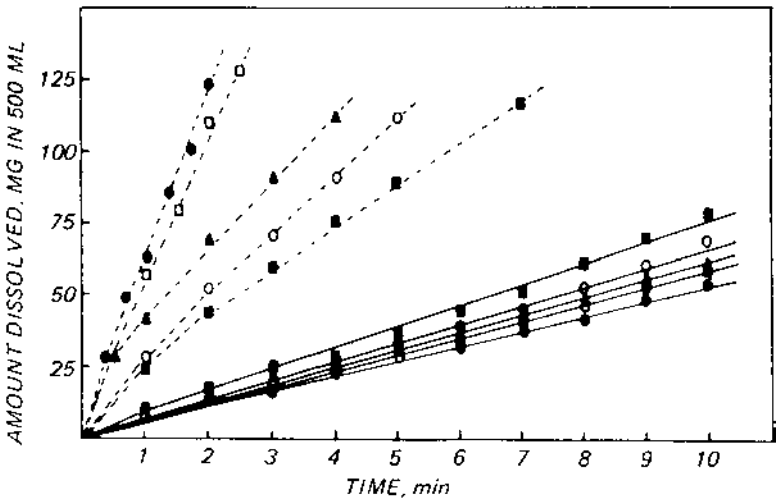


Fig. 12 Effect of particle size on dissolution of phenacetin. ●, particle size 0.11–0.15 mm; □, particle size 0.15–0.21 mm; ▲, particle size 0.21–0.30 mm; ○, particle size 0.30–0.50 mm; ■, particle size 0.50–0.71 mm; —, dissolution medium, 0.1 *N* HCl; - - -, dissolution medium, 0.1 *N* HCl containing 0.2% tween 80. (From Ref. 17.).

Copyright © 2002 Marcel Dekker, Inc.

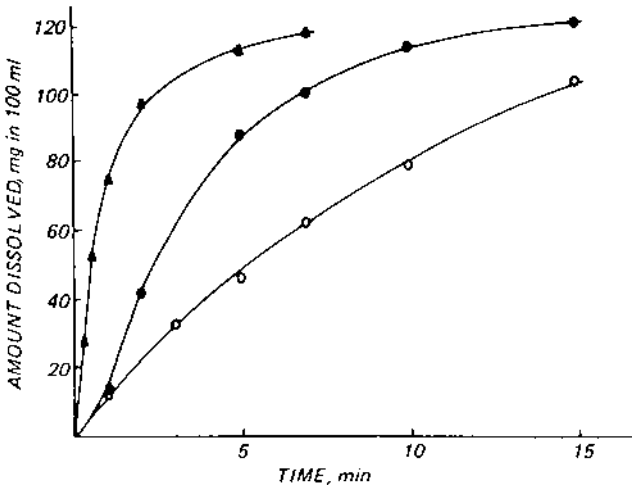


Fig. 13 Rate of dissolution of phenacetin from powder, granules, and tablets in diluted gastric juice (Surface Tension, 42.7 dyn/cm; pH 1.85). ○, Phenacetin Powder; ▲, Phenacetin Granules; ●, Phenacetin Tablets. (From Ref. 17.).

step in the dissolution process. Disintegration is a particularly important step in the dissolution of drugs from coated dosage forms. Tablets or pellets of drugs contained within tablets or capsules are coated for several reasons, as will be discussed in Chapter 10. Rationales for the use of coating include the need to protect a drug during storage or from the very low pH of the stomach. Dosage forms may be coated so that they release their active ingredients slowly for prolonged action. Some drugs are coated to protect the patient's GIT from local irritation. Whatever the reason, these coats are made from materials having various degrees of hydrophilicity that may or may not break down and allow their active ingredients to dissolve. Some poorly formulated coated tablets do not break down at all, and these dosage forms can be recovered intact in the feces. Pharmacists must be aware of such pharmaceutical failures. They must assure themselves that the particular dosage forms they are dispensing to their patients meet all bioavailability standards, as discussed in Chapter 3.

As illustrated in Fig. 9, after a dosage form disintegrates into large particles, these large particles must deaggregate to yield fine particles. Hence, deaggregation may be a rate-limiting step in the dissolution process. Chloramphenicol plasma level versus time curves obtained after the oral administration to healthy volunteers of two 250 mg capsules of each of four different brands of chloramphenicol are plotted in Fig. 15. The in vitro deaggregation rate of the same four brands of chloramphenicol increased in the order $D < B \simeq C < A$. Thus, the in vivo rate and extent of availability of chloramphenicol from these capsule dosage forms correlates well with their in vivo deaggregation rate.

Table 1 summarizes in vitro studies on the effect of pH, a potential patient variable, on the dissolution rate of phenytoin from two different brands of capsules containing 100 mg of sodium phenytoin. Preparation

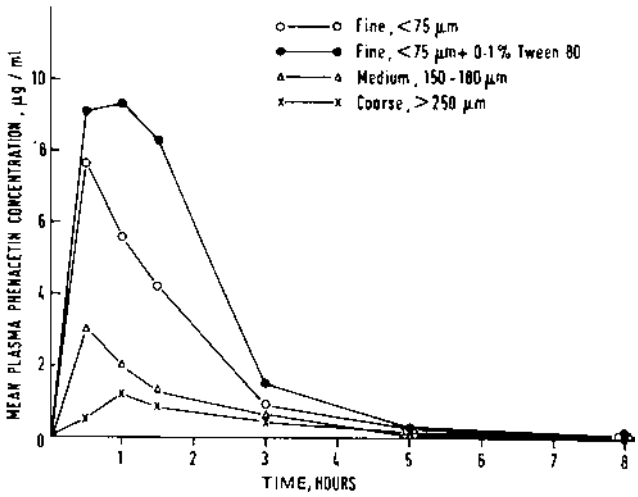


Fig. 14 Mean plasma phenacetin concentrations in six adult volunteers following administration of 1.5-g doses (in aqueous suspensions containing 200 mg of phenacetin per milliliter). (From Ref. 18.).

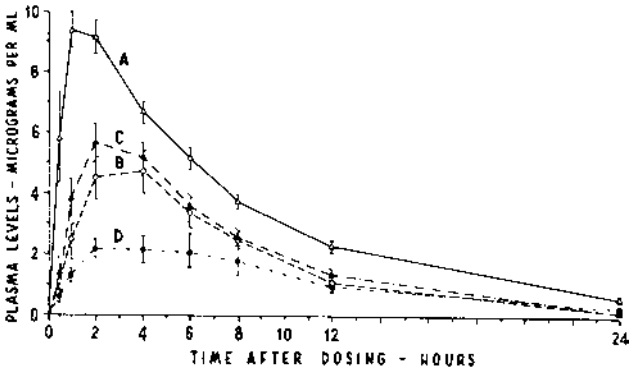


Fig. 15 Mean plasma levels for groups of ten human subjects receiving single 0.5-g doses (as two 250-mg capsules) of chloramphenicol: preparations A, B, C, or D. Vertical lines represent one standard error on either side of the mean. (From Ref. 19.).

Copyright © 2002 Marcel Dekker, Inc.

Table 1 In Vitro Dissolution Characteristics[a] of Two Commercial Phenytoin Sodium Capsule

Solvent system	Initial pH	Time for 50% dissolution (min); average of five capsules (SD)	
		Preparation A	Preparation B
175 mL 0.1 *N* HCl and 175 mL 0.1 *N* NaOH	~ 7	42.4 (5.6)	12.3 (3.2)
175 mL 0.1 *N* HCl for 30 min, then addition of 175 mL 0.1 *N* NaOH	~ 1	34.9[b] (5.6)	80.0[b] (26.2)
175 mL 0.01 *N* HCl for 30 min, then addition of 175 mL 0.01 *N* NaOH	~ 2	32.6[b] (10.4)	25.2[b] (8.6)

[a]Modified Levy-Hayes method, 55 rpm, 17°C.
[b]Time after addition of NaOH; less than 5% dissolved in HCl solution.
Source: Ref. 20.

A retained its compact capsule shape no matter what the initial pH of the dissolution medium. Moreover, the dissolution rate of phenytoin from preparation A was pH-independent. Preparation B, however, broke down into fine particles. When the initial pH was neutral, these particles rapidly dissolved. As the initial pH was lowered, these fine particles precipitated. Since these precipitated particles then dissolved very slowly, it would appear that the precipitate was not the freely soluble sodium phenytoin, but the poorly soluble acid form of phenytoin. Since capsule A did not disintegrate or deaggregate, most of the sodium phenytoin was not exposed to the lower pH media and, as a result, was not converted to its very slowly dissolving acid form.

Effect of Manufacturing Processes

Various manufacturing processes can affect dissolution by altering the effective surface area of drug particles. Each of the individual processes mentioned here is discussed in more detail in Chapter 9. The effect of adding hydrophilic granulating agents to a dosage form has been discussed earlier (see Sec. III.A).

Lubricating agents are often added to capsule or tablet dosage forms so that the powder mass or the finished dosage forms will not stick to the processing machinery. When the hydrophilic lubricating agent sodium lauryl sulfate was added, 325-mg salicylic acid tablets dissolved in 0.1 *N* HCl more rapidly than did control tablets containing no lubricant, as shown in Fig. 16. If, however, the hydrophobic lubricant magnesium stearate was added, the dissolution rate decreased. Most lubricants, and all the effective lubricants, are very hydrophobic, and they act by particle coating. Thus they must be properly formulated to avoid reducing dissolution rate and bioavailability. Once again, increasing hydrophilicity of a dosage form enhanced its dissolution rate in an aqueous medium, but increasing its hydrophobicity decreased its dissolution rate.

Disintegrating agents, such as starch, tend to swell with wetting. In swelling the starch can break apart the dosage form. As seen in Fig. 17, again for 325-mg salicylic acid tablets, as more starch was added to the tablets the drug dissolved more quickly because the dosage form disintegrated more quickly.

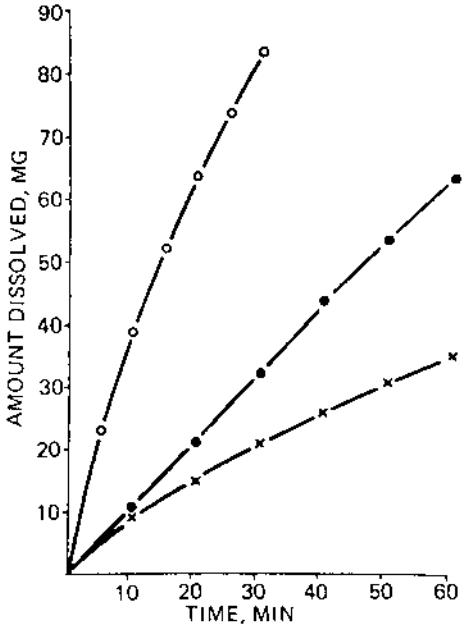


Fig. 16 Effect of lubricant on dissolution of rate of salicylic acid contained in compressed tablets. ×, 3% magnesium stearate; ●, no lubricant; ○, 3% sodium lauryl sulfate. (From Ref. 21.).

Copyright © 2002 Marcel Dekker, Inc.

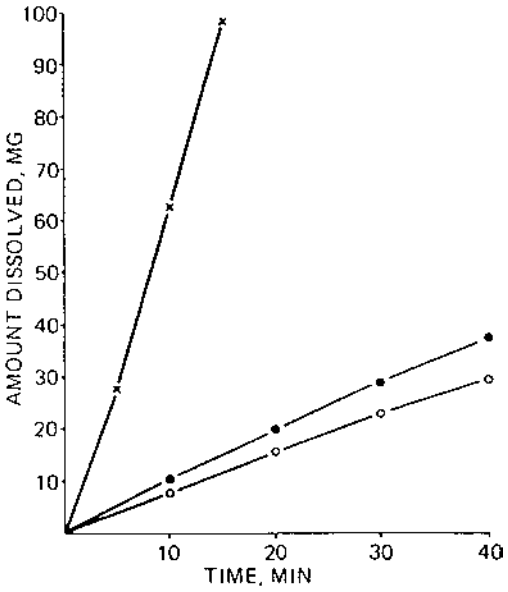


Fig. 17 Effect of starch content of granules on dissolution rate of salicylic acid contained in compressed tablets. ○, 5%; ●, 10%; ×, 20% starch in granules. (From Ref. 22.).

The possible effects on dissolution rate of the force used in compressing the drug-diluent mixture into a tablet dosage form have been summarized in Fig. 18. As compression force is increased, the particles may be more tightly bound to one another. (Part I of Fig. 18 best represents this possibility.) On the other hand, it is also possible that higher pressures may fracture the particles so that they break into yet smaller particles (see part III of Fig. 18). Depending on which of these two extremes is dominant for a given formulation, any of the combinations illustrated in Fig. 18 is possible. Furthermore, a sum of any of them may result. Thus, the effect of the compression force on the dissolution rate of a tablet dosage form would appear to be unpredictable.

The packing density, as illustrated in Fig. 19, may affect the rate of release of a drug (not identified in these studies) from a capsule dosage form [23]. For the rapidly dissolving formulation (see Fig. 19a), packing density had no effect on release rate. For the slowly dissolving formulation (see Fig. 19b), however, increasing packing density decreased dissolution rate. It is probable that further studies of the effect of packing density on dissolution rate for different drugs from different formulations would results in both increasing and decreasing dissolution rates.

From the very few studies discussed here of the many that have been published, it would appear that manufacturing processes can determine the dissolution rate of a drug from its final dosage form. Whether changes in these manufacturing variables are beneficial or detrimental to the ultimate bioavailability of a drug depends on the physicochemical properties of the drug and its dosage form. Pharmacists should be aware of the possible effects that "inert" ingredients and manufacturing methods, which are usually carefully guarded trade secrets, may have on the bioavailability of the drug products they select to dispense to their patients.

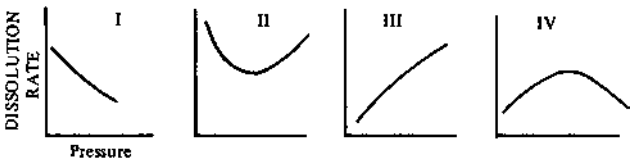


Fig. 18 Effect of compression pressure on dissolution rate. See the text for an explanation. (From Ref. 17.).

B. Saturation Solubility of the Drug, C_s

The next term in Eq. (1) that can be manipulated is C_s, the saturation solubility of the drug. This variable can be influenced by both patient and pharmaceutical variables. The patient variables include the changes in pH as well as the amounts and types of secretions along the GIT. Additionally, both the physical and chemical properties of a drug molecule can be modified to increase or decrease its saturation solubility.

Salt Form of the Drug

Since the salt form of a drug is more soluble in an aqueous medium, the dissolution rate for the salt form of a drug should be greater than the dissolution rate of the nonionized form of the drug. However, the solubility of the salt depends on the counterion; generally, the smaller the counterion, the more soluble the salt. If dissolution is the rate-determining step in absorption, then, for a series of salts and the nonionized form of a drug, it can be anticipated that the rate of availability of the drug will increase as solubility increases. The weakly acidic drug *p*-aminosalicylic acid (PAS) is available as the potassium, calcium, or sodium salt. The solubility of nonionized PAS is 1 g/600 mL of water; of KPAS, 1 g/10 mL; of CaPAS, 1 g/7 mL; and of NaPAS, 1 g/2 mL. Healthy adult volunteers took, on a fasting stomach with 250 mL of water, tablets containing either 4 g of PAS or tablets containing 4 g of

Copyright © 2002 Marcel Dekker, Inc.

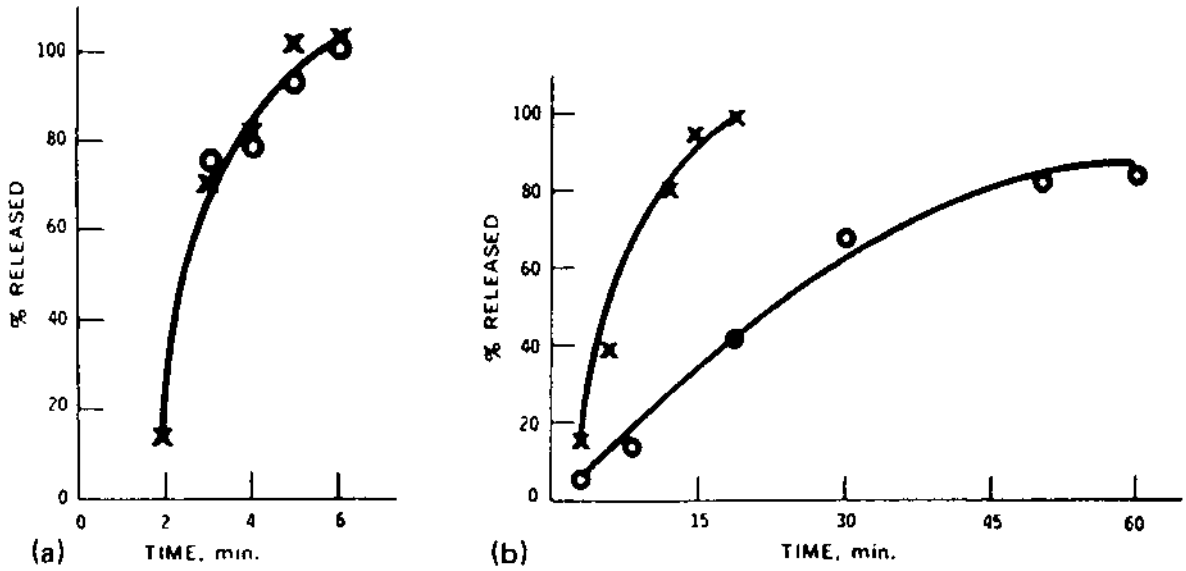


Fig. 19 Influence of packing density on dissolution of drug from capsules in simulated gastric fluid. ×, regular packing: 355 mg/No. 2 capsule; ◯, dense packing: 400 mg/No. 2 capsule. (a) Drug, $CaHPO_4$; (b) Drug, $CaHPO_4$, 5% magnesium stearate. (From Ref. 23.).

one of the salts (containing 2.6–2.8 g of PAS) [24]. Mean plasma concentrations (corrected to a 4-g dose of PAS) versus time curves are shown in Fig. 20. Although the rates of availability of the salts are not significantly different, their rank order does correlate with solubility. All the salt forms were more rapidly available than the nonionized PAS to a significant extent. Furthermore, the extent of availability of the acid form is only about 77% of that of the salt forms, indicating that the drug may not have completely dissolved before the dosage form moved through the GIT. Since commercial tablets were used, it is probable that other dosage form variables (as discussed in Sec. III.A) were not held constant. Thus, the decrease in rate and extent of availability of the PAS may not be exclusively due to differences in solubility.

pH Effect

As indicated, the ionized form of a drug will be more soluble than the nonionized form in the aqueous fluids of the GIT. The classic studies on the beneficial effects of changing nonionized drugs into salt forms were reported by Nelson for tetracycline [25], and Nelson et al. for tolbutamide [26]. Table 2 combines portions of the data from each study. Urinary excretion of the drug or its metabolite was taken as the in vivo measure of the relative absorption rate for the salt and the nonionized

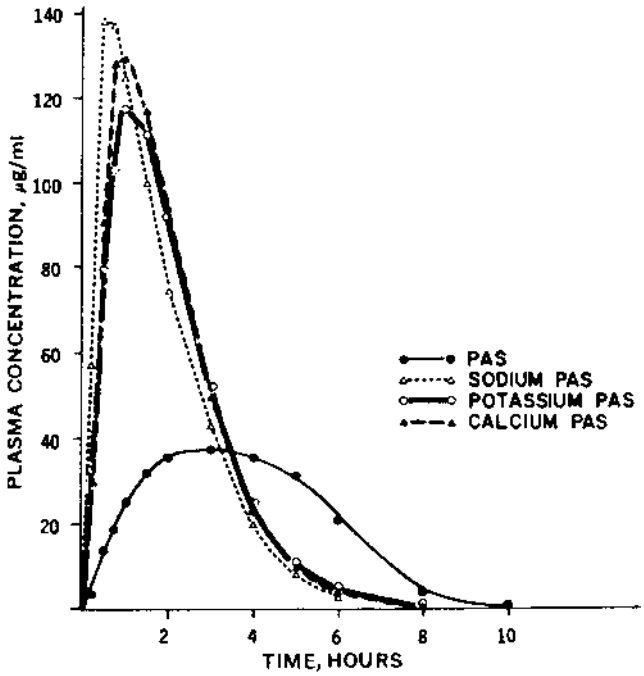


Fig. 20 Mean plasma concentrations of unchanged drug from 12 subjects following administration of four different preparations of *p*-aminosalicylic acid (PAS). Data were corrected to 70 kg of body weight and to a dose equivalent of 4 g of free acid. (From Ref. 24.).

Copyright © 2002 Marcel Dekker, Inc.

Table 2 Correlation of Dissolution Rates with Biological Measurements for Tolbutamide and Tetracycline Absorption in Humans

Drug as nondisintegrating pellet	In vitro dissolution rate[a] ($mg\,cm^{-2}h^{-1}$)		Average amount excreted (mg) to time indicated[d]			Lowering of blood sugar level (mg/100 mL) after 1 h
	0.1 *N*HCl[b] or simulated gastric fluid[a]	pH 7.2 buffer[b] or simulated neutral intestinal fluid[c]	1 h	2 h	3 h	
Tolbutamide	0.21 (N → N)	3.1 (N → I)	5	7	12	5.2
Sodium tolbutamide	1069 (I → N)	868 (I → I)	21	65	117	9.1
Tetracycline	2.6 (N → I)	0.001 (N → N)	0.2	1.5	3.3	
Tetracycline HCl	4.1 (I → I)	7.8 (I → N)	3.0	12.0	20.4	

[a]The N → N, etc. designations for the dissolution data are explained in the text.
[b]Tolbutamide study.
[c]Tetracycline study.
[d]Tolbutamide excretion measured as the carboxytolbutamine metabolite.
Source: Refs. 25 and 26.

form of each drug. No comparison can be made between the two drugs, and they are combined here only to illustrate that the same principles hold for both positively and negatively charged drug ions. Note that the salt forms of the drug dissolve much faster than the nonionized forms (or zwitterion for tetracycline) in all media and that more of the salt forms of the drug are absorbed and subsequently excreted in each period. For a dissociable drug, either an acid or a base, such as tolbutamide or tetracycline, the pH of the GI fluids will determine whether the drug is ionized or nonionized. The dependence of the dissolution rate of a drug on pH is not evident in the Nernst-Brunner as written in Eq. (1). That equation may be rewritten as

$$\frac{dQ}{dt} = \frac{DS}{h}(C_h - C_g) \tag{2}$$

where Q, t, D, S, h, and C_g are defined as before, and $C_h =$ saturation solubility of the drug in the boundary layer, h, at any particular pH; $C_h = C_s$ when the drug is dissolving in an aqueous solution in which it is totally nonionized. Then, for a weakly acidic drug:

$$\mathrm{AH} \overset{K_a}{\rightleftharpoons} \mathrm{A^-} + \mathrm{H^+}; \quad K_a = \frac{[\mathrm{A^-}][\mathrm{H^+}]}{[\mathrm{AH}]}$$

$$C_h = [\mathrm{AH}] + [\mathrm{A^-}]$$

$$C_h = [\mathrm{AH}]\left(1 + \frac{K_a}{[\mathrm{H^+}]}\right) \text{ but, } \quad C_s = [\mathrm{AH}]$$

and, therefore,

$$C_h = C_s\left(1 + \frac{K_a}{[\mathrm{H^+}]}\right)$$

so

$$\frac{dQ}{dt} = \frac{DS}{h}\left\{C_s\left(1 + \frac{K_a}{[\mathrm{H^+}]}\right) - C_g\right\} \tag{3}$$

Therefore, as pH increases, the dissolution rate of a week *acid* increases. Similarly, for a weak base:

$$\mathrm{BH^+} \overset{K_a}{\rightleftharpoons} \mathrm{B} + \mathrm{H^-}; \quad K_a = \frac{[\mathrm{B}][\mathrm{H^+}]}{[\mathrm{BH^+}]}$$

$$C_h = [\mathrm{BH^+}] + [\mathrm{B}]$$

$$C_h = [\mathrm{B}]\left(1 + \frac{[\mathrm{H^+}]}{K_a}\right) \text{ but, } \quad C_s = [\mathrm{B}]$$

and, therefore

$$C_h = C_s\left(1 + \frac{[\mathrm{H^+}]}{K_a}\right)$$

so

$$\frac{dQ}{dt} = \frac{DS}{h}\left\{C_s\left(1 + \frac{[\mathrm{H^+}]}{K_a}\right) - C_g\right\} \tag{4}$$

Thus, as the pH increases, the dissolution rate of a weak *base* decreases. Referring to Table 2, we can see that, for the weak acid tolbutamide, the dissolution rate increases as pH is increased, as predicted by Eq. (3). Additionally, for the weak base tetracycline, as predicted by Eq. (4), the dissolution rate decreases as pH is increased. Thus far, the more rapid dissolution of the salt forms of these drugs and the direction of change of the dissolution rate with pH have been accounted for with Eqs. (1) to (4). However, there are six possible dissolution rate

Copyright © 2002 Marcel Dekker, Inc.

comparisons that can be made for each set of nonionized drug and its salt in the two buffers, as discussed by Benet [15]. The letters in Table 2 indicate similar measurements in the two studies. Thus N → N indicates a nonionized solid drug dissolving in a medium in which the dissolved solute in the bulk solution will be nonionized; N → I indicates a nonionized solid drug dissolving in a medium in which the dissolved solute in the bulk solution will be ionized; I → I indicates an ionized solid drug dissolving in a medium in which the dissolved solute will be ionized; and I → N indicates an ionized solid drug becoming nonionized solute after dissolution.

Process I → I should be faster than process N → N, since the solubility of the dissolving substance will be much greater for the salt than for the nonionized molecule. A similar explanation can be used for (I → N)>(N → N) and (I → I)>(N → N) and (I → I)>(N → I), but it must be remembered that according to Eqs. (1), (3), and (4), C_s is the saturation solubility of the drug in the diffusion layer, not in the bulk solution. It is believed that the dissolving solid acts as its own buffer and changes the pH of the liquid environment immediately surrounding the solid particle; thus, the dissolution rate should be governed by the solubility of the drug in the buffered diffusion layer. The I → N and N → N comparison is especially significant in the oral administration of weakly acidic drugs and their salts, since the acidic region of the stomach is the first solvent medium encountered following normal oral dosing. Frequently, administering the sodium or potassium salt of an acidic drug actually speeds up absorption by increasing the effective surface area of the solid drug according to the following hypothesized process. The salt acts as its own buffer in the diffusion layer and goes into solution in this layer. However, when the salt molecules diffuse out of the layer and encounter the bulk solution, they precipitate out as very fine nonionized prewetted particles. The large surface area thus precipitated favors rapid dissolution when additional fluid becomes available for one of the following reasons: (a) dissolved particles are absorbed; (b) more fluid accumulates in the stomach; or (c) the fine particles are emptied into the intestine. The classic study of Lee et al. [27] comparing serum levels of penicillin V following administration of the salt and free acid to dogs is explained by this phenomenon also. However, in at least three cases—aluminum acetylsalicylate [28], sodium warfarin [29], and the pamoate salt of benzphetamine [30]—administration of the salt slowed dissolution of the drug and subsequent absorption as compared with the nonionized form. This decrease appears to be due to precipitation of an insoluble particle or film on the surface of the tablet, rather than in the bulk solution. Precipitation of an insoluble particle or film onto the surface of the tablet decreases the effective surface area by preventing deaggregation of the particles.

The comparison of I → N and N → I may also be explained by the buffered pH in the diffusion layer and leads to an interesting comparison between a process under kinetic control versus one under thermodynamic control. Because the bulk solution in process N → I favors formation of the ionized species, a much larger quantity of drug could be dissolved in the N → I solvent if the dissolution process were allowed to reach equilibrium. However, the dissolution rate will be controlled by the solubility in the diffusion layer; accordingly, faster dissolution of the salt in the buffered diffusion layer (process I → N) would be expected. In comparing N → I and N → N, or I → N and I → I, the pH of the diffusion layer is identical in each set, and the differences in dissolution rate must be explained either by the size of the diffusion layer or by the concentration gradient of drug between the diffusion and the bulk solution. It is probably safe to assume that a diffusion layer at a different pH than that of the bulk solution is thinner than a diffusion layer at the same pH because of the acid-base interaction at the interface. In addition, when the bulk solution is at a different pH than that of the diffusion layer, the bulk solution will act as a sink and C_g can be eliminated from Eqs. (1), (3), and (4). Both a decrease in the h and C_g terms in Eqs. (1), (3), and (4) favor faster dissolution in processes N → I and I → N as opposed to N → N and I → I, respectively.

Although the explanation for (N → I)>(N → N) and (I → N)>(I → I) is self-consistent for a nonionized drug and its salt form and reflects the experimentally observed values in Table 2, Nelson [31] studied a series of weak organic acids and found (I → I)>(I → N) for the sodium salt of four of these compounds. For example, the dissolution rate of sodium benzoate in pH 6.83 buffer was 1770 mg/100 min cm^2 versus 980 mg/100 mm cm^2 in a pH 1.5 solution. Corresponding values of I → I and I → N were 820 versus 200 for sodium phenobarbital, 2500 versus 1870 for sodium salicylate, and 810 versus 550 for sodium sulfathiazole. The acid forms of these drugs all showed the expected (N → I)>(N → N) relationship, and we cannot yet explain the salt data [33]. As stated earlier, the preceding explanation of the data in Table 2 is presented with reference to a specific theory of dissolution. Although this theory may not be acceptable to some, it does provide a basis for understanding the general principles that dictate the (I → N) ≈ (I → I) >(N → I)>(N → N) relationship observed in dissolution rate measurements for a nonionized drug and its salt form.

Copyright © 2002 Marcel Dekker, Inc.

The pH of a solution affects not only the active ingredients of a dosage form, but also the inert ingredients. Returning to Table 1, the pH dependence of two different commercial capsules of sodium phenytoin were studied [20]. For one capsule, dissolution rate was independent of the initial pH of the dissolution media. For the second capsule, dissolution rate was very much dependent, as has been discussed, on the initial pH. Similar studies have been made using three commercial dosage forms of nitrofurantoin (Table 3). Nitrofurantoin is a weak acid with $pK_a = 7.2$. Equation (3) predicts that as the pH of the dissolution medium increases, the dissolution half-life of nitrofurantoin should decrease. For the suspension of microcrystalline nitrofurantoin and for the capsule, containing 100 mg of nitrofurantoin in the macrocrystalline form, the experimental results follow this prediction. For the tablet, containing 100 mg of nitrofurantoin in the microcrystalline form, the dissolution half-life increases with increasing pH. However, the tablet did not disintegrate at pH 7.20. Thus, for the tablet, it would appear that the physicochemical properties of the dosage form, rather than physicochemical properties of the drug, determined the rate of release of the drug from its dosage form. Interestingly, both the capsule and tablet, which must both disintegrate, are less rapidly available than the suspension. (See the dissolution scheme shown in Fig. 9.)

It has been mentioned that the pH of the bulk fluids may not reflect the pH of the stationary diffusion layer. As the active or inert ingredients of a dosage form dissolve, they may alter the pH of the microenvironment of the stagnant diffusion layer, without significantly changing the pH of the bulk fluids. Moreover, it is possible to intentionally add buffering agents to a capsule of tablet dosage forms. The small amount of buffer may not alter the pH of the bulk fluids of the GIT but can buffer the pH in the diffusion layer to a pH favoring rapid dissolution of the active ingredient. Such an effect was hypothesized when dissolution rates were compared for buffered versus plain (no alkaline additives) tablets of aspirin [33]. Later, more extensive studies on the effects of a variety of buffers on the dissolution of aspirin tablets were carried out [34]. Hydrophilic buffers and buffers that released carbon dioxide increased the dissolution rate; however, hydrophobic buffers decreased the dissolution rate, possibly by effectively waterproofing the tablets.

Table 3 Effect of pH on the Dissolution Half-Life of Nitrofurantoin from Commercial Dosage Forms at 37°C

	Mean dissolution half-life[a] (min)	
Commercial dosage form	pH 1.12	pH 7.20
Aqueous suspension[b]	12.5 (1.2)	2.64 (0.34)
Compressed tablet[c]	77.9 (19.0)	167.0 (35)
Gelatin capsule[d]	212.0 (44)	160.0 (24)

[a]Determined from log-normal probability plots of individual rate data. Mean of five determinations (SD).
[b]Furadantin suspension containing microcrystalline (< 10 μm) drug.
[c]Furadantin tablets containing microcrystalline (< 10 μm) drug.
[d]Macrodantin capsules containing microcrystalline (74–177 μm) drug.
Source: Ref. 32.

In summary, the effect of pH on the dissolution rate of a drug from an oral dosage form depends on (a) the pH of the GI fluids, a patient variable; (b) the acid or base strength of the drug, a pharmaceutical variable; as well as (c) the physicochemical properties of the dosage form, another pharmaceutical variable. Furthermore, by intentionally designing the dosage form such that it buffers the diffusion layer, we can control a patient variable by a pharmaceutical variable.

Pharmacists are probably the only health professionals who, by training, understand pH effects on drug solubility and transport. They must be aware of why different drugs are prepared as different salt forms and how changes in the acid-base environment of the GIT will influence availability. Usually, drugs in a salt form will be more quickly available and, often, available to a greater extent. However, there are instances when the drug may not be formulated in the salt form, even when a proper, stable formulation of the salt may be produced. Such an example was shown in Table 2. Sodium tolbutamide dissolves faster than tolbutamide when both in vitro dissolution data and in vivo urinary excretion measurements are compared. As would be expected, the pharmacological effect (i.e., lowering of the blood sugar level) is also more pronounced after 1 hour for the salt form. However, the product is actually formulated containing the acidic nonionized tolbutamide, because the manufacturer does not want to induce such a rapid decrease in blood sugar, which might lead to diabetic coma.

Solvate Formation

Another variable that influences the saturation solubility of a drug molecule is its degree of solvation. Since the anhydrous, hydrated, and alcoholated forms of a drug have slightly different solubilities, they may well have different dissolution rates and, therefore, different rates of absorption. However, these differences may not be clinically significant [35].

Copyright © 2002 Marcel Dekker, Inc.

Polymorphism

Yet another property of a drug that may affect its saturation solubility and, hence, its dissolution rate is its crystalline state. Many drugs exhibit polymorphism [36]; that is, they are available in an amorphous or several different crystalline states. Chloramphenicol palmitate is available in at least two different forms, A and B. The B form is the more soluble. Figure 21 shows the mean serum level of chloramphenicol versus time curves obtained after oral administration to male volunteers of chloramphenicol palmitate suspensions containing the equivalent of 1.5 g of chloramphenicol [37]. The fraction of the B form of chloramphenicol palmitate ranges from 0% in suspension M to 100% in suspension L. Since the suspensions were identical except for the crystalline form, it would appear that the increase in rate and extent of availability with an increasing percentage of the B form is due to the increasing rate of dissolution. A warning, however, is necessary. In general, the more soluble polymorph is the least stable thermodynamically. Thus, older dosage forms may contain more of the more stable, but less soluble, polymorph. Aging may significantly affect the bioavailability of drugs exhibiting polymorphism. Pharmacists must, for this and many other reasons, be aware of the storage requirements of all drugs dispensed. They must also counsel their patients on the proper storage of their medicines and the importance of discarding out-of-date drugs.

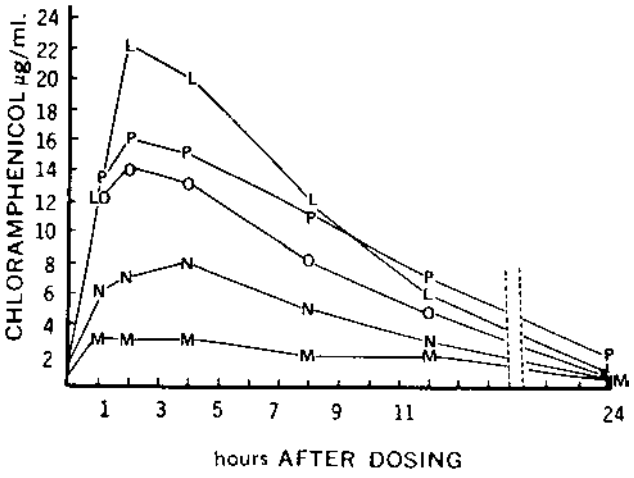


Fig. 21 Comparison of mean blood serum levels obtained with chloramphenicol palmitate suspensions containing varying ratios of A and B polymorphs, following a single oral dose equivalent to 1.5 g of chloramphenicol. Percentage of polymorph B in the suspension: M, 0%; N, 25%; O, 50%; P, 75%;L, 100%. (From Ref. 37.).

Complexation

Several studies have examined the effects of complex formation on the rate and extent of drug availability. A drug may complex with both absorbable and nonabsorbable excipients in a dosage form. This complexation may occur within the dosage form or in the solubilizing fluids, and the resulting complex may be more or less soluble than the drug itself. The rate and extent of availability of promazine from a suspension of the drug adsorbed on either activated charcoal or the clay attapulgite were compared with the availability of the drug from an aqueous solution [38]. Healthy adults took, on an empty stomach, 45 mL of either the solution or one of the suspensions with 4 oz (0.12 L) of water. The rate of availability of promazine from the attapulgite suspension was less than from the solution, but the extent of absorption was the same. When compared with the solution, the promazine-activated charcoal suspension showed both a decreased rate and a decreased extent of availability. However, the association constant for the promazine-attapulgite complex was greater than for the promazine-activated charcoal complex. Thus, it appears that it is not the magnitude of the association constant of the complex, but the rate at which the complex dissociates, that determines whether absorption of the drug is as rapid or as complete as in the absence of complex formation.

In addition to possible complexation with other ingredients in the dosage form, a drug may complex with the natural components of the GIT. The mucin, enzymes, bile, and physiological surfactants found in the mucosal fluids may interact with a given drug. The influence of these interactions on the availability of the drug will depend on the physicochemical properties of the drug and those of the endogenous compound, as discussed in Chapter 2. It seems probable that both increased and decreased drug solubilities will be encountered. Thus, for drugs for which the rate of availability is dissolution rate-limited, such interactions may be beneficial or detrimental. Once again, patient variables may control the availability of the drug from its dosage form.

Solid-Solid Interactions

Preparing a solid solution of a drug is one additional way of controlling its dissolution rate. For this chapter, the term *solid solution* will be used to describe any solid system in which one component is dispersed at the

Copyright © 2002 Marcel Dekker, Inc.

molecular level within another [39]. In deriving Eq. (1) it was assumed that C_s is independent of particle size. However, for extremely small particles, such as those found in solid solutions, the saturation solubility (C_s) of the drug may increase as particle size decreases. This increase in solubility can be described by the Kelvin equation [40]:

$$C_s^{\text{micro}} = C_s \exp\left(\frac{2\gamma M}{r\rho RT}\right) \tag{5}$$

where

C_s^{micro} = saturation solubility of the microscopic particle
C_s = saturation solubility of drug (macroparticles)
γ = interfacial tension between drug particles and the solubilizing fluids
M = Molecular weight of the drug
r = radius of the microscopic drug particle
R = ideal gas constant
ρ = density of the microscopic drug particle
T = absolute temperature

Figure 22 depicts the dissolution rates of griseofulvin, an insoluble antifungal agent, and griseofulvin–succinic acid samples as described by Goldberg et al. [41]. Several conclusions can be drawn from these experiments. The first is that physically reducing the particle size of the griseofulvin by micronization increased the rate of dissolution in accordance with Eq. (1) but did not increase the concentration of drug above its equilibrium solubility. Second, physical mixtures of griseofulvin and succinic acid, at ratios corresponding to those in the solid solution and eutectic mixture, had dissolution rates similar to griseofulvin alone. Therefore, succinic acid does not itself increase the solubility of griseofulvin (by a complexation mechanism, for example). Finally, the solid solution and the eutectic mixture were very rapidly soluble. Both gave supersaturated solutions, as predicted by Eq. (5), because of the extremely small—approaching molecular—size of the griseofulvin particles in these states. Solid solutions and eutectic mixtures have been used, at least in research laboratories, to increase the rate of dissolution of drugs, presumably by decreasing the particle size of the drug molecules [42]. However, since the drug is in an unstable thermodynamic state, it is critical that the stability of the formulation be maintained. Otherwise, the preparation may revert to a state of having even less favorable dissolution properties than those of a micronized formulation, as shown by the dissolution characteristics of the physical mixtures in Fig. 22. Pharmacists should be aware of the potential problems of such formulations and may, once again, take an active role in advising patients of proper storage conditions.

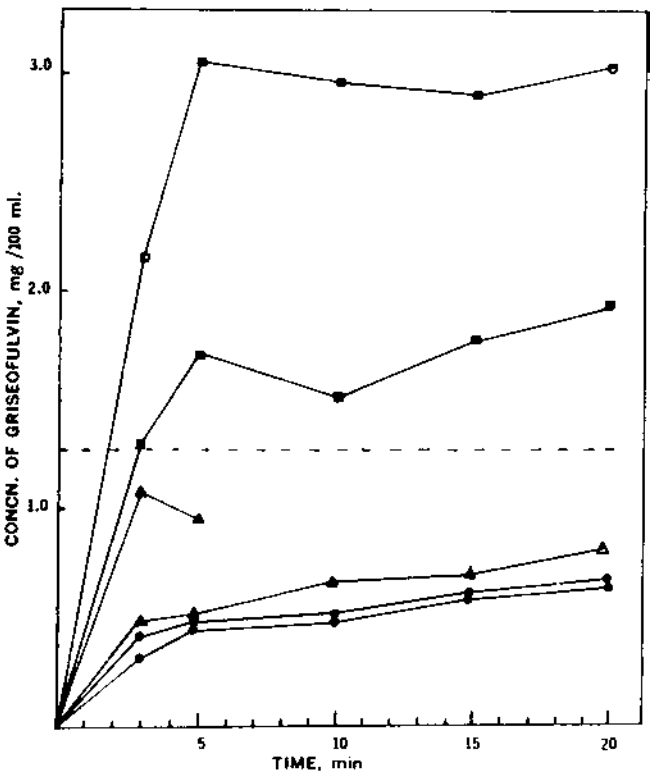


Fig. 22 Dissolution fates of various griseofulvin and griseofulvin-succinic acid samples as determined by the oscillating bottle method. ●, griseofulvin, crystalline; ▲, griseofulvin, micronized; ■, eutectic mixture; ○, physical mixture at eutectic composition; □, solid solution; △, physical mixture at solid solution composition. The dashed line indicates the equilibrium solubility of griseofulvin in water. (From Ref. 41.).

C. Concentration of the Dissolved Drug in Bulk Solution, C_g

In Eq. (1), the saturation solubility of the drug in the diffusion layer does not of itself determine dissolution rate. The determinant is, rather, the difference between C_s and C_g, the concentration of dissolved drug in the bulk fluids of the GIT. Thus, the driving force for dissolution is the concentration gradient, $C_s - C_g$. It is usually assumed that C_g is much smaller than C_s, meaning that dissolution occurs under sink conditions. If a drug is absorbed through the GI membrane very slowly, drug concentration in the GIT fluids may

Copyright © 2002 Marcel Dekker, Inc.

build up. This buildup would, by decreasing the gradient, decrease the dissolution rate. Additionally, the bulk fluids may not be identical with the solubilizing fluids of the diffusion layer. A drug may be more soluble in the boundary layer and then precipitate in the bulk fluids, especially if the pH differs between these two sites. However, these precipitated particles should be quite small, and thus may rapidly redissolve (see the dissolution scheme presented in Fig. 9). On the other hand, the drug may be more soluble in the bulk fluids than in the boundary layer because of a difference in pH or by complexation with other components. Here, the gradient may decrease and dissolution may slow down or even stop. However, the volume of the bulk fluids is much larger than the volume of the boundary layer. Thus, a large absolute amount of drug in the bulk fluids may still give a small value of C_g.

Figure 3 illustrates a situation in which this may not be true. When 250 mL of water was taken with erythromycin tablets, the extent of absorption was much greater than when the tablets were taken with only 20 mL of water. In the latter case, dissolution probably did not occur under sink conditions. Hence, the dissolution rate decreased, and it appears that not all of the erythromycin had a chance to dissolve in the GIT. Note than the dissolution was not, however, the rate-determining step in absorption, since the time to reach the peak concentration was the same in all situations.

Usually, an increase in C_g that would affect the dissolution rate would occur only when another process, such as membrane transport or stomach emptying, becomes the rate-limiting step in drug absorption. As a general rule pharmacists should advise patients to take their oral medications with a full glass of water to ensure that dissolution occurs under optimal conditions.

D. Diffusion Coefficient Divided by the Thickness of the Stationary Diffusion Layer, D/h

Although it is possible to control the dissolution rate of a drug by controlling its particle size and solubility, the pharmaceutical manufacturer has very little, if any, control over the D/h term in the Nernst-Brunner equation, Eq. (1). In deriving the equation it was assumed that h, the thickness of the stationary diffusion layer, was independent of particle size. In fact, this is not necessarily true. The diffusion layer probably increases as particle size increases. Furthermore, h decreases as the "stirring rate" increases. In vivo, as GI motility increases or decreases, h would be expected to decrease or increase. In deriving the Nernst-Brunner equation, it was also assumed that all the particles were spherical and of the same size. In fact, particles in pharmaceutical systems are neither spherical nor uniform. Furthermore, pharmaceutical systems are usually mixtures of different compounds, the particle sizes of which are not identical. Thus pharmaceutical systems are polydisperse and multiparticulate [43]. Their size distribution in terms of number of particles tends to be skewed toward the smaller particles. Furthermore, as dissolution proceeds, the particles become smaller. Hence h can vary considerably initially and throughout the dissolution process.

The other virtually uncontrollable term in Eq. (1) is D, the diffusion coefficient of the drug. For a spherical, ideal drug molecule in solution,

$$D = \frac{kt}{6\pi\eta r}$$

where

k = Boltzmann's constant
T = absolute temperature
r = radius of molecule in solution
η = viscosity of the solution

Both k and 6π are constants. The radius r is a property of the drug molecule not subject to manipulation. The viscosity η of the GI fluids can vary. Increasing the viscosity will decrease dissolution. In addition, increasing the viscosity of the stomach contents will slow gastric emptying, thereby delaying delivery of the drug to the absorption site. Increasing the temperature of the GI fluids tends to increase diffusion. Patients might be advised to take oral dosage forms with warm liquids. However, extremely hot liquids will delay stomach emptying. In summary, D and h are largely uncontrollable factors, although they may influence drug availability.

IV. GASTROINTESTINAL MEMBRANE TRANSPORT

Once a drug is in solution at the absorption site, it must move through the GIT membrane and then into the general circulation. In Chapter 2 Fick's law was used to describe this transport:

$$\left(\frac{dQ_b}{dt}\right)_{g\to b} = D_m A_m R_{m/\text{aq}} \left[\frac{(C_g - C_b)}{X_m}\right] \tag{6}$$

Paradoxically, the larger the value of C_g, the more quickly the drug will more through the membrane. This C_g is the same concentration of drug in the bulk fluids that was minimized in Eq. (1) to increase the rate

Copyright © 2002 Marcel Dekker, Inc.

of dissolution. A second paradox is the relation between $R_{m/aq}$ and C_g. The bulk fluids of the GIT are aqueous. Thus, C_g will increase as the water solubility of the drug increases; $R_{m/aq}$ will, however, decrease. Thus, to be absorbed, a drug cannot be so lipid-soluble that it will not dissolve in the aqueous fluids of the GIT, nor so water-soluble that it will not penetrate the lipid GI membrane. Recently, Lipinski and colleagues presented the empirically derived Rule of 5 (Table 4) [44]. These guidelines allow drug developers to select as drug candidates molecules with acceptable hydrophilicity and acceptable lipophilicity, that is, molecules likely to have good oral availability. Although diffusion across the GI membrane is one of the major rate-limiting steps in oral absorption, it has been extensively covered in Chapter 2 and will not receive further treatment here.

The GI wall is not, however, an inert barrier. Some drugs may be actively transported into the enterocytes, as discussed in Chapter 2. Additionally, it is now known that there are efflux transporters in these cells. The most studied is P-glycoprotein (Pgp), which is a relatively nonselective carrier for a wide variety of drugs [45]. In addition, the enterocytes also contain drug-metabolizing enzymes, particularly cytochrome P4503A4 (CYP3A4), which metabolizes a significant number of marketed drugs [46]. To further complicate the picture, some drugs are substrates of both Pgp and CYP 3A4 [45]. Thus, drug molecules that are absorbed from the lumen of the GI tract may be pumped back out and/or metabolized before they reach the blood. This will limit the overall availability of these drugs. As if this were not enough of a challenge, both Pgp and/or CYP3A4 can be both upregulated and/or down-regulated by other drugs, herbal products, and food [45–48]. Pharmacists can be invaluable resources for other clinicians and patients in integrating the multitude of potential interactions and advising them about how to maximize drug effectiveness.

Table 4 The Rule of 5: Drugs That Have More Than 2 Strikes Are Likely to Be Poorly Absorbed

1. There are more than 5 H-bond donors (expressed as sum of OHs and NHs)
2. There are more than 10 H-bond acceptors (expressed as the sum of Ns and Os)
3. Molecular weight is > 500
4. log P is > 5
5. Compounds that are substrates for biological transporters are exceptions to the rule

Source: Ref. 44.

V. MOVING THE DRUG AWAY FROM THE SITE OF ABSORPTION

The fourth basic process that may influence drug absorption is moving the drug away from the site of absorption. Drugs that have crossed the GI membrane are primarily removed as a function of blood flow [50]. It can be seen in Eq. (6) that if there were no blood flow, the concentration of drug in the blood, C_b, would quickly approach C_g, and net transfer of drug across the GIT would cease. Thus, a decreased blood flow might decrease the rate or removal of passively absorbed drugs [51,52]. Decreased flow could possibly also interfere with active transport systems owing to the reduction of the supply of oxygen to the tissues. Winne and Ochsenfahrt [52] and Winne [53] have developed models and derived equations for GI absorption considering blood flow and countercurrent exchanges, respectively. For the following theoretical discussion, a simplified equation is presented as a modification of Eq. (6) [50]:

$$\left(\frac{dQ_b}{dt}\right)_{g\to b} = \frac{C_g - C_b}{1/P_m A_m + 1/\alpha\ \mathrm{BF}} \tag{7}$$

where

$P_m = D_m R_{m/aq}/X_m$
α = fraction of blood flowing through the capillaries near the GI membrane
BF = GI blood flow

The other terms are defined in Chapter 2. The denominator of Eq. (7) can be interpreted as the resistance of the region between the GI lumen and the blood pool. Winne and Remischovsky [54] divided this resistance into two parts (first and second terms of the denominator): (a) the resistance to transport of the region between the gastrointestinal lumen and the capillary blood (mainly resistance of the membrane), and (b) the resistance to drainage by blood. Figure 23 indicates the influence of blood flow on the rate of intestinal absorption for eight substances from the jejunum of the rat. The absorption of highly permeable materials [those with a small first term in the denominator of Eq. (7)], such as very lipid-soluble or pore-diffusable substances, should be blood flow–limited. Conversely, the absorption rate of drugs characterized by low membrane permeability [those with a large first term in the denominator of Eq. (7)] may be independent of blood flow. From Fig. 23 it may be seen that the absorption of freely permeable tritiated water is very sensitive to blood flow, but that ribitol, a sugar that penetrates the GI

Copyright © 2002 Marcel Dekker, Inc.

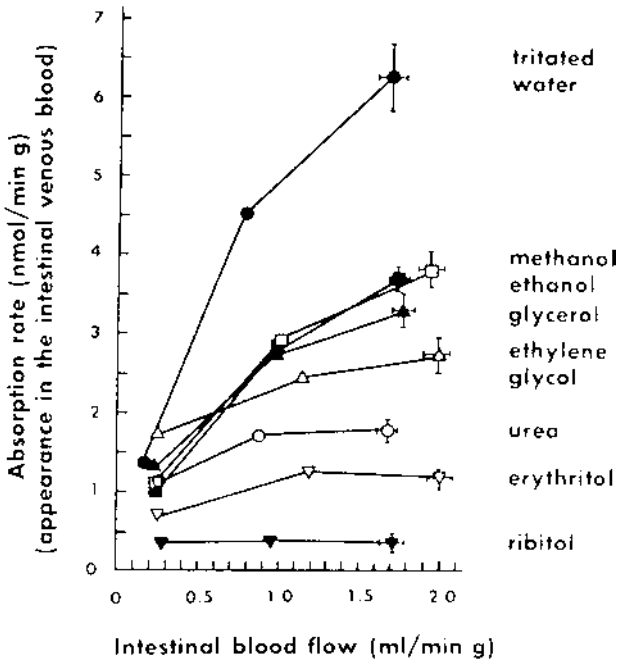


Fig. 23 Dependence of intestinal absorption on blood flow as reported by Winne and Remischovsky. All data are corrected to a concentration of 50 nmol/mL in the solution perfusing jejunal loops of rat intestine. Bracketed points indicate the 50% confidence intervals. (From Ref. 54.).

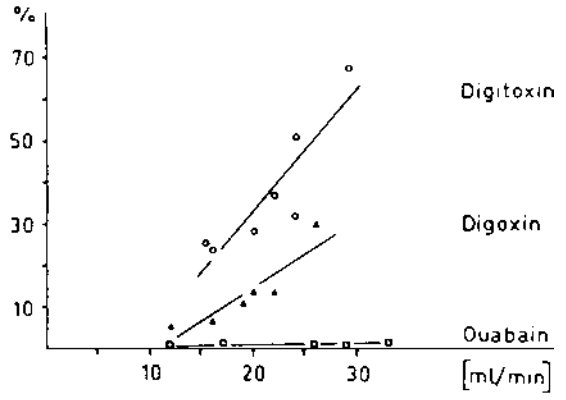


Fig. 24 Total amount of radioactivity absorbed during 1 hr, plotted versus portal blood flow. Ordinate: absorption as a percentage of the amount infused into the duodenum. Abscissa: mean portal blood flow in milliliters per minute. Each point represents one guinea pig. (From Ref. 56).

membrane with great difficulty, is essentially unaffected by changes in the intestinal blood flow in the range studied. As would be expected from Eq. (7), the absorption rate of intermediate substances, such as urea, appear to be flow-limited at low blood flow rates, but then become insensitive to blood flow at higher rates. Winne and coworkers [51–54] have reported a blood flow dependence for several relatively small drug molecules. Crouthamel et al. [55] also noted a decrease in the absorption rate of sulfaethidole and haloperidol as a function of decreased mesenteric blood flow rates, using an in situ canine intestinal preparation. Haas and coworkers [56], using a guinea pig model, found a strong correlation between spontaneously varying portal blood flow and the amount of digitoxin and digoxin absorbed following intraduodenal infusion of the drug. As can be seen in Fig. 24, digitoxin, the most lipophilic of the three cardiac glycosides, showed the most pronounced effects as a function of the blood flow. Digoxin also showed increasing absorption rates as a function of blood flow. However, ouabain, the most hydrophilic of the three, showed no dependence on blood flow. These results are consistent with the predictions of Eq. (7) for the effects of blood flow on the absorption rate of drugs.

These animal studies should indicate to the pharmacist that blood flow can, under certain circumstances, be an important patient variable that may affect the absorption of drugs. Patients in heart failure would generally be expected to have a decreased cardiac output and, therefore, a decreased splanchnic blood flow. This could lead to a decreased rate of absorption for drugs when the blood flow rates in Eq. (7) become rate-limiting. In addition, redistribution of cardiac output during cardiac failure may lead to splanchnic vasoconstriction in patients [57]. Other disease states and physical activity can also decrease blood flow to the GIT [2–4]. Thus, the pharmacist must be aware of the possible effect of blood flow rate, especially alterations in the rate, on the availability of drugs.

VI. PHYSICOCHEMICAL VERSUS BIOCHEMICAL FACTORS INFLUENCING DRUG ABSORPTION AND DRUG AVAILABILITY

As described in Chapter 2, the extent of oral drug availability (F_{oral}) may be described as the product of absorption availability (F_{abs}), the extent of availability through the gastrointestinal membranes (F_G), and the availability through the liver (F_H):

Copyright © 2002 Marcel Dekker, Inc.

$$F_{\text{oral}} = F_{\text{abs}} F_G F_H \quad (8)$$

First-pass loss in the gut will be due to metabolic processes within the intestinal membrane (due primarily to CYP3A and phase 2 enzymes present in the intestine such as glucuronosyltransferases, sulfotransferases, and glutathione S-transferases). First-pass hepatic loss will be due to liver metabolism as well as biliary excretion. Although transporters present in the intestinal epithelium can effect F_G, their consequences are probably reflected as a result of changes in intestinal metabolism. Absorption transporters should decrease intestinal metabolism by speeding the drug through the intestinal membrane and allowing less access of drug to intestinal enzymes, while efflux transporters would increase intestinal metabolism since the drug molecule may be repeatedly absorbed by passive processes increasing access to intestinal enzymes [49].

The physicochemical processes described in detail in this chapter (disintegration, dissolution, saturation solubility, and diffusion) will primarily affect the F_{abs} term in Eq. (8). However, the physicochemical processes, together with gastric emptying and gastrointestinal motility, can also affect the rate of absorption, as described earlier in this chapter. For a drug susceptible to GI metabolism, F_G could be expected to increase if absorption rate could increase to the point that intestinal metabolsm could be saturated, leading to further increases in F_G. In a similar manner a fast absorption rate could also possibly lead to saturation of hepatic enzymes during first pass resulting in an increase in F_H.

VII. SUMMARY

Drug availability following oral dosing may be thought of as the result of four basic steps: (a) getting the drug to its absorption site, (b) getting the drug into solution, (c) moving the dissolved drug through the membranes of the GIT, and (d) moving the drug away from the site of absorption into the general circulation. Although steps a, c, and d were discussed briefly, these topics are also found in Chapter 2 and 5. Step b, the combination of factors influencing the dissolution rate of a drug from its dosage form, served as the major topic of this chapter. Dissolution rate was discussed in terms of the parameters found in the Nernst-Brunner equation. Although this equation [Eq. (1)] is derived in terms of a specific model for dissolution, which, in fact, may not accurately describe the physical process, we believe that the treatment gives pharmacists a point of reference from which they can predict and interpret availability date that are rate-limited by the dissolution step. Two terms in the Nernst-Brunner equation are of major importance and are susceptible to manipulation by the pharmaceutical manufacturer in preparing the dosage form, by the patient in the manner in which he or she takes the drug and stores it between drug dosing, and by the biological system, specifically the GIT of the patient, through interactions with the dosage form and the drug. These two variables are the surface area of the drug particles and the saturation solubility of the particular chemical form of the drug. We believe that pharmacists must be aware of these variables and how they can change in each of the three situations listed previously before they can make rational judgments about which of many pharmaceutical alternatives should be dispensed under a specific set of conditions. Knowledge of these variables and their potential for change also allows pharmacists to make a rational guess about possible drug-related factors that may be responsible for inefficacious drug treatment. Under these conditions, pharmacists may be able to recommend an alternative formulation that will prove to be efficacious.

REFERENCES

1. D. E. Cadwallader, *Biopharmaceutics and Drug Interactions*, 3rd Ed., Raven Press, New York, 1983.
2. W. H. Bachrach, Physiology and pathologic physiology of the stomach, Ciba Clin. Symp., 11, 1–28 (1959).
3. H. W. Davenport, Gastric digestion and emptying: Absorption, in *Physiology of the Digestive Tract*, 3rd Ed., Year Book Medical Publishers, Chicago, 1971, pp. 163–171.
4. L. Z. Benet and B. Hoener, Pathological limitations in the application of rate control systems, in *Proceeding of the 2nd International Conference of Drug Absorption Rate Control in Drug Therapy*, (L. Prescott, ed.), Edinburgh, 1983, pp. 155–165.
5. G. D. Kozloski, J. M. De Vito, J. C. Kisicki, and J. B. Johnson, The effect of food on the absorption of methotrexate sodium tablets in healthy volunteers, Arthritis Rheum., 35, 761–764 (1992).
6. P. G. Welling, H. Huang, P. F. Hewitt, and L. L. Lyons, Bioavailability of erythromycin stearate: Influence of food and fluid volume, J. Pharm. Sci., 67, 764–766 (1978).
7. T. R. Bates, J. A. Sequeria, and A. V. Tembo, Effect of food on nitrofurantoin absorption, Clin. Pharmacol. Ther., 16, 63–68 (1974).
8. J. W. Fara, Physiological limitations: Gastric emptying and transit of dosage forms, in *Proceedings of the 2nd International Conference on Drug Absorption, Rate Control in Drug Therapy* (L. Prescott, ed.), Edinburgh, 1983, pp. 144–150.

Copyright © 2002 Marcel Dekker, Inc.

9. C. P. Dooley, C. Di Lorenzo, and J. E. Valenzuela, Variability of migrating motor complex in humans, Dig. Dis. Sci., 37, 723–728 (1992).
10. V. Manninen, J. Melin, A. Apajalahti, and M. Karesoja, Altered absorption of digoxin in patients given propantheline and metoclopramide, Lancet, 1, 398–400 (1973).
11. C. G. Regardh, P. Lundborg, and B. A. Persson, The effect of antacid, metoclopramide and propantheline on the bioavailability of metoprolol and atenolol. Biopharm. Drug Dispos., 2, 79–87 (1981).
12. J. A. Antonioli, J. L. Schelling, E. Steininger, and G. A. Borel, Effect of gastrectomy and of an anticholinergic drug on the gastrointestinal absorption of a sulfonamide in man, Int. J. Clin. Pharmacol., 5, 212–215 (1971).
13. W. H. Barr and S. Riegelman, Intestinal drug absorption and metabolism. I. Comparison of methods and models to study physiological factors in vitro and in vivo intestinal absorption, J. Pharm. Sci., 59, 154–163 (1970).
14. R. R. Scheline, Metabolism of foreign compounds by gastrointestinal microorganisms, Pharmacol, Rev., 25, 451–523 (1973).
15. L. Z. Benet, Biopharmaceutics as a basis for the design of drug products, in *Drug Design*, Vol. 4 (E. J. Ariens, ed.), Academic Press, New York, 1973, pp. 1–35.
16. W. Nernst and E. Brunner, Z. Phys. Chem., 47, 52–102 (1904).
17. P. Finholt, Influence of formulation on dissolution rate, in *Dissolution Technology* (L. J. Leeson and J. T. Carstensen, eds.), Academy of Pharmaceutical Sciences, American Pharmaceutical Association, Washington, DC, 1974, pp. 106–146.
18. L. F. Prescott, R. F. Steel, and W. R. Ferrier, The effects of particle size on the absorption of phenacetin in man, Pharmacol. Ther., 11, 496–504 (1970).
19. A. J. Glazko, A. W. Kinkel, W. C. Alegnani, and E. L. Holmes, An evaluation of the absorption characteristics of different chloramphenicol preparations in normal human subject, Clin. Pharmacol. Ther., 9, 472–483 (1968).
20. K. Arnold, N. Gerber, and G. Levy, Absorption and dissolution studies on sodium diphenlyhydantoin capsules, Can. J. Pharm. Sci., 5, 89–92 (1970).
21. G. Levy and R. H. Gumtow, Effect of certain tablet formulation factors on dissolution rate of the active ingradient. III. Tablet lubricants, J. Pharm. Sci., 52,1139–1141 (1963).
22. G. Levy, J. M. Antkowiak, J. A. Procknal, and D. C. White, Effect of certain tablet formulation factors on dissolution rate of the active ingredients. II. Granule size, starch concentration, and compression pressure, J. Pharm. Sci., 52, 1047–1051 (1963).
23. J. C. Samyn and W. Y. Jung, In vitro dissolution from several experimental capsule formulations, J. Pharm. Sci., 59, 169–175 (1970).
24. S. H. Wan, P. J. Pentikainen, and D. L. Azarnoff, Bioavailability of aminosalicylic acid and its various salts in humans. III. Absorption from tablets, J. Pharm. Sci., 63, 708–711 (1974).
25. E. Nelson, Influence of dissolution rate and surface on tetracycline absorption, J. Am. Pharm. Assoc. Sci. Ed., 48, 96–103 (1959).
26. E. Nelson, E. L. Knoechel, W. E. Hamlin, and J. G. Wagner, Influence of the absorption rate of tolbutamide on the rate of decline of blood sugar levels in normal humans, J. Pharm. Sci., 51, 509–514 (1962).
27. C. C. Lee, R. O. Froman, R. C. Anderson, and K. C. Chen, Gastric and intestinal absorption of potassium penicillin V and the free acid, Antibiot. Chemother., 8, 354–360 (1958).
28. G. Levy and B. A. Sanai, Comparison of the gastrointestinal absorption of aluminum acetylsalicylic and acetylsalicylic acid in man, J. Pharm. Sci., 51, 58–62 (1962).
29. R. A. O'Reilly, E. Nelson, and G. Levy, Physicochemical and physiologic factors affecting the absorption of warfarin in man, J. Pharm. Sci., 55, 435–437 (1966).
30. W. I. Higuchi and W. E. Hamlin, Release of drug from a self-coating surface: Benzphetamine pamoate pellet, J. Pharm. Sci., 52, 575–579 (1963).
31. E. Nelson, Comparative dissolution rates of weak acids and their sodium salts, J. Am. Pharm. Assoc. Sci. Ed., 47, 297–299 (1958).
32. T. R. Bates, J. M. Young, C. M. Wu, and H. A. Rosenberg, pH-dependent dissolution rate of nitrofurantoin from commercial suspensions, tablets and capsules, J. Pharm. Sci., 63, 643–645 (1974).
33. G. Levy, J. R. Leonards, and J. A. Procknal, Development of in vitro dissolution tests which correlate quantitatively with dissolution rate-limited drug absorption in man, J. Pharm. Sci., 54, 1719–1722 (1966).
34. K. A. Javaid and D. E. Cadwallader, Dissolution of aspirin from tablets containing various buffering agents, J. Pharm. Sci., 61, 1370–1373 (1972).
35. J. W. Poole, G. Owen, J. Silverio, J. N. Freyhot, and S. B. Roseman, Physicochemical factors influencing the absorption of the anhydrous and trihydrate forms of ampicillin, Curr. Ther. Res., 10, 292–303 (1968).
36. J. Hableblian and W. McCone, Pharmaceutical application of polymorphism, J. Pharm. Sci., 58, 911–929 (1969).
37. A. J. Aguiar, J. Krc. A. W. Kinkel, and J. C. Samyn, Effect of polymorphism on the absorption of chloramphenicol palmitate, J. Pharm. Sci., 56, 847–853 (1967).
38. D. L. Sorby, Effect of adsorbents on drug absorption. I. Modification of promazine absorption by activated attapulgite and activated charcoal, J. Pharm. Sci., 54, 677–683 (1965).
39. K. C. Kwan and D. J. Allen, Determination of the degree of crystallinity in solid-solid equilibria, J. Pharm. Sci., 58, 1190–1193 (1969).

Copyright © 2002 Marcel Dekker, Inc.

40. E. N. Hiestand, W. I. Higuchi, and N. F. H. Ho, Theories of dispersion techniques, in *Theory and Practice of Industrial Pharmacy*, 2nd Ed. (L. Lachman, H. A. Lieberman, and J. L. Kanig, eds.), Lea & Febiger, Philadelphia, 1976, p. 159.
41. A. H. Goldberg, M. Gibaldi, and J. L. Kanig, Increasing dissolution rates and gastrointestinal absorption of drugs via solid solutions and eutectic mixtures. III. Experimental evaluation of griseo-fulvin-succinic acid solid solution, J. Pharm. Sci., 55, 487–492 (1966).
42. A. Goldberg, Methods of increasing dissolution rates, in *Dissolution Technology* (I. J. Leeson and J. T. Carstensen, eds.), Academy of Pharmaceutical Sciences, American Pharmaceutical Association, Washington, DC, 1974, pp. 147–162.
43. L. Z. Benet, Theories of dissolution: Multi-particulate systems, in *Dissolution Technology* (I. J. Leeson and J. T. Carstensen, eds.), Academy of Pharmaceutical Sciences, American Pharmaceutical Association, Washington, DC, 1974, pp. 29–57.
44. C. A. Lipinski, F. Lombardo, B. W. Dominy, and P. J. Feeney, Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings, Adv. Drug Del. Rev., 23, 3–25 (1997).
45. V. J. Wacher, L. Salphati, and L. Z. Benet, Active secretion and enterocytic drug metabolism barriers to drug absorption, Adv. Drug Del. Rev., 20, 99–122 (1996).
46. W. E. Evans and M. V. Relling, Pharmacogenomics: translating functional genomics into rational therapeutics, Science, 286, 487–491 (1999).
47. A. Johne, J. Brockmoller, S. Bauer, A. Maurer, M. Langheinrich, and I. Root, Pharmacokinetic interaction of digoxin with an herbal extract from St. John's wort (*Hypericum perforatum*). Clin. Pharmacol. Ther., 66, 338–345 (1999).
48. S. C. Piscitelli, A. H. Burstein, D. Chaitt, R. M. Alfaro, and J. Falloon, Indinavir concentrations and St. John's wort, Lancet, 355, 547–548 (2000).
49. V. J. Wacher, J. A. Silverman, Y. Zhang, and L. Z. Benet, Role of P-glycoprotein and cytochrome P450 3A in limiting oral absorption of peptides and peptidomimetics, J. Pharm. Sci., 11, 1322–1330 (1998).
50. L. Z. Benet, A. Greither, and W. Meister, Gastrointestinal absorption of drugs in patients with cardiac failure, in *The Effect of Disease States on Drug Pharmacokinetics* (L. Z. Benet, ed.), Academy of Pharmaceutical Association, Washington, D.C., 1976, pp. 33–50.
51. L. Ther and D. Winne, Drug absorption, Annu. Rev. Pharmacol., 11, 57–70 (1971).
52. D. Winne and H. Ochsenfahrt, Die formale Kinetic der Resorption unter Berücksichtigung der Darmdurchblutung, J. Theor. Biol., 14, 293–315 (1967).
53. D. Winne, The influence of villous countercurrent exchange on intestinal absorption, J. Theor. Biol., 53, 145–176 (1976).
54. D. Winne and J. Remischovsky, Intestinal blood flow and absorption of nondissociable substances, J. Pharm. Pharmacol., 22, 640–641 (1970).
55. W. G. Crouthamel, L. Diamond, L. W. Dittert, and J. T. Doluisio, Drug absorption. VII. Influence of mesenteric blood flow on intestinal drug absorption in dogs, J. Pharm, Sci., 64, 661–671 (1975).
56. A. Haas, H. Lullman, and T. Peters, Absorption rates of some cardiac glycosides and portal blood flow, Eur. J. Pharmacol., 19, 366–370 (1972).
57. J. Ferrer, S. E. Bradley, H. O. Wheeler, Y. Enson, R. Presig, and R. M. Harvey, The effect of digoxin on the splanchnic circulation in ventricular failure, Circulation, 32, 524–537 (1965).

Copyright © 2002 Marcel Dekker, Inc.

Chapter 5

The Effect of Route of Administration and Distribution on Drug Action

Svein Øie and Leslie Z. Benet

University of California, San Francisco, California

I. THE DOSE-EFFICACY SCHEME

When a health practitioner administers (or "inputs") a dose of drug to a patient, usually the ultimate goal is solely directed to the usefulness of the drug under abnormal conditions. That is, the drug must be efficacious and must be delivered to its site of action in an individual experiencing a particular physiological anomaly or pathological state. Pharmaceutical scientists, on the other hand, concent\ate their attention to solving problems inherent in drug delivery to deliver the optimal dose to the site(s) of action.

The general pathway a drug takes from residence in a dosage form until its clinical utility is depicted in Fig. 1. Ideally, the drug should be placed directly at the site of action, as illustrated by the stippled arrow in Fig. 1, to maximize the effect and minimize side effects relating to unwanted responses at sites other than the target tissue. However, delivery directly to the site of action is, more often than not, impractical or not possible. Instead, we have to settle for the most convenient routes of delivery. This is illustrated by the solid arrows in Fig. 1. That is, the drug is placed directly in the vascular systems or in close proximity to some biological membrane through which the drug can traverse to reach body fluids or the vascular system. The delivery system is generally designed to release the drug in a manner that is conducive to this passage through the membrane. Previous chapters have discussed how drug-delivery systems may be optimized in terms of dissolution in the fluids surrounding the membrane to allow the desired rate of passage through the membrane. Subsequent chapters will deal with specific drug-delivery systems and their optimization. Once the drug has passed through the membrane and into the blood stream adjacent to the site of absorption, a general distribution of the drug will take place throughout the biological system. As pointed out in Chapter 3, the degree of dilution (referred to as the apparent volume of distribution) will dictate the initial concentration of drug in the general circulation, as sampled from a peripheral vein. Usually, the dose of the drug administered to the patient was chosen to give sufficiently high blood levels so that an adequate quantity of the drug would reach the site of action. The rate of input needed to achieve adequate levels of the drug at the site of action is not only influenced by the distribution and general elimination in the body, but may also be modified by the loss processes that are unique to a specific route of administration. This chapter will deal primarily with the distribution and loss processes that result uniquely from the physiological parameters inherent in the use of a particular route of administration.

Unfortunately, no drug is yet so specific that it interacts with only the target site in the target tissue and will not give rise to hyperclinical activity. Too much drug at the wrong place or too high a concentration at the right place may result in unwanted or toxic effects. Thus the practitioner must determine the usefulness of any dose of a drug from a particular drug-delivery

Copyright © 2002 Marcel Dekker, Inc.

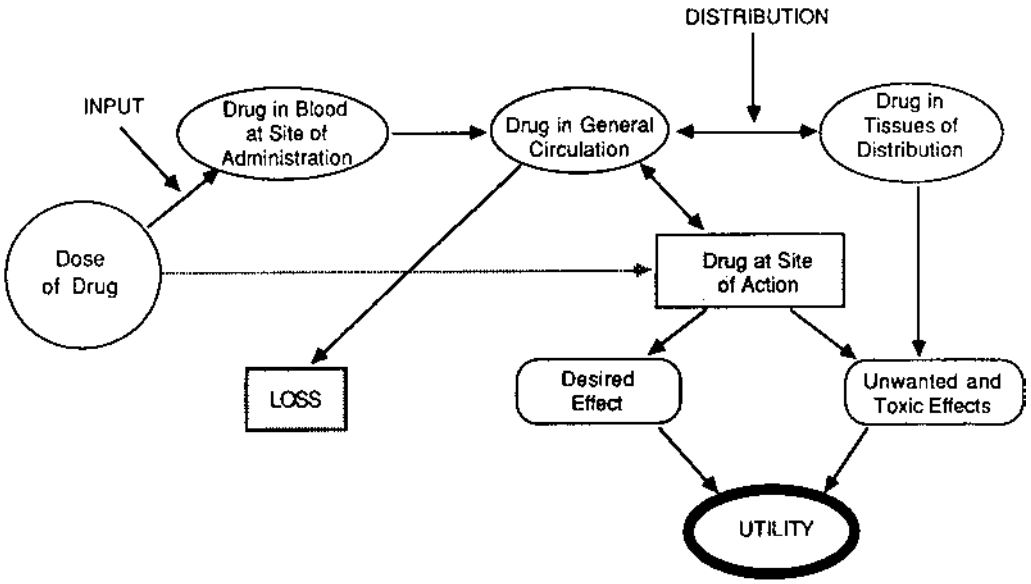


Fig. 1 A schematic representation of the dose-efficacy relationship for a drug.

system by balancing the efficacy achieved from the clinical effect against the toxic reactions observed.

Most drug-delivery systems achieve the required drug levels at the site of action as a result of attaining adequate blood levels in the general circulation (see Fig. 1, solid arrows). This process is followed because of the ease with which present drug-delivery systems can "input" drugs into the general circulation and the inherent difficulties in delivering the drug selectively to a relatively inaccessible site (e.g., pituitary gland). In addition, for many compounds the exact site of action is still unknown. However, when the site of drug action is sufficiently defined, Fig. 1 illustrates the advantage of delivering the drug directly to the site of action. By direct administration to the active site, a lower dose could be used to achieve the clinical effect because the drug no longer is diluted or eliminated en route. As a result, drug concentrations at unwanted sites of action could be kept to a minimum; in addition, clinically effective levels at the site of action might be attained much more rapidly, since the process of distribution throughout the entire body could be avoided. One should not forget that, in addition to the obvious clinical advantage of direct administration, there is also an economical one. By delivering the drug to the site of action, the amount of drug needed is much smaller than by more traditional delivery methods. This is particularly important for many of the newer recombinant compounds that can be very expensive. Much work is currently being carried out in an attempt to achieve selectiveness such as that described in Fig. 1.

II. PHYSIOLOGICAL CONSIDERATIONS FOR THE VARIOUS ROUTES AND PATHWAYS OF DRUG INPUT

A. Drug Input at or Close to the Site of Action

Figure 2 illustrates a number of sites where drug-delivery systems have historically been used to input drug directly to its site of action [1,2]. Various classic dosage forms were developed to take advantage of these input sites: eye, ear, and nose drops; inhalation, oral, topical, and vaginal aerosols; topical solutions, creams, and ointments; and rectal solutions, enemas, and suppositoires. Each of the sites for local drug administration requires specific formulation to allow the drug to remain at the site of application for a sufficient length of time to allow the drug to penetrate through the particular membrane(s) so that it can reach the actual site of action adjacent to the site of application. For example, some opthalmic preparations may be given to elicit a superficial anti-infective effect, such as treatment of an inflammation of the conjuctiva. Thus, only topical effects are desired, and there is no need for the drug to penetrate into the eyeball. Formulation of such products would be quite different than formulation of a drug-delivery system for which the drug must be absorbed into the in-

Copyright © 2002 Marcel Dekker, Inc.

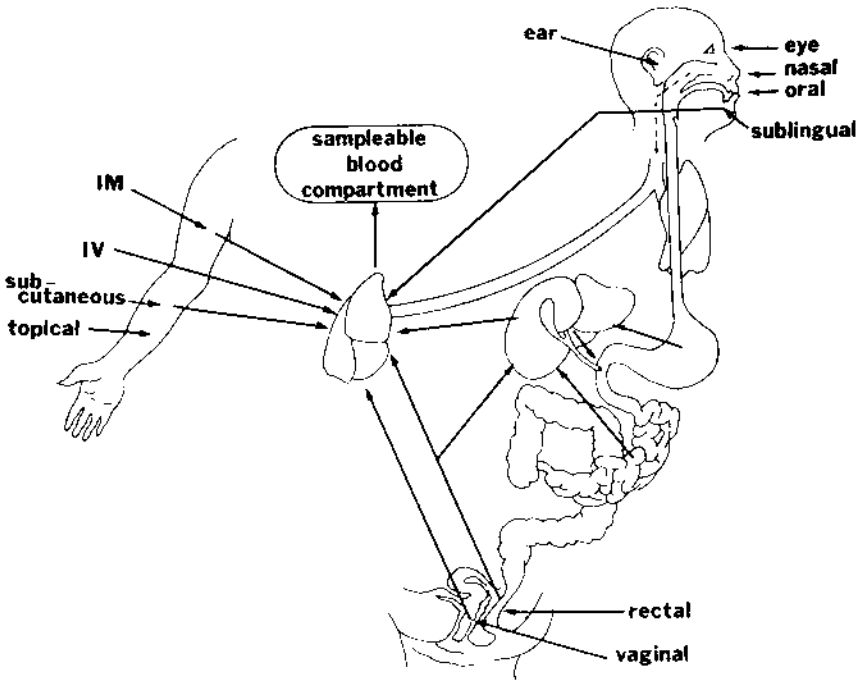


Fig. 2 Various routes and pathways by which a drug may be "input" into the body. The position of one lung is distorted to emphasize that the lungs are in an excellent position for cleansing the blood. The diagram is especially useful in explaining the first-pass effect following oral dosing, for which drug absorbed from the small intestine or stomach must first pass through the liver and, therefore, is subject to metabolism or biliary excretion before reaching the sampleable blood. (From Ref. 2.)

terior of the eye to produce a response, such as miotics, mydriatics, anti-inflammatory drugs that act in the anterior segment of the eye, and, occasionally, drugs for treatment of infections. A detailed description of the factors involved in the development of such opthalmic preparations, as well as the particular physiological characteristics of the eye, will be presented in Chapter 13. Similar types of design problems arise for many of the other sites that are traditionally treated by direct local application. One of the most difficult problems facing the formulator is that the behavior of the diseased tissue may be different from that for healthy individuals, and it may also change over the course of treatment. For example, diseased skin is often more permeable than healthy skin; therefore, the drug may disappear faster from the site of administration that desired, and the effect will be less than expected. Should the formulation be designed to accommodate this phenomenon, one must be mindful of the fact that as the pathological condition improves, the absorption may also change.

Although the classic dosage forms mentioned earlier can be used to put drug directly into the site of action, many of them have a degree of "messiness" that prevents good patient acceptance and adherence. Not only is there an initial psychological barrier that must be overcome, but the general public has an aversion to taking drugs by routes other than oral. There is, in addition, a general dislike for sticky creams, drippy drops, greasy ointments, and the like. Over the last two decades much work has been directed toward developing more acceptable delivery systems than the traditional ones. Emphasis has been placed on long-acting drug-delivery systems that may be more convenient, since they would only require self-administration once a week or possibly at even longer intervals.

A large number of new devices have been developed, and new ones are constantly being investigated. Plastic disks for placement in the eye (similar to a contact lens) that slowly release drug into the humoral fluid; drug-impregnated plastic rings or loops that when placed in the uterus will release controlled amounts of contraceptive agents; bioadhesive tablets or disks that can be placed buccally, nasally, or vaginally for local release; hydrogels for slow release in the eye are examples of such new delivery systems that input drug directly to the site of action.

Copyright © 2002 Marcel Dekker, Inc.

In a more ambitious move, many groups have also embarked on site-specific delivery to less accessible sites than those given in Fig. 2. Although numerous experimental systems have been designed, few have reached the clinical stage. The simplest and most direct method when a specific target organ can be located is cannulation (direct access port). A catheter is placed in an appropriate artery or vein. If a vein is used, the catheter has to reach the organ, otherwise the drug will be flowing away from the target tissue, be diluted with blood from the rest of the body, and not be different than a systematic intravenous administration. Catheters can also be placed in the peritoneum, the bladder, and the cerebrospinal fluid. A drug can now be administered directly into the desired tissues at a rate that can be well controlled. Although catheter delivery is a direct method, it is limited in that it is essentially restricted to inpatient use. Use of implants in the desired tissue, or a drug carrier (e.g., liposomes, nano-particles, and such) that will either home in on the desired tissue by specific receptors or release their content at the desired site by an external stimulus (e.g., magnetic fields, light, current), are drug systems currently being explored for target-specific delivery.

Although the method of direct delivery is a very attractive one, it also has its regulatory problems. Benet [1] has noted that assessing the bioavailability of this system can create difficulties because the manufacturer may not be able to devise a control procedure that can measure drug concentration at the site of action. For example, the extent and rate of availability of an orally administered drug can easily be assessed by measuring blood levels, whereas for a drug input into a site of action, significant blood levels would indicate distribution away from that site. Frequently, significant blood levels of a drug that is administered at a site of action (such as a topical preparation, an eye drop, a nasal insufflation, or an antiobiotic that acts on intestinal flora) indicate either a poor drug-delivery system or substantial overdosing. For this class of drug-delivery system, clinical efficacy necessarily has to serve as the best measurement of drug availability and dosage form efficacy.

B. Drug Input into the Systemic Circulation

The overwhelming majority of existing drugs are, however, given by general routes; that is, by routes that do not deliver the drug directly to the site of action. These modes of drug input rely on a passive delivery of drug through distribution by the vascular system. The most commonly accepted method is oral administration. As will be discussed later, oral administration is not ideal, as one needs to be concerned about whether the drug can be destroyed in the stomach, in the gastrointestinal fluid, in its passage through the gut wall, through the liver, or simply not be absorbed in time before it is expelled from the gastrointestinal tract. Several alternative routes of delivery are being used or are being developed to diminish these potential losses. The advantages and problems inherent in the individual routes of administration will now be discussed.

Parenteral Administration: Intravascular

Of the routes of input depicted in Fig. 2, intravenous (IV) administration yields one of the fastest and most complete drug availabilities. However, intra-arterial injections might be employed when an even faster and more complete input of drug to a particular organ is desired. By administering the drug through an artery, the total drug delivered will enter the organ or tissue to which the artery flows. Intravenously administered drug will first be diluted in the venous system as the venous blood is pooled in the superior and inferior vena cava. It then enters the heart and is subsequently pumped to the lung before it can enter the arterial system and reach the target organ(s). In addition, the fraction of the drug reaching a desired site is dependent on the fraction of the arterial blood flow reaching that site. Additional drug can reach the target tissue only by being recirculated from the other organs. In comparison with intra-arterial administration, IV administration reaches the target slower, and initially at a lower concentration. Although intra-arterial injections appear superior, they are infrequently used because they are considered much more dangerous than IV administration. Intra-arterial administration has been associated with patient discomfort, bleeding, and thrombosis.

In addition to the dilution factor resulting from mixing with larger volumes of blood after intravenous administration, one also needs to consider the possibility of temporary or permanent loss of drug during its passage through the lung. The position of the lungs in Fig. 2 has been distorted to emphasize the point that the lungs are in an excellent strategic position for cleansing the blood, since all of the blood passes through the lungs several times a minute. Apart from their respiratory function and the removal of carbon dioxide from the pulmonary circulation, the lungs serve other important cleansing mechanisms, such as filtering emboli and circulating leukocytes, as well as excretion of volatile substances. The lungs also have metabolic capacity [3] and may serve as a metabolic site for certain drugs [4] or as an excretory route for

Copyright © 2002 Marcel Dekker, Inc.

compounds with a high vapor pressure. The lungs can also act as good temporary storage site for a number of drugs, especially basic compounds, by partitioning of the drug into lipid tissues, as well as serving a filtering function for particulate matter that may be given by IV injection. Accumulation of lipophilic compounds and filtering of any compounds in solid form can be viewed as a temporary clearing or dilution of the drug, as it will eventually leach back into the vascular system. Thus the lung serves as a dampening or clearing device, which is not present following intra-arterial injection. Drugs given by the IV route may, therefore, not necessarily be completely available to the site of action, since a certain fraction of the drug could be eliminated by the lung before entering into the general circulation [5]. This might be called a "lung first-pass effect."

The foregoing concepts may be visualized by referring to Fig. 3. In this figure one can readily see the difference between intra-arterial and intravenous administration of drugs. Let us assume that compartment n is the target tissue. Administration into any vein (i.e., into any of the efferent arrows on the left-hand side of the figure) would lead the drug to the heart and, from there, to the lung. Drug that enters the lung can leave by only one of two routes, as illustrated in Fig. 4: by the blood that leaves the lung, or by being eliminated. The result is that there is a competition between the two routes for the drug, and the greater the ability of the lung to eliminate the drug in comparison with the pulmonary blood flow, the more drug will be extracted. If we assume that the pulmonary blood flow in Q_P and that the intrinsic elimination clearance of the organ is CL_{Int}, and no plasma protein binding occurs ($Cu = C_{\text{out}}$), then the extraction ratio can be expressed as

$$E = \frac{C_{\text{out}} CL_{\text{Int}}}{(C_{\text{out}} CL_{\text{Int}}) + (C_{\text{out}}\, Q_P)} = \frac{CL_{\text{Int}}}{CL_{\text{Int}} + Q_P} \qquad (1)$$

In perfusion models, as depicted in Fig. 3, it is assumed that distribution into and out of the organ is perfusion rate–limited such that drug in the organ is in equilibrium with drug concentration in the emergent blood [6]. The intrinsic clearance of an organ is different from the value we normally think of as the clearance of the organ. The *clearance of the organ* is defined as the rate of loss in relation to the incoming concentration, whereas the *intrinsic clearance* is defined as the rate of loss in relation to the organ concentration (or exiting concentration). In addition, it is also clear that, of the

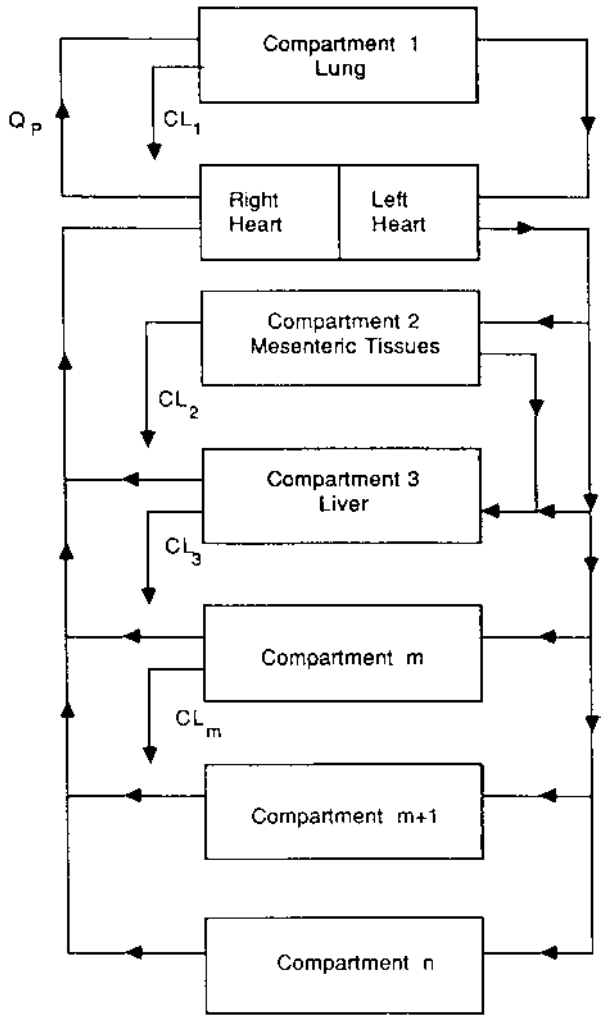


Fig. 3 The body depicted as a physiological perfusion model. Compartment m must be considered as a summation of the individual tissues that metabolize the drug and compartment m + 1 through n as noneliminating tissues.

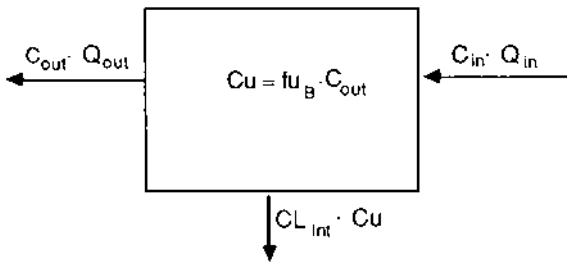


Fig. 4 Flow model of an eliminating organ. The drug enters the body by the organ blood flow, CL_{in} and is immediately mixed in the organ. The drug leaves the organ by either being eliminated ($CL_{int}Cu$) or by the exiting blood flow.

Copyright © 2002 Marcel Dekker, Inc.

drug that escapes elimination in the lung, only a small fraction goes to compartment n, while the rest is distributed to other organs. Drugs that enter these organs will be exposed to elimination in these organs and must necessarily recirculate through the heart and lungs before they again have the opportunity to reach compartment n.

Parenteral Administration: Depot

The other parenteral routes depicted in Fig. 2, intramuscular (IM) and subcutaneous (SC) injections, may also be considered in terms of Fig. 3. Drug absorbed from the IM and SC sites into the venous blood will return to the heart and pass through the lungs before being distributed to the rest of the body. However, there will be an initial lag between the time when the drug is injected and when it enters the circulation. Thus, the kinetics for drugs administered by these parenteral routes would be expected to show a decreased *rate* of availability and may also show a decreased *extent* of availability in comparison with intravenous administration, if loss processes take place at the site of injection. For example, we could consider that drug is now injected directly into compartment m in Fig. 3 and that this compartment is the muscle. The rate at which the drug leaves the muscle will depend primarily on blood flow in relation to the size (apparent volume of distribution) of the organ.

Evans and coworkers [7] measured resting human muscle blood flow through the gluteus maximus, vastus lateralis, and deltoid muscles. Deltoid muscle blood flow was significantly greater than gluteus muscle blood flow, with vastus being intermediate. Because the two sites most commonly used for IM injections are the deltoid and the gluteus muscles, we might expect to see differences in rates of drug absorption following injection to these sites. Lidocaine is one drug that has been investigated for its effect in response to the site of injection [8,9]. Deltoid injection gave higher peak levels than lateral thigh injection, which, in turn, gave higher levels than gluteal injection. Schwartz et al. [9] demonstrated that therapeutic plasma levels for a particular lidocaine dose were reached only when the deltoid injection site was used. Evans et al. [7] concluded, "This demonstrates that the site of injection can influence the plasma level achieved and that the deltoid muscle should be used to achieve therapeutic blood levels as rapidly as possible." Likewise, if a sustained or a prolonged release is desired, this would more readily be achieved by injection into a lower blood flow muscle, such as the gluteus.

Loss processes may also account for a decrease in the extent of availability following an IM injection. This can be visualized by assuming that the dose is injected into compartment m, which, as depicted in Fig. 3, is capable of eliminating the drug. As shown in Fig. 4, the drug can leave the tissue only by one of two routes: either by the blood leaving the organ, or by being eliminated by metabolism in the muscle. In addition, the drug that leaves the site of administration will also be subject to the additional distribution and elimination in the lung, similar to intravenous administration. In other words, drug given by intramuscular administration may not only be further delayed in its distribution to the target organ, but may also show a decreased extent of distribution to the organ, in comparison with the intravenous dose. For example, degradation can take place in the muscle, as shown by Doluisio et al. [10] for ampicillin. These workers found that only 77–78% of an IM does of ampicillin sodium solution was absorbed, as compared with the IV solution. The most likely explanation is that the drug may have been decomposed chemically or enzymatically at the injection site. In addition, temporary losses may also occur. For example, intramuscular doses of phenytoin result in a marked decreased rate and extent of absorption in comparison with IV or oral doses. Wilensky and Lowden [11] demonstrated that this could be due to precipitation of the drug as crystals in the muscle. Although these crystals eventually dissolve, the drug is essentially lost during a normal dosing interval.

Oral Administration

First-Pass Effect. Metabolism in the Gastrointestinal Fluids and Membranes. When a dosage form is administered by the oral route, drug particles come in contact with varying pH solutions, different enzymes, mucus, gut flora, and bile all of which may contribute to decreasing the extent of availability of degradation, binding, or sequestering mechanisms. These factors, as well as the possibility of drug metabolism in the intestinal membrane itself, have been well covered in Chapter 2 and will not be discussed here

Hepatic Metabolism: Linear Systems. As depicted in both Figs. 2 and 3, drug that is absorbed from the gastrointestinal tract must pass through the liver before reaching the sampleable circulation and the rest of the body. Thus, if a drug is metabolized in the liver or excreted into the bile, some of the active drug absorbed from the gastrointestinal tract will be inactivated by

Copyright © 2002 Marcel Dekker, Inc.

hepatic processes before the drug can reach the general circulation and be distributed to its site(s) of action. An exception would be if the liver itself were the target organ, as we then would have to contend with only losses in the gastrointestinal tract and in the gut wall before reaching the site of action.

For many drugs, the fraction of the dose eliminated on the first pass through the liver is substantial. The fraction eliminated is often referred to as the hepatic extraction ratio, designated herein as E_H. Many drugs are known or suspected to have a high hepatic extraction ratio. A short list of some of the better-known compounds is given in Table 1. The hepatic first-pass phenomenon is not restricted to any particular pharmacological or chemical group of drug substances, and the foregoing list includes acids, bases, and neutral compounds.

The available fraction (F) of an oral dose appearing in the sampleable blood circulation will, therefore, be governed by not only the extent of drug absorbed from the gastrointestinal tract, as discused in Chapter 4, but also by the fraction metabolized in the gut membranes (E_G) and the fraction metabolized or excreted into the bile following passage through the liver (E_H), where E_G and E_H are the extraction ratios for the gut and liver, respectively. If it is assumed that a drug is completely absorbed from the gastrointestinal tract and not degraded during passage through the gut membranes, this would be equivalent to an injection of the drug into the hepatic portal vein. Under these conditions the unmetabolized fraction (F_H) of an oral dose appearing in the sampleable blood circulation would equal $1 - E_H$. If the drug elimination in the liver follows first-order kinetics, Rowland [12] has shown that F_H, the fraction of the oral dose that is available following the liver first pass, may be related to liver blood flow (Q_H) and the hepatic clearance for an IV does of the drug ($CL_{H,\text{iv}}$).

$$F_H = 1 - E_H = 1 - \frac{CL_{H,\text{iv}}}{Q_H} \tag{2}$$

Table 1 Drugs Suspected to Have High Hepatic Extraction Ratios

Acetylsalicylic acid	Alprenolol	Aldosterone
Cocaine	Desipramine	Doxorubicin
Fluorouracil	Isoproteranol	Imipramine
Lidocaine	Lorcainide	Morphine
Nitroglycerin	Prazepam	Propranolol

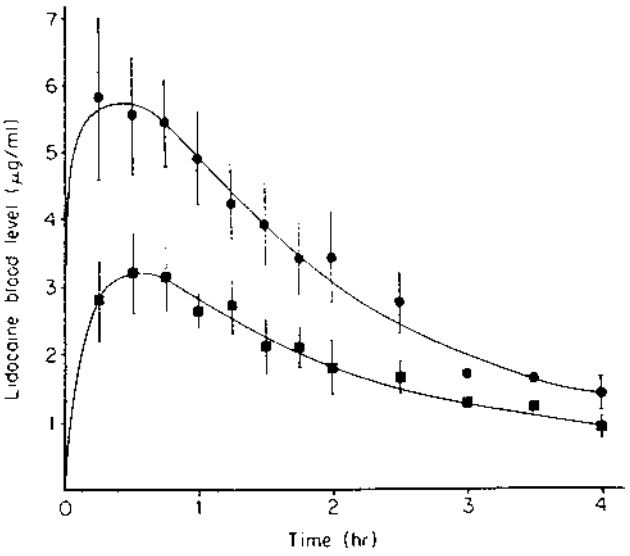


Fig. 5 Mean blood levels of lidocaine hydrochloride in five beagles after exponential IV infusion through a peripheral vein (●) and in the portal vein (■). Vertical bars represent standard errors of the mean. (From Ref. 13.)

This hepatic first-pass effect, as discussed by Boyes et al. [13], is well illustrated in Fig. 5. Identical doses of lidocaine hydrochloride were injected into dogs (beagles) by an exponential infusion process. The upper curve describes the mean levels found when the drug was infused into a peripheral vein, whereas the lower curve describes the lidocaine levels following hepatic portal vein infusion (thereby eliminating any effects caused by gastrointestinal degradation either in the fluids of the tract or in the intestinal membrane). The area under the curve (AUC) measurements for the two curves show that the extent of bioavailability following portal vein infusion is only 60% of that found following infusion into a peripheral vein (i.e., $F_H = 0.6$). This "oral" availability could be predicted for a drug such as lidocaine for which drug elimination occurs predominantly in the liver when the hepatic clearance—calculated by dividing the dose by the area under the peripheral vein infusion curve—and the hepatic blood flow are substituted into Eq. (2). An even greater first-pass effect is found for oral administration of lidocaine to humans, $F_H = 0.25$–0.48 [14].

Thus, if the hepatic clearance for a drug is largely relative to the hepatic blood flow, the extent of availability for this drug will be low when it is given by a route that yields first-pass effects. The decrease in availability is a function of only the anatomical site

Copyright © 2002 Marcel Dekker, Inc.

form which absorption takes place, and no amount of dosage form redesign can improve the availability. Of course, therapeutic blood levels can be reached by this route of administration if larger doses are given, but the health practitioner and pharmacist must be aware that levels of the drug metabolite may increase significantly over that seen following IV administration.

Lidocaine is not normally administered by the oral route, but many of the drugs listed earlier are routinely given orally. For these drugs, analysis of Eq. (2) leads to the conclusion that small variations in plasma or blood clearance of a drug throughout a population may yield significant differences in availability when the drug is given by a route subject to significant first-pass effects. For example, data from a study of Shand et al. [15] for oral and IV administration of propranolol in five men are shown in Table 2. Although the clearance following an IV dose of the drug (column 6) varies by only 67% from the smallest to the largest, the oral availability (column 5) varies by 275%. As would be expected from Eq. (2), the oral availability decreases as the IV clearance increases.

Hepatic Metabolism: Nonlinear Systems. The discussion of the first-pass effect has thus far assumed linear first-order kinetics. Under this condition, the hepatic extraction will be independent of the rate of drug availability. That is, no matter when a drug molecule is absorbed form the gastrointestinal tract and at whatever dose administered, the hepatic extraction and the extent of availability [see Eq. (2)] for that drug will remain constant. This would not be true for a drug for which saturation of the hepatic enzymes is a possibility. Under such a condition, the extraction ratio would vary depending on the concentration of drug in the hepatic portal vein. If the concentration were high, the hepatic enzymes would become partially saturated, and a large amount of drug would pass through the liver without being metabolized (i.e., E_H would be low). Saturable metabolism during the first passage through the liver is not an unusual occurrence even for a drug that, after intravenous administration, fails to show saturable metabolism. The main reason is that drug absorbed from the gastrointestinal tract is diluted to only a small degree before it enters the liver (in the portal blood only), and the concentrations can be relatively high. On the other hand, intravenously administered drug is generally quickly distributed to individual tissues, and the concentration reaching the liver during the elimination process is sufficiently low that no saturation of the metabolism is achieved. However, if the concentration in the hepatic portal vein were low, either because of administration of a low dose or because of very slow absorption of the drug, then the enzymes would not be saturated, and most of the drug could be metabolized on passage through the liver (i.e., E_H would approach 1). In a similar manner, metabolism in the gastrointestinal membranes could also be saturated or not, depending on the dose and the rate of absorption (i.e., E_G would vary from 0 to 1).

Salicylamide appears to be a drug for which the first-pass extraction may be dose-dependent. When a 300-mg oral dose of salicylamide is administered as a solution, the area under the plasma concentration time curve is less than 1% of the area seen following a similar IV dose [16]. Even following a 1-g oral dose, most of the drug is found in the systemic circulation as the inactive glucuronide and sulfate conjugates. However, when a 2-g dose was given to a particular subject (Fig. 6), the area under the plasma concentration time curve for unmetabolized salicylamide increased dramatically (>200 times increase over that seen for a 1-g dose). This very large increase is probably related to saturation of enzymes in the gastrointestinal mucosa as well as in the liver. Saturation of the hepatic enzymes not only increases the extent of oral availability, but a

Table 2 Peak Plasma Levels and Areas Under Plasma Concentration Time Curves Following Oral and Intravenous Administration to Men

Subject	Propranolol, 80 mg fasting orally		Propranolol, 10 mg IV		
	Peak (ng/ml)	Area (ng/ml · hr)	Area (ng/ml · hr)	$\frac{AUC_{Oral}}{AUC_{IV}} \cdot \frac{10}{80} \cdot 100$	Clearance, IV (ml/min)
OF	212	1400	292	60	570
DS	100	480	220	30	756
GY	94	510	200	32	833
JC	45	290	183	20	909
JF	36	220	175	16	950

Source: After Ref. 15.

Copyright © 2002 Marcel Dekker, Inc.

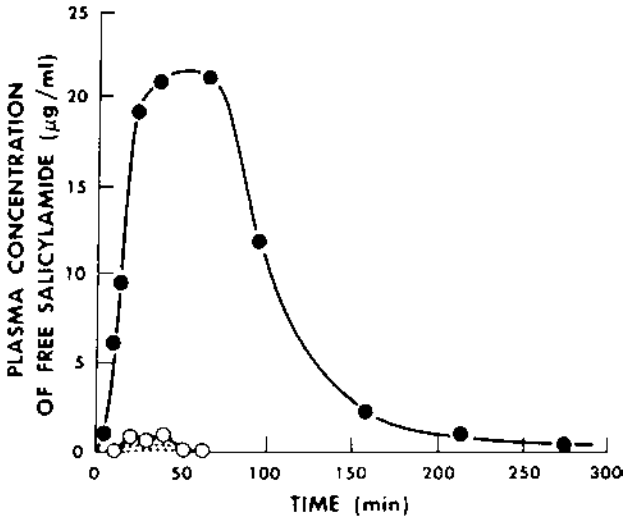


Fig. 6 Comparison of plasma concentrations of intact salicylamide when given as 1-g (○) and a 2-g (●) dose in solution to the same subject. Dotted lines show the plasma concentrations of a 0.5-g dose and 0.3-g dose. (From Ref. 16.)

sufficient amount of unchanged drug is allowed to enter the systemic circulation, thereby decreasing the rate of drug metabolism during the postabsorptive phase as well. This phenomenon is due to enzyme saturation. Thus, the extraction ratio [see Eq. (1)] and, thereby, the drug clearance, change continuously during both the absorption and elimination processes. Under these conditions AUC measurements can no longer be used to determine the extent of drug availability (as described in Chapter 3 for linear systems.)

The effects of stomach emptying on drug availability are discussed in Chapters 2 and 4, relative to processes within the gastrointestinal system. In addition, Benet [1] has interpreted data for *p*-aminobenzoic acid (PABA) as reflecting an example for which the rate of drug absorption modified by changes in stomach emptying causes changes in the extent of drug availability owing to saturation of first-pass metabolism. Table 3 represents a situation in which the drug PABA is completely absorbed from the gastrointestinal tract, yet there is a decrease in the extent of drug available to the systemic circulation as a function of food [17].

The data for the oral solution, single-dose studies suggest that all of the drug is being absorbed, since the total amount of PABA found in the urine (both as unchanged drug and acetyl metabolite) equals the dose administered. It appears that the decreasing fraction of the acetyl metabolite found in the urine with increasing doses would indicate that a saturable metabolic process was operable. For the second series of studies listed in Table 3, the oral solution dose remains constant at 1 g, while increasing amounts of sweet cream are added to the solution. With increasing amounts of fat, there is a concomitant increase in the percentage of dose excreted as the acetyl metabolite. This may be explained by assuming that the fat decreases the rate of stomach emptying, causing the drug to be emptied from the stomach more slowly and to be absorbed over a longer time period. If plasma concentration were maintained at lower levels by slowing absorption, the metabolic site would not reach the same degree of saturation and a greater fraction of the metabolite should appear in the urine. Even if gastric absorption of PABA did occur, owing to retention of the drug in the stomach, the absorption rate should be significantly lower than that seen for the oral solutions without fat, thereby maintaining plasma concentrations of PABA at lower levels. The prolonged administration of the smaller oral and IV doses in the third part of Table 3 yields high levels of metabolite in the urine, which is consistent with the saturable enzyme hypothesis (i.e., when plasma concentrations are maintained at a low level, extensive metabolism will occur).

Biliary Excretion. The effects of significant hepatic extraction as a result of biliary secretion, with or without metabolism, would be expected to follow the same principles just outlined for hepatic metabolism. In fact, a whole class of compounds that serve as biliary contrast agents for radiological examination depend on significant first-pass biliary secretion to be effective.

Several studies in rats have shown that certain acidic and basic compounds can be actively secreted into the bile. Thus, one might expect to see saturation of the biliary excretion process, although data in humans describing this phenomenon have not, as yet, been reported for orally dosed drugs.

Oral Dosing Without a First-Pass Effect. If a drug is not metabolized in the gut wall or the liver and if the drug is not subject to biliary excretion, there will be no first-pass effect following oral dosing. In addition, with drugs for which the hepatic clearance is significantly less than hepatic blood flow [see Eq. (2)], the hepatic extraction will be negligible (i.e., $F_H \approx 1$). Most orally administered drugs used today fall into this latter category. Examination of Fig. 2 indicates that another form of oral administration, sublingual dosing, may avoid first-pass metabolism as well as the degradation

Copyright © 2002 Marcel Dekker, Inc.

Table 3 Extent of Urinary Excretion of *p*-Aminobenzoic Acid (PABA) and Its Acetyl Metabolite as a Function of Route of Administration and Ingestion of Fat

Route	Dose Na-PABA to 61-kg man (g)	Total PABA in urine as percentage of dose in 24 hr	Acetyl-PABA in urine as percentage of total PABA excreted in 24 hr
Oral solution			
Single dose	1	103	51
	2	103	47
	4	102	36
	8	102	30
Fat added to oral dose			
60 g sweet cream	1	95	76
90 g sweet cream	1	104	83
120 g sweet cream	1	99	90
Prolonged administration			
10 oral doses given every 30 min	0.365	95	97
Intravenous infusion, 270 min	0.365	90	95
Intravenous bolus	1	102	51

Source: Ref. 17.

process that may occur in the gastrointestinal fluids. This route of administration has been used predominantly in dosing organic nitrates in patients experiencing angina.

Rectal Administration

As can be seen in Fig. 2, the first-pass effect can be partially avoided by rectal administration. The capillaries in the lower and middle sections of the rectum drain into the inferior vena cava, thus bypassing the liver. However, suppositories tend to move upward in the rectum into a region where veins (such as the superior hemorrhoidal vein) drain predominantly into the portal circulation [18]. In addition, there are extensive anastomoses between the middle and superior hemorrhoidal veins. Thus, Schwarz [19] has suggested that only about 50% of a rectal dose can be assumed to bypass the liver and its first-pass extraction. Again referring to Fig. 2, we can see that absorption of drugs from the vagina would bypass the first-pass effect, although almost no drugs where systemic levels are desired have been formulated using this route of administration.

Other Routes of Administration

It is clear from the foregoing discussion that none of the methods traditionally used for systemic administration are ideal. Intravenous administration requires hospital staff and, therefore, is rarely used in an outpatient setting. Intramuscular and subcutaneous doses may be metabolized at the site of administration, may be significantly delayed, and may run the risk of infections and meet with public reluctance for self-administration, although subcutaneous administration has been met with great success in insulin-dependent diabetics. Oral administration must run through a series of elimination sites before the drug can enter the general circulation. Rectal administration can also be exposed to hepatic first-pass and must also contend with a public reluctance in its usage. Because of these factors, several other routes for systemic administration of drugs have been explored over the years. Significant acceleration in these areas has occurred as a response to the increased interest in the use of peptide and protein drugs made by recombinant techniques. Peptides and proteins are notoriously prone to degradation in the gastrointestinal tract, and the only modes of administration have been parenteral administrations. This limits the use by outpatients, and great efforts are being made to develop alternative routes of delivery.

Nasal Administration. The nasal mucosa is relative permeable to small molecular weight compounds. The most notorious example is cocaine. Cocaine that is snorted is both rapidly and extensively absorbed. Small peptides have also been successfully administered nasally, although the bioavailability is low. However, where the availability is not critical, nasal administration of peptides has been successful. The best example is

Copyright © 2002 Marcel Dekker, Inc.

calcitonin, which shows an activity after nasal administration similar to that seen after intravenous administration, although the plasma concentrations achieved after nasal delivery were lower [20]. This is a compound that has a relatively large therapeutic index for which a variable bioavailability is not critical.

During the absorption process through the nasal lining, the drug has to cross not only lipophilic barriers, but it must also pass through the nasal epithelia, which have a significant capability to metabolize drugs. Nasally administered drug can, in addition to this loss by metabolism, also be removed by mucous flow and ciliary movement and be swallowed. Nasal administration can, under circumstances during which significant amounts are swallowed, be thought of as being similar to administering the drug in part as an oral dose. To increase the absorption by the nasal epithelia, the use of bioadhesive drug-delivery forms and many exciting absorption enhancer techniques are being explored. Chemicals that disrupt the lipophilic membranes as well as opening the tight junction between the cells have resulted in a dramatic increase in the availability. The opening of the tight junctions is particularly interesting, as the compounds pass between the cells, thereby avoiding exposure to the metabolizing enzymes in the nasal epithelia. Whether these absorption enhancement techniques are the ways of the future as a means to reduce first-pass elimination, or if, in the end, they will be judged to invoke too much tissue damage for continuous use, still needs to be evaluated.

Transdermal Absorption. If transdermal delivery of drugs to the systemic circulation is to be successful, it must mimic subcutaneous injection in terms of yielding minimal first-pass skin elimination. The main barrier to absorption is the thick lipophilic keratin layer of the skin. Few nonlipophilic compounds penetrate the skin to a sufficient degree that this mode of delivery can be used for systematic absorption without modification. Highly lipophilic drugs, on the other hand, penetrate the skin with relative ease, although the absorption usually takes a long time. Several drugs are successfully given by dermal administration (e.g., nitroglycerin, scopolamine, nicotine, progesterone). To increase the absorption of hydrophilic compounds, including peptides and proteins, strategies similar to those described previously for nasal absorption have been studied (i.e., use of hydration and chemical enhancers). In addition, use of nonchemical enhancement methods (i.e., iontophoresis) is also being explored. This technique uses a low electric charge to force fluid and solutes to cross the skin. Although it was introduced more than 200 years ago [21], it is only recently that the method appears to have reached a practical stage [22]. These methods not only allow the drug to bypass the gastrointestinal tract and the liver, they also may be employed to achieve very attractive long-term sustained delivery.

Pulmonary Inhalation. Although aerosol preparations now serve primarily as a convenient drug-delivery system that can input drug directly to its site of action, new interest has recently been generated in this delivery system as a potential route for systemic administration of drug. The lung has a relatively large surface area and is relatively permeable to lipophilic compounds and, to some degree, even to protein [23]. Several barriers to absorption by the lungs do exist. The barrier to absorption is greatest in the upper bronchi and decreases in the alveoli. In the upper bronchi the mucus is relatively thick, the surface area small, and ciliar movement tends to move impacted particles up the bronchi and into the esophagus. The particle size of therapeutic inhalation aerosols determines the site of deposition in the lungs and, thus, the clinical effectiveness of a particular formulation. Particles that are too large will impact in the upper bronchi, and those too small will not readily impact on the wall of the lung and, therefore, will simply be exhaled. Particle sizes in the order of 0.3–1 μm are usually considered to be most effective. Sciarra [24] has suggested that almost all drugs given by IV injection can be reformulated into a suitable aerosol, provided that the drug is capable of being deposited in the respiratory tract and is nonirritating. However, the total availability of the pulmonary route will, to a large degree, depend on how much drug is deposited in the lung and, again, how much metabolized during first pass. It is expected that the first-pass bioavailability will be lower than that seen after intravenous administration because the drug has to pass the epithelial cell layer before it can reach the general circulation. From intravenous administration, only the amount of drug that is actually taken up by the lung tissue will be exposed to metabolism and exhalation.

III. DRUG DISTRIBUTION

Figure 1 indicates that distribution will take place as the drug reaches the general circulation. This will dilute the drug and influence the levels at the site of

Copyright © 2002 Marcel Dekker, Inc.

action. Thus, an understanding of drug distribution is critical in designing appropriate drug dosage regimens. This has led to the determination of "apparent" volumes of distribution (as discussed in Chapter 3), which can be used to related the amount of drug in the body (or in a hypothetical compartment) to a measured plasma or blood concentration. The volume of distribution is a function of four major factors: (a) the size of the organs into which the drug distributes; (b) the partition coefficient of drug between the organ and the circulating blood; (c) the blood flow to the distributing organs; and (d) the extent of protein binding of the drug both in the plasma and in various tissues.

A. Organ Size, Blood Flow, and Partition Coefficient

A particular organ in the body may act as a site of distribution or as a site of both distribution and elimination. The relative importance of the various organs as storage or elimination sites depends on how fast the drug gets to each organ and how much space or volume is available to hold the drug. Table 4 presents a compilation of the volumes and blood flows of the different regions of the human body for a standard man, as compiled as Dedrick and Bischoff [25] using the mean estimates of Mapleson [26].

The various regions of the body are listed in decreasing order relative to blood flow per unit volume of tissue (adrenals highest and bone cortex lowest). This value essentially describes how fast a drug can be delivered to a body region per unit volume of tissue and reflects the relative rates at which tissues may be expected to come to equilibrium with the blood. How much drug can be stored or distributed into a tissue will depend on the size of the tissue (volume) and the ability of the drug to concentrate in the tissue (i.e., the partition coefficient between the organ and blood, $K_{O/B}$). For

Table 4 Volumes and Blood Supplies of Different Body Regions for a Standard Man[a]

Tissue	Reference letters	Volume (liters)	Blood flow (ml/min)	Blood flow (ml/100 ml)	Volume of blood in equilibrium with tissue (ml)
Adrenals	A	0.02	100	500	62
Kidneys	B	0.3	1240	410	765
Thyroid	C	0.02	80	400	49
Gray matter	D	0.75	600	80	371
Heart	E	0.3	240	80	148
Other small glands and organs	F	0.16	80	50	50
Liver plus portal system	G	3.9	1580	41	979
White matter	H	0.75	160	21	100
Red marrow	I	1.4	120	9	74
Muscle	J	30	300/600/1500	1/2/5	185/370/925
Skin				1/2/5	18/37/92
Nutritive	K	3	30/60/150		
Shunt	L		1620/1290/300	54/43/10	
Nonfat subcutaneous	M	4.8	70	1.5	43
Fatty marrow	N	2.2	60	2.7	37
Fat	O	10.0	200	2.0	123
Bone cortex	P	6.4	≈ 0	≈ 0	≈ 0
Arterial blood	Q	1.4			
Venous blood	R	4.0			
Lung parenchymal tissue	S	0.6			
Air in lungs	T	2.5 + half Tidal volume			1400[b] 999/795/185[c]
Total		70.0[d]	6480		5400

[a]Standard man = 70-kg body weight, 1.83-m^2 surface area, 30–39 years old.
[b]Arterial blood.
[c]Skin-shunt venous blood.
[d]Excluding the air in the lung.

Copyright © 2002 Marcel Dekker, Inc.

example, the blood flow per unit volume of thyroid gland (see Table 4) is one of the highest in the body, whereas the gland itself is quite small. Thus, if partition of the drug between the thyroid and blood were approximately 1, we would expect to see that the drug in the tissue would rapidly come into equilibrium with that in the blood but that relatively little drug would be found in the thyroid. However, for certain drugs containing iodine moieties, $K_{O/B}$ is enormous, and a significant amount of drug will distribute into this small gland relatively rapidly. In addition, Table 4 lists the volume of blood contained within each tissue and believed to be in equilibrium with the tissue. Thus, the volume of the thyroid in Table 4 is considered to be 20 mL of tissue and 49 mL of blood. Note that total volume of all tissues in column 3 is 70 L, including the 5.4 L of blood volume. This blood volume is broken down into 1.4 L of arterial blood (listed in the last column as being in equilibrium with the air in the lungs) and 4 L of venous blood (in equilibrium with tissues A through O). Since different muscle masses throughout the body receive different blood flows (discussed in Sec. II.B, with reference to drug input following IM injection), Mapleson [26] lists only a range for this tissue as well as for the skin. Dedrick and Bischoff [25] suggest an average blood flow value of 3.25 mL/(100 mL tissue × min) for tissues J and K, corresponding to average total flows of 980 and 98 mL/min for muscle and skin, respectively. Note that total blood flow in column 4 corresponds to the cardiac output, 6.48 L/min.

When discussing drug distribution, if it often convenient to lump various tissue regions into general categories. For example, following an IV bolus injection, the heart, brain, liver, and kidneys achieve the highest and earliest drug concentrations, with equilibrium between these tissues and blood being rapidly achieved. Thus, Dedrick and Bischoff [25] have combined these tissues and other well-perfused regions (see A through H in Table 4) into a well-perfused compartment that they designate as viscera. Similarly, regions I, J, K, and M are lumped into a less well-perfused compartment, the lean tissues, whereas poorly perfused regions N and O are designated as the adipose compartment. Blood flows, volumes, and such, for these lumped compartments can be calculated by summing the appropriate terms in Table 4. By using a perfusion model containing these three lumped compartments and a blood compartment, Bischoff and Dedrick [27] were able to describe thiopental concentrations in various tissues as shown in Fig. 7.

Levels in the dog liver, representative of the visceral tissues, are already at a maximum by the time the first sample is taken, since very rapid equilibrium is achieved between these tissues and the blood. Drug uptake into the less well-perfused skeletal muscle, representative of the lean tissue, is slower—peaking at about 20 minutes, but still achieving apparent distribution equilibrium between 1 and 2 hours. Uptake into the poorly perfused adipose tissue in even slower. In fact, peak levels in this tissue have not even been reached by the time the last samples are taken.

Since the site of action for the barbiturates is the brain, we might expect the pharmacological action to correspond to the time course of thiopental concentrations in the viscera, which, in turn, would be reflected by blood levels, since a rapid equilibrium is attained between viscera and blood. Although the pharmacological action may terminate quickly (within an hour) owing to decreased blood levels, traces of the drug may be found in the urine for prolonged periods (days) owing to accumulation in the fatty tissues. Note

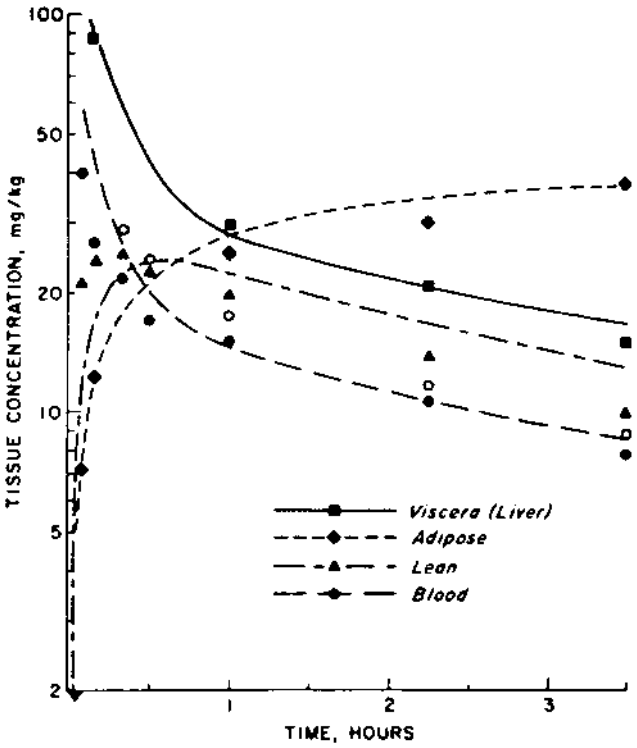


Fig. 7 Thiopental concentrations in various tissues following 25-mg/kg IV bolus doses. Solid symbols indicate data in dogs; the open circles are from data in humans. Lines correspond to predicted values in various tissues using a perfusion model containing compartments corresponding to the blood, viscera, lean, and adipose tissues. (From Ref. 26.)

Copyright © 2002 Marcel Dekker, Inc.

that the partition coefficient for the drug between tissues and blood is greater than 1 for all three tissue groups (i.e., at distribution equilibrium tissue concentrations greater than blood concentrations) but that the $K_{O/B}$ for adipose tissue is very large. This can be deducted from the fact that drug continues to distribute into adipose tissue, even when the concentration in the fat is significantly greater than the blood concentration. At 3½ hours, most of the lipid-soluble thiopental left in the body is in the fat, and at later times this percentage may even increase before distribution equilibrium is reached. At these later times the removal rate of the barbiturate from the body will be controlled by the slow movement of drug out of the fatty tissue as a result of the high partition into the fat and the low blood flow to this region.

B. Protein Binding

One of the major determinants of drug distribution is the extent of protein binding. Many drugs bind to plasma proteins, mainly to albumin and α_1-acid glycoprotein (orosomucoid), but sometimes to lipoproteins, various globulins, and specific binding proteins as well. Up to the present time, most studies of drug protein binding have examined the interaction between drugs in the plasma and plasma albumin and plasma α_1-acid glycoprotein. Total plasma albumin for a 70-kg man is about 120 g. Total interstitial albumin is approximately 156 g [28]. Thus almost 60% of total albumin in the body is found outside the plasma. The total extracellular α_1-acid glycoprotein is approximately 3 g, of which approximately 45% is located extravascularly [29]. α_1-Acid glycoprotein has also been found intracellularly and exists as a membrane-bound form [30], but its total cellular amount is unknown. The extracellular amount of α_1-acid glycoprotein can increase significantly in many diseases (inflammation, infections, cancers) as well as during trauma. The increase is variable and usually averages a twofold change, although increases up to fivefold have been reported. Binding to tissue components is also assumed to be important for the overall distribution, although direct evidence is difficult to obtain. However, albumin and α_1-acid glycoprotein represent only approximately 1% of the "dry tissue" in the body, of which a large fraction is proteins. One might, therefore, expect that the binding of drugs to tissue would affect drug distribution significantly and frequently much more than the binding of drugs to plasma proteins. The binding of drugs to plasma proteins has been studied extensively, primarily because the experiments can be easily carried out. Tissue-binding studies do not have this advantage, and thus knowledge of the qualitative and quantitative aspects of the binding of drugs to tissue components is poorly understood. The partition coefficients between body organs and blood, $K_{O/B}$, discussed in the previous section will be considerably influenced by binding. It is probably for this reason that attempts to correlate drug distribution, drug action, and membrane transport with oil/water partition coefficients have succeeded only infrequently, necessitating the use of different organic solvents in different correlations to obtain at least an approximate rank order correlation.

Effect on Distribution

The extent of distribution to an organ is usually expressed as the apparent volume of distribution of the organ, V_i. Because the blood concentration is used as a reference, the apparent volume of an organ, V_i, can usually be expressed in the following way:

$$V_i = V_T K_{O/B} = V_T\left(\frac{fu}{fu_T}\right) \tag{3}$$

where V_T is the physical volume of the tissue, fu is the unbound fraction of drug in the blood and fu_T is the unbound fraction of drug in the tissue. This relationship indicates that the partition coefficient $K_{O/B}$ essentially reflects the strength of binding to the tissues in comparison with the binding to blood proteins, which will be particularly true for hydrophilic compounds. Decreased binding to proteins in a tissue will increase the value of fu_T, as the fraction of drug in the tissue that is not bound is now increased and the apparent volume of the tissue decreases. Similarly, if the binding to blood proteins is decreased, the relative affinity for the tissue increases and the apparent volume of distribution is increased. Often, volumes of distribution and protein binding are defined in terms of plasma measurements, rather than in blood, as used here. Such an approach is correct as long as one defines explicitly the reference fluid of measurement. We prefer to use blood measurements throughout, because, as described in the next section, the relationship of clearance to organ elimination capacity must be defined in terms of the total flow to each organ.

Effect on Clearance

The *clearance* of an organ can be viewed as the volume of blood completely cleared of drug per unit time. Given this definition, the blood clearance of an organ

Copyright © 2002 Marcel Dekker, Inc.

will be equal to the blood flow through the organ multiplied by the extraction ratio (equal to the fraction of the blood completely cleared of drug). From Eq. (1) we, therefore, obtain

$$CL = QE = Q\left(\frac{CL_{\text{Int}}}{Q + CL_{\text{Int}}}\right) = \frac{Q\,CL_{\text{Int}}}{Q + CL_{\text{Int}}} \tag{4}$$

Equation (1) was derived assuming no protein binding, whereas, in fact, CL_{Int} should be defined in terms of the maximum ability of the organ to remove *unbound* drug because only unbound drug can bind to enzymes and excretory molecules. By including the unbound fraction of drug in Eq. (4) one obtains

$$CL_{\text{organ}} = \frac{Q\,fu\,CL_{\text{Int}}}{Q + fu \cdot CL_{\text{Int}}} \tag{5}$$

where *fu* is the fraction of drug unbound in the blood. [Note that all the discussions in this chapter have related clearances and flows to blood concentrations (i.e., drug concentrations in the actual transporting fluid). Blood concentrations must be used when any physiological interpretation is placed on the clearance process. However, it is easier to carry out studies measuring plasma concentrations and plasma protein binding. These measured parameters may be converted to blood values if one knows the blood/plasma concentration ratio.]

When the intrinisic clearance of an organ is very high compared with blood flow, such that $fu\ CL_{\text{Int}} \gg Q$, the extraction ratio approaches 1 and $CL_{\text{organ}} \to Q$. Thus, according to Eq. (5), if a drug is cleared exclusively in the liver with an extraction ratio approaching unity, organ clearance will be very sensitive to changes in liver blood flow and essentially independent of blood binding [31]. Likewise, for drugs eliminated in the kidney by active-transport processes, elimination appears to be independent of the extent of blood binding. Under these conditions, kidney blood flow becomes the rate-limiting step.

However, for many drugs, blood flow is significantly greater than the intrinsic organ clearance ($Q \gg fu\,CL_{\text{Int}}$) and then Eq. (5) reduces to

$$CL_{\text{organ}} = fu\,CL_{\text{Int}} \tag{6}$$

indicating that organ clearance is dependent on binding in blood and is proportional to the unbound fraction of drug in blood. This phenomenon is valid for both hepatic and renal elimination. For example, the glomerulus in the kidneys will filter approximately 120 mL of plasma per minute. High molecular weight compounds (i.e., most of the plasma proteins) will not be filtered. Likewise, drugs bound to these proteins will be retained. In other words, one could view the filtration process as being filtration of plasma water and of the small molecular weight compounds dissolved therein. Plasma water contains only unbound drug. Increases in binding will lower the fraction of the drug in plasma in unbound form and, therefore, decrease the filtration clearance. For hepatic elimination, data for several compounds have been published over the years that show a correlation between the clearance and the unbound fraction in plasma [32]. We expect similar relationships between the organ clearance and the unbound fraction to be also valid for other organs.

Effect on Availability

Should changes in protein binding have a significant effect on first-pass availability? Combining Eqs. (2) and (5), we obtain:

$$F = 1 - E = 1 - \frac{CL}{Q} = 1 - \left[\frac{fu\,CL_{\text{Int}}}{Q + fu \cdot CL_{\text{Int}}}\right] = \frac{Q}{Q + fu \cdot CL_{\text{Int}}} \tag{7}$$

This indicates that the first-pass availability is a function of organ flow, protein binding, and intrinisic clearance of the organ. When $fu\ CL_{\text{Int}} \gg Q$ (i.e., when we have relatively large extraction ratios), the first-pass bioavailability is equal to

$$F = \frac{Q}{fu\,CL_{\text{Int}}} \tag{8}$$

Under this circumstance, the first-pass bioavailability is inversely proportional to the unbound fraction, and changes in the binding are expected to have a significant effect. It is also clear that changes in both the blood flow and the intrinsic clearance of the first-pass organ may have a significant effect when the extraction ratio is high ($fu\ CL_{\text{Int}} \gg Q$). On the other hand, if $Q \gg fu\,CL_{\text{Int}}$, then Eq. (7) simply says that the first-pass bioavailability is approximately 1 (i.e., little or no drug is eliminated in a first pass), and changes in binding, blood flow, and intrinsic clearance are not expected to have any effect on *F*.

Combined Kinetic Effect of Binding Alterations

Changes in binding can occur in many situations. In some diseases, the concentrations of binding proteins may be altered, and accumulation of endogenous inhibitors of binding may occur. For example, in renal

Copyright © 2002 Marcel Dekker, Inc.

failure the albumin level may decrease, resulting in decreased binding. In renal failure, accumulation of waste products also occurs, and this, in turn, may further suppress the binding. In other diseases (e.g., inflammations), the level of α_1-acid glycoprotein increases, which may lead to increased drug binding. Concomitant administration of compounds that compete for the same binding sites may also decrease the binding and increase the unbound fraction. These changes will lead to changes in the apparent volume of distribution, in clearance, and in the first-pass availability, as discussed earlier, but will they lead to changes in the time required for the drug to reach the target organ and in the activity of the drug? The answer is not simple, as it depends on the actual values of clearance and volume of distribution, as well as on whether the target tissue is among the highly or poorly perfused tissues.

To begin answering this question we must realize that the effect of drugs in the body is related to the unbound and not the total concentration in the body [33]. Only non–protein-bound drug can interact with receptors and, therefore, elicit an effect. Therefore, it is important to look at how the *unbound* concentration changes. Let us assume that a situation of decreased binding exists in blood and that no changes in the tissue binding have occurred. If a dose is given, we expect no changes in the first-pass bioavailability if the extraction ratio in the first-pass tissue is low. If the extraction ratio is high, on the other hand, we expect a significant decrease in the first-pass bioavailability if the extraction [see Eq. (8)] of either an IV administration (if first-pass in the lung is significant) or an oral administration (first-pass effect both in the liver and lung). When it comes to other routes of administration, we need to be more cautious. For example, let us look at dermal administration. If the major first-pass elimination occurs before the drug reaches the blood (i.e., in its passage through the epithelial cells), we do not expect the blood-binding changes to be important. On the other hand, if the major first-pass elimination relates to cell downstream from where the drug enters the blood, changes in the binding will cause changes predictable from Eq. (8).

Will the drug be delayed on its way to the target tissue in the foregoing example? A decreased blood binding is equivalent to increasing the $K_{O/B}$ value. Distribution into various tissues, including the target tissue, therefore, is expected to be more extensive and swifter. However, because the drug usually has to pass the lungs before it can reach the target tissue, high $K_{O/B}$ values in the lung will significantly reduce the amount available to other tissues when the $K_{O/B}$ value increases for the lung. Consequently, a delay of distribution here can occur. On the other hand, if the $K_{O/B}$ value in the lung is low, the amount sequestered by the lung will be too small to delay the distribution. In this situation, no delay is expected in distribution of drug to the target tissue, if the target tissue is highly perfused. However, if the target tissue is a poorly perfused tissue, a delay may still occur. In this situation, a higher $K_{O/B}$ value will mean that the highly perfused tissues can take up more of the initial drug presented to them. In turn, the diffusion out of the hihgly perfused tissues will be slower, and may take longer before the redistribution to poorly perfused tissues is completed. Therefore, if the effect is in the poorly perfused tissues, the effect will lag.

On the continuous administration, the drug's effect is dependent on the average unbound steady-state concentration. The unbound concentration is dependent on the average rate of dosing, bioavailability, clearance, and degree of binding in plasma. Let us, for the moment, assume that the bioavailability and dosing rate remain constant. The change in the steady-state unbound concentration now depends on whether the extraction ratio is low or high (or whether $Q \gg fu\,CL_{\text{Int}}$ or $fu\,CL_{\text{Int}} \gg Q$ in addition to the alteration of the binding. For a low-extraction ratio compond, the average unbound concentration, Cu_{ss} is

$$Cu_{\text{ss}} = \frac{F(\text{rate of dosing})}{CL_{\text{Int}}} \tag{9}$$

For a high-extraction ratio, the value is

$$Cu_{\text{ss}} = \frac{F(\text{rate of dosing})}{Q/fu} \tag{10}$$

Under the assumptions of constant bioavailability and dosing rate, a high-extraction ratio compound is expected to have an increased unbound steady-state concentration when the plasma binding is decreased, and a low-extraction ratio compound is not affected by binding changes. Now, however, we also need to evaluate whether the first-pass availability is affected by a change in the binding and adjust our expectation accordingly. Under special circumstances for which the first-pass organ is the major metabolizing organ, some simplified concepts can be established. For a low-extraction ratio compound in this situation, neither the first-pass bioavailability nor the clearance relative to unbound drug will be affected, and the result is that there is no overall effect on the average unbound steady-state concentration and the activity of the drug. For a high-extraction ratio compound

Copyright © 2002 Marcel Dekker, Inc.

both the clearance realative to unbound drug [Eq. (5)] and the first-pass bioavailability [see Eq. (7)] will increase proportional to the increase in the unbound fraction in plasma, and the result is again that the average unbound concentration is not affected by a change in the plasma binding. If the first-pass organ is different from the major eliminating organ, the extraction ratio in these organs may be different, and their dependence on binding may differ (i.e., the unbound clearance and bioavailability may be affected to a different extent). If saturation of the first-pass takes place, there is also likely to be different effect of binding on the first-pass bioavailability and unbound clearance, even if the first-pass organ is the major eliminating organ.

Changes in tissue binding have effects only on the tissue distribution and not on clearance. Tissue-binding changes, therefore, are not expected to affect the first-pass bioavailability, but are expected to alter the distribution [See Eq. (3)] and $K_{O/B}$ value. If the value of $K_{O/B}$ is reduced, we expect a smallar sequestering of drug in its passage through the body, higher initial concentrations will reach the target tissue, and the effect will occur more swiftly and, initially, more potently. But because the apparent volume of distribution is smaller, the half-life will also be smaller:

$$t_{1/2} = \frac{0.693V}{CL} \tag{11}$$

and the concentration and the effect will fall off faster. Documentation for such changes is difficult to obtain, because we cannot measure the tissue binding directly, and we can make only inferences from overall changes in the kinetic parameters and changes in plasma-binding values [34].

A third possibility—that the binding in plasma and tissue changes to the same degree—is not expected to affect the apparent volume of distribution to the individual organs [see Eq. (3)], and the value of $K_{O/B}$ will not be changed. The clearance, on the other hand, is affected only by changes in binding in blood and will change as described in Sec. III.B, the section describing the effect of protein-binding changes on clearance. Assuming that the first-pass bioavailability is not affected, this will mean that sequestering of drug will not increase, and the total concentration reaching the target tissue should not be significantly affected. However, because the unbound fraction in plasma is increased (decreased), the unbound concentration will be increased (decreased), the unbound concentration initially reaching the target tissue will be higher (lower), and the initial pharmacological activity will be higher (lower). The duration of this change will depend on the half-life of the drug, which will be altered according to Eq. (11).

IV. SUMMARY

The extent and time course of drug action can be markedly affected by the route of drug administration into the patient as well as the pattern of drug distribution within the patient. Drug efficacy can be improved, and drug toxicity probably decreased, if the drug can be administered directly to its site of action. However, several factors prevent direct application of drugs to the site of action, including incomplete knowledge about the action site and also poor patient adherence owing to the inconvenience of using direct application formulations. Because of these factors, most drug products have been formulated as oral, solid dosage forms.

Drugs that are rapidly cleared by hepatic processes will show a decreased extent of availability following oral administration owing to metabolism of the drug on its first pass through the liver. The magnitude of this first-pass effect will depend on the blood flow to the liver and the intrinsic clearing ability of the liver (i.e., the ability of the organ to eliminate the drug independently of the rate at which drug is brought to the organ). This first-pass elimination by metabolic or biliary excretion processes can be excluded if the drug is absorbed from a sublingual site. The rectal administration of drugs eliminates approximately one-half the first-pass metabolism. Absorption through nasal, dermal, and other sites may also give rise to lower first-pass elimination and higher bioavailability than oral administration if the metabolic capacity of these sites for the drug in question is small.

Drug distribution in the patient will depend on the blood blow to various sites in the body as well as on the partition coefficient of the drug between the blood and distributive organs. Protein binding, both in the blood and in the tissues, will markedly affect this distribution. However, free drug concentrations are generally believed to be the effective determinant in drug therapy. Often a redistribution owing to changes in protein binding will have little effect on the therapeutic efficacy, since, although total drug distribution changes, the average unbound concentrations at steady state in blood remain essentially similar. An understanding of the effects of the route of administration as well as the distribution of the drug within the body is critical to the pharmacist in planning appropriate drug dosage regimens.

Copyright © 2002 Marcel Dekker, Inc.

REFERENCES

1. L. Z. Benet, Biopharmaceutics as a basis for the design of drug products, in *Drug Design*, Vol. 4, (E. Ariens, Ed.), Academic Press, New York, 1973, pp. 1–35.
2. L. Z. Benet, Input factors as determinants of drug activity: route, dose, dosage regimen, and the drug delivery system, in *Principles and Techniques of Human Research and Therapeutics*, Vol. 3, (F. G. McMahon, Ed.), Futura, New York, 1974, pp. 9–23.
3. T. E. Gram, The metabolism of xenobiotics by the mammalian lung, in *Extrahepatic Metabolism of Drugs and Other Foreign Compounds* (T. E. Gram, Ed.), S.P. Medical and Scientific Books, New York, 1980, pp. 159–209.
4. J. R. Vane, The role of the lungs in the metabolism of vasoactive substances, in *Pharmacology and Pharmacokinetics* (T. Teorell, R. L. Dedrick, and P. G. Condliffe, Eds.), Plenum Press, New York, 1974, pp. 195–207.
5. W. L. Chiou, Potential pitfalls in the conventional pharmacokinetic studies: Effects of the initial mixing of drug in blood and the pulmonary first-pass elimination, J. Pharmacokinet. Biopharm., 7, 527–536 (1979).
6. M. Rowland, L. Z. Benet, and G. G. Graham, Clearance concepts in pharmacokinetics, J. Pharmacokinet. Biopharm., 1, 123–136 (1973).
7. E. F. Evans, J. D. Proctor, M. J. Fraktin, J. Velandia, and A. J. Wasserman, Blood flow in muscle groups and drug absorption, Clin. Pharmacol. Ther., 17, 44–47 (1975).
8. L. S. Cohen, J. E. Rosenthal, D. W. Horner, Jr., J. M. Atkins, O. A. Matthews, and S. F. Sarnoff, Plasma levels of lidocaine after intramuscular administration, Am. J. Cardiol., 29, 520–523 (1972).
9. M. L. Schwartz, M. B. Meyer, B. G. Covino, R. M. Narange, W. Sethi, A. J. Schwartz, and P. Kemp, Antiarrhythmic effectiveness of intramuscular lidocaine: Influence of different injection sites, J. Clin. Pharmacol., 14, 77–83 (1974).
10. J. T. Doluisio, J. C. LaPiana, and L. W. Dittert, Pharmacokinetics of ampicillin trihydrate, sodium ampicillin, and sodium dicloxacillin following intramuscular injection, J. Pharm. Sci., 60, 715–719 (1971).
11. A. J. Wilensky and J. A. Lowden, Inadequate serum levels after intramuscular administration of diphenylhydantoin, Neurology, 23, 318–324 (1973).
12. M. Rowland, Influence of route of administration on drug availability, J. Pharm. Sci., 61, 70–74 (1972).
13. R. N. Boyes, J. H. Adams, and B. R. Duce, Oral absorption and disposition kinetics of lidocaine hydrochloride in dogs, J. Pharmacol. Exp. Ther. 174, 1–8 (1970).
14. M. Rowland, Effect of some physiologic factors on bioavailability of oral dosage forms, in *Dosage Form Design and Bioavailability* (J. Swarbrick, Ed.), Lea & Febiger, Philadelphia, 1973, pp. 181–222.
15. D. G. Shand, E. M. Nuckolls, and J. A. Oates, Plasma propranolol levels in adults with observations in four children, Clin. Pharmacol. Ther., 11, 112–120 (1970).
16. W. H. Barr, Factors involved in the assessment of systemic or biologic availability of drug products, Drug Inf. Bull., 3, 27–45 (1969).
17. M. Drucker, S. H. Blondheim, and L. Wislicki, Factors affecting acetylation in vivo of *para*-aminobenzoic acid by human subjects, Clin. Sci., 27, 133–141 (1964).
18. A. G. De Boer, D. D. Breimer, H. Mattie, J. Pronk, and J. M. Gubbens-Stibbe, Rectal bioavailability of lidocaine in man: Partial avoidance of "first-pass" metabolism, Clin. Pharmacol. Ther., 26, 701–709 (1979).
19. T. W. Schwarz, in *American Pharmacy*, 6th ed., (J. B. Sprowls, Jr. and H. M. Beal, Eds.), J. B. Lippincott, Philadelphia, 1966, pp. 311–331.
20. A. E. Pontiroli, M. Alberetto, and G. Pozza, Intranasal calcitonin and plasma calcium concentrations in normal subjects, Br. Med. J., 290, 1390–1391 (1985).
21. Y. W. Chien and K. Banga, Iontophoretic (transdermal) delivery of drugs: Overview of historic development, J. Pharm. Sci., 78, 353–354 (1989).
22. D. Parasrampuria and J. Parasrampuria, Percutaneous delivery of proteins and peptides using iontophoretic techniques, J. Clin. Pharm. Ther., 16, 7–17 (1991).
23. D. T. O'Hagan and L. Illum, Absorption of peptides and proteins from the respiratory tract and the potential for development of locally administered vaccine, Crit. Rev. Ther. Drug Carrier Syst., 7, 35–97 (1990).
24. J. J. Sciarra, Aerosols, in *Prescription Pharmacy*, 2nd ed. (J. B. Sprowls, Jr., Ed.), J. B. Lippincott, Philadelphia, 1970, pp. 280–328.
25. R. L. Dedrick and K. B. Bischoff, Pharmacokinetics in applications of the artifical kidney, Chem. Eng. Progr. Symp. Ser., 64, 32–44 (1968).
26. W. W. Mapleson, An electric analogue for uptake and exchange of inert gases and other agents, J. Appl. Physiol., 18, 197–204 (1963).
27. K. B. Bischoff and R. L. Dedrick, Thiopental pharmacokinetics, J. Pharm. Sci., 57, 1347–1357 (1968).
28. J. G. Wagner, *Fundamentals of Clinical Pharmacokinetics*. Drug Intelligence Publishers, Hamilton, IL, 1975, pp. 24–26.
29. F. Bree, G. Houin, J. Barre, J. L. Moretti, V. Wirquin, and J.-P. Tillement, Pharmacokinetics of intravenously administered ^{125}I-labelled human alpha-1-acid glycoprotein, Clin. Pharmacokinet, 11, 336–342 (1986).
30. C. G. Gahmberg and L. C. Anderson, Leucocyte surface origin of human alpha-1-acid glycoprotein (orosomucoid), J. Exp. Med., 148, 507–521 (1978).
31. T. W. Guenthert and S. Øie, Effect of plasma protein binding on quinidine kinetics in the rabbit, J. Pharmacol. Exp. Ther., 215, 165–171 (1980).

Copyright © 2002 Marcel Dekker, Inc.

32. T. F. Blaschke, Protein binding and kinetics of drugs in liver disease, Clin. Pharmacokinet, 2, 32–44 (1977).
33. S. Øie and J.-D. Huang, Binding, should free drug levels be measured? in *Topics in Pharmaceutical Sciences 1983* (D. D. Breimer and P. Speiser, Eds.), Elsevier, Amsterdam, 1983, pp. 51–62.
34. B. Fichtl, Tissue binding of drugs: Methods of determination and pharmacokinetic consequences, in *Plasma Binding of Drugs and Its Consequences* (F. Belpaire, M. Bogaert, J. P. Tillement, and R. Verbeeck, Eds.), Academia Press, Ghent, 1991, pp. 149–158.

Copyright © 2002 Marcel Dekker, Inc.

Chapter 6

Chemical Kinetics and Drug Stability

J. Keith Guillory and Rolland I. Poust

University of Iowa, Iowa City, Iowa

I. INTRODUCTION

In the rational design and evaluation of dosage forms for drugs, the stability of the active components must be a major criterion in determining their suitability. Several forms of instability can lead to the rejection of a drug product. First, there may be chemical degradation of the active drug, leading to a substantial lowering of the quantity of the therapeutic agent in the dosage form. Many drugs (e.g., 5-fluorouracil, carbamazepine, digoxin, and theophylline) have narrow therapeutic indices, and they need to be carefully titrated in individual patients so that serum levels are neither too high as to be potentially toxic nor too low as to be ineffective. In these cases it is of paramount importance that the dosage form reproducibly deliver the same amount of drug.

Second, although chemical degradation of the active drug may not be extensive, a toxic product may be formed in the decomposition process [1]. Dearborn [2] described several examples in which the products of degradation are significantly more toxic than the original therapeutic agent. Thus, the conversions of tetracycline to epianhydrotetracycline, arsphenamine to oxophenarsine, and *p*-aminosalicylic acid to *m*-aminophenol in dosage forms give rise to potentially toxic agents that, when ingested, can cause undesirable effects. Nord et al. [3] have reported that the antimalarial chloroquine can produce toxic reactions that are attributable to the photochemical degradation of the substance. Phototoxicity has also been observed to occur following administration of chlordiazepoxide and nitrazepam [4]. Another example of an adverse reaction due to a degradation product was provided by Neftel et al. [5], who showed that infusion of degraded penicillin G led to sensitization of lymphocytes and formation of antipenicilloyl antibodies.

Third, instability of a drug product can lead to a decrease in its bioavailability rather than loss of drug or formation of toxic degradation products. This reduction in bioavailability can result in a substantial lowering in the therapeutic efficacy of the dosage form. This phenomenon can be caused by physical and/or chemical changes in the excipients in the dosage form, independent of whatever changes the active drug may have undergone. A more detailed discussion of this subject is given in Sec. II.B.

Fourth, there may be substantial changes in the physical appearance of the dosage form. Examples of these physical changes include mottling of tablets, creaming of emulsions, and caking of suspensions. Although the therapeutic efficacy of the dosage form may be unaffected by these changes, the patient will most likely lose confidence in the drug product, which then has to be rejected.

Fifth, while the drug substance itself may retain its potency, excipients such as antimicrobial preservatives, solubilizers, emulsifying or suspending agents may degrade, compromising the integrity of the drug product.

A drug product, therefore, must satisfy stability criteria chemically, toxicologically, therapeutically, and physically. Basic principles in pharmaceutical kinetics can often be applied to anticipate and quantify the

Copyright © 2002 Marcel Dekker, Inc.

undesirable changes so that they can be circumvented by stabilization techniques. Some chemical compounds, called prodrugs [6,7], are designed to undergo chemical or enzymatic conversion in vivo to pharmacologically active drugs. Prodrugs are employed to solve one or several problems presented by active drugs (e.g., short biological half-life, poor dissolution, bitter taste, inability to penetrate through the blood-brain barrier, etc.). They are pharmacologically inactive as such but are converted back in vivo to their parent (active) compounds. Naturally, the rate and extent of this conversion (governed by the same laws of kinetics described in this chapter) are the primary determinants of the therapeutic efficacy of these agents.

In the present chapter, stability problems and chemical kinetics are introduced and surveyed. The sequence employed is as follows: first, an overview of the potential routes of degradation that drug molecules can undergo; then, a discussion of the mathematics used to quantify drug degradation; a delineation of the factors that can affect degradation rates, with an emphasis on stabilization techniques; and finally a description of stability-testing protocols employed in the pharmaceutical industry. It is not the intent of this chapter to document stability data of various individual drugs. Readers are referred to the compilations of stability data [8] and to literature on specific drugs (e.g., Ref. 9 and earlier volumes) for this kind of information.

II. ROUTES BY WHICH PHARMACEUTICALS DEGRADE

Since most drugs are organic molecules, many pharmaceutical degradation pathways are, in principle, similar to reactions described for organic compounds in standard organic chemistry textbooks. On the other hand, different emphases are placed on the types of reactions that are commonly encountered in the drug product stability area as opposed to those seen in classical organic chemistry. In the latter case, reactions generally are described as tools for use by the synthetic chemist; thus, the conditions under which they are carried out are likely to be somewhat drastic. Reactive agents (e.g., thionyl chloride or lithium aluminum hydride) are employed in relatively high concentrations (often >10%), and exaggerated conditions, such as refluxing or heating in a pressure bomb are common. Reactions are effected in relatively short periods of time (hours or days). In contrast, the active drug components in pharmaceuticals are usually present in relatively low concentrations. For example, dexamethasone sodium phosphate, a synthetic adrenocorticoid steroid salt, is present only to the extent of about 0.4% in its injection, 0.1% in its topical cream or opthalmic solution, and 0.05% in its opthalmic ointment. The decomposition of a drug is likely to be mediated not by reaction with another active ingredient but by reaction with water, oxygen, or light. Reaction conditions of interest usually are ambient or subambient. Reactions in pharmaceuticals ordinarily occur over months or years as opposed to the hours or days required for completion of reactions in synthetic organic chemistry.

Reactions such as the Diels-Alder reaction and aldol condensations, which are important in synthetic and mechanistic organic chemistry, are of only minor importance when drug degradation is being considered. Pharmaceutical scientists need to refocus their attention on reactions such as hydrolysis, oxidation, photolysis, racemization, and decarboxylation, the routes by which most pharmaceuticals degrade.

A familiarity with reactions of particular functional groups is important if one is to gain a broad view of drug degradation. It is a difficult task to recall degradative pathways of all commonly used drugs. Yet, through the application of functional group chemistry it is possible to anticipate the potential mode(s) of degradation that drug molecules will likely undergo. In the following discussion, therefore, degradative routes are demonstrated by calling attention to the reactive functional groups present in drug molecules. The degradative routes are described, through the use of selected examples, as *chemical* when new chemical entities are formed as a result of drug decomposition and as *physical* when drug loss does not produce distinctly different chemical products. For further information about chemical functional group reactivity, the interested reader should consult a recent review ([10] and the comprehensive text by K. A. Connors [11]).

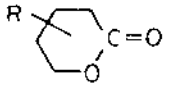
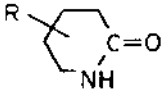
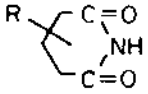
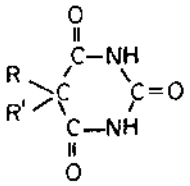
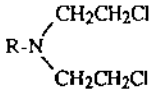
A. Chemical Degradative Routes

Solvolysis

In this type of reaction the active drug undergoes decomposition following reaction with the solvent present. Usually the solvent is water, but sometimes the reaction may involve pharmaceutical cosolvents such as ethyl alcohol or polyethylene glycol. These solvents can act as nucleophiles, attacking the electropositive centers in drug molecules. The most common solvolysis reactions encountered in pharmaceuticals are those involving ''labile'' carbonyl compounds such as esters, lactones, and lactams (Table 1).

Copyright © 2002 Marcel Dekker, Inc.

Table 1 Some Functional Groups Subject to Hydrolysis

Drug type		Examples
Esters	$RCOOR'$	Aspirin, alkaloids
	$ROPO_3M_x$	Dexamethasone sodium phosphate
		Estrone sulfate
	$ROSO_3M_x$	Nitroglycerin
	$RONO_2$	
Lactones		Pilocarpine Spironolactone
Amides	$RCONR'_2$	Thiacinamide Chloramphenicol
Lactams		Penicillins Cephalosporins
Oximes	$R_2C = NOR$	Steroid oximes
Imides		Glutethimide Ethosuximide
Malonic ureas		Barbiturates
Nitrogen mustards		Melphalan

Although all the functional groups cited are, in principle, subject to solvolysis, the rates at which they undergo this reaction may be vastly different. For example, the rate of hydrolysis of a β-lactam ring (a cyclized amine) is much greater than that of its linear analog. The half-life (the time needed for half of the drug to decompose) of the β-lactam in potassium phenethicillin at 35°C and pH 1.5 is about 1 hour. The corresponding half-life for penicillin G is about 4 minutes [12]. In contrast, the half-life for hydrolysis of the simple amide propionamide in 0.18 molal H_2SO_4 at 25°C is about 58 hours [13]. It has been suggested that the antibacterial activity of β-lactam antibiotics arises from a combination of their chemical reactivity and their molecular recognition by target enzymes. One aspect of their chemical reactivity is their acylating power, and while penicillins are not very good acylating agents, they are more reactive than simple, unsubstituted amides [14]. Unactivated or "normal" amides undergo nonenzymatic hydrolysis slowly except under the most extreme conditions of pH and temperature because the N-C(O) linkage is inherently stable, yet when the amine function is a good leaving group (and particularly if it has a pK_a greater than 4.5), amides can be susceptible to hydrolysis at ordinary temperatures. (For a review on this subject see Ref. 15.) Acyl-transfer reactions in

Copyright © 2002 Marcel Dekker, Inc.

peptides, including the transfer to water (hydrolysis), are of fundamental importance in biological systems where the reactions proceed at normal temperatures and enzymes serve as catalysts.

The most frequently encountered hydrolysis reaction in drug instability is that of the ester, but curtain esters can be stable for many years when properly formulated. Substituents can have a dramatic effect on reaction rates. For example, the *tert*-butyl ester of acetic acid is about 120 times more stable than the methyl ester, which, in turn, is approximately 60 times more stable than the vinyl analog [16]. Structure-reactivity relationships are dealt with in the discipline of physical organic chemistry. Substituent groups may exert electronic (inductive and resonance), steric, and/or hydrogen-bonding effects that can drastically affect the stability of compounds. A detailed treatment of substituent effects can be found in a review by Hansch et al. [17] and in the classical reference text by Hammett [18].

A dramatic decrease in ester stability can be brought about by intramolecular catalysis. This type of facilitation is effected mostly by neighboring groups capable of exhibiting acid-base properties (e.g., $-NH_2$, -OH, -COOH, COO^-, etc.). If neighboring-group participation leads to an enhanced reaction rate, the group is said to provide anchimeric assistance [19]. For example, the ethyl salicylate anion undergoes hydrolysis in alkaline solution at a rate 10^6 times greater than the experimental value for the uncatalyzed cleavage of ethyl *p*-hydroxybenzoate. The rate advantage is attributed to intramolecular general base catalysis by the phenolate anion [20]. Self-associated insulin degrades at a rate approximately 2.5 times greater than the monomer at pH 2 and 3 due to intramolecular catalysis by the unionized carboxyl terminus of the A chain [21].

Oxidation

Oxidation reactions are important pathways of drug decomposition. In pharmaceutical dosage forms, oxidation is usually mediated through reaction with atmospheric oxygen under ambient conditions, a process commonly referred to as autoxidation. Oxygen is, itself, a diradical, and most autoxidations are free radical reactions. A free radical is a molecule or atom with one or more unpaired electrons. Of considerable importance to pharmaceutical scientists is a reliable method for determining and controlling oxygen concentration in aqueous solutions [22]. Sensitivity to oxidation of a drug can be ascertained by investigating its stability in an atmosphere of high oxygen tension. Usually a 40% oxygen atmosphere allows for a rapid evaluation. Results are compared against those obtained under inert (nitrogen or argon) or ambient atmospheres. A thorough review of autoxidation and of antioxidants has been published [23].

The mechanisms of oxidation reactions are usually complex, involving multiple pathways for the initiation, propagation, branching, and termination. Many oxidation reactions are catalyzed by acids and bases [23]. Often autoxidation reactions are initiated by trace amounts of impurities such as metal ions or hydroperoxides. Thus, ferric ion catalyzes the degradation reaction and decreases the induction period for the oxidation of the compound procaterol [24]. As little as 0.0002 M copper ion has been shown to increase the rate of vitamin C oxidation by a factor of 10^5 [25]. Hydroperoxides contained in polyethylene glycol suppository bases have been implicated in the oxidation of codeine to codeine-*N*-oxide [26]. Peroxides apparently are responsible for the accelerated degradation of benzocaine hydrochloride in aqueous cetomacrogol solution [27] and of a corticosteriod in polyethylene glycol 300 [28,29]. Butyl *p*-hydroxybenzoate, formulated in a suspension containing poly(ethylene glycol) and polysorbate 80, was found to turn yellow under ambient conditions if not adequately protected from oxygen. The degradation product, oxalamidine, may be formed by the reaction of butamben with an oxidation product of poly(-ethylene glycol) [30].

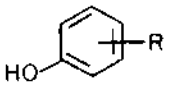
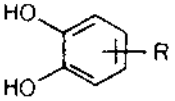
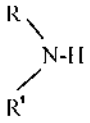
The formation of peroxides and formaldehyde in the high-purity polyoxyethylene surfactants in toiletries has been shown to lead to contact dermatitis [31]. Peroxides in hydrogenated castor oil can cause autoxidation of miconazole [32]. Oxidative decomposition of the polyoxyethylene chains occurs at elevated temperature, leading to the formation of ethylene glycol, which may then be oxidized to formaldehyde. When polyethylene glycol and poloxamer were used to prepare solid dispersions of bendroflumethiazide, a potent, lipophilic diuretic drug, the drug reacted with the formaldehyde to produce hydroflumethiazide [33].

In the case of nucleotides, oxidation mainly involves the reactions of reactive oxygen species such as the hydroxyl radical ($\cdot OH$), hydrogen peroxide (H_2O_2), superoxide ($\cdot O_2^-$), and/or singlet oxygen ($^1O_2$). They are generated by ionizing irradiation, photolysis, and/or the metal-catalyzed reduction of oxygen or decomposition of peroxides [34].

A list of some functional groups that are subject to autoxidation is shown in Table 2. The products of oxidation are usually electronically more conjugated; thus, the appearance of, or a change in,

Copyright © 2002 Marcel Dekker, Inc.

Table 2 Some Functional Groups Subject to Autoxidation

Functional group		Examples
Phenols		Phenols in steroids
Catechols		Catecholamines (dopamine, isoproterenol)
Ethers	$R-O-R'$	Diethylether
Thiols	RCH_2SH	Dimercaprol (BAL)
Thioethers	$R-S-R'$	Phenothiazines (chlorpromazine)
Carboxylic acids	$RCOOH$	Fatty acids
Nitrites	RNO_2	Amyl nitrite
Aldehydes	$RCHO$	Paraldehyde
Amines		Morphine Clozapine (to N-oxides)

color in a dosage form is suggestive of the occurrence of oxidative degradation.

Photolysis

Normal sunlight or room light may cause substantial degradation of drug molecules. The energy from light radiation must be absorbed by the molecules to cause a photolytic reaction. If that energy is sufficient to achieve activation, degradation of the molecule is possible. Saturated molecules do not interact with visible or near-ultraviolet light, but molecules that contain π-electrons usually do absorb light throughout this wavelength range. Consequently, compounds such as aromatic hydrocarbons, their heterocyclic analogs, aldehydes, ketones, etc., are most susceptible to photolysis. In general, drugs that absorb light at wavelengths below 280 nm have the potential to undergo decomposition in sunlight, and drugs with absorption maxima greater than 400 nm have the potential for degradation in both sunlight and room light. Fortunately, most pharmaceuticals can be protected adequately from photodegradation by the use of appropriate packaging.

A dramatic example of photolysis is the photodegradation of sodium nitroprusside in aqueous solution. Sodium nitroprusside, $Na_2Fe(CN)_5NO{\cdot}2\ H_2O$, is administered by intravenous infusion for the management of acute hypertension. If the solution is protected from light it is stable for at least one year; if exposed to normal room light, it has a shelf life of only 4 hours [35].

Photolysis reactions often are associated with oxidation because the latter category of reactions frequently can be initiated by light. The photooxidation of phenothiazines with the formation of N- and S-oxides is typical. But photolysis reactions are not restricted to oxidation. In the case of sodium nitroprusside, it is believed that degradation results from loss of the nitro-ligand from the molecule, followed by electronic rearrangement and hydration. Photo-induced reactions are common in steroids [36]; an example is the formation of 2-benzoylcholestan-3-one following irradiation of cholest-2-en-3-ol benzoate. Photoadditions of water and of alcohols to the electronically excited state of steroids have also been observed [37].

Phototoxicity caused by drugs is a problem that is being observed with greater frequency. The primary

Copyright © 2002 Marcel Dekker, Inc.

reactions responsible for initiating phototoxicity can usually be described in terms of basic photochemical reaction patterns. These often involve generation of singlet state excited oxygen. Phototoxic drugs include furosemide, acetazolamide, triamterine, thiocolchicoside, and chlorthalidone [38].

A guideline for evaluation of the photostability of new drug substances and dosage forms was published in 1997 [39]. Subsequently, a joint study was undertaken by the U.S. Food and Drug Administration (FDA) and Pharmaceutical Research and Manufacturers of America to evaluate the ICH guideline. A draft chapter for the United States Pharmacopeia, based on the guideline and the joint study, has been published in Pharmacopeial Forum [40].

Dehydration

The preferred route of degradation for prostaglandin E_2 and tetracycline is the elimination of a water molecule from their structures. The driving force for this type of covalent dehydration is the formation of a double bond that can then participate in electronic resonance with neighboring functional groups. In physical dehydration processes, such as those occurring in theophylline hydrate and ampicillin trihydrate [41], water removal does not create new bonds but often changes the crystalline structure of the drug. Since it is possible that anhydrous compounds may have different dissolution rates compared to their hydrates [42,43], dehydration reactions involving water of crystallization may potentially affect the absorption rate of the dosage form.

Racemization

The racemization of pharmacologically active agents is of interest because enantiomers often have significantly different absorption, distribution, metabolism, and excretion, in addition to differing pharmacological actions [44]. The best known racemization reactions of drugs are those that involve epinephrine, pilocarpine, ergotamine, and tetracycline. In these cases, the reaction mechanism appears to involve an intermediate carbonium ion or carbanion, which is stabilized electronically by the neighboring substituent group. For example, in the racemization of pilocarpine [45], a carbanion is produced and stabilized by delocalization to the enolate. In addition to the racemization reaction, pilocarpine is also degraded through hydrolysis of the lactone ring.

Most racemization reactions are catalyzed by acid and/or by base. A notable exception is the "spontaneous" racemization of the diuretic and antihypertensive agent, chlorthalidone, which undergoes facile S_N1 solvolysis of its tertiary hydroxyl group to form a planar carbonium ion. Chiral configuration is then restored by nucleophilic attack (S_N2) of a molecule of water on the carbonium ion with subsequent elimination of a proton [46].

When configurational stability is truly low and results in half-lives of racemization or epimerization, which are on the order of minutes or hours, the phenomenon has pharmacological significance. On the other hand, when the half-lives are on the order of months or years, the phenomenon has pharmaceutical significance and may shorten the shelf life of drugs [47].

Incompatibilities

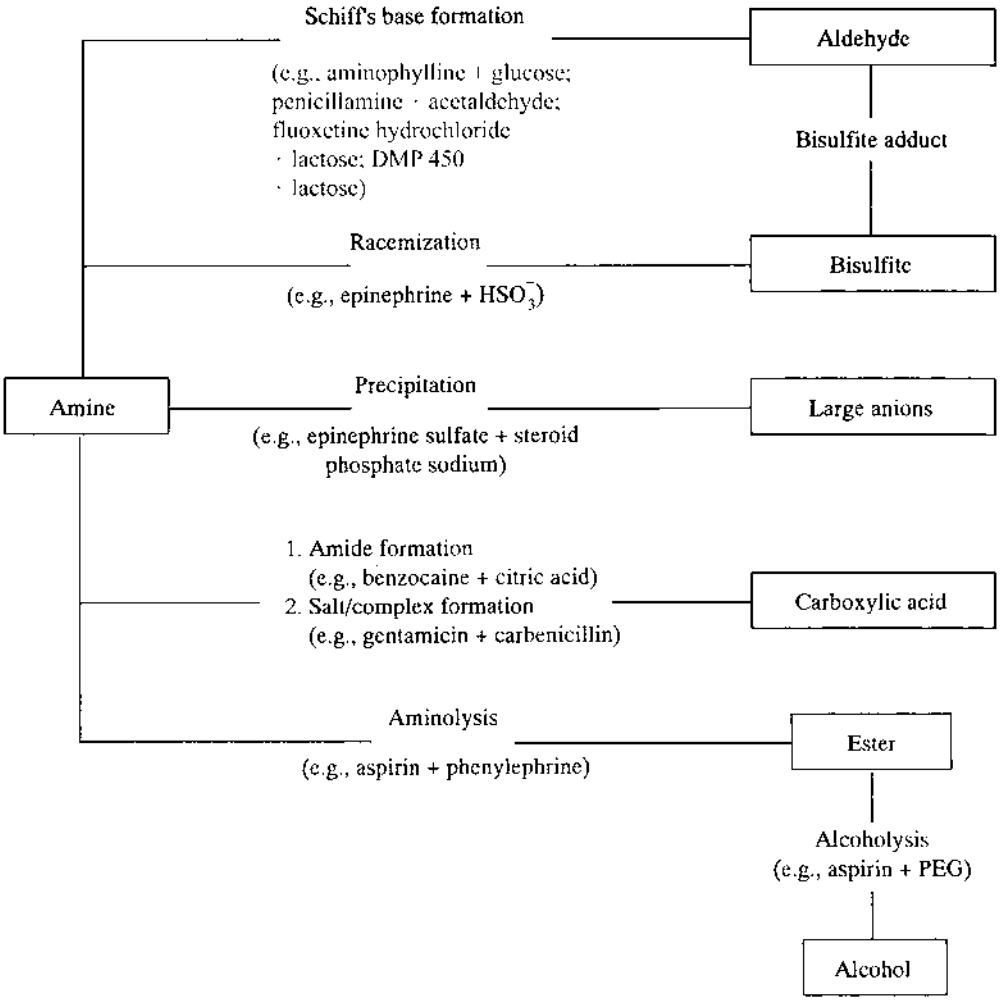
Chemical interactions between two or more drug components in the same dosage form, or between active ingredient and a pharmaceutical adjuvant, occur frequently. An example of drug-drug incompatibility is the inactivation of cationic aminoglycoside antibiotics, such as kanamycin and gentamycin, by anionic penicillins in IV admixtures. The formation of an inactive complex between these two classes of antibiotics occurs not only in vitro but apparently also in vivo in patients with severe renal failure [48]. Thus, when gentamycin sulfate was given alone to patients on long-term hemodialysis, the biological half-life of gentamycin was greater than 60 hours. But, when carbenicillin disodium (CD) was given with gentamycin sulfate (GS) in the dose ratio CD/GS = 80:1, the gentamycin half-life was reduced to about 24 hours.

Incompatibilities have also been observed in solid dosage forms. A typical tablet contain binders, disintegrants, lubricants and fillers. Compatibility screening for a new drug should consider two or more excipients from each class. Serajuddin et al. have developed a drug-excipient compatibility screening model to predict interactions of drug substances with excipients [49].

Many pharmaceutical incompatibilities are the result of reactions involving the amine functional group. A summary of the potential interactions that can occur between various functional groups is given in Table 3. The Maillard reaction between simple amines or amino acids, peptides or proteins, and carbonyl compounds, such as the "glycosidic" hydroxyl group of reducing sugars, is typical and often results in mottling of tablets. It is a variant of the Schiff's base reaction. One would think that pharmaceutical scientists would realize by now that combinations of lactose and primary or secondary amines will lead to Amadori rearrangement products, which are responsible for browning reactions; however, this topic remains the subject of many reports [50,51].

Copyright © 2002 Marcel Dekker, Inc.

Table 3 Some Potential Drug Incompatibilities



The Maillard reaction is likely to take on additional significance with the introduction of many new protein and peptide pharmaceuticals. For example, Tarelli et al. have demonstrated that lysine vasopressin undergoes rapid glycation in the presence of reducing sugars in both aqueous and solid formulations and that the N-terminal adduct can form rapidly even at $-20°C$ [52]. A textbook that deals with the consequences for the chemical and life sciences of the Maillard reaction has been published [53].

Other Chemical Degradation Reactions

Other chemical reactions, such as hydration, decarboxylation, or pyrolysis, are also potential routes for drug degradation. For example, cyanocobalamin may absorb about 12% of water when exposed to air, and *p*-aminosalicyclic acid decomposes with evolution of carbon dioxide to form *m*-aminophenol when subjected to temperatures above 40°C. The temperature at which pyrolytic decomposition of terfenadine occurs has been used as a criterion for determining which of several tablet excipients will be preferable with respect to long-term stability of the drug substance [54].

B. Physical Degradative Routes

Polymorphs are different crystal forms of the same compound [55]. They are usually prepared by crystallization of the drug from different solvents under diverse conditions. However, exposure to changes in temperature, pressure, relative humidity, and comminution which are encountered in processes such as drying, granulation, milling, and compression may also lead to polymorphic transformations.

Copyright © 2002 Marcel Dekker, Inc.

Steroids, sulfonamides, and barbiturates are notorious for their propensity to form polymorphs [56]. Yang and Guillory [57] attempted to correlate the frequency of occurrence of polymorphism in sulfonamides with certain aspects of chemical structure. They found that sulfonamides that did not exhibit polymorphism have somewhat higher melting points and heats of fusion than those that were polymorphic. The absence of polymorphism in sulfacetamide was attributed to the stronger hydrogen bonds formed by the amide hydrogen in this molecule. These stronger hydrogen bonds were not readily stretched or broken to form alternate crystalline structures.

Since polymorphs differ from one another in their crystal energies, the more energetic ones will seek to revert to the most stable (and the least energetic) crystal form. When several polymorphs and solvates (substances that incorporate solvent in a stoichiometric fashion into the crystal lattice) are present, the conditions under which they may interconvert can become quite complex, as is true of fluprednisolone [58].

Polymorphs may exhibit significant differences in important physicochemical parameters such as solubility, dissolution rate, and melting point [59]. Thus the conversion from one polymorph to another in a pharmaceutical dosage form may lead to a drastic change in the physical characteristics of the drug. A well-known example of this phenomenon is the conversion of a more soluble crystal form (form II) of cortisone acetate to a less soluble form (form V) when the drug is formulated into an aqueous suspension [60]. This phase change leads to caking of the cortisone acetate suspension.

The phenomenon of pseudopolymorphism is also observed, i.e., compounds can crystallize with one or more molecules of solvent in the crystal lattice. Conversion from solvated to nonsolvated, or hydrate to anhydrous, and vice versa, can lead to changes in solid-state properties. For example, a moisture-mediated phase transformation of carbamazepine to the dihydrate has been reported to be responsible for whisker growth on the surface of tablets. The effect can be retarded by the inclusion of Polyoxamer 184 in the tablet formulation [61].

Another physical property that can affect the appearance, bioavailability, and chemical stability of pharmaceuticals is degree of crystallinity. Amorphous materials tend to be more hygroscopic than their crystalline counterparts. Also, there is a substantial body of evidence that indicates that the amorphous forms of drugs are less stable than their crystalline counterparts [62]. It has been reported, for example, that crystalline cyclophosphamide [63] is much more stable than the amorphous form. Paradoxically, M. J. Pikal and D. R. Rigsbee have reported that in the case of biosynthetic human insulin, the amorphous form is more stable [64].

Vaporization

Some drugs and pharmaceutical adjuvants possess sufficiently high vapor pressures at room temperature that their volatilization constitutes a major route of drug loss. Flavors, whose constituents are mainly ketones, aldehydes, and esters, and cosolvents (low molecular weight alcohols) may be lost from the formulation in this manner. The most frequently cited example of a pharmaceutical that "degrades" by this route is nitroglycerin, which has a vapor pressure of 0.00026 mm at 20°C and 0.31 mm at 93°C [65]. Significant drug loss to the environment can occur during patient storage and use. In 1972, FDA issued special regulations governing the types of containers that may be used for dispensing sublingual nitroglycerin tablets [66].

Reduction of vapor pressure and, hence, of volatility of drugs such as nitroglycerin can be achieved through dispersion of the volatile drug in macromolecules that can provide physicochemical interactions. The addition of macromolecules such as polyethylene glycol, polyvinylpyrrolidone, and microcrystalline cellulose allows for preparation of "stabilized" nitroglycerin sublingual tablets [67,68]. A β-cyclodextrin-nitroglycerin tablet is currently being marketed in Japan to achieve the same purpose.

Another aspect of nitroglycerin instability has been observed by Fusari [68]. When conventional (unstabilized) nitroglycerin sublingual tablets are stored in enclosed glass containers, the high volatility of the drug gives rise to redistribution of nitroglycerin among the stored tablets. Interestingly, this redistribution leads to an increase in the standard deviation of the drug contents of the tablets, rather than the reverse. This migration phenomenon results in a deterioration in the uniformity of the tablets upon storage.

"Aging"

The most interesting and perhaps the least-reported area of concern regarding the physical instability of pharmaceutical dosage forms is generally termed "aging." This is a process through which changes in the disintegration and/or dissolution characteristics of the dosage form are caused by subtle, and sometimes unexplained, alterations in the physicochemical

Copyright © 2002 Marcel Dekker, Inc.

properties of the inert ingredients or the active drug in the dosage form [69]. Since the disintegration and dissolution steps may be the rate-determining steps in the absorption of a drug, changes in these processes as a function of the "age" of the dosage form may result in corresponding changes in the bioavailability of the drug product.

An example of this phenomenon was provided by deBlaey and Rutten-Kingma [70], who showed that the melting time of aminophylline suppositories, prepared from various bases, increased from about 20 minutes to over an hour after 24 weeks of storage at 22°C. Like the dissolution time for solid dosage forms, the melting time for suppositories can be viewed as an in vitro index of drug release. Thus, an increase in melting time can perceivably lead to a decrease in bioavailability. The mechanism responsible for this change appeared to involve an interaction between the ethylenediamine in aminophylline and the free fatty acids present in the suppository bases. Interestingly, no increase in melting time was detected when the suppositories were stored at 4°C, even up to 15 months.

"Aging" of solid dosage forms can cause a decrease in their in vitro rate of dissolution [71], but a corresponding decrease in in vivo absorption cannot be assumed automatically. For example, Chemburkar et al. [72] showed that when a methaqualone tablet was stored at 80% relative humidity for 7–8 months, the dissolution rate as measured by in vivo absorption was not affected. This lack of in vitro dissolution–in vivo absorption correlation for the aged product was observed even though the particular dissolution method (that of the resin flask) was shown by the same workers to be capable of discriminating between the absorption of several trial dosage forms of the same drug.

Adsorption

Drug-plastic interaction is increasingly being recognized as a major potential problem when intravenous solutions are stored in bags or infused via administration sets that are made from polyvinyl chloride (PVC). For example, up to 50% drug loss can occur after nitroglycerin is stored in PVC infusion bags for 7 days at room temperature [73]. This loss can be attributed to adsorption rather than chemical degradation because the drug can be recovered from the inner surface of the container by rinsing with a less polar solvent (methanol in this case). Similarly, more than 40% of a dose of quinidine gluconate was lost when the drug was administered with a conventional PVC intravenous administration set. Drug loss was reduced by using a winged intravenous catheter and shorter tubing [74]. A diverse array of drugs, including diazepam [75], insulin [76], isosorbide dinitrate [77], and others [78], has shown substantial adsorption to PVC. The propensity for significant adsorption is related to the oil/water partition coefficient of the drug since this process depends on the relative affinity of the drug for the hydrophobic PVC (dielectric constant of about 3) and the hydrophilic aqueous infusion medium.

Physical Instability in Heterogeneous Systems

The stability of suspensions, emulsions, creams, and ointments is dealt with in other chapters. The unique characteristics of solid-state decomposition processes have been described in reviews by D. C. Monkhouse [79,80] and in the monograph on drug stability by J. T. Carstensen [81]. Baitalow et al. have applied an unconventional approach to the kinetic analysis of solid-state reactions [82]. The recently published monograph on solid-state chemistry of drugs also treats this topic in great detail [83].

III. QUANTITATION OF RATE OF DEGRADATION

Before undertaking a discussion of the mathematics involved in the determination of reaction rates is undertaken, it is necessary to point out the importance of proper data acquisition in stability testing. Applications of rate equations and predictions are meaningful only if the data utilized in such processes are collected using valid statistical and analytical procedures. It is beyond the scope of this chapter to discuss the proper statistical treatments and analytical techniques that should be used in a stability study. Some perspectives in these areas can be obtained by reading the comprehensive review by Meites [84], the paper by P. Wessels et al. [85], and the section on statistical considerations in the stability guidelines published by FDA in 1987 [86] and in the more recent *Guidance for Industry* published in June 1998 [87].

A. Kinetic Equations

Consider the reaction

$$a\mathrm{A} + b\mathrm{B} \rightarrow m\mathrm{M} + n\mathrm{N} \tag{1}$$

where A and B are the reactants, M and N the products, and a, b, m, and n the stoichiometric coefficients describing the reaction. The rate of change of the concentration C of any of the species can be expressed

Copyright © 2002 Marcel Dekker, Inc.

by the differential notations $-dC_A/dt$, $-dC_B/dt$, dC_M/dt, and dC_N/dt. Note that the rates of change for the reactants are preceded by a negative sign, denoting a decrease in concentration relative to time (rate of disappearance). In contrast, the differential terms for the products are positive in sign, indicating an increase in concentration of these species as time increases (rate of appearance). The rates of disappearance of A and B and the rates of appearance of M and N are interrelated by equations that take into account the stoichiometry of the reaction:

$$-\frac{1}{a}\frac{dC_A}{dt} = -\frac{1}{b}\frac{dC_B}{dt} = \frac{1}{m}\frac{dC_M}{dt} = \frac{1}{n}\frac{dC_N}{dt} \tag{2}$$

The Rate Expression

The rate expression is a mathematical description of the rate of the reaction at any time t in terms of the concentration(s) of the molecular species present at that time. By using the hypothetical reaction $aA + bB \rightarrow$ products, the rate expression can be written as

$$-\frac{dC_A}{dt} = -\frac{dC_B}{dt} \propto C_{A(t)}^{a} C_{B(t)}^{b} \tag{3}$$

Equation (3) in essence states that the rate of change of the concentration of A at time t is equal to that of B and that each of these changes at time t is proportional to the product of the concentrations of the reactants raised to the respective prowers. Note that $C_{A(t)}$ and $C_{B(t)}$ are time-dependent variables. As the reaction proceeds, both $C_{A(t)}$ and $C_{B(t)}$ will decrease in magnitude. For simplicity, these concentrations can be denoted simply by C_A and C_B, respectively.

$$-\frac{dC_A}{dt} = -\frac{dC_B}{dt} = kC_A^{a} C_B^{b} \tag{4}$$

where k is a proportionality constant, commonly referred to as the reaction rate constant or the specific rate constant. The format for rate expressions generally involves concentration terms of only the reactants and very rarely those of the products. The latter occurs only when the products participate in the reaction once it has been initiated.

The order of the reaction, n, can be defined as $n = a + b$. Extended to the general case, the order of a reaction is the numerical sum of the exponents of the concentration terms in the rate expression. Thus if $a = b = 1$, the reaction just described is said to be second-order overall, first-order relative to A, and first-order relative to B. In principle, the numerical value of a or b can be integral or fractional.

Special attention is directed to those instances in which the rate of reaction is apparently independent of the concentration of one of the reactants, even though this reactant is consumed during the reaction. For example, in the reaction between an ester and water (hydrolysis) in a predominantly aqueous environment, the theoretical rate expression for the ester can be written in terms of the concentrations of the ester (C_E) and water (C_W):

$$-\frac{dC_E}{dt} = kC_E C_W \tag{5}$$

If the initial concentration of the ester is 0.5 M or less, complete hydrolysis of the ester will bring about a corresponding decrease in the concentration of water of 0.5 M or less. Since the initial water concentration is 1000/18, which is about 55 M for an aqueous solution, the loss of water through reaction is insignificantly small and C_W can be considered a constant throughout the entire course of the reaction. Thus, in practice,

$$-\frac{dC_E}{dt} = k_\pi C_E \tag{6}$$

where $k_\pi = kC_W$. The reaction is thus apparently first-order relative to ester and zero-order relative to water; the overall reaction is known as a pseudo-first-order reaction and k_π the pseudo-first-order constant.

This type of kinetics is observed whenever the concentration of one of the reactants is maintained constant, either by a vast excess initial concentration or by rapid replenishment of one of the reactants. Thus, if one of the reactants is the hydrogen ion or the hydroxide ion, its concentration, though probably small when compared with that of the drug, can be kept constant throughout the reaction by using buffers in the solution. Similarly, the concentration of an unstable drug in solution can be maintained invariant by preparing a drug suspension, thus providing excess solid in equilibrium with the drug in solution.

Simple Reactions

It is obvious that to quantify the rate expression, the magnitude of the rate constant k needs to be determined. Proper assignment of the reaction order and accurate determination of the rate constant is important when reaction mechanisms are to be deduced from the kinetic data. The integrated form of the reaction equation is easier to use in handling kinetic data. The integrated kinetic relationships commonly used for zero-, first-, and second-order reactions are summarized in Table 4. [The reader is advised that basic kinetic

Copyright © 2002 Marcel Dekker, Inc.

Table 4 Rate Expressions for Zero-, First-, and Second-Order Reactions

			Second-order	
	Zero-order	First-order	$a = b = c_0$	$a \neq b$
Differential rate expression	$-\frac{dc}{dt} = k$	$-\frac{dc}{dt} = kc$	$-\frac{dc}{dt} = kc^2$	$-\frac{dc}{dt} = kc_a c_b$
Integrated rate expression	$k = \frac{c_0 - c}{t}$	$k = \frac{1}{t}\ln\frac{c_0}{c}$	$\frac{1}{c} - \frac{1}{c_0} = kt$	$k = \frac{1}{t(a-b)}\ln\frac{b(a-x)}{a(b-x)}$
$t_{1/2}$	$\frac{c_0}{2k}$	$\frac{0.693}{k}$	$\frac{1}{c_0 k}$	(i) When $x = 0.5a$ $\frac{1}{k(a-b)}\ln\frac{0.5ab}{a(b-0.5a)}$ (ii) When $x = 0.5b$ $\frac{1}{k(a-b)}\ln\frac{b(a-0.5b)}{0.5ab}$
$t_{90\%}$	$\frac{c_0}{10k}$	$\frac{0.105}{k}$	$\frac{0.11}{c_0 k}$	(i) When $x = 0.1a$ $\frac{1}{k(a-b)}\ln\frac{0.9ab}{a(b-0.1a)}$ (ii) When $x = 0.1b$ $\frac{1}{k(a-b)}\ln\frac{b(a-0.1b)}{0.9ab}$

theory is also extensively exploited in pharmacokinetics; for further information on this subject, see Chapter 3.] The concentration symbols in Table 4 are defined as follows: c is the concentration of the drug at any time t and c_0 is the initial concentration. In the last column describing a second-order reaction in which the reactants A and B do not have the same initial concentrations, these are designated as a and b, respectively; x is the concentration reacted at time t.

In a reaction of either zero-order or second-order, the time to reach a certain fraction of the initial concentration, for example, $t_{1/2}$ or $t_{90\%}$ [the time required for the drug concentration to decrease to 90% of its original value (i.e., 10% degradation)] is dependent on c_0. This is illustrated in Fig. 1, in which a zero-order reaction (Fig. 1a) and a second-order reaction (Fig. 1b) are plotted with two initial concentrations. It is readily seen that for a zero-order reaction, the $t_{1/2}$ increases

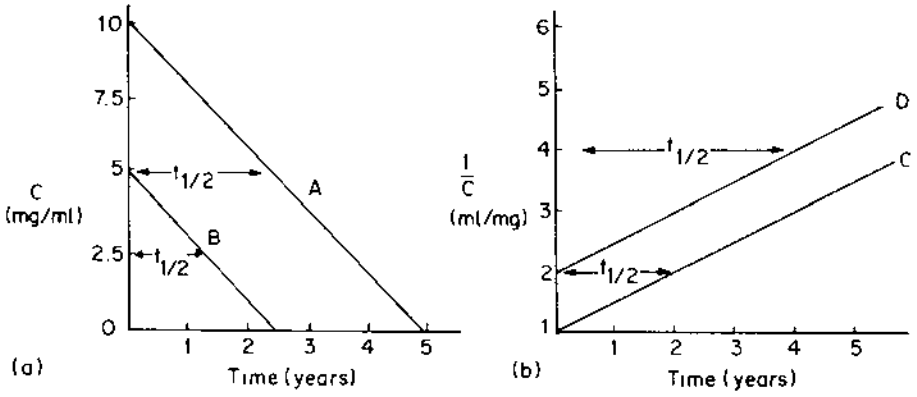


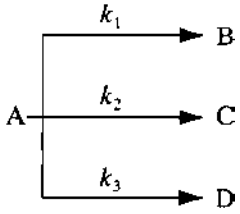
Fig. 1 Effect of initial concentration on the half-life of (a) a zero-order and (b) a second-order reaction. In (a), $k = 2$ mg/year-mL; curve A, initial concentration $c_0 = 10$ mg/mL, $t_{1/2} = 2.5$ years; curve B, $c_0 = 5$ mg/mL, $t_{1/2} = 1.25$ years. In (b), $k = 0.5$ mL/mg-year; curve C, $c_0 = 1$ mg/mL, $t_{1/2} = 2$ years; curve D, $c_0 = 0.5$ mg/mL, $t_{1/2} = 4$ years.

Copyright © 2002 Marcel Dekker, Inc.

with a higher initial concentration. Conversely, for a second-order reaction, $t_{1/2}$ decreases with increasing initial concentration. For a reaction obeying first-order kinetics, the $t_{1/2}$ or $t_{90\%}$ is independent of c_0.

Complex Reactions

Parallel First-Order Reactions. In many instances, the active drug may degrade through more than one pathway:



(7)

If the concentration of the active drug, A, can be monitored, the composite rate constant, $k' = k_1 + k_2 + k_3$, can easily be determined from the relationship $[A] = [A]_0 e^{-k't}$, where $[A]_0$ is the initial concentration and [A] is the concentration at time t. If the concentrations of A cannot be determined because of assay difficulties, it is still possible to determine k' by monitoring one of the degradation products. For example, if the concentrations of B can be assayed as a function of time, and the concentration of B at time infinity, $[B]_\infty$, is also determined, the following relationships can be derived:

$$[B] = \frac{k_1}{k'}[A]_0(1 - e^{-k't}) \tag{8}$$

$$[B]_\infty = \frac{k_1}{k'}[A]_0 \tag{9}$$

$$\ln\left(1 - \frac{[B]}{[B]_\infty}\right) = -k't \tag{10}$$

Approach to Equilibrium Through First-Order Reactions. This type of reaction can be represented by Eq. (11):

$$A \underset{k_2}{\overset{k_1}{\rightleftharpoons}} B \tag{11}$$

The concentrations of A and B as a function of time can be derived:

$$[A] = \frac{k_2}{k_1 + k_2}[A]_0 + \frac{k_1}{k_1 + k_2}[A]_0(e^{-(k_1+k_2)t}) \tag{12}$$

$$[B] = \frac{k_1}{k_1 + k_2}[A]_0(1 - e^{-(k_1+k_2)t}) \tag{13}$$

The combined constants $(k_1 + k_2)$ can be obtained through Eq. (14), and the individual rate constants k_1 and k_2 can now be calculated through Eq. (15):

$$\ln\,([A] - [A]_\infty) = \ln\,[B]_\infty - (k_1 + k_2)t \tag{14}$$

$$k_1[A]_\infty = k_2[B]_\infty \tag{15}$$

Fractional Order. In the decomposition of pure solids, the kinetics of reactions can often be more complex than simple zero- or first-order processes. Carstensen [88] has reviewed the stability of solids and solid dosage forms as well as the equations that can be used in these cases. In addition to zero- and first-order kinetics, solid-state degradations are often described by fractional-order equations.

More complicated reactions schemes, including first-order reversible consecutive processes and competitive consecutive reactions, are considered in a textbook by Irwin [89]. Professor Irwin's textbook also includes computer programs written in the BASIC language. These programs can be used to fit data to the models described.

B. Energetics of Reactions

According to the transition state theory, the reaction between two molecules, A and B, to form products C and D proceeds through a transition state, X:

$$A + B \overset{K^\ddagger}{\rightleftharpoons} X \rightarrow C + D \tag{16}$$

Here $K^\ddagger$ is a thermodynamic equilibrium constant that can be expressed as a function of the activities [Eq. (17)] or of the activity coefficients γ_X, γ_A, and γ_B [Eq. (18)]:

$$K^\ddagger = \frac{a_X}{a_A a_B} \tag{17}$$

$$K^\ddagger = \frac{[X]}{[A][B]}\frac{\gamma_X}{\gamma_A \gamma_B} \tag{18}$$

The rate of the reaction $-d[A]/dt$ is proportional to the concentration of the transition state

$$-\frac{d[A]}{dt} = k'[X] \tag{19}$$

where k' is a proportionality constant. Combining Eqs. (18) and (19) yields

$$-\frac{d[A]}{dt} = k'K^\ddagger[A][B]\frac{\gamma_A \gamma_B}{\gamma_X} \tag{20}$$

Copyright © 2002 Marcel Dekker, Inc.

If the activity coefficients are assumed to be unity, the specific rate constant k is then identical to $k'K^{\ddagger}$. It can be shown that

$$k' = \frac{k_B T}{h} \tag{21}$$

where k_B is Boltzmann's constant, h is Planck's constant, and T is the absolute temperature. Thus,

$$k = \frac{k_B T}{h} K^{\ddagger} \tag{22}$$

and

$$\ln k = \ln \frac{k_B}{h} + \ln T + \ln K^{\ddagger} \tag{23}$$

Differentiating relative to T, we obtain

$$\frac{d \ln K^{\ddagger}}{dT} = \frac{d \ln k}{dt} - \frac{1}{T} \tag{24}$$

Since

$$\frac{d \ln K^{\ddagger}}{dT} = \frac{\Delta H^{\ddagger}}{RT^2} \tag{25}$$

where $\Delta H^{\ddagger}$ is the enthalpy of activation, Eq. (26) can be obtained by combining Eqs. (24) and (25):

$$\frac{d \ln k}{dT} = \frac{\Delta H^{\ddagger} + RT}{RT^2} \tag{26}$$

The classic Arrhenius equation is given by Eq. (27), where E_a is the energy of activation:

$$\frac{d \ln k}{dT} = \frac{E_a}{RT^2} \tag{27}$$

On comparing Eqs. (26) and (27), it follows that

$$\Delta H^{\ddagger} = E_a - RT \tag{28}$$

The other thermodynamic parameters, $\Delta G^{\ddagger}$ and $\Delta S^{\ddagger}$, the free energy and entropy of activation, respectively, can also be obtained from the foregoing relationships:

$$\Delta G^{\ddagger} = -RT \ln K^{\ddagger} = -RT \ln \frac{kh}{k_B T} \tag{29}$$

and

$$\Delta S^{\ddagger} = -\frac{\Delta G^{\ddagger} + \Delta H^{\ddagger}}{T} = R \ln \frac{kh}{k_B T} + \frac{E_a - RT}{T} \tag{30}$$

The magnitudes of the thermodynamic parameters, $\Delta H^{\ddagger}$ and $\Delta S^{\ddagger}$, sometimes provide evidence supporting proposed mechanisms of drug decomposition. The enthalpy of activation is a measure of the energy barrier that must be overcome by the reacting molecules before a reaction can occur. As can be seen from Eq. (28), its numerical value is less than the Arrhenius energy of activation by the factor RT. At room temperature, RT is only about 0.6 kcal/mol. The entropy of activation can be related to the Arrhenius frequency factor (i.e., the fraction of molecules possessing the requisite energy that actually reacts). This parameter includes steric and orientation requirements of the reactants, the transition state, and the solvent molecules surrounding them. For unimolecular reactions, $\Delta S^{\ddagger}$ has a value of near zero or slightly positive. For bimolecular reactions, $\Delta S^{\ddagger}$ is more negative. For example, in the hydrolysis of esters and anhydrides, the entropy of activation is on the order of -20 to -50 entropy units, reflecting a transition state in which several solvent molecules are immobilized for solvation of the developing charges [90].

IV. THE ARRHENIUS EQUATION AND ACCELERATED STABILITY TESTING

The purpose of stability testing is to assess the effects of temperature, humidity, light, and other environmental factors on the quality of a drug substance or product. The data produced are used to establish storage conditions, retest periods, shelf loss, and to justify overages included in products for stability reasons. The most useful equation relating temperature and reaction rate is the Arrhenius equation. This equation (27) may be integrated and rewritten as Eqs. (31) and (32).

$$k = \mathrm{A} e^{-E_a/RT} \tag{31}$$

$$\ln \frac{k_1}{k_2} = \frac{E_a}{R} \left(\frac{1}{T_2} - \frac{1}{T_1} \right) \tag{32}$$

where E_a is a constant and the subscripts 1 and 2 denote the two different temperature conditions. A plot of $\ln k$ as a function of $1/T$, referred to as the *Arrhenius plot*, is linear according to Eq. (31) if E_a is independent of temperature. It is thus possible to conduct kinetic experiments at elevated temperatures and obtain estimates of rate constants at lower temperatures by extrapolation of the Arrhenius plot. This procedure, commonly referred to as accelerated stability testing, is most useful when the reaction at ambient temperatures is too slow to be monitored conveniently and when E_a is relatively high. For example, for a reaction with an E_a of 25 kcal/mol, an increase from 25 to 45°C brings about a 14-fold increase in the reaction rate constant. In comparison, a rate increase of just threefold is obtained for the same elevation in temperature when E_a is 10 kcal/mol. The magnitude of E_a for a reaction can be obtained from the slope of its Arrhenius plot.

Copyright © 2002 Marcel Dekker, Inc.

Hydrolysis reactions typically have an E_a of 10–30 kcal/mol, while oxidation and photolysis reactions have smaller energies of activation [91].

The elevated temperatures most commonly used are 40, 50, and 60°C in conjunction with ambient humidity. Occasionally, higher temperatures are used. The samples stored at the highest temperatures are examined weekly for physical and chemical changes. If a substantial change is seen, samples stored at lower temperatures are examined. If there is no change after 30 days at 60°C, the stability prognosis is excellent. Corroborative evidence must be obtained by monitoring the samples stored at lower temperatures for longer durations.

An underlying assumption of the Arrhenius equation is that the reaction mechanism does not change as a function of temperature (i.e., E_a is independent of temperature). Since accelerated stability testing of pharmaceutical products normally employs a narrow range of temperature (typically, 35°C to at most 70°C), it is often difficult to detect nonlinearity in the Arrhenius plot from experimental data, even though such nonlinearity is expected from the reaction mechanism [92]. Thus, even complex biological processes may show Arrhenius behavior within certain temperature ranges; Laidler [93] cited such phenomena as the frequency of flashing of fireflies and the rate of the terrapin's heartbeat as examples.

Non-Arrhenius behavior has been observed in pharmaceutical systems [94]. This may be attributed to the possible evaporation of solvent, multiple reaction pathways, change in physical form of the formulation, and so on [95] when the temperature of the reaction is changed. An interesting case of non-Arrhenius behavior is the increased rate of decomposition of ampicillin upon freezing. Savello and Shangraw [96] showed that for a 1% sodium ampicillin solution in 5% dextrose, the percentage of degradation at 4 hours is approximately 14% at -20°C, compared to 6% at 0°C and 10% at 5°C. This decrease in stability in frozen solutions is observed most frequently when the reaction obeys second- or higher-order kinetics. For example, the formation of nitrosomorpholine from morpholine and nitrite obeys third-order kinetics [97], and the rate of nitrosation is drastically enhanced in frozen solutions (Fig. 2). A marked acceleration in the hydrolytic degradation of methyl, ethyl, and *n*-propyl 4-hydroxybenzoates in the frozen state has also been reported by Shiva et al. [98]. These authors found that although pseudo-first-order conditions found in the liquid state are also observed in the frozen state, the rate of reaction under frozen state conditions showed very much less dependency on the initial hydroxide ion concentration.

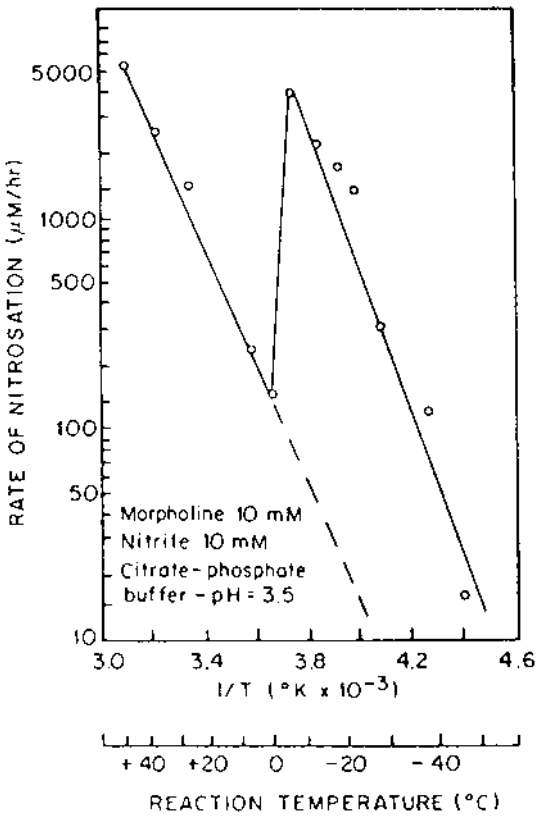


Fig. 2 Effect of temperature on the rate of nitrosation of morpholine with nitrite in citrate-sodium phosphate buffer for temperatures above and below the freezing temperature. (From Ref. 97.). Reproduced with the permission of the Copyright holder, the American Chemical Society, © 1975.

The mechanism for rate enhancement in frozen solutions has been reviewed by Pincock [99]. In reactions following second- or higher-order kinetics, an increase in rate may be brought about by concentration of the reactants in the liquid phase, the solute molecules being excluded from the ice lattice when the solution freezes. In some cases an increase in rate may be due to a change in pH upon freezing. Fan and Tannenbaum [97] reported that citrate–sodium hydroxide and citrate–potassium phosphate buffers do not change pH upon freezing, but citrate–sodium phosphate buffer at pH 8 decreases to pH 3.5 and sodium hydrogen phosphate at pH 9 decreases to pH 5.5 upon freezing. A possible explanation for this phenomenon is now available [100]. Monosodium phosphate forms supersaturated solutions on cooling that become amorphous, with no precipitation of the salt. The disodium and monopotassium salts, on the other hand, readily precipitated when the initial

Copyright © 2002 Marcel Dekker, Inc.

solution concentration was >0.2 M. The possibility of a pH change and rate acceleration should be considered when evaluating the stability of freeze-dried products. Proteins are particularly sensitive to changes in pH, folding or unfolding to varying degrees in response to changes in pH. Proteins tend to be most stable at their isoelectric point due to electrostatic interactions [101], but when a solution is adjusted to the optimum pH for stability at room temperature using buffers, that pH may not be maintained throughout the lyophilization cycle, and the protein may aggregate or undergo denaturation.

Nonlinearity in Arrhenius plots frequently is observed in following the temperature dependence of protein degradation. Degradation mechanisms in proteins often change with temperature [102]. At 60°C, the aggregation of interleukin 1β (1L-1β) in aqueous solution follows apparent first-order behaviour to 30% drug remaining, but at or below 55°C the aggregation deviates from apparent first-order and becomes biphasic (slow and fast) [103].

Considerable interest has been generated in the use of accelerated stability testing based on a single condition of elevated temperature and humidity. For Abbreviated New Drug Applications (ANDAs) the FDA stability guidelines [86,87] suggest that a tentative expiration date of 24 months may be granted for a drug product if satisfactory stability results can be documented under a stressed condition of 40°C *and* 75% relative humidity. The simplicity of such a guideline is naturally attractive because a substantial saving in time can be obtained in advancing a drug product to the marketplace [104].

V. ENVIRONMENTAL FACTORS THAT AFFECT REACTION RATE

A rational way to develop approaches that will increase the stability of fast-degrading drugs in pharmaceutical dosage forms is thorough study of the factors that can affect such stability. In this section, the factors that can affect decomposition rates are discussed; it will be seen that under certain conditions of pH, solvent, presence of additives, and so on, the stability of a drug may be drastically affected. Equations that may allow prediction of these effects on reaction rates are discussed.

A. pH

The pH of a drug solution may have a very dramatic effect on its stability. Depending on the reaction mechanism, a change of more than 10-fold in rate constant may result from a shift of just one pH unit. When drugs are formulated in solution, it is essential to construct a pH versus rate profile so that the optimum pH for stability can be located. Many pH versus rate profiles are documented in the literature, and they have a variety of shapes. The majority of these pH versus rate profiles can be rationalized using an approach in which the reaction of each molecular species of the drug with hydrogen ion, water, and with hydroxide ion is analyzed as a function of pH. The discussion that follows is divided according to the ionization capability of the drug.

1. When the drug is nonionizable in water, three hydrolytic pathways are available [Eq. (33)]: it can degrade by specific acid catalysis represented by the first kinetic term in Eq. (33), water hydrolysis (second term), and specific base catalysis (third term):

$$-\frac{dc}{dt} = k_1[H^+]c + k_2 c + k_3[OH^-]c \tag{33}$$

Equation (33) can be rearranged to give

$$-\frac{dc}{c\,dt} = k_{obs} = k_1[H^+] + k_2 + k_3[OH^-] \tag{34}$$

Note that k_1 and k_3 are second-order constants, whereas k_2 is a pseudo-first-order constant. The pH versus rate profile can be constructed by considering, in turn, that one of the three kinetic terms is predominating, thus:

(a) When $k_1[H^+] >> k_2 + k_3[OH^-]$,

$$k_{obs} = k_1[H^+] \text{ and } \log k_{obs} = \log k_1 - \text{pH} \tag{35}$$

(b) When $k_2 >> k_1[H^+] + k_3[OH^-]$,

$$k_{obs} = k_2 \text{ and } \log k_{obs} = \log k_2 \tag{36}$$

(c) When $k_3[OH^-] >> k_1[H^+] + k_2$,

$$k_{obs} = k_3[OH^-] \text{ and}$$

$$\log k_{obs} = \log k_3 + \text{pH} \tag{37}$$

Equations (35) through (37) are plotted and shown in Fig. 3a. The lines are stippled to indicate that the relative positions of the lines are not fixed, but are dependent on the relative magnitudes of the rate constants. For example, when $k_3[OH^-] > k_2 >> k_1[H^+]$, a $\log k_{obs}$ versus pH profile such as the one depicted in Fig. 3b may result. On the other hand, if $k_1[H^+]$ and $k_3[OH^-]$ are both much greater than k_2, a $\log k_{obs}$ versus pH profile may resemble the curve shown in Fig. 3c. When $k_1[H^+] > k_2 >> k_3[OH^-]$, the $\log k_{obs}$ versus pH profile

Copyright © 2002 Marcel Dekker, Inc.

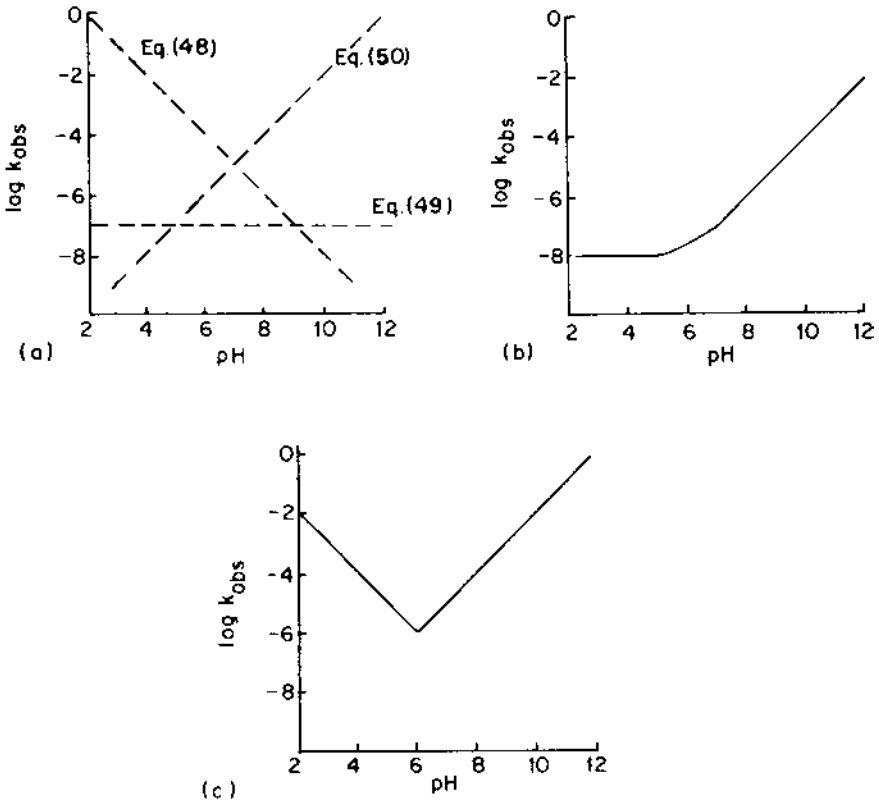


Fig. 3 The $\log k_{obs}$ versus pH profiles of nonionizable drugs.

will be a mirror image of Fig. 3b; and when $k_2 >> k_1[H^+] + k_3[OH^-]$, the rate constant will be pH-independent.

2. When the drug is either monoacidic or monobasic, an equation similar to Eq. (34) can be written. Here, however, three kinetic terms are written for the acidic form of the drug, HA, and three terms for the basic form, A (electronic charges on HA and A are not designated here because either HA or A can be charged):

$$k_{obs} = k_1[H^+]f_{HA} + k_2 f_{HA} + k_3[OH^-]f_{HA} + k_4[H^+]f_A + k_5 f_A + k_6[OH^-]f_A$$

where

$$f_{HA} = \frac{[HA]}{[HA] + [A]} = \frac{[H^+]}{[H^+] + K_a} \tag{39}$$

and

$$f_A = \frac{[A]}{[HA] + [A]} = \frac{K_a}{[H^+] + K_a} \tag{40}$$

Again, Eq. (38) can be analyzed by considering each individual term as a function of pH. Since the magnitudes of both f_{HA} and f_A are dependent on the relative magnitudes of K_a and H^+, the kinetic terms can be evaluated under three conditions: (a) when $[H^+] >> K_a$, (b) when $[H^+] = K_a$, and (c) when $[H^+] << K_a$ (Table 5). The log k_{obs} versus pH profile for each kinetic term is shown in Fig. 4, using a hypothetical pK_a of 6 and the condition that $k_1 = 10^7\, k_2 = k_3 = K_4 = 10^7\, k_5 = k_6 = 1$. Compared with the curves shown in Fig. 3a, the profiles in Fig. 4 show one break each in the lines, with a change of slope of 1 unit at the breaks. It is also seen that term (b) is equivalent to term (d) and that term (c) is equivalent to term (e), as far as their dependency on pH is concerned (see Table 5 and Fig. 4). These terms, therefore, are kinetically equivalent and are indistinguishable from each other in a rate expression. Equation (38), then, can be reduced to a combination of only four terms. The shape of the overall log k_{obs}/pH profile of any drug is determined by the relative magnitudes of the four kinetic terms over the pH range considered. Each $\log k_{obs}$ versus pH profile of a

Copyright © 2002 Marcel Dekker, Inc.

Table 5 Kinetic Expressions for Each Term in Eq. (38)

	$\log k_{obs}$		
Logarithm of kinetic term	When $[H^+] \gg K_a$	When $[H^+] = K_a$	When $K_a \gg [H^+]$
$\log k_1[H^+]f_{HA}$	$\log k_1 - pH$	$\log \frac{k_1 K_a}{2}$	$\log \frac{k_1}{K_a} - 2\ pH$
$\log k_2 f_{HA}$	$\log k_2$	$\log \frac{k_2}{2}$	$\log \frac{k_2}{K_a} - pH$
$\log k_3[OH^-]f_{HA}$	$\log k_3 K_w + pH$	$\log \frac{k_3 K_w}{2K_a}$	$\log \frac{k_3 K_w}{K_a}$
$\log k_4[H^+]f_A$	$\log k_4 K_a$	$\log \frac{k_4 K_a}{2}$	$\log k_4 - pH$
$\log k_5 f_A$	$\log k_5 K_a + pH$	$\log \frac{k_5}{2}$	$\log k_5$
$\log k_6[OH^-]f_A$	$\log k_6 K_w K_a + 2\ pH$	$\log \frac{k_6 K_w}{2K_a}$	$\log k_6 K_w + pH$

monoacidic or monobasic drug can be adequately described by a combination of no more than four terms. Figure 5 illustrates this principle by showing the log k_{obs} versus pH profiles of idoxuridine [105] and acetylsalicylic acid [106]. The hydrolysis of idoxuridine (see Fig. 5a) as a function of pH can be rationalized by the equation $k_{obs} = k_2 f_{HA} + k_5 f_A^- + k_6[OH^-]f_A^-$ (three kinetic terms), whereas the hydrolysis of acetylsalicylic

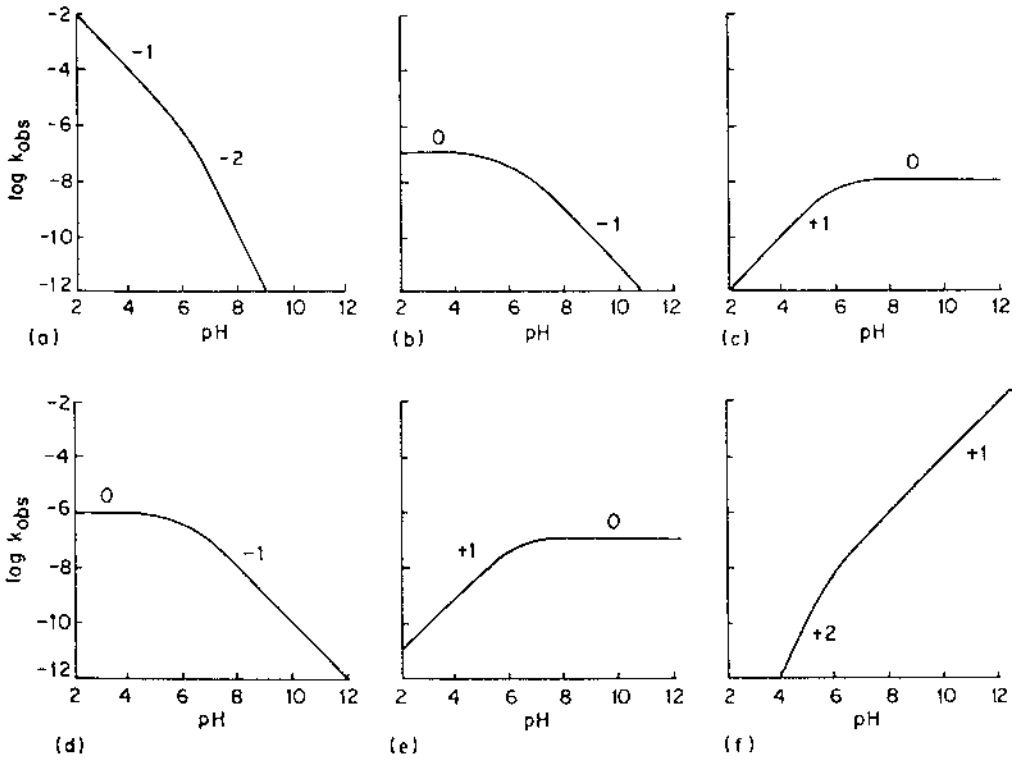


Fig. 4 The $\log k_{obs}$ versus pH profile for each kinetic term in Eq. (38): $k_1 = 10^7$, $k_2 = k_3 = k_4 = 10^7$, $k_5 = k_6 = 1$; $K_a = 10^{-6}$. Each number next to the curve indicates the slope of that portion of the curve.

Copyright © 2002 Marcel Dekker, Inc.

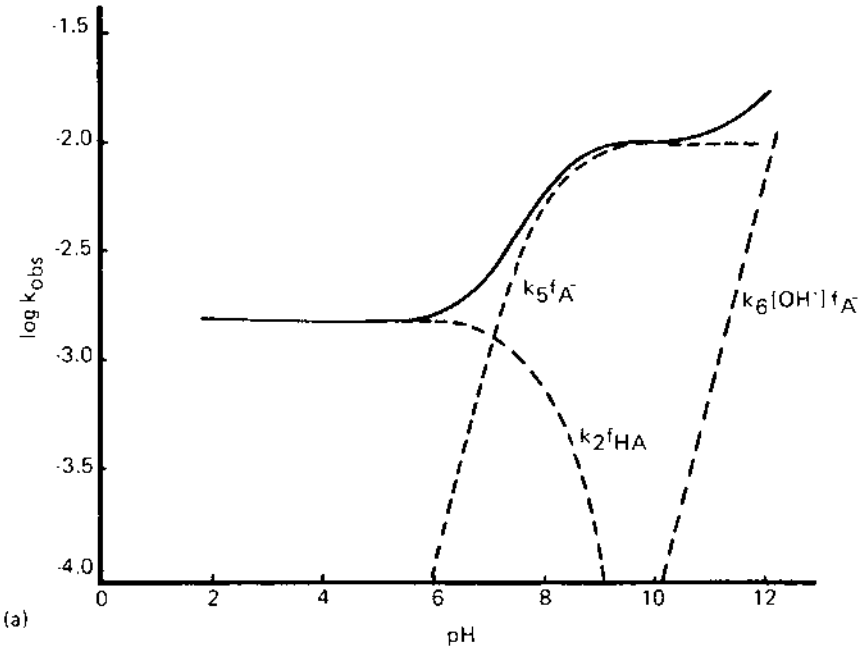


Fig. 5 (a) The pH versus log hydrolysis rate constant profile of idoxuridine at 60°C (solid line). The dashed lines indicate the individual contributions of each kinetic term. (b) pH versus log hydrolysis rate constant profile of acetylsalicylic acid (solid line). The dashed lines indicate the individual contribution of each kinetic term. Adapted from Refs. 105 and 106.

acid (see Fig. 5b) can be decided only by using all four kinetic terms ($k_{obs} = k_1[H^+]f_{HA} + k_4[H^+]f_{A^-} + k_5 f_{A^-} + k_6[OH^-]f_{A^-}$).

Some compounds exhibit pH behavior in which a bell-shaped curve is obtained with maximum instability at the peak [107]. The peak corresponds to the intersection of two sigmoidal curves that are mirror images. The two inflection points imply two acid and base dissociations responsible for the reaction. For a dibasic acid (H_2A) for which the monobasic species (HA^-) is most reactive, the rate will rise with pH as [HA^-] increases. The maximum rate occurs at $pH = (pK_1 + pK_2)/2$ (the mean of the two acid dissociation constants). Where an acid and base react, the two inflections arise from the two different molecules. The hydrolysis of penicillin G catalyzed by 3,6-bis(dimethylaminomethyl)catechol [108], is a typical example. For a systematic interpretation of pH-degradation profiles, see the review papers by van der Houwen et al. [109] and Connors [110].

B. Solvent

In many pharmaceutical dosage forms, it may be necessary to incorporate water-miscible solvents to solubilize the drug. These solvents are generally low molecular weight alcohols such as ethanol, propylene glycol, and glycerin or polymeric alcohols such as the polyethylene glycols. Solvent effects can be quite complicated and difficult to predict. In addition to altering the activity coefficients of the reactant molecules and the transition state, changes of the solvent system may bring about concomitant changes in such physicochemical parameters as pK_a, surface tension, and viscosity, thus indirectly affecting the reaction rate. In some cases, an additional reaction pathway may be generated, or there may be a change in the product mix.

The angiotensin-converting enzyme inhibitor, moexipril, undergoes hydrolysis as well as a cyclization reaction leading to the formation of diketopiperazines. In mixed solvent (75–90% ethanol) systems the hydrolysis reaction is suppressed, but the rate of the cyclization reaction increases by 5.5 to 29-fold [111]. In the presence of increasing concentrations of ethanol in the solvent, aspirin degrades *via* an extra route forming the ethyl ester of acetylsalicylic acid [112]. On the other hand, a solvent change may bring about stabilization of a compound. The hydrolysis of barbiturates occurs 6.7-fold faster in water than in 50% ethanol and 2.6-fold faster in water than in 50% glycerol [113].

Copyright © 2002 Marcel Dekker, Inc.

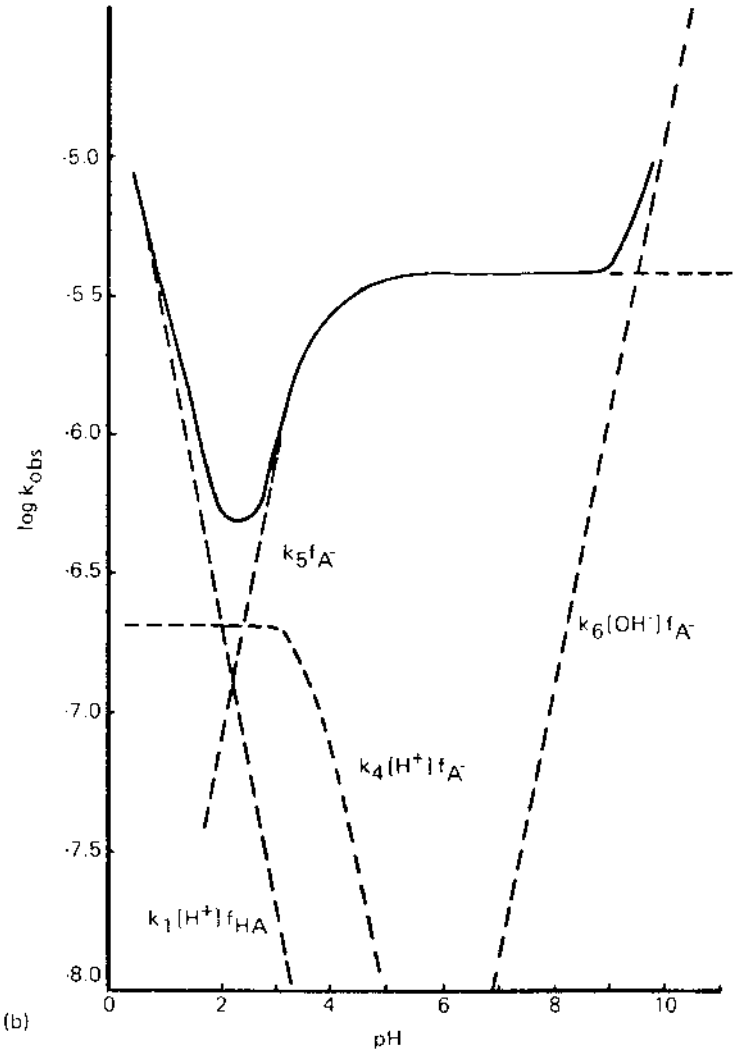


Fig. 5 (b)

Many approaches have been used to correlate solvent effects. The approach used most often is based on the electrostatic theory, the theoretical development of which has been described in detail by Amis [114]. The reaction rate is correlated with some bulk parameter of the solvent, such as the dielectric constant or its various algebraic functions. The search for empirical parameters of solvent polarity and their applications in multiparameter equations has recently been intensified, and this approach is described in the book by Reichardt [115] and more recently in the chapter on medium effects in Connor's text on chemical kinetics [110].

Although the solvent effect on reaction rate could, in principle, be large, the limited availability of nontoxic solvents suitable for pharmaceutical products has rendered this stabilization approach somewhat impractical in most circumstances.

C. Solubility

As mentioned earlier in this chapter, penicillins are very unstable in aqueous solution by virtue of hydrolysis of the β-lactam ring. A successful method of stabilizing penicillins in liquid dosage forms is to prepare their insoluble salts and formulate them in suspensions. The reduced solubility of the drug in a suspension decreases the amount of drug available for hydrolysis. An example of improved stability of a

Copyright © 2002 Marcel Dekker, Inc.

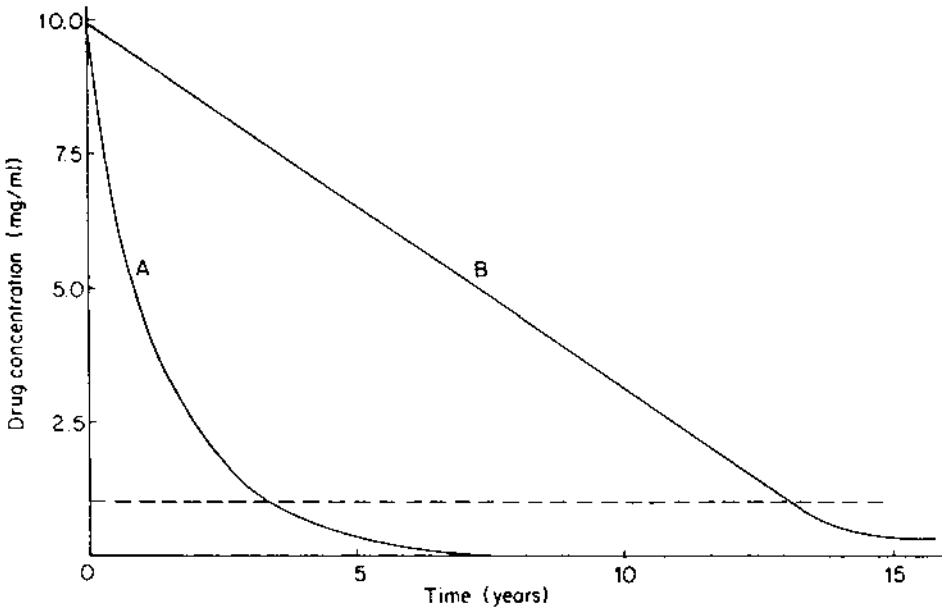


Fig. 6 Solubility effects on drug stability: curve A, drug formulated as 10 mg/mL solution ($t_{1/2} = 1$ year); curve B, drug formulated as a suspension with a saturated solubility of 1 mg/mL ($t_{1/2} = 7.3$ years).

suspension over that of a solution is illustrated in Fig. 6, in which a hypothetical drug is formulated as a 10 mg/mL solution (curve A) and as a suspension containing the same total amount of drug but with a saturated solubility of 1 mg/mL (curve B). It is seen that the drug in solution undergoes first-order degradation with a half-life of 1 month. In the suspension, the drug degrades through zero-order kinetics until there is no more excess solid present, after which point first order kinetics is operative.

D. Additives

Buffer Salts

In many drug solutions, it is necessary to use buffer salts in order to maintain the formulation at the optimum pH. These buffer salts can affect the rate of drug degradation in a number of ways. First, a primary salt effect results because of the effect salts have on the activity coefficient of the reactants. At relatively low ionic strengths, the rate constant, k_μ, is related to the ionic strength, μ, according to

$$\ln k_\mu = \ln k_0 + 1.02\, z_A z_B \sqrt{\mu} \qquad (41)$$

where k_0 is rate constant at $\mu = 0$, z_A and z_B are the ionic charges of reactants A and B, respectively, and the constant 1.02 is applicable to aqueous solutions at 25°C. According to Eq. (41), a plot of ln k_μ as a function of $\sqrt{\mu}$ should yield a slope approximately equal to the product $z_A z_B$, which in theory should always be an integer because both z_A and z_B are integral numbers. In practice, however, the slope is often fractional. The sign of the slope is sometimes informative in identifying the reactants that participate in the rate-limiting step in the reaction mechanism. But one must avoid drawing such conclusions in instances where a choice is to be made between kinetically equivalent rate terms [110].

Buffer salts also can exert a secondary salt effect on drug stability. From Table 5 and Fig. 5 it is clear that the rate constant for an ionizable drug is dependent on its pK_a. Increasing salt concentrations, particularly from polyelectrolytes such as citrate and phosphate, can substantially affect the magnitude of the pK_a, causing a change in the rate constant. (For a review of salt effects, containing many examples from the pharmaceutical literature see Ref. 116.)

Lastly, buffer salts can promote drug degradation through general acid or general base catalysis. In these cases, the rate expression will contain additional kinetic terms describing the applicable reactions between different molecular species of the drug and buffer components. The efficiency of general acid or base catalysis by the buffer components is often described by the Brønsted relationship:

Copyright © 2002 Marcel Dekker, Inc.

$$k_A = G_A K_A^{\alpha}$$

and

$$k_B = G_B K_B^{\beta} \tag{42}$$

where k_A and k_B are the catalytic constants for general acid and base catalysis, respectively, K_A and K_B are the acid and base ionization constants, respectively, and G_A, G_B, α, and β are constants characteristic of the reaction, the solvent, and the temperature [117, 118].

Surfactants

Addition of surface-active agents may accelerate or decelerate drug degradation. Because micellar catalysis may provide a model for enzyme reactions, acceleration of rate due to the presence of surfactants is well documented [119]. Millionfold accelerations of reaction rates have been reported [120]. Stabilization of drugs through the addition of surfactants, is, by comparison, less frequently reported. An example in which both effects are observed is especially rare. The hydrolysis of aspirin in the plateau region (pH 6–8) is inhibited by the presence of micelles of cetyltrimethylammonium bromide and cetylpridinium chloride, while in the region where the normal base-catalyzed reaction occurs (pH > 9), the reaction is catalyzed by micelles of these same surfactants. The mechanism of hydrolysis in the plateau region has been shown to involve intramolecular general base catalysis by the adjacent ionized carboxyl group both in the presence and absence of micelles. This reaction is inhibited in the presence of micelles because the substrate molecules are solubilized into the micelle and water is less available in this environment than in normal aqueous solutions [121].

Complexing Agents

Higuchi and Lachman [122] pioneered the approach of improving drug stability by complexation. They showed that aromatic esters can be stabilized in aqueous solutions in the presence of xanthines such as caffeine. Thus, the half-lives of benzocaine, procaine hydrochloride, and tetracaine are increased by approximately two- to fivefold in the presence of 2.5% caffeine. This increase in stability is attributed to the formation of a less reactive complex between caffeine and the aromatic ester. Professor K. A. Connors has written a comprehensive textbook that describes methods for the measurement of binding constants for complex formation in solution—along with discussions of pertinent thermodynamics, modeling statistics, and regression analysis [123]. The various experimental methods useful for measuring equilibrium constants are also discussed.

A good deal of attention recently has been directed towards the use of derivatives of cyclodextrin for the solubilization and stabilization of pharmaceuticals [124–126]. One cautionary note—complexation may adversely affect the dissolution an/or permeability characteristics of the drug, thereby possibly decreasing drug bioavailability.

Antioxidants and Chelating Agents

Antioxidants and chelating agents are used to protect drugs against autoxidation. Mechanistically, some antioxidants such as ascorbic acid, ascorbyl palmitate, sodium bisulfite, sodium metabisulfite, sodium sulfite, acetone sodium bisulfite, sodium formaldehyde sulfozylate, thioglycerol, and thioglycolic acid act as reducing agents. They are easily oxidized, preferentially undergo autoxidation, thus consuming oxygen and protecting the drug or excipient. They are often called oxygen scavengers because their autoxidation reaction consumes oxygen. They are particularly useful in closed systems where the oxygen cannot be replaced once it is consumed [23]. Primary or true antioxidants act by providing electrons or labile H^+, which will be accepted by any free radical to terminate the chain reaction. In pharmaceuticals, the most commonly used primary antioxidants are butylated hydroxytoluene (BHT), butylated hydroxyanisole (BHA), the tocopherols (vitamin E) and propyl gallate. Chelating agents act by forming complexes with the heavy metal ions that often are required to initiate oxidation reactions. The chelating agents used most often are ethylenediamine tetraacetic acid derivatives and salts, citric acid, and tartaric acid.

E. Light and Humidity

The mathematical relationship between light intensity and drug degradation is much less developed than those describing pH and temperature effects. Part of the reason, perhaps, is that light effects on stability can be substantially avoided by using amber containers, which shield off most of the ultraviolet light. Another deterrent to careful study of the kinetics of these reactions is the fact that multiple products of reaction often are obtained. In the case of nifedipine, for example, six degradation products were isolated from tablets exposed for 30 days under exposure to normal room light [127]. Regulatory authorities usually require a statement on the photostability of products and

Copyright © 2002 Marcel Dekker, Inc.

the means of protection, if required. Often both daylight and artificial light sources are employed for tests on drug substances [128].

Humidity is a major determinant of drug product stability in solid dosage forms. Elevation of relative humidity usually decreases stability, particularly for those drugs highly sensitive to hydrolysis [129]. In addition, increased humidity also can accelerate the "aging" process [49,71,72]. Humidity does not always affect drug stability adversely. Cyclophosphamide in lyophilized cakes containing mannitol or sodium bicarbonate was found to undergo rapid ($t_{90} \approx 15$ days) degradation in the solid state. The cyclophosphamide was found to be in the amorphous state in these formulations. However, on exposure to high humidity the cyclophosphamide was converted to the crystalline monohydrate form, which exhibited greatly improved stability [63]. Reviews dealing with the effects of moisture on the physical and chemical stability of drugs are available [130,131]. As peptides and proteins have become more important as therapeutic agents, the role residual moisture plays in their stabilization has attracted a good deal of attention [132, 133].

VI. STABILITY TESTING IN THE PHARMACEUTICAL INDUSTRY

Stability testing of drug substances and drug products begins as part of the drug discovery/synthesis development/preformulation effort and ends only with the demise of the compound or commercial product. Activities include testing of drug substance, of compatibility with excipients, of preclinical formulations, of phase I formulations and modifications, of the final NDA (commercial) formulation, and of postapproval formulation changes. The regulatory basis for the various aspects of stability testing is established in 21 CFR 211.137, 211.160, 211.170, 211.190, 314.50, 314,70, and 314.81 [134–140]. In addition, FDA guidelines for submitting stability data were published several years ago [86]. More recently, FDA has issued a draft guidance titled "Stability Testing of Drug Substances and Drug Products" [87], which covers all aspects of drug substance and drug product stability testing.

While the Code of Federal Regulations establishes the statutory basis for carrying out stability testing, the FDA and the Expert Working Group of the International Conference on Harmonization (ICH) of Technical Requirements for the Registration of Pharmaceuticals for Human Use have published guidances for conducting the actual studies [86,87, 141,142]. These guidances provide definitions of key terms and principles used in the stability testing of drug substances and drug products. Certain aspects of these guidances will be discussed later in this chapter.

Another valuable source of information on the establishment and maintenance of stability testing programs can be found in Carstensen and Rhodes [81].

A. Resources

Personnel

The number and types of personnel are dictated by the size of the program, the functions contained within the program, and the nature of the program. Some companies maintain separate development and commercial product programs; other integrate the two. It is important, both for operating efficiency and regulatory compliance, that all personnel, regardless of their function, receive adequate and well-documented training in both cGMPs and in the technical aspects of their jobs.

Education and Experience. The program is generally headed by a professional with several years of experience in the company. Experience in some aspect of formulation development or in the stability testing program itself may be as important as the education level of the person heading the program. People with a bachelor's degree in pharmacy, chemistry, or related science as well as people with masters or doctoral level degrees have led successful programs. If the size of the program warrants, intermediate-level scientists, usually at the bachelor's level, may assume responsibility for specific functions within the program, *e.g.*, chemical and physical testing, documentation, etc. Technicians generally have a high-school education or equivalent with clerical and /or scientific experience or interest. Clerical and data entry personnel have traditional training and experience in their respective areas.

Organization. Functions are often divided into testing (chemical, physical and biological), documentation, and clerical/computer operations, if the information system is computerized. Successful manual or paper systems are possible without the aid of computers. However, custom-designed software or commercially available database programs can also be programmed to automate the program. The documentation function usually consists of one or more persons who prepare the stability sections or regulatory documents. Persons specifically trained in technical writing or scientists with an interest and talent in document preparation generally perform well in this capacity.

Copyright © 2002 Marcel Dekker, Inc.

Facilities

Storage Chambers. Several chambers capable of accurately maintaining different temperatures and combinations of temperature and humidity are needed. For most drug products, these include, as outlined in recent FDA [87] and ICH [142] guidances, $-15°C \pm 5°C$, $5°C \pm 3°C$/ambient humidity, $25°C \pm 2°C/60\%RH \pm 5\%$, $30°C \pm 2°C/60\%\ RH \pm 5\%$, $40°C \pm °C/75\%RH \pm 5\%$. For liquid products stored in semi-permeable containers subject to water loss, exposure to lower humidities, *e.g.*, $25° \pm 2°C/40\%RH \pm 5\%$, $30°C \pm 2°C/40\%RH \pm 5\%$, $40°C \pm 2°C/15\%RH \pm 5\%$, is needed. Also, a high-intensity light cabinet and a cycling chamber capable of cycling both temperature and humidity are needed. These chambers should be calibrated periodically according to a standard operating procedure, and records of these calibrations should be maintained in a logbook for each chamber.

For drug products normally stored at room temperature, long-term stability testing should be done at 25°C/60%RH or 25°C/40%RH. Stress testing should be done at 40°C/75%RH or 40°C/15%RH for 6 months. If "significant change" occurs at these stress conditions, then the product should be tested at an intermediate condition, *i.e.*, 30°C/75%RH or 30°C/40%RH. "Significant change" is defined in the guidelines.

Storage chambers should be validated with respect to their ability to maintain the desired conditions, and, if so equipped, the ability to sound an alarm if a mechanical or electrical failure causes the temperature to deviate from preestabilished limits. They should also be equipped with recording devices, which will provide a continuous and permanent history of their operation. Logbooks should be maintained and frequent readings or mercury-in-glass, National Institute of Science and Technology–traceable thermometers recorded.

Bench Space. Adequate laboratory bench, desk, and file space are needed for physical, chemical, and microbiological testing, for documentation, and for storing records, respectively.

Equipment

Chemical Testing. Adequate instrumentation for a variety of different test methods should be available. Most stability-indicating chemical assays are performed by high-performance liquid chromatography. Occasionally, gas chromatography, infrared spectrophotometry, or spectrofluorimetry are used. Test methods should be validated [143–146] and stability indicating, *i.e.*, able to distinguish the active ingredient from its degradation products so that the active can be accurately measured. Also, methods are needed for identifying and quantitating degradation products present at levels of 0.1% or greater.

Biological Testing. A portion of the laboratory may be reserved for biological testing, or this work can be done by the company's microbiology laboratory. The ability to perform sterility, pyrogen, LAL, preservative challenge, and bioburden tests is needed to support the stability program. As is the case for chemical assays, test methods should be validated and operator familiarity should be documented.

Physical Testing. Equipment and trained personnel should be available for performing such tests as pH, tablet hardness, etc. One important and sometimes overlooked aspect of physical testing is the recording of product appearance. Carefully defined descriptions of appearance and standard descriptions of changes in appearance should be developed, especially when there is a high probability that the person who made the observation at the previous sampling time will not be the person making the observation at the next sampling time. Some companies maintain samples at a lower-than-label storage condition, *e.g.*, refrigeration, to use as standards, assuming that minimal or no appearance change will occur at this condition. The same argument for standard nomenclature applies to other test parameters which are subjective in nature.

Computers. A certain number of personal computers are necessary for report generation and regulatory submission preparation. In addition, these may be useful for record keeping, depending on the type of stability information system that the company chooses to use. Alternatively, if the information system is intended to be accessible (read only) to many users, it may be more efficient to develop a local area network of mini-computers. The size of the database will help determine the nature of the software/hardware configuration used for this function.

B. Program

Scope and Goals

Activities encompassed by the stability program include sample storage of either development or production batches (or both), data collection and storage/retrieval, physical, chemical, and microbiological testing, document preparation of regulatory

Copyright © 2002 Marcel Dekker, Inc.

submissions, and package evaluation. In certain companies, personnel in separate departments may perform some of these functions, *e.g.*, regulatory document preparation. Nonetheless, the function is part of the company's overall stability program.

Protocols

FDA stability guidelines [87] and ICH guidelines [142] are rather detailed regarding sampling times, storage conditions, and specific test parameters for each dosage form. Generally, samples stored at the product label storage condition—controlled room temperature for most products—are tested initially and after 3, 6, 9, 12, 18, and 24 months and annually thereafter. Accelerated testing is generally done more frequently and for a shorter duration, *e.g.*, at 1, 2, 3, and 6 months. Three batches should be tested to demonstrate batch-to-batch uniformity. The number three represents a compromise between a large number desired for statistical precision and the economics of maintaining a manageable program. Generally real-time data obtained at the label storage conditions on the final formulation in the final packaging configuration(s) are needed for an NDA. Supportive data obtained from drug substance stability studies, preformulation studies, and investigational formulations tested during clinical trials and formulation development may be used to supplement primary stability data. Requirements for the IND are less defined, the only requirement being that there should be adequate data to support the clinical batch(es) for the duration of the trials.

There are instances, especially in the case of solid, oral dosage forms, where several package types and configurations are desired by marketing and three or more strengths are needed for flexibility in dosing. In these situations, it may be feasible to apply the principles of bracketing and/or matrixing to reduce the amount of testing. Bracketing refers to reduced testing of either an intermediate dosage strength or package size when the formulation characteristics of all strengths are virtually identical or when the same container/closure materials are used for all package sizes. Matrixing refers to reduced testing regardless of strength or container in situations where there are similarities in formulation or container/closure. Bracketing and matrixing are acceptable only when the product is chemically and physically very stable and does not interact with the container/closure. Demonstration of this chemical and physical stability must be documented by preformulation, drug substance stability, and early formulation stability data. Although not as common, it may be possible to utilize bracketing and matrixing with other types of dosage forms. In all cases, discussions of such strategies with FDA prior to implementation are imperative.

Documentation

The need for adequate documentation of laboratory operations is established not only by good science but also by regulatory requirements [137].

Documentation of all facets of the operation is necessary. This includes validation and periodic calibration of storage chambers, instrumentation, and computer programs. Logbooks for the storage chambers and instruments are also necessary. Standard operating procedures are needed for, among other things, the stability program itself, use of instrumentation, documentation of experiments and their results, determination of expiration dates, investigation of specification failures, and operation of a computerized record-keeping system.

Many companies have developed or purchased computer software for the purpose of storing stability data for a large number of studies. Examples of commercially available systems are "SLIM" [147] and "Stability System" [148]. These systems can perform other functions as well, including work scheduling, preparation of summaries of selected or all studies in the system, tabulation of data for individual studies, label printing, statistical analysis and plotting, and search capabilities. Such systems should be validated to keep pace with current regulatory activity [149].

C. Regulatory Concerns

cGMP Compliance

Current Good Manufacturing Practices [135] establish the requirements for maintaining a stability program and require that most pharmaceutical dosage forms have an established expiration date supported by test data [134]. There are few allowable exceptions.

FDA Stability Guidelines

The guidelines under which stability programs operate and corresponding documentation is prepared were issued in 1987 [86]. A draft revision was issued in 1998 [87]. Although the agency emphasizes that these are guidelines and not regulations, it is generally prudent to follow specific recommendations as indicated in the guidelines. Deviations or omissions should be addressed, and the reasons should be supported with data where applicable.

Copyright © 2002 Marcel Dekker, Inc.

Regulatory Submissions

An easy-to-read stability summary document will go a long way toward rapid approval of any regulatory submission. Such a document should include a number of items. A clear statement of the objective(s) of the studies included in the submission and the approach that was taken to achieve the objective(s) is critical. This statement of objective(s) should accompany basic information including product and drug substance names, dosage forms and strengths, and type(s) and container/closure systems. Although the objective is usually stated in the summary letter accompanying the submission, a brief reminder to the reviewing chemist is helpful.

A discussion of each of the parameters tested in the course of the evaluation, including test methods and specifications for each, should then follow. These parameters should follow those recommended in the stability guidelines [87] for the specific dosage form. It is especially important to provide a rationale for those parameters not studied. Next should come the study design itself, which should include a list of batch identification number, size, and date of manufacture as well as packaging configuration, storage conditions, and sampling times for each batch. The strategy and rationale for any bracketing or matrixing should also be presented.

The actual data, including replicates, mean, and range, in tabular form should follow, accompanied by a brief discussion of the data. It is important to explain any out-of-specification data. Statistical analyses for all parameters, which lend themselves to such analyses along with conclusions, should be incorporated into the document at this point. These statistical analyses should be accompanied by the results of experiments conducted to determine the "poolability" of batches, or commonality of slopes and intercepts of individual batches. Graphs of these data should be included as part of the documentation.

Protocols for these batches and a commitment to continue them along with a "tentative" expiry date should also be included. Approval of these protocols will allow extension of the expiry date without a special supplement as long as the data remain within specifications. These data will ultimately be reported to FDA as part of periodic reports following NDA approval. Protocols intended for use on commercial batches should also be submitted.

Finally, the three-part commitment to mount studies for the first three production batches and a statistically determined number (at least one) each year, to update current studies in annual reports, and to withdraw any lots not meeting specifications should appear in the submission. Statistical sampling of production batches is usually based on $\log N$, $\sqrt{N}$, etc., where N is the number of batches produced per year. These batches are generally spread over various package types and manufacturing campaigns. There should be a standard operating procedure to handle specification deviations including confirmation of the result, cause-and-effect investigation, impact analysis, final report to management, and field alert or batch recall notice to FDA.

Annual Product Review

Once a product gains FDA approval for marketing, the sponsor should maintain a readily retrievable profile of commercial batches. This includes individual batch release data and stability data. These data should be compiled throughout the year and tabulated prior to the anniversary of NDA approval for submission in the annual product report to FDA. By maintaining an ongoing database, which is reviewed as new information is added, changing trends in the data can be observed and management notified if any of these trends are unfavorable.

REFERENCES

1. T. L. Paal and E. Liptak-Csekey, Acta Pharm. Jugosl. 40, 199 (1990).
2. E. H. Dearborn, in *The Dating of Pharamaceuticals* (J. J. Windheuser and W. L. Blockstein, Eds.), University Extension, University of Wisconsin, Madison, WI, 1970, p. 29.
3. K. Nord, J. Karlsen, and H. H. Tønnesen, Int. J. Pharm., 72, 11 (1991).
4. P. J. G. Cornelissen, G. M. J. Beijersbergen van Henegouwen and K. W. Gerritsma, Int. J. Pharm, 1, 173 (1978).
5. K. A. Neftel, M. Walti, H. Spengler, and A. L. deWeck, Lancet, 986 (1982).
6. V. J. Stella, T. J. Mikkelson, and J. D. Pipkin, in *Drug Delivery Systems* (R. L. Juliano, Ed.), Oxford University Press, New York, 1980.
7. M. D. Taylor, Adv. Drug Delivery Rev., 19, 131 (1996).
8. K. A. Connors, G. L. Amidon, and V. J. Stella, *Chemical Stability of Pharmaceuticals: A Handbook for Pharmacists*, 2nd ed., Wiley, New York, 1986.
9. H. G. Brittain, Ed., *Analytical Profiles of Drug Substances and Excipients*, Vol. 27, Academic Press, Inc., San Diego, 2000.
10. F. Pellerin, D. Baylocq and N. Chanon, Rev. Anal. Chem. 11, 171 (1992).

Copyright © 2002 Marcel Dekker, Inc.

11. K. A. Connors, *Reaction Mechanisms in Organic Analytical Chemistry*, Wiley Interscience, New York, 1970.
12. M. A. Schwartz, A. P. Granatek, and F. H. Buckwalter, J. Pharm. Sci., 51, 523 (1962).
13. V. K. Krieble and K. A. Holst, J. Am. Chem. Soc., 60, 2976 (1938).
14. L. A. Casey, R. Galt, and M. I. Page, J. Chem. Soc. Perkin Trans. 2, 23 (1993).
15. R. S. Brown, A. J. Bennet, and H. Slebocka-Tilk, Acc. Chem. Res., 25, 481 (1992).
16. H. B. Mark, Jr., and G. A. Rechnitz, in *Chemical Analysis*, Vol. 24 (P. J. Elving and I. M. Kolthoff, Eds.), Wiley-Interscience, New York, 1970.
17. C. Hansch, A. Leo, and R. W. Taft, Chem. Rev., 91, 165 (1991).
18. L. P. Hammett, *Physical Organic Chemistry*, 2nd ed., McGraw-Hill, New York, 1970.
19. B. Capon and S. P. McManus, *Neighboring Group Participation*, Plenum, New York, 1976.
20. M. N. Khan and S. K. Gambo, Int. J. Chem. Kinetics, 17, 419 (1985).
21. R. T. Darrington and B. D. Anderson, Pharm. Res. 11, 784 (1994).
22. R. E. Lindstrom, S. N. Patel, and P. K. Wilkerson, J. Parenter. Drug Assoc., 34, 5 (1980).
23. D. M. Johnson and L. C. Gu, Autooxidation and antioxidants, in *Encyclopedia of Pharmaceutical Technology*, J. Swarbrick and J. C. Boylan, Eds., Marcel Dekker, New York, 1988, pp. 415–449.
24. T. M. Chen and L. Chafetz, J. Pharm. Sci., 76, 703 (1987).
25. P. Finholt, H. Kristiansen, L. Kyowezynski, and T. Higuchi, J. Pharm. Sci., 55, 1435 (1966).
26. J. Schulz and K.-H. Bauer, Acta Pharm. Technol., 32, 78 (1986).
27. R. Hamburger, E. Azaz, and M. Donbrow, Pharm. Acta Helv., 50, 10 (1975).
28. J. W. McGinity, J. A. Hill, A. L. La Via, J. Pharm. Sci., 64, 356 (1975).
29. J. W. McGinity, T. R. Patel, A. H. Naqvi, and J. A. Hill, Drug Dev. Comm., 2, 505 (1976).
30. E. J. Ginsburg, D. A. Stephens, P. R. West, A. M. Buko, D. H. Robinson, L. C. Li, and A. R. Bommireddi, J. Pharm. Sci., 89, 776 (2000).
31. M. Bergh, K. Magnusson, J. L. G. Nilsson, and A. T. Karlberg, Contact Dermatitis, 39, 14 (1998).
32. M. Nishikawa and K. Fuji, Chem. Pharm. Bull., 39, 2408 (1991).
33. R. Frontini and J. B. Mielck, Int. J. Pharm., 114, 121 (1995).
34. D. Pogocki and C. Schöneich, J. Pharm. Sci., 89, 443 (2000).
35. M. J. Frank, J. B. Johnson, and S. H. Rubin, J. Pharm. Sci., 65, 44 (1976).
36. K. Thoma and R. Kerker, Pharm. Ind., 54, 551 (1992).
37. J. A. Waters, Y. Kondo, and B. Witkop, J. Pharm. Sci., 61, 321 (1972).
38. F. Vargas, H. Mendez, J. Sequera, J. Rojas, G. Fraile, and M. Velasquez, Tox. Subst. Mech., 18, 53 (1999).
39. Fed. Reg. 62(95), 27115–27122 (1997).
40. Pharmacopeial Forum, 26, 384 (2000).
41. E. Shefter, H.-L. Fung, O. Mok, J. Pharm. Sci., 62, 791 (1973).
42. E. Shefter and T, Higuchi, J. Pharm. Sci., 52, 781 (1963).
43. K. R. Morris, Structural aspects of hydrates and solvates, in *Polymorphism in Pharmaceutical Solids* (H. G. Brittain, Ed.), Marcel Dekker, New York, 1999, p. 132.
44. F. Jamali, R. Mehon, F. M. Pasutto, J. Pharm. Sci., 78, 695 (1989).
45. M. A. Nunes and E. Brochmann-Hanssen, J. Pharm. Sci., 63, 716 (1974).
46. G. Severin, Chirality, 4, 111 (1992).
47. M. Reist, B. Testa, and P.-A. Carrupt, Enantiomer, 2, 147 (1997).
48. L. J. Riff and G. G. Jackson, Arch. Intern. Med., 130, 887 (1972).
49. A. T. M. Serajuddin, A. B. Thakur, R. N. Ghoshal, M. G. Fakes, S. A. Ranadive, K. R. Morris, and S. A. Varia, J. Pharm. Sci., 88, 696 (1999).
50. R. D. Vickery and M. B. Maurin, J. Pharm. Biomed. Anal., 20, 385 (1999).
51. D. D. Wirth, S. W. Baertschi, R. A. Johnson, S. R. Maple, M. S. Miller, D. K. Hallenbeck and S. M. Gregg, J. Pharm. Sci., 87, 31 (1998).
52. E. Tarelli, P. H. Corran, B. R. Bingham, H. Mollison, and R. Wait, J. Pharm. Biomed. Anal. 12, 1355 (1994).
53. R. Ikan, Ed., *The Maillard Reaction: Consequences for the Chemical and Life Sciences*, John Wiley & Sons, New York, 1996.
54. M. D. Santos-Buelga, M. J. Sanchez-Martin, and M. Sanches-Camazano, Thermochim. Acta, 210, 255 (1992).
55. J. Haleblian and W. McCrone, J. Pharm. Sci., 58, 911 (1969).
56. M. Kuhnert-Brandstätter, *Thermomicroscopy in the Analysis of Pharmaceuticals*, Pergamon Press, Oxford, 1971, pp. 37–42.
57. S. S. Yang and J. K. Guillory, J. Pharm. Sci., 61, 26 (1972).
58. J. Haleblian, R. T. Koda, and J. A. Biles, J. Pharm. Sci., 60, 1485 (1971).
59. H. G. Brittain, Ed., *Polymorphism in Pharmaceutical Solids*, Marcel Dekker, Inc., New York, 1999.
60. T. J. Macek, U.S. Patent, 2,671,750, March 9, 1954.
61. S. Luthala, Acta Pharm. Nordica, 4, 271 (1992).
62. S. R. Byrn, R. R. Pfeiffer and J. G. Stowell, *Solid-State Chemistry of Drugs*, 2nd ed., SSCI, Inc., West Lafayette, IN, 1999, p. 256.
63. T. R. Kovalcik and J. K. Guillory, J. Parenter. Sci. Tech., 42, 29 (1998).

Copyright © 2002 Marcel Dekker, Inc.

64. M. J. Pikal and D. R. Rigsbee, Pharm. Res., 14, 1379 (1997).
65. S. Budavari, Ed., *The Merck Index*, 11th ed., Merck & Co., Rahway, NJ, 1989.
66. Fed. Reg., 37, 15959 (1972).
67. H.-L. Fung, S. K. Yap, and C. T. Rhodes, J. Pharm. Sci., 63, 1810 (1974).
68. S. A. Fusari, J. Pharm. Sci., 62, 2021 (1973).
69. Z. Chowhan, Pharm. Tech., 6 (9) 47 (1982).
70. C. J. deBlaey and J. J. Rutten-Kingma, Pharm. Acta Helv., 51, 186 (1976).
71. S. T. Horhota, J. Burgio, L. Lonski, and C. T. Rhodes, J. Pharm. Sci., 65, 1746 (1976).
72. P. B. Chemburkar, R. D. Smyth, J. D. Buehler, P. B. Shah, R. S. Joslin, A. Polk, and N. H. Reavey-Cantwell, J. Pharm. Sci., 65, 529 (1976).
73. B. L. McNiff, E. F. McNiff, and H.-L. Fung, Am. J. Hosp. Pharm., 36, 173 (1979).
74. D. Darbar, S. DellOrto, G. R. Wilkinson, and D. M. Roden, Am. J. Health-Syst. Pharm., 53, 655 (1996).
75. W. A. Parker, M. E. Morris, and C. A. Shearer, Am. J. Hosp. Pharm., 36, 505 (1979).
76. J. I. Hirsch, J. H. Wood, and R. B. Thomas, Am. J. Hosp. Pharm., 38, 995 (1981).
77. P. A. Cossum and M. S. Roberts, Eur. J. Clin. Pharmacol., 19, 181 (1981).
78. E. A. Kowaluk, M. S. Roberts, H. D. Blackburn, and A. E. Pollack, Am. J. Hosp. Phar., 38, 1308 (1981).
79. D. C. Monkhouse and L. Van Campen, Drug Dev. Ind. Pharm., 10, 1175 (1984).
80. D. C. Monkhouse, Drug Dev. Ind. Pharm., 10, 1373 (1984).
81. J. T. Carstensen and C. T. Rhodes, *Drug Stability, Principles and Practices*, 3rd ed., Marcel Dekker, Inc., New York, 2000.
82. F. Baitalow, H.-G. Schmidt, and G. Wolf, Thermochim. Acta 337, 111 (1999).
83. S. R. Byrn, R. R. Pfeiffer, and J. G. Stowell, *Solid-State Chemistry of Drugs*, 2nd ed., SSCI, Inc., West Lafayette, IN, 1999.
84. L. Meites, CRC Crit. Rev. Anal. Chem., 8, 55 (1979).
85. P. Wessels, M. Holz, F. Erni, K. Krummen, and J. Ogorka, Drug Dev. Ind. Pharm., 23, 427 (1997).
86. Guideline for Submitting Documentation for the Stability of Human Drugs and Biologics, February, 1987, Center for Drugs and Biologics, Food and Drug Administration, Rockville, MD.
87. Guidance for Industry: Stability Testing of Drug Substances and Drug Products: Draft Guidance, June, 1998, Center for Drug Evaluation and Research and Center for Biologics Evaluation and Research, U. S. Department of Health and Human Services, Food and Drug Administration, Rockville, MD.
88. J. T. Carstensen, J. Pharm. Sci., 63, 1 (1974).
89. W. J. Irwin, *Kinetics of Drug Decomposition, Basic Computer Solutions*, Elsevier Science, Amsterdam, 1990.
90. W. P. Jencks, *Catalysis in Chemistry and Enzymology*, McGraw-Hill, New York, 1969, p. 513.
91. L. Lachman, P. DeLuca, and M. J. Akers, Kinetic principles and stability testing, in *The Theory and Practice of Industrial Pharmacy*, 3rd ed. (L. Lachman, H. A. Lieberman, and J. L. Kanig, Eds.), Lea & Febiger, Philadelphia, 1986, p. 766.
92. H.-L. Fung and S.-Y. P. King, in *Pharm Tech Conference '83 Proceedings*, Aster Publishing, Springfield, OR, 1983.
93. K. J. Laidler, J. Chem. Educ., 49, 343 (1972).
94. M. J. Pikal, A. L. Lukes, and J. E. Lang, J. Pharm. Sci., 66, 1312 (1977).
95. A. J. Woolfe and H. E. C. Worthington, Drug Dev. Commun., 1, 185 (1974).
96. D. R. Savello and R. F. Shangraw, Am. J. Hosp. Pharm., 28, 754 (1971).
97. T.-Y. Fan and S. R. Tannenbaum, J. Agric. Food Chem., 21, 967 (1973).
98. R. Shija, V. B. Sunderland, and C. McDonald, Int. J. Pharm., 80, 203 (1992).
99. R. E. Pincock, Acc. Chem. Res., 2, 97 (1969).
100. N. Murase, P. Echlin, and F. Franks, Cryobiology, 28, 364 (1991).
101. T. Chen, Drug Dev. Ind. Pharm., 18, 1311 (1992).
102. W. Wang, Int. J. Pharm., 185, 129 (1999).
103. L. C. Gu, E. A. Erdös, and H.-S. Chang, Pharm. Res. 8, 485 (1991).
104. H.-L. Fung and S.-Y. P. King, in *Pharm Tech Conference '83 Proceedings*, Aster Publishing, Springfield, OR, 1983.
105. L. J. Ravin, C. A. Simpson, A. F. Zappala, and J. J. Gulesich, J. Pharm. Sci., 53, 106 (1964).
106. E. R. Garrett, J. Am. Chem. Soc., 79, 3401 (1957).
107. J. I. Wells, *Pharmaceutical Preformulation*, Ellis Horwood Ltd., West Sussex, England, 1988.
108. M. A. Schwartz, J. Pharm. Sci., 53, 1433 (1964).
109. O. A. G. J. van der Houwen, M. R. de Loos, J. H. Beijnen, A. Bult, and W. J. M. Underberg, Int. J. Pharm., 155, 137 (1997).
110. K. A. Connors, *Chemical Kinetics: The Study of Reaction Rates in Solution*, VCH Publishers, Inc., New York, 1990, pp. 273–292.
111. L. Gu and R. G. Strickley, Int. J. Pharm., 60, 99 (1990).
112. E. R. Garrett, J. Org. Chem., 26, 3660 (1961).
113. K. Thoma and M. Struve, Pharm. Ind., 47, 1078 (1985).
114. F. S. Amis, *Solvent Effects on Reaction Rates and Mechanisms*, Academic Press, New York, 1966.
115. C. Reichardt, *Solvents and Solvent Effects in Organic Chemistry*, 2nd ed. VCH Verlagsgesellschaft mbH, Weinheim, 1988.
116. J. T. Carstensen, J. Pharm. Sci., 59, 1140 (1970).
117. B. G. Cox, A. J. Kresge and P. E. Sørensen, Acta Chem. Scand., A24, 202, (1988).

Copyright © 2002 Marcel Dekker, Inc.

118. A. J. Kresge, Chem. Soc. Rev., 2, 475 (1973).
119. C. A. Bunton and G. Savelli, in *Advances in Physical Organic Chemistry*, Vol. 22 (V. Gold and D. Bethell, Eds.) Academic Press, Orlando, FL, 1986, pp. 213–309.
120. S. Otto, J. B. F. N. Engberts, and J. C. T. Kwak, J. Am. Chem. Soc., 120, 9517 (1998).
121. T. J. Broxton, Aust. J. Chem., 35, 1357 (1982).
122. T. Higuchi and L. Lachman, J. Am. Pharm. Assoc. Sci. Ed., 44, 521 (1955).
123. K. A. Connors, *Binding Constants: The Measurement of Molecular Complex Stability*, John Wiley and Sons, New York, 1987.
124. J. V. Stella and R. A. Rajewski, Pharm. Res., 14, 556 (1997).
125. D. O. J. Thompson, Crit. Revi. Ther. Drug Carrier Syst., 14, 1 (1997).
126. T. Loftsson, Drug Stability, 1, 22 (1995).
127. N. Hayase, Y. Itagaki, S. Ogawa, S. Akutsu, and Y. Abiko, Y., J. Pharm. Sci. 83, 532 (1994).
128. N. H. Anderson, D. Johnson, M. A. McLelland, and P. Munden, J. Pharm. Biomed. Anal., 9, 443 (1991).
129. D. Genton and U. W. Kesselring, J. Pharm. Sci., 66, 676 (1977).
130. J. T. Carstensen, Drug Dev. Ind. Pharm. 14, 1927 (1988).
131. C. Ahlneck and G. Zografi, Int. J. Pharm., 62, 87 (1990).
132. T. Chen, Drug. Dev. Ind. Pharm., 18, 1311 (1992).
133. M. J. Hageman, Drug. Dev. Ind. Pharm., 14, 2047 (1988).
134. Code of Federal Regulations, Title 21, Food and Drugs, Part 211, Current good manufacturing practice for finished pharmaceuticals, Subpart G, §211.137 Expiration Dating.
135. Code of Federal Regulations, Title 21, Food and Drugs, Part 211, Current good manufacturing practice for finished pharmaceuticals, Subpart I, §211.166 Stability Testing.
136. Code of Federal Regulations, Title 21, Food and Drugs, Part 211, Current good manufacturing practice for finished pharmaceuticals, Subpart I, §211.170 Reserve Samples.
137. Code of Federal Regulations, Title 21, Food and Drugs, Part 211, Current good manufacturing practice for finished pharmaceuticals, Subpart J, §211.194 Laboratory Records.
138. Code of Federal Regulations, Title 21, Food and Drugs, Part 314, Applications for FDA approval to market a new drug or antibiotic drug, Subpart B, §314.50 Content and format of an application.
139. Code of Federal Regulations, Title 21, Food and Drugs, Part 314, Applications for FDA approval to market a new drug or antibiotic drug, Subpart B, §314.70 Supplements and other changes to an approved application.
140. Code of Federal Regulations, Title 21, Food and Drugs, Part 314, Applications for FDA approval to market a new drug or antibiotic drug, Subpart B, §314.81 Other postmarketing reports.
141. ICH Expert Working Group, Q1A Stability Testing of New Drug Substances and Products, International Conference on Harmonisation of Technical Requirements for the Registration of Pharmaceuticals for Human Use, 1994.
142. ICH Expert Working Group, Q1A (R) Stability Testing of New Drug Substances and Products (draft), International Conference on Harmonisation of Technical Requirements for the Registration of Pharmaceuticals for Human Use, 2000.
143. Guideline for Submitting Samples and Analytical Data for Methods Validation, February 1987, Center for Drugs and Biologics, Food and Drug Administration, Rockville, MD.
144. ICH Expert Working Group, Q2B Validation of Analytical Procedures: Methodology, International Conference on Harmonisation of Technical Requirements for the Registration of Pharmaceuticals for Human Use, 1996.
145. E. Debesis, J. P. Boehlert, T. E. Givand, and J. C. Sheridan, Pharm. Tech., 6(9), 120–137 (1982).
146. United States Pharmacopeia 24, <1225>, Validation of Compendial Methods, 2000, pp. 2149–2152.
147. Stability Lab Information Manager, Metrics, Inc., P.O. Box 4035, Greenville, NC 27836.
148. Stability System, ScienTek Software, Inc., P.O. Box 323, Tustin, CA 92781.
149. R. F. Tetzlaff, Pharm. Tech., 16(5), 70 (1992).

Copyright © 2002 Marcel Dekker, Inc.

Chapter 7

Preformulation

Jens T. Carstensen

University of Wisconsin, Madison, Wisconsin

I. INTRODUCTION

Historically, preformulation evolved in the late 1950s and early 1960s as a result of a shift in emphasis in industrial pharmaceutical product development. Up until the mid-1950s, the general emphasis in product development was to develop elegant dosage forms and organoleptic considerations far outweighed such (as yet unheard of) considerations as whether a dye used in the preparation might interfere with stability or with bioavailability. In fact, pharmacokinetics and biopharmaceutics were in their infancy, and although stability was a serious consideration, most analytical methodology was such that even gross decomposition often went undetected.

It was, in fact, improvement in analytical methods that spurred the first programs that might bear the name "preformulation." Stability-indicating methods would reveal instabilities not previously known, and reformulation of a product would be necessary. When faced with the problem of attempting to sort out the component of incompatibility in a 10-component product, one might use many labor hours. In developing new products, therefore, it would be logical to check, ahead of time, which incompatibilities the drug exhibited (testing it against common excipients). This way the disaster could be prevented in advance.

A further cause for the birth of preformulation was the synthetic organic programs started in many companies in the 1950s and 1960s. Pharmacological screens would show compounds to be promising, and pharmacists were faced with the task of rapid formulation. Hence they needed a fast screen (i.e., a preformulation program) to enable them to formulate intelligently. The latter adverb implies that some of the physical chemistry had to be known, and this necessitated determination of physicochemical properties, a fact that is also part of preformulation. The approach was so logical, indeed, that it eventually became part of official requirements for INDs and NDAs [1]:

> New drug substances in Phase I submission. For the drug substance, the requirement includes a description of its physical, chemical, or biological characteristics. We in the reviewing divisions regard stability as one of those characteristics. The requirement for NDA submissions . . . stability information is required for both the drug substance and drug product. A good time to start to accumulate information about the appropriate methodology and storage stations for use in dosage form stations for use in dosage form stability studies, therefore, is with the unformulated drug substance. . . . Stress storage conditions of light, heat and humidity are usually used for these early studies, so that the liable structures in the molecule can be quickly identified. . . . If degradation occurs, the chemical reaction kinetics of the degradation should be determined. . . . Physical changes such as changes from one polymorph to another polymorph should be examined. . . . With the drug substance stability profile thus completed, the information should be submitted in the IND submission.

II. TIMING AND GOALS OF PREFORMULATION

The goals of the program are, therefore, (1) to establish the necessary physicochemical parameters of a new drug substance, (2) to determine its kinetic rate profile,

Copyright © 2002 Marcel Dekker, Inc.

(3) to establish its physical characteristics, and (4) to establish its compatibility with common excipients. To view these in their correct perspective, it is worthwhile to consider where (i.e., at what time) in an overall industrial program preformulation takes place. The following events take place between the birth of a new drug substance and its eventual marketing (most investigational drug substances never make it to the marketplace for one reason or another):

1. The drug is synthesized and tested in a pharmacological screen.
2. The drug is found sufficiently interesting to warrant further study.
3. Sufficient quantity is synthesized to (a) perform initial toxicity studies, (b) do initial analytical work, and (c) do initial preformulation.
4. Once past initial toxicity, phase I (clinical pharmacology) begins and there is a need for actual formulations (although the dose level may not yet be determined).
5. Phase II and III clinical testing then begins, and during this phase (preferably phase II) an order of magnitude formula is finalized.
6. After completion of the above, an NDA is submitted.
7. After approval of the NDA, production can start (product launch).

III. PHYSICOCHEMICAL PARAMETERS

Physicochemical studies are usually associated with great precision and accuracy and in the case of a new drug substance would include (1) pK (if the drug substance is an acid or base), (2) solubility, (3) melting point and polymorphism, (4) vapor pressure (enthalpy of vaporization), (5) surface characteristics (surface area, particle shape, pore volume), and (6) hygroscopicity. Unlike usual physicochemical studies, an abundance of material is usually not at hand for the first preformulation studies: in fact, at the time the function starts, precious little material is supplied, and, therefore, the formulator will often settle for good estimates rather than attempting to generate results with four significant figures.

There is another good reason not to aim "too high" in the physicochemical studies of the first sample of drug substance. In most cases the synthesis is only a first scheme; in later scale-up it will be refined, and in general the first small samples contain some small amounts of impurities, which may influence the precision of the determined constants. But it is necessary to know, grosso modo, important properties such as solubility, pK, and stability.

A. pK_a and Solubility

One important goal of the preformulation effort is to devise a method for making solutions of the drug. Frequently, the drug is not sufficiently soluble in water itself to allow for the desired concentrations, for example, for injection solutions. Solubilities are determined by exposing an excess of solid to the liquid in question and assaying after equilibrium has been established. This usually is in the range of 60–72 hours, and to establish that equilibrium indeed has been established, sampling at earlier points is necessary. Unstable solutions pose a problem in this respect and are dealt with in more detail later. Solubilities cannot be determined by precipitative methods (e.g., by solubilizing an acid in alkali and then lowering the pH to the desired pH) because of the so-called meta-stable (solubility) zone [2]. In this chapter drug substances are subdivided into two categories: (1) ionizable substances and (2) (virtually) nonionizable substances.

Ionizable Substances

For substances that are carboxylic acids (HA) it is advantageous to determine the pK_a since this property is of importance in a series of considerations. For carboxylic acid the species A^- usually absorbs in the ultraviolet (UV) region, and its concentration can be determined spectrophotometrically [3]; HA, on the other hand, will absorb at a different wavelength.

The molar absorbances of the two species at a given wavelength are denoted ε_0 and ε_- (and it is assumed that at the wavelength chosen $\varepsilon_0 < \varepsilon_-$) and it can be shown that if the solution is m_0 molar in total A, then

$$A^-/HA = (\varepsilon - \varepsilon_0 m_0)/(m_0\varepsilon_- - \varepsilon) \quad (1)$$

so that ratio A^-/HA can be determined in a series of buffers of different pH. Hence the pK can be found as the intercept by plotting pH as a function of $\log[(A^-)/(HA)]$ by Henderson-Hasselbach:

$$pH = pK + \log(A^-/HA) \quad (2)$$

If several buffer concentrations are used, extrapolation can be carried out to zero ionic strength, and the pK_a can be determined. For initial studies, however, a pK in

Copyright © 2002 Marcel Dekker, Inc.

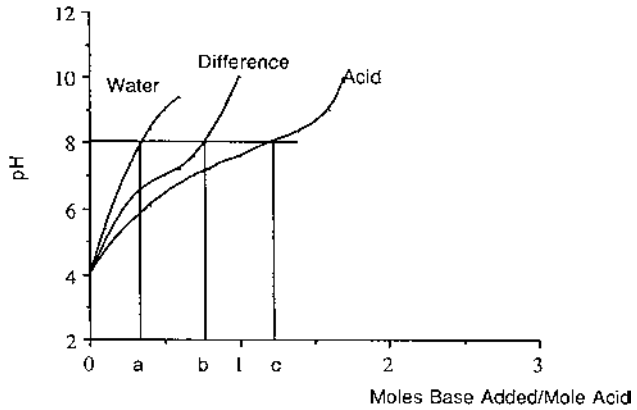


Fig. 1 Typical titration curves. The "water" curve indicates the amount of alkali needed to "titrate" the water, and the "acid" curve is a conventional titration curve. The difference curve is the horizontal difference between the "acid" and the "water" curve, and is the adjusted titration curve, e.g., point b is c−a. The pK is the point of inflection, which is also the point at where half of the acid is neutralized.

the correct range (i.e., ± 0.2 unit) will suffice, so that the determination above can be done at one buffer concentration only.

The conventional approach is, of course, to do titrations (Fig. 1) that will yield graphs of fraction neutralized as a function of pH. Usually, the water is titrated as well [4], and what is presented in Fig. 1 is the "difference." The pK is then the pH at half neutralization (which is also the inflection point). The pH-solubility curve can now be constructed simply by determining the solubility of HA (at low pH) and A^- (e.g., of NaA) at high pH (e.g., at pH 10). It is noted that at a given pH the amount in solution in a solubility experiment is:

$$S = S_{HA} + C_{A^-} \tag{3}$$

where S denotes solubility. The last term can be determined from knowledge of the pH and use of Eq. (2).

For drugs that are amines, the free base is frequently poorly soluble, and in this case the pK is often estimated by performing the titration in a solvent containing some organic solvent (e.g., ethanol). By doing this at different organic solvent concentrations (e.g., 5, 10, 15, 20%), extrapolation can be carried out to 0% solvent concentration to estimate the aqueous pK.

Solubility profiles paralleling pH curves have been reported by Granero et al. [5].

Nonionizable Substances

For hydrophobic, (virtually) nonionizable substances [i.e., those that show no ionic species of significance in the pH range 1 to 10 (e.g., diazepam)], solubility can usually be improved by addition of nonpolar solvents. Aside from solubility, stability is also affected by solvents in either a favorable or a nonfavorable direction [6]. Theoretical equations for solubility in water [7] and in binary solvents [8] have been reported in literature, but in general the approach in preformulation is pseudoempirical. Most often the solubility changes as the concentration of nonpolar solvent C_2, increases. For binary systems it may simply be a monotonely changing function [9], as shown in Fig. 2. The solubility is usually tied to the dielectric constant, and in a case such as that shown by the squares, the solubility is often log-linear when plotted as a function of inverse dielectric constant, E, that is,

$$\ln S = -e_1/(\varepsilon + e_2) \tag{4}$$

where ε is the dielectric constant and the e-terms are constants [3].

Frequently, however, the solubility curve has a maximum (as shown by circles in Fig. 2, when plotted as both a function of C_2 and ε [10]. In either case it is possible to optimize solubility by selection of a solvent system with a given value of ε; that is, once the curve has been established, the optimum water/solvent ratio for another solvent can be calculated from known dielectric constant relationships [11].

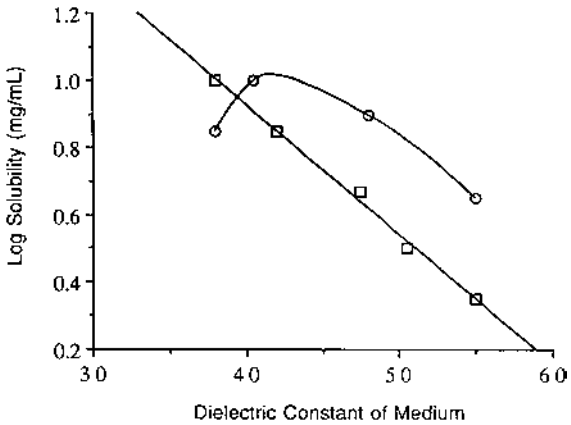


Fig. 2 Squares: Solubility of 7-chloro-1,3-hydro-5-phenyl-2H-1, 4-benzodiazepine-2-one-4-oxide in aqueous propylene glycol. (Data from Ref. 9) Circles: Solubility of another benzodiazepine. (Unpublished data.)

Copyright © 2002 Marcel Dekker, Inc.

Ternary Systems and Optimization

Frequently, *ternary* solvent systems are resorted to. Examples are water–propylene glycol–benzyl alcohol and water–propylene glycol–ethanol. In such cases the solubility profile is usually presentable by a ternary diagram [12]. This type of diagram usually demands a fair amount of work; that is, solubility of the drug substance in many solvent compositions must be determined. A priori, it would therefore seem that they would be out of place in a situation where only limited quantities of drug are available. However, their principle gives some validity to optimization procedures.

The diagram can be one of two types, as shown in Figs. 3 and 4. In the first type, the solubility may be assumed to be of the type

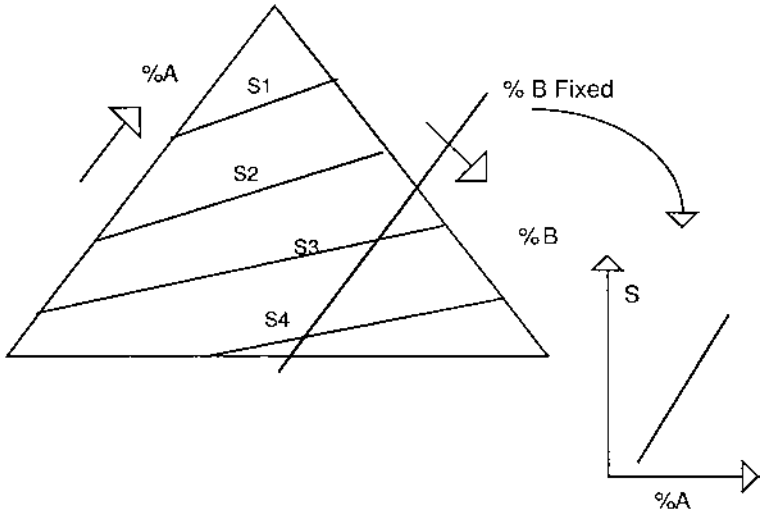


Fig. 3 Ternary diagram of solubility of a compound in a ternary mixture with linear solubility response. (Inset) Concentration of drug in compositions with constant concentration of B. The composition of the solute is the constant concentration of B, the concentration of A in the abscissa, and the complement concentration of the third component. The drug solubility response is linear in the A concentration in this case.

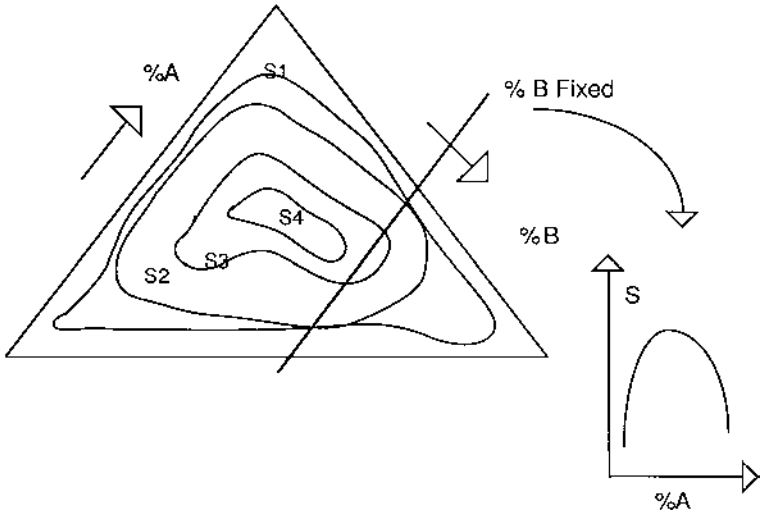


Fig. 4 Ternary diagram and tie line concentration in a nonlinear system.

Copyright © 2002 Marcel Dekker, Inc.

$$S = a_{10} + a_{11}C_1 + a_{12}C_2 \quad (5)$$

where C denotes concentrations of nonaqueous solvents. An example of this is shown on the right of the figure. In Eq. 5 the subscripts to C denote the two nonaqueous solvents. Hence, three solubility experiments would determine the relationship (with zero degrees of freedom). It is usual to do at least five and determine possible curvature [i.e., inclusion of more terms in Eq. (5)].

In the second case in Fig. 4, each tie line will give a parabolic-type curve as shown on the right of the figure. Hence at a given concentration of C_2 the solubility can be approximated by

$$S = b_{10} + b_{11}C_1 + b_{12}C_1^2 \quad (6)$$

where, in the simplest case,

$$b_{10} = C_{20} + C_{21}C_2 \quad (7)$$

Hence optimization can be achieved by five (or more) experiments, with zero (or in general $n-5$) degrees of freedom.

The effects of pH and of mixed solvent systems have been described by Tongaree et al. [13].

Prediction of Solubility

It is advantageous with a new drug substance to be able to estimate what its solubility will be prior to carrying out dissolution experiments. There are several systems of solubility prediction, most notably those published by Amidon and Yalkowsky [14–16] in the 1970s. Their equation for solubility of *p*-aminobenzoates in polar and mixed solvents is a simplified two-dimensional analog of the Scatchard-Hildebrand equation and is based on the product of the interfacial tension and the molecular surface area of the hydrocarbon portion of a molecule.

More recently Bodor et al. [17] developed a semi-empirical solubility predictor based on 20 variables (S = molecular surface in Å^2, I_a = indicator variable for alkanes, D = calculated dipole moment in Debyes; Q_n = square root of sum of squared charges on oxygen atoms; Q_o = square root of sum of squared charges on oxygen atoms; V = molecular volume in Å^2, S_2 = square of molecular surface; C = constant; MW = molecular weight, $\{O\}$ = ovality of molecule, A_{bh} = sum of absolute values of atomic charges on hydrogen atoms, A_{bc} = sum of absolute values of atomic charges on hydrogen atoms, A_m = indicator variable for aliphatic amines, and N_h = number of N−H single bonds in the molecule.)

The aqueous solubilities, W, of 331 compounds were found to follow the following equation (with tolerances omitted):

$$\begin{aligned}\log W = {} & -56.039 + 0.32235D - 0.59143I_a \\ & + 38.443Q_n^4 - 51.536Q_n^2 + 18.244Q_n \\ & + 34.569Q_o^4 - 31.835Q_o^2 + 15.061Q_o \\ & + 1.9882A_m + 0.15689N_h + 0.00014102S^2 \\ & + 0.40308S - 0.59335A_{bc} + -0.42352V \\ & + 1.3168A_{bh} + 108.80\{O\} - 61.272\{O\}^2\end{aligned} \quad (8)$$

Of the parameters listed, only the ovality and the indicator value for the alkanes I_a are unfamiliar entities that are obtained from literature [17].

Dissolution

The importance of dissolution will be discussed in more detail later. A short note on the topic is, however, necessary at this point. According to Noyes-Whitney [18];

$$dm/dt = VdC/dt = -kA(S - C) \quad (7)$$

where

m = mass not dissolved
V = liquid volume
t = time
k = so-called intrinsic dissolution rate constant (cm/s)
A = surface area of the dissolving solid

Many criticisms have been voiced against Eq. (7), but in general it is correct, and it will be assumed to be so in the following. If an experiment is carried out with constant surface (as, e.g., using a Wood's apparatus [19]) or with smaller amounts, making a small pellet and encasing it in wax and exposing only one face to a dissolution medium, or if an excess of solid prevails throughout the dissolution experiment, then Eq. (7) may be integrated to give

$$\ln[1 - (C/S)] = -(kA/V)t \quad (8)$$

or

$$C = S[1 - \exp(-\{kA/V\}t)] \quad (9)$$

A typical curve following Eq. (9) is shown as "C" in Fig. 5.

Copyright © 2002 Marcel Dekker, Inc.

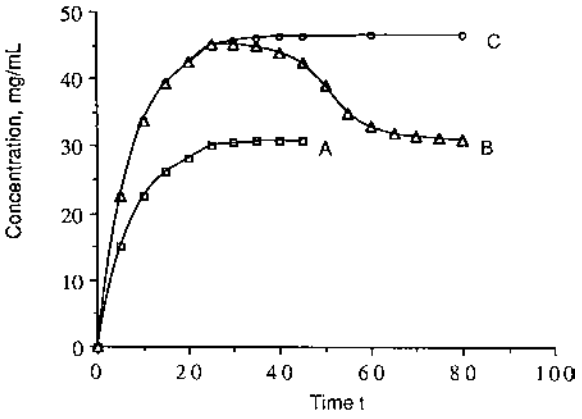


Fig. 5 Dissolution profiles obtained from the solubility determination of two polymorphic forms of the same drug substance. A is the stable form with solubility 31 mg/mL. B is the profile of the metastable form with solubility 46 mg/mL. This solubility (circles) is not achieved in many instances, and precipitation of the stable form occurs at a point beyond the solubility of A, and the trace becomes B. C is the hypothetical profile of the metastable form.

Solubility of Unstable Compounds

Quite often a compound is rather unstable in aqueous solution. Hence the long exposure to liquid required for traditional solubility measurements will cause decomposition, and the resulting solubility results will be unreliable. In this particular case a method known as Nogami's method may be used. If a solution experiment is carried out as a dissolution experiment with samples taken at equal time intervals, δ, it can be shown [20] that when the amount dissolved at time $t + \delta$ is plotted versus the amount dissolved at time t, a straight line will ensue (Fig. 6). When the concentration at time $t + \delta$ is plotted versus the concentration, at t the following relationship holds:

$$C(t + \delta) = S[1 - \exp(-k\delta)] + \exp(-k\delta)C(t) \qquad (10)$$

hence such a plot (as shown in Fig. 6) will give k from the slope, and inserting this in the intercept expression will give S. The advantage of the method is that it can be carried out in a short period of time and will reduce the effect of decomposition; the disadvantages is that it is not as precise as ordinary solubility determinations.

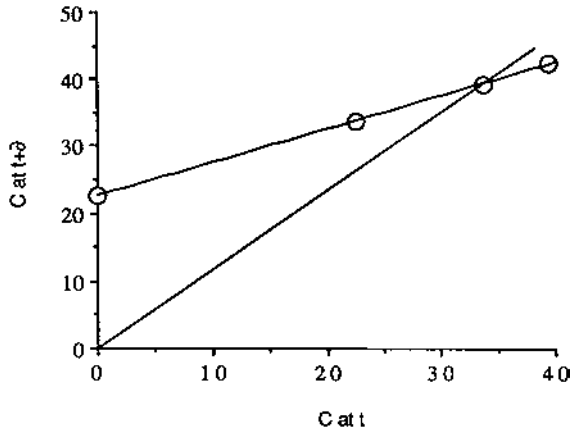


Fig. 6 The Nogami method applied to the data in Fig. 5.

Solubility of Metastable Polymorphs

Polymorphism is an important aspect of the physical properties of a drug substance. One of the characteristics of a metastable polymorph (to be discussed in some detail at a later point) is that it is more soluble than its stable counterpart. The solubility profile of the polymorph will be as shown in Fig. 5; A is the stable form, with solubility of 31 mg/mL. B is the profile of the metastable form, with solubility of 46 mg/mL. This solubility (circles) is usually not achieved, and precipitation of the stable form occurs at a point beyond the solubility of A, and the trace becomes B.

In such cases, the Nogami method can be applied to the early points curve (Fig. 6) and the solubility, S', of the polymorph can be assessed. One of the important aspects of metastable polymorphs in pharmacy is exactly their higher solubility, since the dissolution rate will also be higher [Eq. (7)]. Hence the bioavailability will be increased where this is dissolution rate limited [21].

Polymorphism

Solids exist as either amorphous compounds or crystalline compounds [22]. In the latter, the molecules are positioned in lattice sites. Usually, amorphous substances decompose by first-order kinetics, and as such distinguish themselves from crystalline compounds [23].

A lattice is a three-dimensional array, and there are eight systems known. Inorganic substances are usually defined by one crystal system by the so-called radius-ratio rule [22], but organic compounds often have the capability of existing in more than one crystal form, a phenomenon referred to as polymorphism.

Powder diffraction x-ray analysis and DSC are the methods most often used to distinguish between different polymorphs. Single crystal x-ray analysis [24] will give information about the position of molecular groups within the crystal and thus actually defines the

Copyright © 2002 Marcel Dekker, Inc.

difference between the different forms. Often, however, crystals cannot be grown to an adequate size to allow this to be carried out. Stephenson [25] reported that recent advances in crystallographic computing have made it possible to solve by powder diffraction methods structures that have not been possible to solve by single-crystal methods.

If a compound exhibits polymorphism, one of the forms will be more stable (physically) than the other forms; that is, of n existing forms $n-1$ forms will possess thermodynamic tendency to convert to the nth, stable form (which then has the lowest Gibbs energy; it should be noted that in the preformulation stage it is not known whether the form on hand is the stable polymorph or not).

One manner in which different polymorphs are created is by way of recrystallizing them from different solvents, and at a point in time when sufficient material (and this need not be very much) is available, the preformulation scientist should undertake recrystallization from a series of solvents.

Knowledge of polymorphic forms is of importance in preformulation because suspension systems should never be made with a metastable form (i.e., a form other than the stable crystal form). Conversely, a metastable form is more soluble than a stable modification, and this can be of advantage in dissolution [Eq. (9)]. There are two types of polymorphism, a fact illustrated in the following discussion.

If the vapor pressure or solubility of a compound is plotted as a function of temperature, a plot such as that shown in Fig. 7 will result. Here form I is the form that is stable at temperature of 20°C. If the compound exists

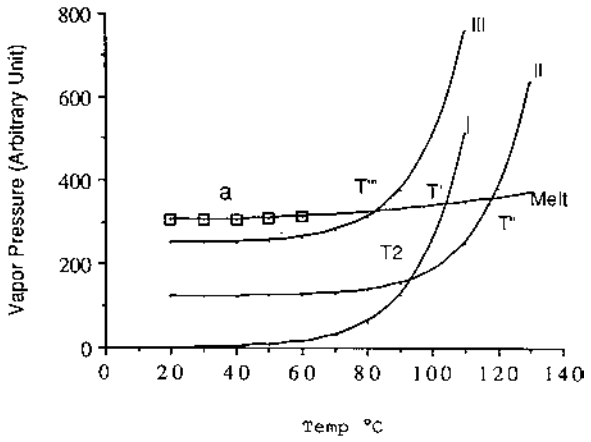


Fig. 7 Dependence of vapor pressure and solubility for an enantiotropic pair (I/II), a monotropic pair (I/III), and an amorphous compound (a).

in the two forms I and II, the phenomenon is referred to as an entiotropic system, since, on heating to the temperature T_2, form I will transform into form II. Form II may exist below temperature T_2, but perturbations (e.g., presence of moisture) will convert it to form I, and the energy involved in the transformation will be

$$E = \mathrm{RT}\ \ln[S_2/S_1] \tag{11}$$

where S_2 and S_1 are solubilities of II and I.

If the compound is present as form I at room temperature and heated up fast it will melt at T', which is lower than the melting point of form II (T'').

A different situation exists if the compound exists as form I and form III. This is referred to as a monotropic system, and here III is unstable relative to I over the whole solid range. In this case, however, the melting point of the "unstable" polymorph is lower than that of the stable (T''' is lower than T').

If a metastable polymorph is kept dry, it may be stable for eons, and it is therefore not referred to as an unstable but, rather, a metastable polymorph. An even more energetic state is, of course, represented by amorphous forms, which may be considered supercooled liquids. Today, polymorphism is checked for in two fashions. Thermal methods will give information as to whether a polymorph is stable, enantiotropic, or monotropic. If the system possesses a transition point, ΔG is zero at this point. In a fashion like the melting process, where ΔG is also zero, the transition is associated with an enthalpy change, which is endothermic. Hence, if a sample of a compound is heated up and a change in enthalpy occurs below the melting point, it is an enantiotropic transformation if the point is reproducible and if it occurs again after the sample has been cooled back down and shows the same enthalpy change upon heating. Otherwise, it is monotropic, in which case, as shown in Fig. 8, there either will be a (lower) melting point (T''') with a single endotherm or there will be a melting followed by a recrystallization into the more stable forms, which then melt [denoted III(Alternate) in Fig. 8]. An amorphous compound has no melting point, and above glass transition temperatures its vapour pressure curve simply continues into that of the melt.

Second, x-ray diffraction will, directly, give spacings in the crystal and reveal differences between samples. Finally, solubility curves can be carried out, and if a nick in the solubility curve is found (Fig. 9), this is a transition temperature. If no nick is found, there is no transition temperature, but if the dissolution curves are as shown in Fig. 5, there is polymorphism, and it may

Copyright © 2002 Marcel Dekker, Inc.

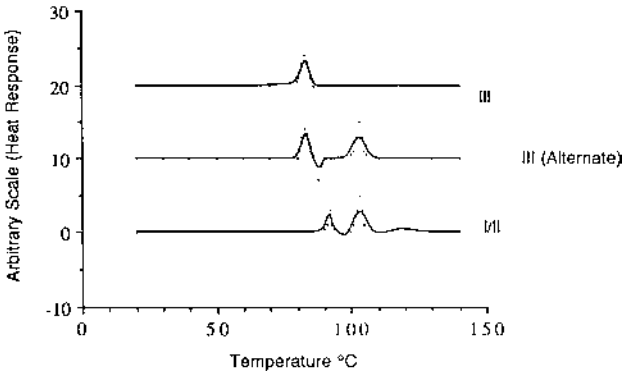


Fig. 8 DSC tracings of the polymorphs shown in Fig. 7.

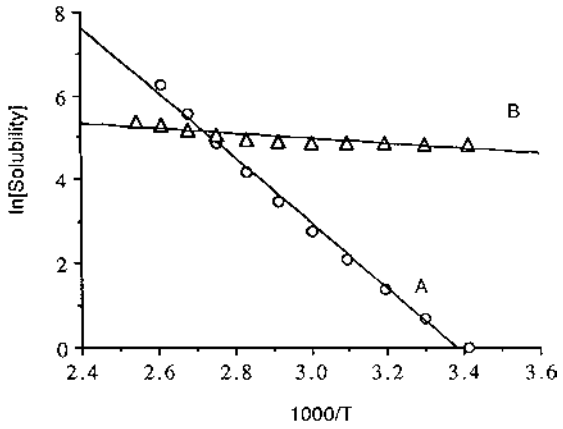


Fig. 9 Solubility curves of an enantiotropic pair.

be concluded that a monotropic system exists, if this is true at all texperatures.

C. Partition Coefficient

Partition coefficients between water and an alkanol (e.g., octanol) should be determined in preformulation programs [26]. The partition coefficient of a compound that exists as a monomer in two solvents is given by

$$K = C_1/C_2 \tag{12}$$

If it exists as an n-mer in one of the phases, the equation becomes

$$K = (C_1)^n/C_2 \tag{13}$$

or

$$\log k = n \log C_1 - \log C_2 \tag{14}$$

The easiest way to determine the partition coefficient is to extract $V_1\,\text{cm}^3$ of saturated aqueous solution with $V_2\,\text{cm}^3$ of solvent, and determine the concentration C_2 in the latter. The amount left in the aqueous phase is $(C_1V_1 - C_2V_2) = M$, so that the partition coefficient is given by

$$K = M/(V_1C_1) \tag{15}$$

If it is assumed that the species is monomeric in both phases, the partition coefficient becomes the ratio of the solubilities, and it is simply sufficient to determine the solubility of the drug substance in the solvent (since it is assumed that the solubility is already known in water):

$$K = S_1/S_2 \tag{16}$$

D. Vapor Pressure

In general, vapor pressures are not all that important in preformulation, but it should always be kept in mind that a substance may have sufficiently high vapor pressure to (a) become lost to a sufficient extent to cause apparent stability problems and content uniformity problems and (b) exhibit a potential for interaction with other compounds and adsorption onto or sorption into package components [27].

Most drug substances are, substantially, not volatile. As an initial screen, it can be determined whether the drug is sufficiently volatile to cause concern by placing a weighed amount of it in a vacuum desiccator and weighing it daily for a period of time. It is better to have a high-vacuum system for this, and the use of a vacuum electrobalance is best for this purpose. A good estimate of the vapor pressure can be obtained [28] by using a pierced thermal analysis cell, placing it on a vacuum electrobalance, and monitoring the weight loss rate. Using a substance with known vapor pressure can then be used for calibration, the loss rates being proportional to the vapor pressures.

Moisture isotherms were, up until recently, not considered part of preformulation. However, with the advent of microcalorimetry, moisture isotherms and surface areas may be determined with mg quantities of drug substance [29,30].

E. Surface Characteristics

The surface characteristics of a batch of a drug substance may greatly influence its properties in processing (flow, dissolution). Crystals may crystallize in

Copyright © 2002 Marcel Dekker, Inc.

different habits (plate, needle, cube), and these may not be due to morphology; depending on crystallization circumstances, they could all be the same crystal form, but of different habit [31].

It is a good practice, during development of a new drug, all the way through to the NDA, to take photomicrographs of each new batch of drug substance delivered to the product development department. In this manner there will exist a permanent reference record, and when deviations from expected behavior occur during the product development sequence, the photomicrograph will be one record that may throw light on the problem. Aside from this, the specific surface area (A'', cm^2/g) of each batch of drug substance should be measured.

Shape and Fractal Dimension

Shape is of great interest and affects many properties, and it is important to have a record of how a shape changes as the synthesis of the raw material changes during the development process. In the simplest form microscopy of all batches used in product development should be carried out to determine the ratio of longest to shortest dimension (average of 10 measurements). This is a type of shape factor.

Shape Factors

A series of definitions have been developed to describe the shape of a particle [32,33].

A commonly used shape factor converts the volume of a particle, v, to its volumetric mean diameter, a_v:

$$v = \alpha_v a_v^3 \tag{17}$$

Similarly, a shape factor may defined which converts the surface area, s, of a particle to its surface mean diameter, a_s:

$$s = \alpha_s a_s^2 \tag{18}$$

Another, more practical, frequently used shape factor is the overall shape factor Γ, given by:

$$\Gamma = SV^{-2/3} = N^{1/3}\alpha_s a_s^2 / \{a_v^2[\alpha_v]^{2/3}\} \tag{19}$$

If a small polydisperse sample of powder is dissolved under sink conditions, then the dimensions, b, of the particle will decrease linearly with respect to time [33–37]:

$$b = b_0 - Kt \tag{20}$$

where b_0 is the initial size of the particle, t is time, and K is a constant. It is noted that the initial distribution function of the powder sample (obtainable from a small sample by electronic counting, for example) can be used to calculate the mass undissolved until the critical time, t^*. At a critical time, t^*, the smallest particle will disappear from the dissolution medium, and up to this point in time (t^*) the total number, N, of particles in the system remains the same [33,34].

If the powder sample is allowed to dissolve at times $t < t^*$, then the mass undissolved will be

$$M = \int_{b_{\min}}^{b_{\max}} \{\{N\rho\, f(b_0)\}(\alpha_{v_0})(b_0 - Kt)^3\} db_0 \tag{21}$$

where ρ is density and the subscript zero denotes initial condition. If the cubed term is expanded, then:

$$M = A_1 - B_1 t + C_1 t^2 - D_1 t^3 \tag{22}$$

The coefficients A_1, B_1, C_1, and D_1 are elaborated on below. An example is shown in Fig. 10.

$$A_1 = N\rho\alpha_{v_0} \int_{b_{\min}}^{b_{\max}} f(b_0) b_0^3\, db_0 = N\rho\alpha_{v_0}\mu_3 \tag{23}$$

This term is obviously the original mass of the powder sample, and μ_3 is the third moment of the probability

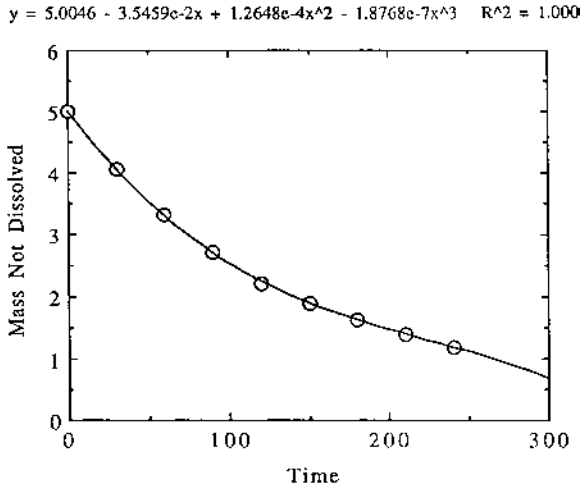


Fig. 10 Dissolution data of an oxalic acid mesh cut, plotted by way of third-order polynomial.

Copyright © 2002 Marcel Dekker, Inc.

distribution function [38]. The third coefficient in Eq. (22) is

$$C_1 = 3N\rho K^2 \alpha_{v_0} \int_{b_{\min}}^{b_{\max}} f(b_0) b_0 \, db_0 = 3K^2 N \rho \alpha_{v_0} \mu_1 \quad (24)$$

where μ_1 is the first moment of the probability density function and also the mean of the distribution [38].

The second coefficient in the expansion is

$$B_1 = 3N\rho K \alpha_{v_0} \int_{b_{\min}}^{b_{\max}} f(b_0) b_0^2 \, db_0 = 3KN\rho\alpha_{v_0}\mu_2 \quad (25)$$

where μ_2 is the second moment of the probability density function [38].

The variance of the powder population is given by:

$$s^2 = \mu_2 - \mu_1^2 \quad (26)$$

The coefficient to the last term is given by:

$$D_1 = N\rho K^3 \alpha_{v_0} \int_{b_{\min}}^{b_{\max}} f(b_0) db_0 = K^3 N \rho \alpha_{v_0} \quad (27)$$

where use has been made of the fact that the probability density function as used is normalized:

$$D_1 = \int_{b_{\min}}^{b_{\max}} f(b_0) db_0 = 1 \quad (28)$$

Eq. (22) may be divided through by $A_1 = M_0$, in which case it takes the form:

$$M/M_0 = 1 - B_2 t + C_2 t^2 - D_2 t^3 \quad (29)$$

and the coefficients with subscript 2 are then the coefficients with subscript 1 divided by M_0.

The coefficients of the terms in t, t^2, and t^3 in Eq. (29) are given by the following equations:

$$B_2 = K\mu_2/\mu_3 \quad (30)$$

$$C_2 = K^2\mu_1/\mu_3 \quad (31)$$

$$D_2 = K^3/\mu_3 \quad (32)$$

In a typical research and development setting, in the event that a new drug candidate is recognized by the drug-discovery group, then the dissolution rate constant K for that compound under specified hydrodynamic conditions can be determined from powder dissolution data and practical size analysis by microscopy.

This can be done via Eqs. (29) through (32). From the dissolution data, the coefficient B_2 is obtained and through the results from microscopy the moments μ_2 and μ_3 can be evaluated. By knowing N, the initial number of particles, and the density of the solid, the average initial volume shape factor for a polydisperse powder can be estimated.

The same problem can be considered in the opposite direction [34]. Knowing the K-value for a compound (e.g., oxalic acid dihydrate) under specified hydrodynamic conditions and the fraction undissolved as a function of time, the moments of the distribution function of a "dimension of significance" can be obtained.

At the critical time, there is a change in slope in the cube root law plot [37,39].

The fraction undissolved data until the critical time can be least-square fitted to a third degree polynomial in time as dictated by Eq. (29). The moments of distribution μ_1, μ_2, and μ_3 can be evaluated from Eqs. (30) through (32), with three equations used to solve for three unknowns. These values may be used as first estimates in a nonlinear least-squares fit program, and the curve will, hence, reveal the best values of both shape factor, size distribution, and K-value.

However, more sophisticated methods may be used to attain a good feel for the shape factor, namely its fractal dimension. This is most conveniently carried out by use of imaging techniques.* The general principle of this is shown in Figs. 11 and 12.

As an example (and this is a hypothetical example only), a particle is shown in Fig. 11 such as might appear on a microscope slide. This particle is gridded out in the form shown in the upper lefthand corner of Fig. 11. The number of squares in which parts of the trace of the particle is located is counted. This number is N, and the length of the grid size is g. The grid size is arbitrarily set equal to one in this example. The grid

*An example of an excellent system is that of Universal Imaging Concepts Image 1 Softward 486 66 Computer (Data Store, INc), Hamamatsu C2400-77H BYW Chip Camera (CCD Camera Control), Sony 19″ PVM1943MD Color Monitor, UP5000 (Sony) Color Printer Nikon Model Labophot-2 40 × ocular × 10 on screen magnification.

Copyright © 2002 Marcel Dekker, Inc.

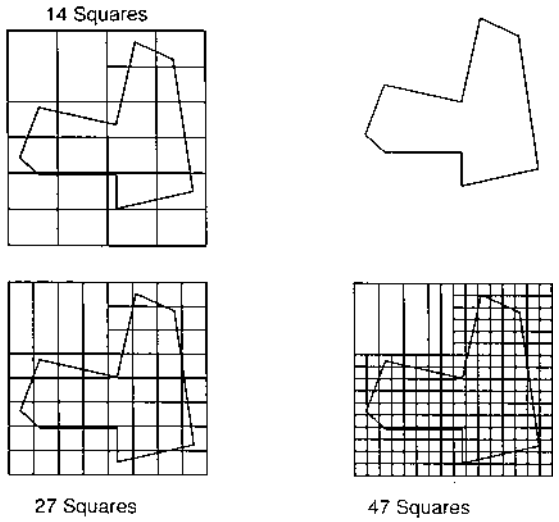


Fig. 11 Dividing the field containing a particle into more and more squares, by dividing the grid length by 2.0 in each step.

length is now halved, and the number of squares counted again, and this is repeated a number of times. The fractal equation is then given by:

$$\ln[N] = -n \ln[g] + q \tag{33}$$

where n and q are constants, and where n, the slope of the line, is the fractal dimension [40]. This dimension is characteristic for shapes, and its determination constitute a record of how shape may change during synthesis.

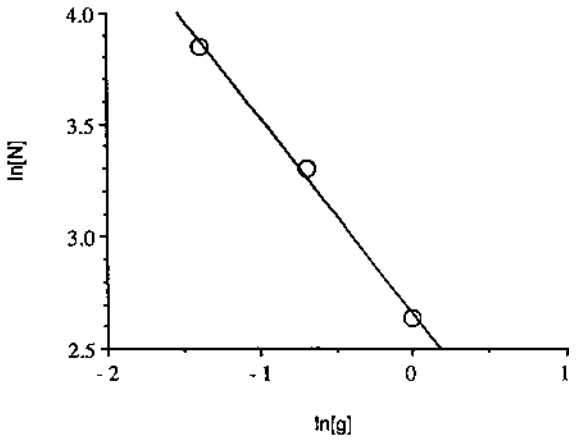


Fig. 12 Fractal graph of the data in Fig. 11. The least-squares fit for the equation is $\ln[N] = 2.6575 - 0.8735 \ln[g]$.

The number of decreases in the so-called "measuring stick" (in this case the grid) is of great importance, and it is the behavior at the smaller end of the measuring stick size that is important. In fact, the small number of iterations leading to Eq. (18) may well be falacious, but at the lower end of the scale the relationship will still be correct (but the true fractal dimension may or will be different from the value found in Fig. 11).

Particle Size and Size Distributions

The size of particles is of great importance, and particle size determinations should be carried out in preformulation as well in formulation functions. For small particle sizes, simple microscopy [41] may be used, but again imaging techniques, particularly with motorized stages, are more representative and much easier to carry out [42].

When characterizing a powder, it is important to know what type of diameter is being described, since various techniques give different types of diameter.

The volume-surface mean diameter, d_{sv}, can be determined by permeamitry (e.g., Fisher Sub-Sieve Sizer) for a fine powder [41]. It is given by

$$d_{sv} = 6V/A'' \tag{34}$$

where V is the solids volume of the sample and A'' is the geometric surface area. A'' can be obtained for a coarse powder by sieve analysis:

$$A'' = \{6/\rho\} \sum w_i/d_i \tag{35}$$

where ρ is the solids density (g/cm^3), d_i the mean diameter of the ith mesh cut, and w_i the weight retained by the ith screen. If the drug substance is not porous,

$$G = A/A'' \tag{36}$$

where G is the rugosity and is a measure of the surface roughness.

Pore Size Distribution

With the advent of mercury intrusion porosimeters, it is advantageous to perform a pore size distribution of investigational batches of a drug [43]. The Washburn equation [44] states that the pressure, P, necessary to intrude a pore is given by

$$\ln P = -q\gamma(1/r) \cos[\theta] \tag{37}$$

Copyright © 2002 Marcel Dekker, Inc.

where

q = a constant
γ = interfacial tension
θ = contact angle
r = radius of the pore being penetrated

Since mercury has a contact angle with most solids of about 140°, it follows that its cosine is negative (i.e., it takes applied pressure to introduce mercury into a pore). In a mercury porosimeter, a solids sample is evacuated in a cell, mercury is then intruded, and the volume, V, is noted (it actually reads out), and the pressure, P, is then increased stepwise. In this fashion it is possible to deduce the pore volume of a particular radius [corresponding to P by Eq. (21)]. A pore size distribution will give the total internal pore area as well, which can be of importance in dissolution.

Hygroscopicity

Hygroscopicity is, of course, an important characteristic of a powder. It can be shown, roughly, for a fairly soluble compound that the hygroscopicity is related to its solubility [45,46], although it has been shown that the heat of solution plays an important part in what is conceived as "hygroscopicity" [47–49]. A hygroscopicity experiment is carried out most easily by exposing the drug substance to an atmosphere of a known relative humidity (e.g., storing it over saturated salt solutions in desiccators). Each solution will give a certain relative humidity (RH), and the test is simply to weigh the powder from time to time and determine the amount of moisture adsorbed (weight gained). This does not work with drug substances that decompose (e.g., effervescent mixtures will start losing weight due to carbon dioxide evolution) [50].

It can be shown that if the air space is sufficiently agitated to prevent vapor pressure gradients, the initial uptake rate (g H_2O/g solid per hour) is related to relative humidity by

$$L = a_{21}[\mathrm{RH} - \mathrm{RH}_0] \tag{38}$$

where RH_0 is the vapor pressure of a saturated solution of the drug substance in water. The latter can be estimated by an ideality assumption; if the solubility is expressed as a mole fraction X_s, the vapor pressure over a saturated solution will be P' given by

$$P' = (1 - X_s)P^* \tag{39}$$

where P^* is water's vapor pressure at that temperature.

The experiments above are rather easy to carry out and should always be part of a preformulation program, since hygroscopicity can be so important that it will dictate whether or not a particular salt should be used. Dalmane, for instance, is a monosulfate and is used as such since the disulfate, desirable in many other respects, is so hygroscopic that it will remove water from a hard-shell capsule and make it exceedingly brittle.

IV. STABILITY

Stability of a drug substance and product is monitored throughout the development and clinical phases. This monitoring requires stability-indicating assay methodology, and this is a subject that is separate from performulation per se. In most instances, the major, feasible decomposition products are identified early [51], and as such it is known if the pathways are hydrolytic, oxidative, or photochemical.

High-performance liquid chromatography (HPLC) is the usual chemical method employed, although "older" methods, such as thin layer chromatography (TLC) can be of use. Microcalorimetry can also be of use [52].

It should again be emphasized that at the onset of a new drug program, there are only small amounts of drug substance at hand. One of the first tasks for the preformulation scientist is to establish the framework within which the first clinical batches can be formulated. To this end it is important to know with which common excipients the drug is compatible. Below, the distinction will be made between solid and liquid dosage forms.

A. Compatibility Test for Solid Dosage Forms

It is important for the formulator of a new drug substance to know with which excipients he can work and with which he cannot. Some pharmaceutical incompatibilities are known to the formulator, e.g., magnesium stearate/aspirin, and glucose/amines [53].

To assess unknown incompatibilities it is customary to make a small mix of drug substance with an excipient [54,55], place it in a vial, place a rubber stopper in the vial, and dip the stopper in molten carnauba wax (to render it hermetically sealed). The wax will harden and form a moisture barrier up to 70°C. A list of common excipients characteristic of this type of test is shown in Table 1. At times it is possible to obtain quantitative relationships of

Copyright © 2002 Marcel Dekker, Inc.

Table 1 Categories for Two-Component Systems

		Worse		Total score	
	Identical	17–27 months at 25°C	10 days at 55°C	25°C	55°C
Drug per se					
Dry	15	4	1	38	31
5% H_2O	9	8	3	49	38
+ Magnesium stearate					
Dry	16	3	1	34	30
5% H_2O	15	4	1	43	35
+ Calcium stearate					
Dry	13	4	3	37	32
5% H_2O	12	5	3	38	35
+ Stearic acid					
Dry	15	5	0	42	31
5% H_2O	7	11	2	60	38
+ Talc					
Dry	14	5	1	38	30
5% H_2O	10	8	2	45	34
+ Acid-washed talc					
Dry	12	8	0	44	31
5% H_2O	10	9	1	49	35
+ Lactose					
Dry	12	5	3	38	32
5% H_2O	9	7	4	65	56
+ $CaHPO_4$, anhydrous					
Dry	12	6	2	46	36
5% H_2O	9	8	3	66	53
+ Cornstarch					
Dry	12	5	3	39	34
5% H_2O	10	5	5	40	37
+ Mannitol					
Dry	10	7	3	39	31
5% H_2O	8	7	5	47	45
+ Terra alba					
Dry	14	6	0	41	28
5% H_2O	11	6	3	50	45
+ Sugar 4x					
Dry	12	6	2	41	34
5% H_2O	9	7	4	63	61

Source: Ref. 54.

excipient characteristics and interaction rates [55,56]. In addition to the test as described, a similar set of samples are set up where 5% moisture is added. A storage period of 2 weeks at 55°C (except for stearic acid and dicalcium phosphate, where 45°C is used) is employed, after which time the sample is observed physically for (1) caking, (2) liquefaction, (3) discoloration, and (4) odor or gas formation. It is then assayed by TLC or HPLC.

It is noted that one of the samples set up is the drug by itself. This is done for several reasons, one of which is that it is now required by FDA for IND submissions [1]. One more reason is that at the onset of a program, the organic synthesis of the compound may lack the refinement it will later have, and it is not uncommon that there will be several weak spots (impurities) on a TLC chromatogram of a compound obtained by initial laboratory synthesis. Hence, in selecting the excipients with which the drug substance is deemed to be compatible, it is customary to use as criteria that (after accelerated exposure of a drug/excipient mix) no new spots have developed and that the intensity of the spots in the drug stored under similar conditions (2 weeks at 55°C) are the same as in the acceptable excipient. This type of program is used by many companies with good success (i.e., the formulas developed based on the findings from the compatibility program are stable).

It should be noted that liquefaction at times occurs because of eutectic formation (e.g., often with caffeine combinations) and that this may not necessarily be associated with decomposition. On the other hand, discoloration (e.g., amines and sugars) usually is.

Finally, the reason for not forcing dicalcium phosphate (a very valuable formulation aid in direct compression) beyond 50°C is that at higher temperatures (actually [57] above 70°C) it converts to the anhydrate, a conversion that is, curiously enough, catalyzed by water. In other words, the dihydrate will be autocatalytic in this respect at elevated temperatures, and it should not be ruled out based on high-temperature findings.

B. Kinetic pH Profiles

In preformulation one task is to establish the stability of the drug substance in both solid and dissolved state. In the latter case it is important, with small samples of drug substance, to assess (a) the effect of buffer type, (b) the effect of buffer concentration, (c) the effect of pH in a practical range, (d) the effect of temperature, and (e) the kinetic salt effect.

At times the solubility of a drug in water is insufficient at room temperature to allow a meaningful kinetic study, in which case it can, at times, be carried out at elevated temperature [58]. If done at several different temperatures, it may be possible to estimate the stability at room temperature by extrapolation. Frequently, a broad screen of stability is performed on the initial small sample used for initial performulation; this is frequently referred to as "forced decomposition studies" [59], in which the drug is exposed to "acid degradation," "base degradation," "aqueous degradation," "drug powder

Copyright © 2002 Marcel Dekker, Inc.

degradation," and "light degradation." More refined studies are eventually needed.

For any compound marketed by a pharmaceutical concern, at one time during its development there should be a concerted effort to establish a very exact pH profile. To do this correctly is a time-consuming undertaking. However, the information that can be gleaned from it is very important with regard to formulations, and it is therefore customary to carry out an approximate kinetic pH profile [60] early in the development stage. This will allow formulation of solutions for injections and for oral products as well, at a pH and using buffers that will give the best stability. Without it, formulation is essentially guesswork.

Most drug decompositions are hydrolyses, where the drug concentration C decreases with time via

$$dC/dt = -k_2 C_{H_2O} C \tag{40}$$

Since the water concentration hardly changes, this (bimolecular) reaction scheme is reduced to the pseudo-first-order expression

$$dC/dt - KC \tag{41}$$

where K is the first-order rate constant. This integrates to the well-known form

$$\ln[C/C_0] = -Kt \tag{42}$$

Hence semilog plotting of concentration versus time (Fig. 13) will give a straight line with a slope from which K is calculated. But most reactions are catalyzed

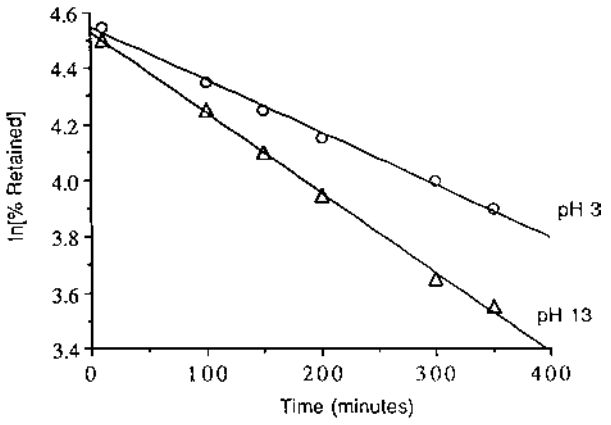


Fig. 13 Pseudo-first-order decompositions of carbuterol at 85°C at an ionic strength of 0.5 (Graphs plotted from data in Refs. 61 and 62.)

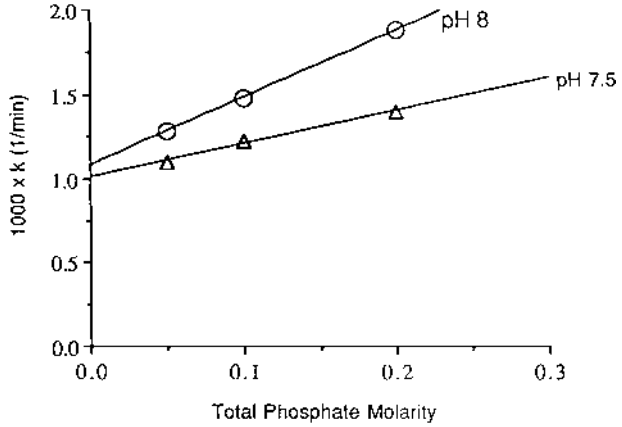


Fig. 14 Buffer concentration dependence of carbuterol at 85°C at an ionic strength of 0.5 (Graphs plotted from data in Refs. 61 and 62.)

by buffers, by hydrogen ions, and by hydroxyl ions, so K will be of the form

$$K = k + k_+(\mathrm{H}) + k_-(\mathrm{OH}) + k_B(B) \tag{43}$$

where B denotes buffer concentration and k is a rate constant. A decomposition experiment is now carried out at, say, five pH values, each using two buffer concentrations. A graph is constructed (Fig. 14) versus (B) at the various pH values. At low pH (where hydroxyl ion concentration can be disregarded), Eq. (43) becomes

$$K = k_+(\mathrm{H}) + k_B(B) \tag{44}$$

so that the plot in Fig. 14 will give

$$K = k_+(\mathrm{H}) \tag{45}$$

as intercept and k_B as slope. The above allows assessment of the effect of the buffer (which is an important point in the buffer selection). Taking the 10 logarithms of Eq. (45) now gives

$$\log K = -\mathrm{pH} + \log(k_+) \tag{46}$$

A similar argument will show that at high pH

$$\log K = \mathrm{pH} - 14 + \log(k_-) \tag{47}$$

This explains why the extremes of a pH profile (Fig.15) often have slopes of plus or minus unity [61,63].

Copyright © 2002 Marcel Dekker, Inc.

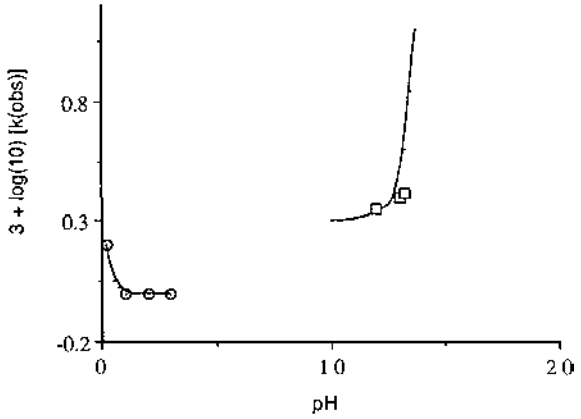


Fig. 15 Kinetic pH profile at low (circles) and high (squares) pH of carbuterol at 85°C. (Graphs plotted from data in Refs. 61 and 62.)

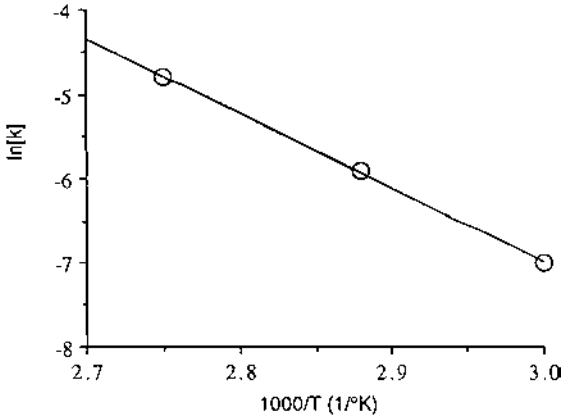


Fig. 16 Arrhenius plotting of carbuterol rate constants at pH 4 of carbuterol. (Graph plotted from data in Refs. 61 and 62.)

The horizontal part is due to the uncatalyzed rate constant, k_+, in Eq. (43). A pH profile can be done at, for example, six pH values, and since there are two kinetic points (times) and two buffer concentrations at each, a total of 24 assays are needed, which is not insurmountable. This number may be minimized and optimized by careful selection of pH and buffer concentrations [60]. Later in the program the pH profile should be repeated but with multiple points and several buffer concentrations, but this is beyond the point of preformulation. An example of a "full" pH profile is one running from pH 1 to pH 11 [61–64].

C. Liquid Compatibilities

The pH profile is the most important part of liquid compatibilities. However, two component systems are set up in aqueous (or other types of) solutions and treated as in Sec. IV.A. This is now required in the stability guidelines [1], which state that "it is suggested that the following conditions . . . be evaluated in studies on solutions or suspensions of bulk drug substances: acidic and alkaline pH, high oxygen and nitrogen atmospheres, and the presence of added substances, such as chelating agents and stabilizers" and it is suggested "that stress testing conditions . . . include variable temperature (e.g., 5, 50, 75°C)."

Aqueous Solution Capability

In general, such studies are carried out by placing the drug in a solution of the additive. These can be (and usually are) a heavy metal (with or without chelating agents present) or an antioxidant (in either oxygen or nitrogen atmosphere). Usually, both flint and amber vials are used, and in many cases an autoclaved condition is included. This will provide answers to questions about susceptibility to oxidation, to light exposure, and to heavy metals. These are important questions as far as injectable compatibilities are concerned. Exposure to various plugs is frequently included at this point so that early injectable preparations can be formulated.

For preparations for oral use, knowledge of the desired dosage form is important, but compatibility with ethanol, glycerin, sucrose, corn syrup, preservatives, and buffers is usually carried out. This type of study also gives an idea of the activation energy, E, of the predominant reaction in solution. The Arrhenius plots (Fig. 16) for compounds in solution are usually quite precise.

Denoting the rate constant by k, absolute temperature by T, the gas constant by R, and a collision factor by Z:

$$k = Z \exp[-E/(RT)] \tag{48}$$

or its more useful logarithmic cousin,

$$\ln k = \ln Z - [E/(R)](1/T) \tag{49}$$

Nonaqueous Liquid

With transdermal dosage forms being of great importance of late, it is advisable to test for compatibilities with "ointment" excipients and with polymers (e.g., ethylvinyl polymer, if that is the desired barrier).

In the case of transdermals, the dosage form is either directly placed in a stirred liquid or it is placed in a

Copyright © 2002 Marcel Dekker, Inc.

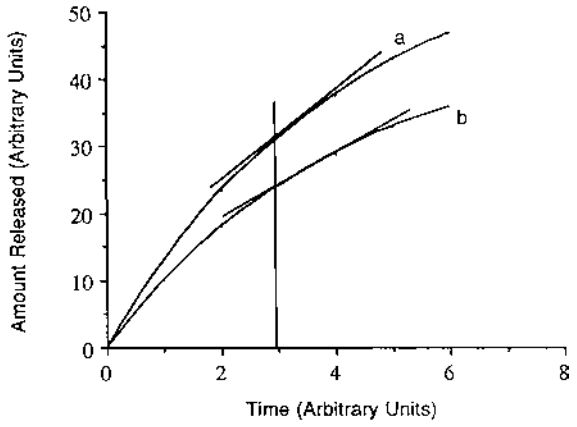


Fig. 17 Slope determination of flux of ointment release and release from ointment + membrane.

cell with an appropriate membrane (e.g., cadaver skin) to estimate the release characteristics of the drug from the ointment [65].

It should be noted here that if the overall flux is J, then

$$1/J = [1/(J_{\text{ointment}})] + [1/(J_{\text{membrane}})] \tag{50}$$

where subscripts refer to the respective phase. J_{membrane} can be obtained from curves such as shown in Fig. 17 where first the overall flux is obtained (with the membrane in place), giving the value of J. Then the release is obtained without the membrane in place, giving J_{ointment}, that is,

$$J = [1/A]\, dm_1/dt \tag{51}$$

and

$$J_{\text{ointment}} = (1/A)\, dm_2/dt \tag{52}$$

J_{membrane} is then obtained as the difference of the reciprocals.

In vivo testing is usually carried out by applying the dosage form to hairless rats followed by subsequent sacrifice. Since the skin consists of a number of layers with differing hydrophilicity, the overall fate of the drug is of importance.

V. DISSOLUTION OF DRUG SUBSTANCE AND DOSAGE FORM

In the time path solid dosage forms (tablets or capsules) must eventually to be manufactured for the clinic (e.g., in phase II). If possible, the drug substance per se is subjected to a dissolution test in a Wood's apparatus [19]. This test is useful, although quite dependent on hydrodynamic conditions. It consists of placing the powder in a type of tablet die, compressing the tablet, and exposing the flat, exposed side of the tablet (with surface area A) to a dissolution liquid (usually water or N/10 HC1) in which it has a solubility S. Under these conditions [66] the intrinsic dissolution rate constant (cm/s) can be obtained by Eq. (8), which under sink conditions* (i.e., where C is 15% of S) becomes

$$C = (SkA/V)t \tag{53}$$

It has been suggested [67] that if k is obtained under sink conditions over a pH range of 1–8 at 37°C in a USP vessel by way of Eq. (53) at 50 rpm, then if the dissolution rate constant (kA/V) is greater than 1 mg min^{-1} cm^{-2}, the drug is prone not to give dissolution-rate–limited absorption problems. On the other hand, if the value is less than 0.1, such problems can definitely be anticipated, and compounds with values of kA/V of 0.1–1 mg min^{-1} cm^{-2} are in a gray area. For compound selectivity it is frequently useful to express dissolution findings in terms of k (i.e., in cm/s).

For small amounts of powder, dissolution of the particulate material can often be assessed (and compared with that of other compounds) by placing the powder in a calorimeter [68] and measuring the heat evolved as a function of time. The surface area must be assessed microscopically (or by image analyzer), and the data must be plotted by a cube root equation [39]:

$$1 - [M/M_0]^{1/3} = -(2kS/\rho r)t \tag{54}$$

where M is mass not dissolved, M_0 the initial amount subjected to dissolution, is true density, S is solubility, and r is the mean "radius" of the particle. The method is simply comparative, not absolute, due to the hydrodynamics being different in the calorimeter than it would be in a dissolution apparatus.

The stability of the drug substance in the dissolution liquid is of importance and has been reported on by Shishoo et al. [69].

*Strictly speaking, sink conditions are when the amount dissolved plotted versus time yields a line which, within experimental error, is linear. When the surface area, A, is constant, then this corresponds to 15% dissolved. When the surface area changes (e.g., during particulate dissolution), then this number may be smaller.

Copyright © 2002 Marcel Dekker, Inc.

A. Biopharmaceutical Aspects

One important aspect of drug dosage form development is, of course, to obtain a dosage form that is absorbed in a desired fashion. In most cases this implies a rapidly and completely absorbed dosage form. This means that it is necessary to test the drug substance itself for in vivo release characteristics. A good indication as to whether a drug may give problems in this respect is a comparison of LD_{50} values by parenteral and oral routes. If the former toxicity is much greater than the latter, there is often an absorption problem. In the following, it is assumed that the problems are dissolution dictated.

B. Semi In Vivo Testing

The general goal is, of course, to submit an IND for the drug and to get it into testing in humans in the clinic. Frequently biological absorption characteristics are checked by such procedures as the everted sac technique [70]. Here a segment of the small intestine of a rat is everted. The ends are then tied off, and physiological fluid containing no drug (placebo liquid) is filled into the sac. It is then placed in a vessel containing a solution of the drug in a buffer solution. The setup is kept at 37°C, and all the while oxygen is being supplied to the solution. After a given interval the contents of the sac are assayed for content of the drug. This can then be repeated for other times. Collections of several samples from the same intestine segment is possible [71], and this greatly facilitates the procedure. Other methods, such as that suggested by Dolusio et al. [72], exist and are used in preformulation efforts. Here rats are anesthetized and the ileal and duodenal ends of their intestines are then cannulized, allowing sampling and liquid introduction.

C. In Vivo Testing

For preformulation purposes, some animal testing is usually performed prior to phase I. This could be in rats, dogs, or other species. The animals are tested by being given a specific dose on a specific regimen (e.g., fasting, single dose, or after each meal), and blood is then collected at various intervals. In this fashion a blood-level curve is obtained (Fig. 18). The ultimate value of this is not an absolute. To extrapolate from one species to another is a dubious undertaking, and the ramifications of this are beyond the scope of the chapter. General conclusions can be drawn and methods are briefly described below.

It is customary to do one set of tests by the parenteral route and at this stage (Fig. 18) to assume or

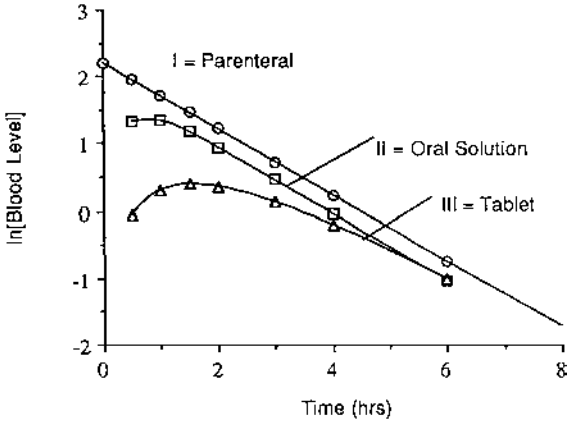


Fig. 18 Blood-level curves of identical doses of a drug by parenteral route (I) and by oral route as a solution (II) and as a tablet (III).

define that 100% of the dose is absorbed [73]. Depending on pathways, this may later be modified, but the correctness of the assumption is of lesser importance at this point, as will be seen shortly. Next, a solution (or if solubility is limited, a suspension) is administered, giving a different blood level (curve II in Fig. 18). Finally, the drug is administered dry (e.g., in a capsule), giving rise to curve III in Fig. 18.

If the areas under the curves are denoted by A, then (based on equal dose) $A\text{II}/A\text{I}$ is the fraction absorbed by oral route. $A\text{III}/A\text{II}$ is the fraction efficiency of the solid dosage form. The reason for this latter is, of course, that the solid dosage form has to dissolve before the drug contained in it is available for absorption. It is the latter ratio that is important to the investigating pharmaceuticist, and therefore the outcome of the parenteral form is actually not a consideration from a formulation point of view. It is critical overall and if it is low, it may, at the point of parenteral data acquisition, be advisable to stop the program and evaluate the possibility of derivatives that would give better availability.

There are large volumes of literature written on this subject, but the simplest manner of evaluating the formula efficiency by way of blood-level data is the following simplified model: It is assumed that upon ingestion the tablets or capsules disintegrate into (N) particles, and that the drug then dissolves from these particles [74,75]. Each particle will release its content in T time units. The number of particles not dislodged from the dosage form at time t is

$$N_t = N_0 \exp(-kt) \tag{55}$$

Copyright © 2002 Marcel Dekker, Inc.

It can be shown that for this type of situation the blood levels, B, will be a function of time by an equation of the type

$$B = A[\exp(-k_1 t) - \exp(-k_2 t)] \tag{56}$$

where k_1 is an elimination rate constant and k_2 is a function of dissolution and absorption rate constants. Figure 18 shows a case in point, where

$$B\text{II} = 7[\exp(-0 - 494t) - \exp(-2 - 813t)] \tag{57}$$

$$B\text{III} = 4.48[\exp(-0.472t) - \exp(-0.85t)] \tag{58}$$

It is noted that the k_1 values are about the same, as they should be, and that the difference, k'', between the k_2 values,

$$k'' = 2.81 - 0.85 = 2.0 \tag{59}$$

is a parameter indicative of the retardation due to disintegration and dissolution. The areas under the curves, Q, are

$$Q = (1/k_1) - (1/k_2) \tag{60}$$

In the cases in Fig. 18, the areas would be 0.942 µg h/cm^3 for the tablet and 1.67 for the solution. The formulation efficiency is $0.94/1.67 = 56\%$ by this criterion, and the tablet formulation could stand some improvement.

REFERENCES

1. R. C. Schultz, Stability of Dosage Forms, FDA-Industry Interface Meeting, Washington, DC, Oct. 7, 1983; Stability Guidelines, Congressional Record, May 7, 1984.
2. N. Rodriguez-Hornedo, Ph.D. thesis, University of Wisconsin, 1984.
3. W. J. W. Underberg and H. Lingeman, J. Pharm. Sci., 72, 553 (1983).
4. T. Parke and W. Davis, Anal. Chem., 25, 642 (1954).
5. G. Granero, M. M. deBertorello and M. C. Brinon, Int. J. Pharm., 190, 41 (1999).
6. S. K. Bakar and S. Niazi, J. Pharm. Sci., 72, 1024 (1983).
7. S. H. Yalkowsky and S. C. Valvani, J. Pharm. Sci., 72, 912 (1983).
8. W. E. Acree and J. H. Rytting, J. Pharm. Sci., 72, 293 (1983).
9. J. T. Carstensen, K. S. Su, P. Maddrell, and H. Newmark, Bull. Parenter. Drug. Assoc., 25, 193 (1971).
10. A. N. Paruta and S. A. Irani, J. Pharm. Sci., 54, 1334 (1964).
11. G. Cave, F. Puisieux, and J. T. Carstensen, J. Pharm. Sci., 68, 424 (1979).
12. D. Sroby, R. Bitter, and J. Webb, J. Pharm. Sci., 52, 1149 (1963).
13. S. Tongaree, D. R. Flanagan, and R. I. Poust, Pharm. Dev. Tech., 4, 571 (1999).
14. S. H. Yalkowsky, G. L. Flynn, and G. L. Amidon, J. Pharm. Sci., 61, 983 (1972).
15. G. L. Amidon, S. H. Yalkoskky, and S. Leung, J. Pharm. Sci., 63, 1858 (1974).
16. S. H. Yalkowsky, G. L. Amidon, G. Zografi, and G. L. Flynn, J. Pharm. Sci., 64, 48 (1975).
17. N. Bodor, Z. Gabanyi, C. -K. Wong, J. Am. Chem. Soc., 111, 3783 (1989).
18. A. Noyes and W. Whitney, J. Am. Chem. Soc., 23, 689 (1897).
19. J. H. Wood, G. Catacalos, and S. Lieberman, J. Pharm. Sci., 52, 296 (1963).
20. H. Nogami, T. Nagai, and A. Suzuki, Chem. Pharm. Bull., 14, 329 (1966).
21. M. Shibata, H. Kokobu, K. Morimoto, K. Morisaka, T. Ishida, and M. Inoue, J. Pharm. Sci., 72, 1436 (1983).
22. J. T. Carstensen, *Advanced Pharmaceutical Solids*, Marcel Dekker, New York, 2001, p. 117, 118.
23. J. T. Carstensen, *Advanced Pharmaceutical Solids*, Marcel Dekker, New York, 2001, p. 227.
24. P. J. Cox and J. L. Wardell, Int. J. Pharm., 194, 147 (2000).
25. G. A. Stephenson, J. Pharm. Sci., 89, 958 (2000).
26. S. H. Yalkowsky, S. C. Valvani, and T. J. Roseman, J. Pharm. Sci., 72, 866 (1983).
27. M. Pikal and A. L. Lukes, J. Pharm. Sci., 65, 1269 (1976).
28. J. T. Carstensen and R. Kothari, J. Pharm. Sci., 70, 1095 (1981).
29. M. Pudipeddi, T. D. Sokoloski, S. P. Duddu, and J. T. Carstensen, J. Pharm. Sci., 85, 381 (1995).
30. M. Ohta, Y. Tozuka, T. Oguchi, and K. Yamamoto, Drug Dev. Ind. Pharm., 26, 643 (2000).
31. J. T. Carstensen, *Pharmaceutics of Solids and Solid Dosage Forms*, Wiley-Interscience, New York, 1977, pp. 6, 41.
32. J. M. Dallavalle, in *Micromeritics*, 2nd. ed., Pitman Publishing Co., New York, 1948, p. 142.
33. J. T. Carstensen and M. N. Musa, J. Pharm. Sci., 61, 223 (1972).
34. J. T. Carstensen, *Advanced Pharmaceutical Solids*, Marcel Dekker, New York, 2001, pp. 79–80.
35. M. V. Dali and J. T. Carstensen, Pharm. Res., 13, 155 (1996).
36. I. C. Edmundson and K. A. Lees, J. Pharm. Pharmacol., 17, 193, (1945).

Copyright © 2002 Marcel Dekker, Inc.

37. J. T. Carstensen and M. Patel, J. Pharm. Sci., 64, 1770 (1975).
38. C. A. Bennett and N. L. Franklin, *Statistical Analysis in Chemistry and the Chemical Industry*, Wiley and Sons Inc., New York, 1961.
39. A. Hixson and J. Crowell, Ind. Eng. Chem., 23, 923 (1931).
40. J. T. Carstensen and M. Franchini, Drug Dev. Ind. Pharm., 18, 85 (1992).
41. B. Kaye, in *The Fractal Approach to Heterogeneous Chemistry*, D. Avnir, (Ed.), John Wiley and Sons, Chichester, NY, 1989, p. 62.
42. J. T. Carstensen, *Pharmaceutics of Solids and Solid Dosage Forms*, John Wiley, New York, 1977, pp. 56, 226, 228.
43. J. T. Carstensen and X. P. Hou, J. Pharm. Sci., 74, 466, (1985).
44. E. H. Washburn, Phys. Rev. 17, 273 (1921).
45. J. T. Carstensen, *Pharmaceutics of Solids and Solid Dosage Forms*, Wiley-Interscience, New York, 1977, pp. 11–15.
46. L. Van Campen, G. Zografi, and J. T. Carstensen, Int. J. Pharm. 5, 1 (1980).
47. L. Van Campen, G. L. Amidon, and G. Zografi, J. Pharm. Sci., 72, 1381 (1983).
48. L. Van Campen, G. L. Amidon, and G. Zografi, J. Pharm. Sci., 72, 1388 (1983).
49. L. Van Campen, G. L. Amidon, and G. Zografi, J. Pharm. Sci., 72, 1394 (1983).
50. J. T. Carstensen and F. Usui, J. Pharm. Sci., 74, 1293 (1984).
51. P. J. Jansen, M. J. Akers, R. M. Amos, S. W. Baertschi, G. G. Cooke, D. Dorman, C. A. J. Kemp, S. R. Maple, and K. A. Mccune, J. Pharm. Sci., 89, 885 (2000).
52. L. Ljunggren, N. Vokova and H. Hansson, Int. J. Pharm., 202, 71 (2000).
53. H. Lucida, J. E. Parkin, and V. B. Sunderland, Int. J. Pharm., 202, 47 (2000).
54. J. T. Carstensen, J. B. Johnson, W. Valentine, and J. Vance, J. Pharm. Sci., 53, 1050 (1964).
55. K. J. Hartauer, G. N. Abuthnot, S. W. Baertschi, R. A. Johnson, W. D. Luke, N. G. Pearson, E. C. Rickard, C. A. Tiongle, P. K. S. Tsang, and R. E. Wiens, Pharm. Dev., Tech., 5, 303 (2000).
56. P. R. Perrier and U. W. Kesselring, J. Pharm. Sci., 72, 1072 (1983).
57. A. D. F. Toy, Inorganic phosphorous chemistry, in *Comprehensive Inorganic Chemistry* (J. C. Bailar, Jr., H. J. Emelius, R. Nyholm, and A. F. Trotman-Dickenson, Eds.), A. Wheaton and Co., Exeter, England, pp. 389–543.
58. Z. Zhivkova and I. Stankova, Int. J. Pharm., 200, 181 (2000).
59. J. E. Bodnar, J. R. Chen, W. H. Johns, E. P. Mariani, and E. C. Shinal, J. Pharm. Sci., 72, 535 (1983).
60. J. T. Carstensen, M. Franchini, and K. Ertel, J. Pharm. Sci., 81, 303 (1992).
61. J. T. Carstensen, *Drug Stability*, Marcel Dekker, New York, 1991, p. 60.
62. L. J. Ravin, E. S. Rattie, A. Peterson, and D. E. Guttman, J. Pharm. Sci., 67, 1528 (1978).
63. C. M. Won, T. E. Molnar, V. L. Windisch, and R. E. McKean, Int. J. Pharm., 190, 1 (1999).
64. M. Yanatan, B. Z. Chowdhry, I. C. Ashurst, M. J. Snowden, C. Davies-Cutting, and S. Gray, Int. J. Pharm., 200, 279 (2000).
65. Y. W. Chien, P. R. Keshary, Y. C. Huang, and P. P. Sarpotdar, Drug. Dev. Ind. Pharm., 72, 968 (1983).
66. J. T. Carstensen, in *Dissolution Technology* (L. Leeson and J. T. Carstensen, eds.), The Academy of Pharmaceutical Sciences, American Pharmaceutical Association, Washington, DC, 1974, p. 5.
67. S. Riegelman, Dissolution Testing in Drug Development and Quality Control, The Academy of Pharmaceutical Sciences Task Force Committee, American Pharmaceutical Association, 1979, p. 31.
68. K. Iba, E. Arakawa, T. Morris, and J. T. Carstensen, Drug Dev. Ind. Pharm., 17, 77 (1991).
69. C. J. Shishoo, S. A. Shah, I. S. Rathod, S. S. Savale, J. S. Kotecha, and P. B. Shah, Int. J. Pharm., 190, 109 (1999).
70. T. H. Wilson and G. Wiseman, J. Physiol., 123, 116 (1954).
71. R. K. Crane and T. H. Wilson, J. Appl. Physiol., 12, 145 (1958).
72. J. T. Dolusio, N. F. Billups, L. W. Dittert, E. J. Sugita, and J. V. Swintosky, J. Pharm. Sci., 58, 1196 (1969).
73. J. T. Carstensen, *Pharmaceutics of Solids and Solid Dosage Forms*, Wiley-Interscience, New York, 1977, pp. 99, 101.
74. J. T. Carstensen, J. L. Wright, K. W. Blessel, and J. Sheridan, J. Pharm. Sci., 67, 48 (1978).
75. J. T. Carstensen, J. L. Wright, K. W. Blessel, and J. Sheridan, J. Pharm. Sci., 67, 982 (1978).

Copyright © 2002 Marcel Dekker, Inc.

Chapter 8

Cutaneous and Transdermal Delivery—Processes and Systems of Delivery

Gordon L. Flynn

The University of Michigan, Ann Arbor, Michigan

I. INTRODUCTION

The skin forms the body's defensive perimeter against what is in reality the biologically hostile environment we humans live in. As such, in the normal course of living, it suffers more physical and chemical insult than any other tissue of the body. It is inadvertently scraped, abraded, scratched, bruised, cut, nicked, and burned. Insects bite it, sting it, and occasionally furrow through it. It is exposed to detergents, solvents, and myriad other chemicals and residues. Bacteria, yeasts, molds, and fungi live on its surface and within its cracks and crevices. It is brushed, smeared, dusted, sprayed, and otherwise anointed with toiletries, cosmetics, and drugs. Any of these exposures can rile the skin and/or provoke allergy. If there is only minor damage associated with such insults, the skin repairs itself in short order without a trace left of the injury. If the insult is severe, its reconstruction takes far longer and may occur with scarring. Such repair is essential in that humans cannot survive for long with an extensively damaged skin because it is the skin that keeps us from losing water and life's other essential chemicals to the external environment. In its intact state the skin is a formidable barrier, resistant to chemicals and tissue-harmful ultraviolet rays and virtually impenetrable to life-threatening microorganisms. To perform these necessary functions, the skin has to be tough and at the same time flexible, for it is stretched and flexed continually as we move around within it. In its healthy state it is thus a remarkable fabric, strong and far more complex than any man-made material [1].

Myriad medicated products are applied to the skin or readily accessible mucous membranes that in some way either augment or restore a fundamental function of the skin or pharmacologically modulate an action in the underlying tissues. Such products are referred to as *topicals* or *dermatologicals*. A *topical delivery* system is one that is applied directly to any external body surface either by inuncting it (spreading and rubbing in a semisolid with the fingers), by spraying or dusting it on, or by instilling it (applying a liquid as drops). Thus, "topical" is frequently used in contexts where the application is to the surface of the eye (cornea and conjunctival membranes), the external ear, the nasal mucosa or the lining of the mouth (buccal mucosa), or even the rectum, the vagina, or the urethral lining. The term "dermatological," on the other hand, describes products that are only to be applied to the skin or the scalp. An "external use only" label is used to denote such restricted use. The distinction between general topical use and external use is not trivial. Mucous membranes allow the rapid absorption of many locally applied substances that do not pass readily through intact, normal skin. This raises the potentials both drugs and excipients have to be locally irritating and/or systemically toxic. Consequently, many drugs can be applied to the external skin with impunity that should not be placed in contact with moist mucosal

Copyright © 2002 Marcel Dekker, Inc.

surfaces. Upon applying an external use only label to a package, the pharmacist warns the patient of such dangers. This chapter mainly deals with dermatological products, but general concepts and drug-delivery rationale are applicable to other modes of topical therapy as well.

The distinctions pharmacists have to make concerning topical dosage forms and their suitability for use obviously go far deeper than merely appreciating the significance of external use only labels. Pharmacological, toxicological, and risk-benefit valuations must be made for every dispensed product. An attending physician usually has such issues in mind when he or she prescribes. But even before the physician or pharmacist sees a product, the manufacturer has to establish its safety and effectiveness to the FDA's satisfaction. As part of the process, the delivery system for the drug is subjected to intense FDA scrutiny. Despite this regulatory oversight, the system isn't failsafe. Consequently, the dispensing pharmacist's lack of input into drug and drug-delivery system development does not abrogate his or her responsibility for assuring that every dispensed pharmaceutical conforms to high standards. The consummate professional thus takes every opportunity evaluate, by literature or by personal observation, how the various products he or she dispenses measure up to established standards. A pharmacist should be continually asking the following questions. Is the drug bioavailable as administered? Are the drug and its dosage form stable? Is the formulation free of contamination? Is the product pharmaceutically elegant? Attribute particulars of course vary by dosage form type and the route of administration. In any case, a "no" to any of the first three questions is reason to remove a product from distribution. Elegance is sometimes sacrificed relative to other function, but only in degree. A goal of this chapter is to elucidate the attributes of dermatological dosage forms that are helpful in evaluating and selecting products.

II. THE STRUCTURE AND FUNCTION OF SKIN

In order to answer questions regarding the therapeutic and cosmetic uses of the myriad dermatological concoctions available, a pharmacist must be knowledgeable about the anatomical structure and physiological functions of the skin and the chemical compositions and physicochemical properties of its constituent tissues. Some understanding of how its properties are affected by disease and damage is a must, as is knowledge of how the skin's physiology and function vary with age, race, environmental conditions, and other factors. Rational approaches to topical therapy rest on having such insights.

A. Skin Functions

General functions of the skin are outlined in Table 1. These functions include containment of tissues and organs, multifaceted protections, environmental sensing, and body temperature regulation. Some skin functions are inextricably entwined. For instance, containment and the barrier functions are to some extent inseparable. Active sweating is accompanied by increased peripheral blood flow, which in turn is tied in with greater nourishment of the cells of the skin as needed to promote their proliferation, differentiation, and specialization.

Let us consider how the skin is structured to better understand how this tissue performs some of its vital functions. Consider the cross section of the skin sketched in Fig. 1. This illustration shows the readily distinguishable layers of the skin, from the outside of the skin inwards: the ~10 μm thin, fully differentiated, devitalized outer epidermal layer called the stratum corneum; the ~100 μm thin live, cellular epidermis; and the ~1000 μm thin (1 mm thin) dermis. Note that all the thicknesses specified here are representative only, for the actual thickness of each stratum varies severalfold from place to place on the body. Dispersed

Table 1 Functions of the Skin

Containment of body fluids and tissues
Protection from harmful external stimuli (barrier functions)
Microbial barrier
Chemical barrier
Radiation barrier
Thermal barrier
Electrical barrier
Reception of external stimuli
Tactile (pressure)
Pain
Thermal
Regulation of body temperature
Synthesis and metabolism
Disposal of biochemical wastes (through secretions)
Intraspecies identification and/or attraction (apocrine secretions)
Blood pressure regulation

Source: Refs. 2 and 3.

Copyright © 2002 Marcel Dekker, Inc.

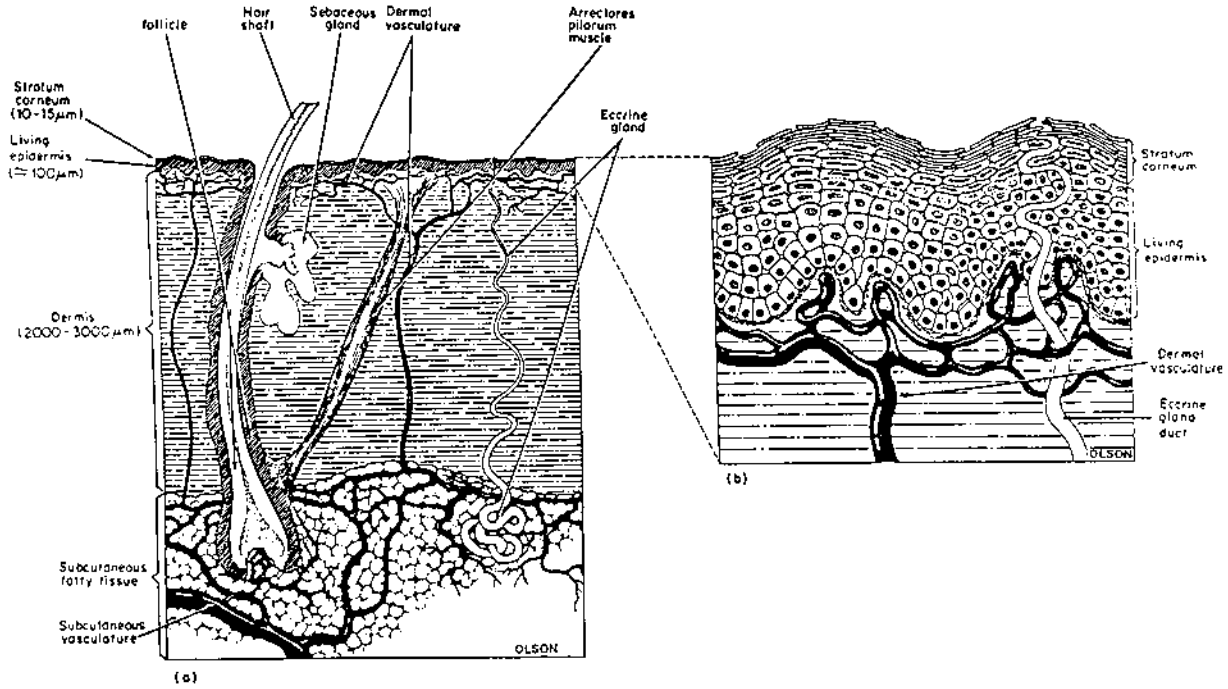


Fig. 1 Sketch of the skin.

throughout the skin, varying in number and size depending on body site, are several glands and appendages. These include (a) hair follicles and their associated sebaceous glands (pilosebaceous glands), (b) eccrine sweat glands, (c) apocrine sweat glands, and (d) nails of the fingers and toes. Each appendage has unique population densities and body distributions. The appendages also exhibit characteristic structural differences from one location to another on the body.

A highly complex network of arteries, arterioles, and capillaries penetrates the dermis from below and extends up to the surface of, but not actually into, the epidermis. A matching venous system siphons the blood and returns it to the central circulation. Blood flow through the vasculature is linked to the production and movement of lymph through a complementary dermal lymphatic system. The dermis is laced with tactile, thermal, and pain sensors.

B. Stratum Corneum

The outermost layer of the skin appearing in the exploded epidermal sketch of Fig. 1(b) represents the stratum corneum (the horny layer). The principal barrier element of the skin, it is an essentially metabolically inactive tissue comprised of acutely flattened, stacked, hexagonal cell–building blocks formed from once-living cells. These cellular building blocks are layered 15–25 cells deep over most of the body [2]. At some sites the cells appear to be stacked in neat columns. At other body locations the cellular arrangement appears random. The stratum corneum exhibits regional differences in thickness, being as thick as several hundred micrometers on the friction surfaces of the body (palms and soles). However, over most of the body, the tissue measures only about 10 μm thick, dimensionally less than a fifth of the thickness of an ordinary piece of paper [2,4]. It is a dense tissue, about 1.4 g/cm^3 in the dry state, a fact that has led it to also be referred to as the stratum compactum.

The stratum corneum is continuously under formation. Microscopic flakes (squamae) dislodged from the surface through wear and tear are replaced with new cells from beneath, with complete turnover of the horny layer occurring roughly every 2 weeks in normal individuals [5]. In humans the cells that give rise to the stratum corneum originate exclusively in the basal layer (also called basement, germinative, or proliferative layer) of the epidermis. Cell division in the basal layer begins an extraordinary process in which daughter cells are pushed outwards, first to form a layer of so-called spinous or prickle cells and then, serially, the granular, lucid, and horny layers. As suggested in Fig. 1, during their transit through the epidermal mass the cells flatten acutely. The protein

Copyright © 2002 Marcel Dekker, Inc.

and lipid components that eventually characterize the fully differentiated horny layer are synthesized in transit, with the initial strands of the structural protein that will eventually fill the cell interior being formed in the basal layer. During their inexorable migration to the stratum corneum, the protein content of the cells expands to the point that massed proteins of several kinds are distinguishable as they merge into the granular layer. Filagrin (*fila*ment agg*r*egating prote*in*), a basic protein that stains deeply, thereby giving the granular layer some of its characteristic histological appearance, is found here. Released in one of the culminating events of formation of the stratum corneum, filagrin induces individual helical strands of soluble protein to twist together into multistranded, slightly helical fibers. These fibers, in turn, spontaneously bundle into larger helical arrays. In culmination, the intracellular space of the fully differentiated horny cell is literally packed with structural protein, namely semi-crystalline α-keratin intermixed with more amorphous β-keratin. Nothing but keratin is visible inside the fully differentiated horny cell by electron microscope. The intracellular space is dense, offering little freedom of movement to organic molecules that may become dissolved within it. Moreover, because of its remarkable ionic character, the intracellular keratin mass borders on being thermodynamically impenetrable to organic molecules.

Lipid synthesized during a keratinocyte's epidermal transit is collected in small vesicles that become visible within the granular layer. These were designated membrane-coating granules long before their content and function were known. As granular cells undergo transformation, the "coating granules" gravitate to the outermost cell membrane and are passed exocytotically into the intercellular space. A lipid "mortar" is formed that seals the horny structure. Virtually all the lipid of the stratum corneum is in the interstitial space, much of it being present in liquid crystalline, bilayer assemblage [6]. The densely packed keratin platelets caulked with intercellular lipid make the stratum corneum, pound for pound, an incredibly resistive moisture barrier. An exoskeleton (infrastructure) of residual cell membranes bound together by desmosomes and tonofibrils acts to separate the keratin and lipid domains. The lipid content of the horny layer represents about 20% of the stratum corneum's dry weight, while the endoskeleton contributes roughly 5% to the weight (see Table 2) [2,5,7].

In its normal state at ordinary relative humidity, the stratum corneum takes up moisture to the extent of 15–20% of its dry weight [2]. The ionic character of keratin is certainly a factor of consequence here. Should the skin become waterlogged, the water content of horny tissue covering the friction surfaces (callused surfaces) can rise to several multiples of the tissue's dry weight. The water content of ordinary stratum corneum appears to be less affected under the same circumstance for it seems that the stratum corneum covering the greater part of the human body has less capacity to imbibe water. Nevertheless, all horny tissue becomes hydrated to some degree when the natural evaporation of water, so-called insensible perspiration, from the skin's surface is held in check upon applying an occlusive dressing. The horny tissue thus becomes more pliable. Consequently, molecules diffuse through it with greater facility. It is likely that some substances may exhibit greater solubility within the hydrated horny mass, adding further to their ease of permeation. Conversely, the stratum corneum becomes brittle when it dries out. Ultra-dry, inelastic horny tissue splits and fissures when stretched, giving rise to the common conditions we know as chapped lips, windburn, and dishpan hands.

The stratum corneum is thus a dense, polyphasic epidermal sheathing made from dehydrated and

Table 2 Composition of the Stratum Corneum

Tissue component	Gross composition	Percentage of dry weight
Cell membrane	Lipid, protein	~5
Intercellular space	Mostly lipid, some protein and polysaccharide	~20
Intracellular space	Fibrous protein (~65–70%), nonfibrous (soluble) protein (~5–10%)	~75
Overall protein	Water soluble (10%), keratin (~65%), cell wall (~5%)	70–80
Overall lipid		10–20
All other		Up to 10
Water (normal hydration)		15–20
Water (fully hydrated)		Upwards of 300

Source: Data from Refs. 2 and 4.

Copyright © 2002 Marcel Dekker, Inc.

internally filamented former cells held together by desmosomes, tonofibrils (intercellular anchors), and interstitial lipid. It has been estimated that it contains 10 times the fibrous material of the living epidermis in roughly one-tenth the space [8]. The stratum corneum is in contact with the living epidermal mass at its undersurface. Its external surface interfaces the environment. Cells at the basement of the stratum corneum contain water at the high thermodynamic activity of the physiological milieu, whereas air at the surface of the skin tends to have a far lower water activity. As a result, under ordinary circumstances, water diffuses out through the skin (down the implied activity gradient) and into the environment. This process is known as insensible perspiration. About 5 mL of water is lost this way per square meter of intact body surface per hour (or 0.5 mg/cm^2/h) [9]. An adult has upwards of 2 m^2 of body surface area, whereas a 2-year-old has about 0.75 m^2 and an infant roughly 0.25 m^2. Consequently, an adult loses about 250 mL (~8 oz) of water per day by insensible perspiration. A small child loses less than half this amount. Such water loss can increase to alarming proportions, as much or more than 100 mL/h/m^2, over skin ravaged by disease or damage. Evaporation rates as the latter over expansive areas of the skin can carry away sufficient body water and heat to gravely dehydrate a patient and lower his or her core body temperature to hypothermic levels.

C. Viable Epidermis

The animate cells of the epidermis make a sharp, upper interface with the lifeless stratum corneum. They also have a well-demarcated, deep interface with the dermis (Fig. 1). When physicochemically considered, the viable epidermis is nothing more than a wedge of tightly massed, live cells. Consequently, the whole of this live, cellular mass is regarded as a singular diffusional field (resistance) in percutaneous absorption models, although, when viewed under microscope, the tissue is clearly multilayered. The identifiable strata, bottom to top, are: (a) the basal layer (s. germinativum), a single layer of cubical or columnar cells that is unremarkable in appearance, (b) the multicellular spinous or prickle layer (s. spinosum) in which the cells exhibit sharp surface protuberances, and (c) the granular layer (s. granulosum), a thin layer that stains to yield a mottled appearance. In some histological displays a fourth, upper transitional and translucent layer is also distinguishable (s. lucidum). These layers reflect the progressive differentiation of the cells that eventuates in their death and placement as "bricks" within the horny structure. As elsewhere in the body, water found in the live epidermis has an activity equivalent to that of a highly dilute, isotonic NaCl solution (0.9% NaCl). The density and consistency of the live epidermal composite are only a little greater than found for water. The interface the viable epidermis makes with the stratum corneum is flat. However, the interface with the dermis is papillose (mounded). Myriad tiny bulges of the epidermis fit with exacting reciprocity over dermal depressions and ridges. It is these ridges that give the friction surfaces of the body their distinctive patterns (e.g., fingerprints). Importantly, since hair follicles and eccrine glands have epidermal origins, cells capable of regenerating the epidermis actually extend well into and through the dermis by way of these tiny glands (Fig. 1). When the skin is superficially injured, surviving cells at the base of these glands regenerate a scar-free surface. Discounting these deep rootages, the epidermis is on the order of 100–250 μm thick [10].

Keratinocytes account for most of the cells of the epidermis. One also finds Langerhans cells of white blood cell progeny here. The latter well-dispersed cells function as antigen-presenting cells (APCs) in the skin's immunological responses. Yet another group of cells, melanocytes, are strategically placed in the epidermis just above the epidermal-dermal junction (Table 3). Acting under the influence of melanocyte-stimulating hormone (MSH), these synthesize and deposit the pigment granules into skin that give the human races their unique skin colorations. Melanocytes are also set into action by ultraviolet radiation. Their activities in this circumstance lead to skin darkening. Migrant macrophages and lymphocytes are occasion-

Table 3 Cells of the Skin

Cell type	Principal function
Cells of the epidermis	
Keratinocytes	Form keratinized structures
Langerhans cells	Antigen presentation
Melanocytes	Pigment synthesis
Macrophages, lymphocytes	Migrant cells, immune responses
Cells of the dermis	
Fibroblasts	Fiber synthesis
Mast cells	Make ground substance, histamine
Blood cells	
Endothelial cells	Form the blood vessels
Nerve cells and endings	Sensors

Source: Data from Refs. 2 and 11.

Copyright © 2002 Marcel Dekker, Inc.

ally found in skin sections. Such cells can be numerous in traumatized skin.

D. Dermis

The dermis appears in Fig. 1 as a nondescript region lying between the epidermis and a region of subcutaneous fat. In reality it is a complex structure held together by a meshwork of structural fibers, collagen, reticulum, and elastin. Most of the space between fibers is filled with a mucopolysaccharidic gel called the ground substance [2]. Approximate proportions of these phases are indicated in Table 4. The dermis ranges from about 1 mm (1000 μm) to about 5 mm in thickness [11]. The upper one fifth or so of the wedge of tissue, the papillary layer by name, is finely structured and the support for the delicate capillary plexus which nurtures the epidermis. The papillary dermis merges into a far coarser fibrous matrix, the reticular dermis. This deepest layer of the true skin is the main structural element of the skin. Of considerable importance, the microcirculation that subserves the skin is entirely located in the dermis. The dermis is also penetrated by sensory nerve endings (pressure, temperature, and pain) and houses an extensive lymphatic network. Fibroblasts, cells that synthesize the structural fibers, are found here [2], and one also finds mast cells scattered about (Table 3). The latter are thought to synthesize the ground substance. They are known to be a source of histamine that is released when the skin is immunologically provoked.

E. Skin's Circulatory System

Arteries entering the skin arise from more substantial vessels located in the subcutaneous connective tissue. These offshoots form a plexus just beneath the dermis [11]. Branches from this subcutaneous network directly supply blood to the hair follicles, the glandular appendages, and the subcutaneous fat. Branches to the upper skin from this deep plexus divide again within the lower dermis, forming a deep subpapillary network. Arterioles reaching the upper dermis out of this plexus are on the order of 50 μm in diameter. They exhibit arteriovenous anastomoses, shunt-like connections that link the arterioles directly to corresponding venules. The dermal arterioles then further branch to form the shallower subpapillary plexus of capillary loops that bring a blood supply up into the papillae at the dermal/epidermal interface. The epidermis itself is avascular.

The veins of skin are organized along the same lines as the arteries in that there are both subpapillary and subdermal plexuses [11]. The main arteriole communication to these is the capillary bed. Copious blood is passed through capillaries when the core body is either feverish or overheated, far more than needed to sustain the life force of the epidermis, and this rich perfusion lends a red coloration to skin. When there is opposite physiological need, the capillary bed is short-circuited as blood is passed directly into the venous drainage by way of the arteriovenous anastomoses. Fair skin noticeably blanches when this occurs. These mechanisms act in part to regulate body temperature and blood pressure.

The vascular surface available for exchange of substances between the blood and local tissue has been estimated to be of the same magnitude as that of the skin, i.e., 1–2 cm^2 per square centimeter of skin. At room temperature about 0.05 mL of blood flows through the skin per minute per gram of tissue. This perfusion increases considerably when the skin is warmed [4,12]. Sufficient blood courses to within 150 μm of the skin's surface to efficiently draw chemicals into the body that have percutaneously gained access to this depth [7]. Blood circulation at this level is turned off by vasoconstrictors (e.g., glucocorticoids) and turned up by vasodilators (e.g., nicotine), respectively. These vasoactivities are so reliable that vasoconstriction (blanching) has become an FDA-sanctioned measure of the penetration of corticosteroids through the skin [13,14]. The relationships between capillary blood flow and local clearances of percutaneously absorbed drugs, including the influences of vasoconstriction and vasodilation, are not well drawn.

The lymphatic system of the skin extends up and into the papillary layers of the dermis. A dense, flat meshwork of lymphatic capillaries is found here [11]. Lymph passes into a deeper network at the lower boundary of the dermis. Serum, macrophages, and lymphocytes readily negotiate through the skin's lymphatic and vascular networks.

Table 4 Composition of the Dermis

Component	Approximate % composition
Collagen	75.0
Elastin	4.0
Reticulin	0.4
Ground substance	20.0

Source: Data from Refs. 2 and 11.

Copyright © 2002 Marcel Dekker, Inc.

F. Skin Appendages

Hair follicles and their associated sebaceous glands (pilosebaceous glands), eccrine glands, apocrine glands, and finger and toenails are all considered skin appendages. Hair follicles are found everywhere within the skin except for the soles of the feet, the palms of the hand, the red portion (vermilion border) of the lips, and the external genitalia. All are formed from fetal epidermal cells. Hair differs markedly in its prominence from place to place over the body. Delicate primary hair is found on the fetus; secondary hair or down covers the adult forehead; terminal hair ordinarily blankets the scalp and is found as pubic and axillary (underarm) hair [2]. A hair (hair shaft) emerges from a follicle, as shown in Fig. 1. Each follicle is set within the skin at a slight angle. Each consists of concentric layers of cellular and noncellular components positioned on a slight angle. Each is anchored to the surrounding connective tissue by an individual strand of smooth muscle, the arrector pilorum. Contraction of this muscle causes the hair to stand upright, merely raising goose pimples on human skin. In animals like cats, hair stands on end and makes the animal more imposing as part of the fight-or-flight response.

The hair shaft is formed continuously by cell division, differentiation, and compaction within the bulb (base) of each active hair follicle, a process that is completed deep in the follicle. Hair, like stratum corneum, is thus a compact of fused, keratinized cells. Collectively, hair follicles occupy about one thousandth of the skin's surface [10,15], a factor that sets a limit on the role that follicular orifices can play as a route of penetration. Each hair follicle possesses one or more flask-like sebaceous glands (Fig. 1). These have ducts that vent into the open space surrounding the hair shaft just below the skin's surface. The cells of sebaceous glands, or sebocytes, are programmed to divide, differentiate, and die. Before they die and disintegrate, they pack themselves full of lipid-containing vesicles. The residue left behind at their death is mixed with other follicular debris below the follicular orifice to form an oily substance called sebum. Sebum is then forced upwards around the hair shaft and onto the skin surface. The follicular outlets for sebum exhibit diameters ranging from 200 to 2000 μm (2 mm) depending on body location [11]. Glands with the largest openings are found on the forehead, face, nose, and upper back. These large follicles contain an almost microscopic hair, if they contain one at all, and are therefore referred to as sebaceous follicles. With about 100 follicles per cm^2 of skin, the follicular route takes up roughly a thousandth of the skin surface.

Eccrine glands (salty sweat glands) are found over the entire body except the genitalia and lips. They appear as tubes extending from the skin surface all the way to the footings of the dermis. Here the tube coils into a ball roughly 100 μm in diameter (Fig. 1) [11]. By anatomical count, there are between 150 and 600 glands per cm^2 of body surface depending on body site [16]. They are particularly concentrated in the palms and soles, attaining densities in these locations well in excess of 400 glands/cm^2. However, since many of these glands remain dormant, estimates of their numbers are appreciably lower if based on actual sweating units. Each gland has an approximately 20 μm diameter orifice at surface of the skin from which its secretions are spilled. In total, these glandular openings represent approximately one ten-thousandth of the skin's surface [10]. Eccrine sweat is a dilute (hypotonic), slightly acidic (pH$\sim$5.0 due to traces of lactic acid) aqueous solution of salt. Its secretion is stimulated when the body becomes overheated through warm temperatures or exercise. Evaporation of the water of the sweat cools the body's surface and thus the body. Since the gland is innervated by the autonomic nervous system, eccrine sweating is also stimulated emotionally (the clammy handshake).

Apocrine glands are found only in the axillae (armpits), in the anogenital region, and around the nipples. Along with other secondary sexual characteristics, these glands develop at puberty. We know that they are innervated emotionally and through concupiscence. In the mature female they exhibit cyclical activities in harmony with the menstrual cycle. Like eccrine glands, they are coiled tubular structures, but the coils are roughly 10 times larger. They extend well into the subcutaneous layer beneath the skin [2,11]. Each gland is paired with a hair follicle; its secretion is vented into the sebaceous duct of the follicle beneath the surface of the skin. Because this secretion is minor in amount and combines with sebum before reaching the skin's surface, its exact chemical make-up remains an enigma. What is not a mystery is that bacterial decomposition of the secretion is responsible for human body odor.

III. SKIN FUNCTIONS

The skin's main physiological roles are outlined in Table 1. Of these, the chemical barrier function is central to the use of topical drugs because deposition

Copyright © 2002 Marcel Dekker, Inc.

of a topical drug into the deeper, living strata of skin is a prerequisite for achieving its pharmacological effect. Degeneration in some of the functions can be pathognomonic of disease. Even where specific functions do not relate materially to the skin's state of health, they are tied in with cosmetic practices and thereby are of interest to pharmacists.

A. Containment

The containment function relates specifically to the ability of the skin to confine underlying tissues and restrain their movements. The skin draws the strength it needs to perform this mechanical role from its tough, fibrous dermis [2]. Ordinarily, the skin is taut even when under resting tension, yet it stretches easily and elastically when the body is in motion, quickly returning to normal contours when the stretching ceases. This extensibility of the skin is attributable to an alignment of collagen fibers under tension and in the direction of a load, which are otherwise nonaligned in the ground tension state. Elastin fibers attached to individual collagen strands relax and, in doing so, restore the irregular order of the restive state. As one ages, the resilience of these dermal fibers decreases and the tensile strength of the tissue increases. Eventually, the skin becomes stretched beyond its ability to elastically restore its initial condition and it folds over itself, or wrinkles. Lost elasticity is advanced through extended exposure to ultraviolet (UV) radiation (sunlight), and thus wrinkling is often pronounced on dedicated sunbathers.

The behavior of the epidermis when distended is also of importance. Obviously, this layer should not be torn or broken when placed under mechanical stress, for an intact epidermis is the body's first line of defense against infection. It is the stratum corneum's role to fend against tearing [2]. Pound for pound, this tissue is actually stronger than the dermal fabric, and, as a rule, it is sufficiently elastic to adjust to stretching. Its pliability, however, is conditional, and it fissures and cracks if stretched when excessively dry. Arid atmospheres alone can produce this condition (windburn). Detergents and solvents, which extract essential, water-sequestering lipids from the stratum corneum, and diseases such as psoriasis associated with a malformed horny structure render the stratum corneum brittle and prone to fissuring.

Although much is still to be learned about the factors that contribute to the pliability of the stratum corneum, it is generally accepted that its elasticity is dependent on a proper balance of lipids, hygroscopic, water-soluble substances, and water, all in conjunction with its keratin proteins. Water is its principal plasticizer, or softening agent, and it takes roughly 15% moisture to maintain adequate pliability. The capacity of the stratum corneum to bind and hold onto water is greatly reduced by extracting it with lipid solvents such as ether and chloroform. Moreover, there is a further significant decrease in the water-binding capacity of callus, a thickened stratum corneum found on the palms and soles, when it is extracted with water after having first been treated with a lipid solvent. The latter observation seems to tell us that amino acids, hydroxy acids, urea, and inorganic ions, cosmetically referred to at the skin's natural moistening factor, and the stratum corneum's lipids both assist the stratum corneum in retaining moisture necessary to plasticize its mosaic, filamented matrix. In effect, the water makes the tissue less crystalline through its interposition between polymer strands.

B. Microbial Barrier

Normal stratum corneum, taken in its entirety, is a dense molecular continuum penetrable only by molecular diffusion. It is virtually an absolute barrier to microbes, preventing them from reaching the viable tissues and an environment suitable for their growth. The outermost stratum corneum is continuously being shed in the form of microscopic scales (natural desquamation) and to a limited depth is laced with tiny crevices. Many microorganisms, pathogens and harmless forms alike, are found in these rifts. Surgeons know well that superficial washing is insufficient to remove these surface microbes, and therefore the surgical scrub is an energetic and intense cleansing with a disinfectant soap. The microorganisms residing on and in the skin can and do initiate infections if seeded into living tissues as a result of abrasive or disease-induced stratum corneum damage. Consequently, antiseptics and antibiotics are widely used to chemically sanitize wounds.

Beyond physical barrier protection, several natural processes lead to skin surface conditions unfavorable to microbial growth. Both sebaceous and eccrine secretions are acidic, lowering the surface pH of the skin below that welcomed by most pathogens. This acid mantle (pH$\sim$5) [16] is moderately bacteriostatic. Sebum also contains a number of short-chain fungistatic and bacteriostatic fatty acids, including propanoic, butanoic, hexanoic, and heptanoic acids [17]. That the skin's surface is dry also offers a level of protection. It comes as no surprise that fungal infections and other skin infections are more prevalent in the skin's folds

Copyright © 2002 Marcel Dekker, Inc.

during warm weather as intensified sweating leaves the skin continually moist in these regions.

Glandular orifices provide possible entry points for microbes. The duct of the eccrine sweat gland is tiny and generally evacuated. Experience tells us that this is not an easy portal of entry, although localized infection is seen occasionally in infants suffering prickly heat. Pilosebaceous glands seem more susceptible to infection, particularly those on the forehead, face, and upper back, referred to as sebaceous follicles. Glands at these specific locations have an almost imperceptible hair surrounded by a massive sebaceous apparatus and are especially prone to occlusion and subsequent infection (acneform pimples and blackheads). Such sebaceous gland infections are usually localized. However, if the infected gland ruptures and spews it contents internally, deep infection is possible. The body defends against this by walling off the lesion (forming a sac, or cyst) and then destroying and eliminating the infected tissue. The destruction of cystic acne is deep, so much so that facial scarring is associated with it. In hair follicles containing prominent hairs, the growing hair shaft acts as a sebum conveyer, which unblocks the orifice. It may be strictly coincidental, but such follicles seem less prone to clogging and infection.

C. Chemical Barrier

The intact stratum corneum also acts as a barrier to chemicals brought into contact with it. Its diffusional resistance is orders of magnitude greater than found in other barrier membranes of the body. Externally contacted chemicals can, in principle, bypass the stratum corneum by diffusing through the ducts of the appendages. The ability of each chemical to breach the skin and the diffusional route or routes it takes are dependent on its own physicochemical properties and the interactions it has within the skin's various conduit regimes. Being central to the effectiveness of dermatological products, exposition of the skin's barrier properties is made in following sections.

D. Radiation Barrier

Ultraviolet wavelengths of 290–310 nm from the UV-B band of radiation constitute the principal tissue-damaging rays of the sun, which are not fully atmospherically filtered. An hour's exposure to the summer sun and its damaging rays can produce a painful burn with a characteristic erythema. The skin has natural mechanisms to prevent or minimize such sun-induced trauma, but it takes time to set these into place. Upon stimulation by ultraviolet rays, particularly longer, lower-energy rays above 320 nm, melanocytes at the epidermal-dermal junction produce the pigment melanin. Melanin's synthesis begins in the corpus of the melanocyte, forming pigment granules that migrate outwards to the tips of the long protrusions of these star-like cells. Adjacent epidermal cells endocytotically engulf these projections. Through this cellular cooperation melanin, which absorbs and diffracts harmful UV rays, becomes dispersed throughout the epidermis and a person tans, with his or her capacity to sunburn declining accordingly. It should be realized that tanning takes time, several days in fact, and is incapable of protecting a person on first exposure. Damaging ultraviolet exposure also stimulates epidermal cell division and thickening of the epidermis (acanthosis). Such thickening, too, takes several days. When effected, it also lends protection to the underlying tissues.

Pharmacists should tell their sun-deprived, fair-skinned patrons not to spend more than 15–20 minutes in the mid-day sun (10:00 a.m. to 3:00 p.m.) on first exposure when traveling to vacation spots such as Florida [18]. This is ample, safe exposure to initiate the tanning response in those who are able to tan. Exposures can be increased incrementally by 15 minutes a day until a 45-minute tolerance is developed, which is generally an adequate level of sun protection in conjunction with the use of sunscreens. It should be obvious that dark-skinned people are already heavily pigmented and thus far less susceptible to burning. Other individuals don't tan at all and must apply sunscreens with high protection factors before sun bathing.

E. Electrical Barrier

Dry skin offers high impedance to the flow of an electrical current [3]. Stripping the skin by successively removing layers of the stratum corneum with an adhesive tape reduces the electrical resistance about sixfold, which tells us that the horny layer is the skin's prime electrical insulator. Its high impedance complicates the measurement of body potentials, as is done in electroencephalograms and electrocardiograms. Consequently, electrodes having large contact areas are used to monitor the brain's and the heart's electrical rhythms. Granular salt suspensions or creams and pastes containing high percentages of electrolytes are placed between the electrode surface and the skin to assure that electrical conductance is adequate to make the measurements.

Copyright © 2002 Marcel Dekker, Inc.

F. Thermal Barrier and Body Temperature Regulation

The body is basically an isothermal system fine-tuned to 37°C (98.6°F). The skin has major responsibility in temperature maintenance. When the body is exposed to chilling temperatures that remove heat faster than the body's metabolic output can replace it, changes take place in the skin to conserve heat. Conversely, when the body becomes overheated, physiological processes come into play that lead to cooling.

The skin's mechanism of heat conservation involves its very complex circulatory system [2,3]. To conserve heat, blood is diverted away from the skin's periphery by way of the arteriovenous anastomoses. The externalmost circulation of the blood is effectively shut down, leading to a characteristic blanching of the skin in fair-skinned individuals. Less heat is irradiated and convectively passed into the atmosphere. Furry mammals have yet another mechanism to conserve body heat. Each tiny arrector pilorum stands its hair up straight, adding appreciable thickness to the insulating air layer entrapped in the fur, reducing heat loss.

When the body is faced with the need to cast out thermal energy, the circulatory processes are reversed and blood is sent coursing through the skin's periphery, maximizing radiative and convective heat losses. This process produces a reddening in light skin, a phenomenon that is particularly noticeable following strenuous exercise. Exercise also leads to profuse eccrine sweating, a process that is even more efficient in heat removal. Watery sweat evaporates, with the heat attending this process (heat of vaporization) cooling the skin's surface. Factors that accelerate the evaporative process, such as the gentle flow of air produced by a fan, accelerate cooling. Low humidity favors evaporation, and one is more comfortable and sweating is less noticeable when the air's moisture content is low. Pharmacists should be aware that eccrine sweating is a vital process not to be tampered with. Coverage of the body with a water-impermeable wrapping, as has occasionally been done in faddish weight control programs, may result in hyperthermia, particularly if there is concurrent exercise. In its extreme, hyperthermia can be fatal.

IV. RATIONALE FOR TOPICALS

One's grasp of topical dosage forms and their functioning can be nicely organized into several broad usage categories. For instance, many products exist to augment the skin barrier (Table 5). Sunscreens and anti-infective drugs obviously do this. The barrier is made pliable and restored in function by emollients. Pastes are sometimes used to directly block out sunlight and at other times to sequester irritating chemicals that would otherwise penetrate into the skin. Even insect repellants add function to the barrier.

Table 5 Barrier Augmentation by Topical Products

Product type	Barrier effect
Sunscreens	Enhance radiation barrier
Topical anti-infectives	Augment microbial barrier
Emollients	Moisturize stratum corneum, restore barrier
Insect repellents	Add a chemical barrier to insects
Poison ivy products	Negate antigens, augment chemical barrier
Diaper rash products	Build up moisture and chemical barrier

A second general purpose of topical application involves the selective access drugs have to epidermal and dermal tissues when administered this way. Penetration of the skin can drench the local tissues with the drug prior to the systemic dissemination and dilution. As a result, the drug's systemic levels are kept low and pharmacologically inconsequential. In contrast, systemic treatment of local conditions bathes highly blood-perfused tissues with the drug first, with the drug's systemic effects or its side effects sometimes overpowering the actions sought for it in the skin.

In a few instances, drugs are applied to the skin to actually elicit their systemic effects. This is called transdermal therapy. Transdermal therapy is set apart from local treatment on several counts. It is only possible with potent drugs that are also highly skin permeable. To be used transdermally, compounds must be free of untoward cutaneous actions as well. When these demanding conditions are met, transdermal therapy offers an excellent means of sustaining the action of a drug. Transdermal delivery also skirts frequently encountered oral delivery problems such as first-pass metabolic inactivation and gastrointestinal (GI) upset. Transdermal therapy is actually an old medical strategy, as compresses and poultices have been used for centuries, although never with certainty of effect. The current effective use of small adhesive patches to treat systemic disease or its symptoms has revolutionized the practice.

Copyright © 2002 Marcel Dekker, Inc.

A. Therapeutic Stratification of the Skin

How does a person best organize his or her thinking relative to these different rationales? One can start by asking what the topical drug is supposed to do. Is it to be applied to suppress inflammation? Eradicate infectious microorganisms? Provide protection from the sun? Stop glandular secretions? Provide extended relief from visceral pain? Regardless of which feat the drug is to perform, the answer to the question directs us to where and sometimes how the drug must act to be effective or to the target for the drug. Once knowing the locus of action, one can then consider its accessibility. Clearly, if the drug cannot adequately access its target, little or no therapeutic benefit will be realized.

Sundry drug targets exist on, within, or beneath the skin. These include: (1) the skin surface itself (external target), (2) the stratum corneum, (3) any one of several levels of the live epidermis, (4) the avascular, upper dermis, (5) any one of several deeper regions of the dermis, (6) one or another of the anatomically distinct domains of the pilosebaceous glands, (7) eccrine glands, (8) apocrine glands, (9) the local vasculature, and, following systemic absorption, (10) any of numerous internal tissues. As these targets become increasingly remote, delivery to them becomes sparser as the result of dilution via tissue distribution, and, consequently, adequacy of delivery becomes less certain. Moreover, the specific properties of these targets and their negotiability are very much determined by the state of health of the skin. Disease and damage alter the barrier characteristics of the skin and therefore target accessibility itself.

Causes of skin damage and/or eruptions are diverse and may alternatively be traced to damage, irritant or allergic reactions, an underlying pathophysiological condition, or an infection. Depending on the problem, the entire skin or only a small part of it may be involved. Moreover, disease may be manifest in one part of a tissue as a consequence of a biochemical abnormality in another. For instance, the cardinal expression of psoriasis is its thickened, silvery, malformed stratum corneum (psoriatic scale), but the disease actually results from maverick proliferation of keratinocytes in the germinal layer of the epidermis. Humankind suffers many skin problems, each unique in expression to the well-trained eye. The names of some common afflictions are listed in Table 6, with indications of the tissue source of the problems. Table 7 adds pathophysiological terms used to describe the expressions of disease to the lexicon. Irrespective of their fundamental tissue origins, most diseases fan out and involve other tissue components. Inflammation and skin eruption are common sequelae. The nature and developing pattern of a skin eruption become the determinants of its diagnosis. There are subtleties, and it often takes a dermatologist to make a proper differential diagnosis.

The pharmacist will, from time to time, be called upon to examine an eruption or condition and make recommendation for treatment. If and only if the condition is unmistakable in origin, delimited in area, and of modest intensity should the pharmacist recommend an over-the-counter remedy for its symptomatic relief. Physicians neither need nor want to see inconsequential cuts, abrasions, or mosquito bites or unremarkable cases of chapped skin, sunburn, or poison ivy eruption, and so on. However, if infection is present and at all deep-seated or if expansive areas of the body are involved, otherwise minor problems can pose a serious threat and physician referral is mandatory. Patients should also be directed to counsel with a physician whenever the origins of a skin problem are in question.

B. Surface Effects

Of the many possible dermatological targets mentioned above, the skin surface is clearly the easiest to access. Surface treatment begins at the fringe of cosmetic practice. Special cosmetics are available to hide unsightly blemishes and birthmarks. These lessen self-consciousness and are psychologically uplifting. Applying a protective layer over the skin is sometimes desirable. For example, zinc oxide pastes are used to create a barrier between an infant and its diaper that adsorbs irritants found in urine, ameliorating diaper rash. These same pastes literally block out the sun and at the same time hold in moisture, protecting the ski enthusiast from facial sunburn and windburn on the high slopes. Transparent films containing ultraviolet light–absorbing chemicals are also used as sunscreens. Lip balms and like products lay down occlusive (water-impermeable) films over the skin, preventing dehydration of the underlying stratum corneum and thereby allaying dry skin and chapping. The actions of calamine lotion and other products of the kind are limited to the skin's surface. The suspended matter in these purportedly binds urushiol, the hapten (allergen) found in poison ivy and oak. However, these may best benefit the patient by drying up secretions, relieving itchiness. In all these instances where the film itself is therapeutic, bioavailability has little meaning.

Copyright © 2002 Marcel Dekker, Inc.

Table 6 Common Afflictions: Brief Outline of Common Dermatological Disorders and Other Common Skin Problems

Skin problems	Examples
I. General involvements	
A. Physical damage	
1. Blunt instrument	Contusion, bruise
2. Sharp instrument	Cut, nick, animal bite
3. Scraping, rubbing	Abrasion, blister
4. Heat	Burns (1°, 2°, 3°), blister
5. Ultraviolet radiation	Sunburn
6. Insects	Mosquito bite, bee sting, ticks, mites (chiggers), lice, crab lice
B. Chemical damage	
1. Contact dermatitis	Poison ivy, poison oak
2. Contact allergy	Cosmetic dermatitis
3. Solvent extraction	"Dishpan hands"
II. Abnormalities of the epidermis	
A. Stratum corneum	
1. Tardigrade sloughing and thickening	Ichthyosis
2. Hyperdryness	Chapping, windburn
3. Hyperproliuerative thickening, abnormal structural organization	Psoriasis
B. Viable epidermis	
1. Cell damage and inflammation	Eczema, general dermatitis
2. Fluid collection	Blister
3. Abnormal cell growth (not division)	Keratosis
4. Thickening of granular layer	Lichen planus
5. Hyperproliferation, incomplete keratinization	Psoriasis
6. Malignancy	Epithelioma
III. Abnormalities of the dermis and dermal epidermal interface	
A. Melanocyte abnormalities	
1. Hyperfunction	Tanning, chloasma, freckles
2. Hypofunction	Vitiligo
3. Abnormal growth	Mole
4. Malignancy	Melanoma
B. Dermal/epidermal interface	
1. Lifting of the epidermis	Dermatitis hypetiformis
2. Overgrowth of papillary layer	Warts
C. Dermis	
1. Vascular reactions	Urticaria, hives
2. Abnormal growth of fibrinocyte	Scar, keloid
3. Abnormal polymerization	Scieroderma, lupus erythematosus
IV. Abnormalities of the glands (appendages)	
A. Hair follicle	
1. Hyperactivity	Hirsutism
2. Hypoactivity	Alopecia, baldness
B. Sebaceous glands	
1. Hyperactivity	Seborrhea
2. Occlusion	Acne, pimples
C. Eccrine sweat gland	
1. Hyperactivity	Hyperhidrosis
2. Occlusion, inflammation	Miliaria (pricky heat, heat rash)
V. Infectious diseases	
A. Bacterial	Carbuncles (boils)
B. Fungal	Athlete's foot, ringworm
C. Viral	Chickenpox, herpes simplex (cold sores)
D. Protozoal	Topical amebiasis

Source: Refs. 16 and 19.

Copyright © 2002 Marcel Dekker, Inc.

Table 7 Pathophysiological Terms: Brief Definitions of Select Pathophysiological Terms

Term	Definition
Acne	Inflammatory disease of the sebaceous glands characterized by papules, comedones, pustules, or a combination thereof
Alopecia	Deficiency of hair
Bulla	Large blister or vesicle filled with serous fluid
Chloasma	Cutaneous discoloration occurring in yellow-brown patches and spots
Comedo (p1. comedones)	Plug of dried sebum in the sebaceous duct; blackhead
Dermatitis	Inflammation of the skin
Dermatitis herpetiformis	Dermatitis marked by grouped erythematous, papular, vesicular, pustular, or bullous lesions occurring in varied combinations
Eczema	An inflammatory skin disease with vesiculation, infiltration, watery discharge, and the development of scales and crusts
Hirsutism	Abnormal, heavy hairiness
Ichthyosis	A disease characterized by dryness, roughness, and scaliness of the skin caused by hypertrophy of the stratum corneum
Infiltration	An accumulation in a tissue of a foreign substance
Keloid	Growth of the skin consisting of whitish ridges, nodules, and plates of dense tissue
Keratosis	Any horny growth
Lichen planus	Inflammatory skin disease with wide, flat papules occurring in circumscribed patches
Lupus erythematosus	A superficial inflammation of the skin marked by disklike patches; with raised reddish edges and depressed centers, covered with scales or crusts
Miliaria	An acute inflammation of the sweat glands, characterized by patches of small red papules and vesicles, brought on by excessive sweating
Nodule	Small node that is solid to the touch
Papilla	Small, nipple-shaped elevation
Psoriasis	A skin disease characterized by the formation of scaly red patches, particularly on the extensor surfaces of the body (elbows, knees)
Pustule	Small elevation of the skin filled with pus
Scleroderma	A disease of the skin in which thickened, hard, rigid, and pigmented patches occur with thickening of the dermal connective tissue layer
Seborrhea	A disease of the sebaceous glands marked by excessive discharge
Urticaria	Condition characterized by the appearance of smooth, slightly elevated patches, whiter than the surrounding skin
Vesicle	Small sac containing fluid; a small blister
Vitiligo	A skin disease characterized by the formation of light-colored (pigment-free) patches

Copyright © 2002 Marcel Dekker, Inc.

Bioavailability does matter with topical antiseptics and antibiotics even though these also act mainly at the skin's surface. These anti-infectives are meant to stifle the growth of surface microflora, and thus formulations that penetrate into the cracks and fissures of the skin where the microorganisms reside are desirable. The extent to which the surface is sanitized then depends on uptake of the anti-infective by the microbes themselves. Slipshod formulation can result in a drug being entrapped in its film and inactivated. For instance, little to no activity is to be expected when a drug is placed in a vehicle in which it is highly insoluble. Ointment bases that contain salts of neomycin, polymyxin, and bacitracin are suspect in this regard in that hydrocarbon vehicles are extremely poor solvents for such drugs. Inunction (rubbing in) may release such drugs, but pharmacist should seek evidence that such formulations are effective before recommending them.

Deodorants are also targeted to the skin surface to keep microbial growth in check. They slow or prevent rancidification of the secretions of apocrine glands found in and around the axillae (armpits) and the anogenital regions. Medicated soaps also belong in this family.

C. Stratum Corneum Effects

The stratum corneum is the most easily accessed part of the skin itself, and there are two actions targeted to this tissue, namely emolliency, the softening of the horny tissue, which comes about through remoisturizing it, and keratolysis, the chemical digestion and removal of thickened or scaly horny tissue. Tissue needing such removal is found in calluses, corns, and psoriasis and as dandruff. Common agents such as salicylic acid and, to a lesser extent, sulfur cause lysis of the sulfhydral linkages holding the keratin of the horny structure together, leading to its disintegration and sloughing.

It has been mentioned that elasticity of the stratum corneum depends on its formation and on the presence of adequate natural lipids, hygroscopic substances, and moisture [19,20]. Simply occluding the surface and blocking insensible perspiration can induce remoisturization (emolliency). However, it is best accomplished by lotions, creams, and/or waxy formulations (e.g., lip balms), which replenish lost lipid constituents of the stratum corneum. The fatty acids and fatty acid esters these contain in part fill the microscopic cracks and crevices in the horny layer, sealing it off, stabilizing its bilayer structures, allowing it to retain moisture. Many emollient products also contain hygroscopic glycols and polyols to replenish and augment natural moisturizing factors of this kind, also assisting the stratum corneum in retaining moisture.

The introduction of moisturizing substances into the stratum corneum is ordinarily a straightforward process. Deposition of keratolytics, on the other hand, is not as easily achieved, as these agents must penetrate into the horny mass itself. Some salicylic acid–containing corn removers are therefore made up as concentrated nonaqueous solutions in volatile solvents. As these volatile solvents evaporate, drug is concentrated in the remaining vehicle and thereby thermodynamically driven into the tissue. These many examples illustrate the fact that when the therapeutic target is at the skin's surface or is the stratum corneum, the therapeutic rationale behind the treatment usually involves enhancing or repairing or otherwise modulating barrier functions (Table 5).

D. Drug Actions on the Skin's Glands

A few products moderate operation of the skin's appendages. These include antiperspirants (as opposed to deodorants), which use the astringency of chemicals such as aluminum chloride to reversibly irritate and close the orifices of eccrine glands [21] to impede the flow of sweat. Astringents also decimate the population of surface microbes, explaining their presence in deodorants. The distinction between antiperspirants and deodorants is legally significant because antiperspirants alter a body function and are regulated as drugs while deodorants are classified as cosmetics. Thus, the antiperspirant action has to be scientifically proven before it can be claimed for a product. Nevertheless, given the similarities in the compositions of deodorants and antiperspirants, they are likely functionally equivalent. Since eccrine glands are mediated by cholinergic nerves, sweating also can be shut off by anticholinergic drugs administered systematically [22] or topically. However, such drugs are too toxic for routine use as antiperspirants even when administered topically.

Acne is a common glandular problem arising from hyperproliferative closure of individual glands in the unique set of pilosebaceous glands located in and around the face and across the upper back. Irritation of cells lining the ducts of such glands initiates the formation of lesions. Sheets of sloughed, sebum-soaked, keratinized cells that grow out from the walls surrounding the sebaceous duct are what clog the duct. Still forming sebum is then trapped behind the obstruction, often bulging out the skin and giving rise to an observable lesion (papule). This may become

Copyright © 2002 Marcel Dekker, Inc.

infected and fill with a purulent exudate (pustule), or, after infection has set in, it may internally rupture and thereby begin the processes that lead to ulceration and scarring. Alternatively, the buildup of concentric sheets of sloughed cells may widen the glandular opening, with melanin in the widening plug darkening to the point of being black (blackhead).

Soap and water is considered a therapeutic treatment in acne when it is used to unblock the pores. Sebum is emulsified and it and other debris is removed. Alcoholic solvents, often packaged as moistened pledgets, are used for the same purpose. With either treatment care must be exercised not to dry out and further irritate the skin. Both local and systemic antibiotics and antiseptics suppress the formation of lesions. It is believed these attenuate the population of anerobic microorganisms that are deep-seated in the gland, the metabolic byproducts of which irritate the lining of the gland, setting off lesion formation. Mild cases of acne improve and clear under the influence of astringents, possibly for the same reason. Retinoids, oral and topical, reset the processes of epidermal proliferation and differentiation. Through such dramatic influences on cell growth patterns, they actually prevent the formation of lesions. However, because of concerns over toxicity, they tend to be used only in the most severe cases of acne and thus are prescribed for those patients whose acne lesions progress to cysts.

Hair is a product of the pilosebaceous apparatus and in this sense is glandular. It often grows out visibly in places where such display is unwanted. It may be shaved, but chemical hair removers (depilatories) along with other products are also used to remove it. In the main, the use of depilatories is cosmetic rather than therapeutic. However, depilatories may be prescribed in hirsutism, where the existence of coarse, dark facial hair is psychologically distressing to the female patient for whom shaving is an anathema. Hair, like the stratum corneum, is composed of layers of dead, keratinized cells. However, its keratin is more susceptible to the action of keratolytics because it is structured in ways that make it more chemically pervious. Thioglycolate-containing, highly alkaline creams generally dissolve hair in short order without doing great harm to surrounding tissues. Facial skin is delicate, however, and depilatories must be used carefully here.

E. Effects in Deep Tissues

Local, Regional, and Systemic Delivery

When the target of therapy lies beneath the stratum corneum, topical drug delivery is more difficult and becomes more uncertain. Therefore, many potentially useful drugs find no place in topical therapy due to their inability to adequately penetrate the skin. Nevertheless, a number of pathophysiological states can be controlled through local administration and subsequent percutaneous absorption. For example, most skin conditions are accompanied by inflammation of the skin; topical corticosteroids and nonsteroidal anti-inflammatory drugs alike are used to provide symptomatic relief in such instances. Corticosteroids are also used in psoriasis where, in addition to suppressing inflammation, they somehow act on the basal epidermal layer to slow proliferation and restore the skin's normal turnover rhythm [23]. Pain originating in the skin can be arrested with locally applied anesthetics. Over-the-counter (OTC) benzocaine and related prescription drugs are used for this purpose. Hydroquinone is applied to the skin to lighten excessively pigmented skin by oxidizing melanin deep within the surface. Another treatment that involves percutaneous absorption is the application of 5-fluorouracil (5-FU) for the selective eradication of premalignant and basal cell carcinomas of the skin [22]. In all of these examples the key to success is the ability we have to get therapeutic amounts of the drugs through the stratum corneum and into the viable tissues.

Systemic actions of some drugs can also be achieved via local application, in which case their delivery is known as transdermal delivery. The application of warmed, soft masses of medicated bread meals and clays over wounds or aching parts of the body dates to antiquity. However, few such plasters and poultices (cataplasms) have survived into modern medicine. Just past the middle of the twentieth century improvements in analytical instrumentation made it possible to measure the exceedingly low circulating levels of drugs that build up in the body during the course of therapy. It was at this time that serious research was begun on the ways drugs might be delivered to lower their risk and extend their duration of action. Novel delivery systems involving nontraditional routes of administration were subsequently conceived, constructed, and put to the test. Transdermal delivery with adhesive patches evolved as one of the innovations.

The possibilities for transdermal delivery might have been seen long ago in the systemic toxicities of certain topically contacted chemicals. As long as a century ago it was known that munitions workers who handle nitroglycerine suffer severe headaches and ringing in the ears (tinnitus). These same effects are experienced to a degree by those taking nitroglycerine to alleviate angina. The association between therapy

Copyright © 2002 Marcel Dekker, Inc.

and the inadvertent percutaneous absorption of nitroglycerine was finally made in the 1970s and a nitroglycerine ointment was introduced, producing peak blood levels comparable to those attained upon sublingual administration of traditional tablet triturates, but levels that were also sustained. Since the permeability of human skin is highly variable and since patient needs themselves vary, patients using nitroglycerine ointments were (and are) instructed to gradually lengthen the ribbon of ointment expressed from the tube and rubbed into the skin until tinnitus and headache were experienced. They were then counseled to back off from this dosage. The lasting nature of nitroglycerine's clinical action when administered in this fashion helped free patients from the fear they had of waking in the middle of the night with a heart attack (angina). Consequently, a nitroglycerine ointment became the first commercially successful, therapeutically proven transdermal delivery system. But ointments are greasy and suffer variability in their dosing, even with dose titration, as a result of the fact that different patients apply semisolids more or less thinly and therefore over lesser or greater areas.

Since about 1980, sophisticated adhesive patches for transdermal delivery of scopolamine (motion sickness), nitroglycerine (anginal symptoms), clonidine (regulation of blood pressure), β-estradiol (menopausal symptoms), fentanyl (cancer pain), nicotine (smoking cessation), and testosterone (hypogonadism) have been introduced into medicine [24–26]. These patches, affixed to an appropriate body location, deliver drug continuously for periods ranging from about half of a day (nitroglycerine) to a week (clonidine, β-estradiol). To achieve a long duration of delivery, a patch must contain a reservoir of its respective drug. In one early type of system, nitroglycerine was formulated into a liquid-filled sponge that was held in intimate contact with the skin by an adhesive band around the periphery of the patch. In more recent patch designs a membrane is placed over the delivery surface of the patch. Adhesive covering this membrane anchors the patch to the skin over the entire contact area of the patch. This interfacing membrane can be turned into a rate-controlling membrane to regulate delivery and prevent dose dumping should the patch be inadvertently placed over a site of inordinately high permeability. Actually, this rate-controlling concept was used in the design of the first patch, the scopolamine transdermal system. The development of transdermal patches for yet other drugs is an active research area.

It is obvious from the above that the skin is a formidable barrier irrespective of whether therapy is to be local, regional or systemic, and the first concern in topical delivery is sufficiency of delivery. With local therapy, the aim is to get enough drug into the living epidermis or its surroundings to effect a pharmacological action there without producing a systemically significant load of the drug. The latter is actually a rare occurrence except when massive areas of application are involved. Regional therapy involves effects in musculature and joints deep beneath the site of application. To be successful, this requires a greater delivery rate because an enormous fraction of the drug that passes through the epidermis is routed systemically via the local vasculature. Indeed, the levels of drug reached in deep local tissues have proven to be only a few multiples higher than those obtained upon systemic administration of the drug [27]. Even more drug has to be delivered per unit area to transdermally effectuate a systemic action.

Factors Affecting Functioning of the Skin Barrier

A matter of considerable consequence in topical delivery is variability in skin permeability between patients, which may be as much as 10-fold. The underlying sources of this high degree of variability are thought to be many and diverse. Humans differ in age, gender, race, and health, all of which are alleged to influence barrier function. Yet, insofar as can be told, a full-term baby is born with a barrier-competent skin, and, barring damage or disease, the skin remains so through life. There is little convincing evidence that senile skin, which tends to be dry, irritable, and poorly vascularized, is actually barrier compromised [28]. However, premature neonates have inordinately permeable skins. The incubators used to sustain such infants provide a humidified environment, which abates insensible perspiration, and a warm one, conditions that not only make the baby comfortable but that also forestall potentially lethal dehydration and hypothermia [29].

Gender, too, affects the appearance of human skin. Nevertheless, there is little evidence that the skin of males and females differs greatly in permeability. However, there are established differences in the barrier properties of skin across the races of humans. While the horny layers of Caucasians and Blacks are of equal thickness, the latter has more cell layers and is measurably denser [30]. As a consequence, black skin tends to be severalfold less permeable [30,31].

Humidity and temperature also affect permeability. It has long been known that skin hydration, however brought about, increases skin permeability. Occlusive

Copyright © 2002 Marcel Dekker, Inc.

wrappings are therefore placed over applications on occasion to seal off water loss, hydrate the horny layer, and increase drug penetration. In the absence of such intervention, the state of dryness of the stratum corneum is determined by the prevailing humidity, explaining why a dry skin condition is exacerbated in the winter months in northern climes.

Temperature influences skin permeability in both physical and physiological ways. For instance, activation energies for diffusion of small nonelectrolytes across the stratum corneum have been shown to lie between 8 and 15 kcal/mole [4,32]. Thus thermal activation alone can double the rate skin permeability when there is a 10°C change in the surface temperature of the skin [33]. Additionally, blood perfusion through the skin in terms of amount and closeness of approach to the skin's surface is regulated by its temperature and also by an individual's need to maintain the body's 37°C isothermal state. Since clearance of percutaneously absorbed drug to the systemic circulation is sensitive to blood flow, a fluctuation in blood flow might be expected to alter the uptake of chemicals. No clear-cut evidence exists that this is so, however, which seems to teach us that even the reduced blood flow of chilled skin is adequate to efficiently clear compounds from the underside of the epidermis.

Above all else, the health of the skin establishes its physical and physiological condition and thus its permeability. Consequences attributable to an unhealthy condition of skin can be subtle or exaggerated. Broken skin represents a high permeability state, and polar solutes are several log orders more permeable when administered over abrasions and cuts. Irritation and mild trauma tend to increase the skin's permeability even when the skin is not broken, but such augmentation is far less substantial. Sunburn can be used to illustrate many of the barrier-altering events that occur in traumatized skin. Vasodilation of the papillary vasculature with marked reddening of the skin is among the first signs that a solar exposure has been overdone. In its inflamed state, the skin becomes warm to the touch. After a day or two, epidermal repair begins in earnest and the tissue is hyperproliferatively rebuilt in its entirety. It doubles in thickness and a new stratum corneum is quickly laid down [34]. Because the newly formed stratum corneum's anchorage to existing tissues is faulty, the preexisting horny layer often eventually peels. Of more importance, hyperplastic repair leads to a poorly formed horny structure of increased permeability to water (as measured by transepidermal water loss) and presumably other substances. Given these events surrounding irritation, since many chemicals found in the workplace and home are mildly irritating, including the soaps we use to bathe and the detergents we use to clean house and cloths, is it really any wonder that the permeability of human skin is so demonstrably variable?

Some chemicals have prompt, destructive effects on the skin barrier. Saturated aqueous phenol, corrosive acids, and strong alkalis instantly denature the stratum corneum and destroy its functionality even as their corrosive actions stifle the living cells beneath. Though the stratum corneum may appear normal following such damage, the skin may be only marginally less permeable than denuded tissue [35]. Furthermore, permeability remains high during the full duration of wound repair and until a competent stratum corneum is laid down over the injured surface. Other chemicals are deliberately added to formulations to raise the permeability of skin and improve drug delivery. For obvious reasons, these are referred to as skin penetration enhancers. More will be said of these later.

Thermal burning produces comparably high states of permeability immediately following burning providing that the surface temperature of the skin is raised above 80°C, a temperature on the lower side of temperatures able to denature keratin [36]. However, burning temperatures below 75°C, though fully capable of deep tissue destruction in seconds, leave the structure of the stratum corneum itself relatively unscathed. Burn wounds of this kind remain impermeable until tissue repair and restructuring processes get under way and the necrotic tissue with its horny capping is sloughed. The increased permeability of severely burned skin immediately following the trauma can be highly consequential in terms of drug delivery. In the instance of deep burns obtained at lower than keratin-denaturing temperatures, topical delivery of antibiotics and antiseptics into a wound, as may be necessary to control wound sepsis, remains difficult for as long as the stratum corneum over the wound stays in place, and aggressive use of antiseptics is warranted. In the other extreme there is risk of toxic systemic accumulation of the antiseptics and antibiotics, particularly in major burns covering 20% or more of the body surface area. Conservative treatment is warranted. Since burns are rarely well characterized with respect to their permeability dimension, attending physicians and pharmacists need to monitor antiseptic usage carefully to control wound sepsis without poisoning the patient. Finally, if not surgically debrided, the necrotic tissue is eventually walled off, enzymatically digested, loosened, and sloughed, producing a denuded, open, granulating surface.

Copyright © 2002 Marcel Dekker, Inc.

Small full-thickness wounds are closed and sealed off quickly by reconstruction of the epidermis from the edges of the wound. Large full-thickness injuries take too long to heal by this process and require grafting. All such wounds remain highly permeable until covered over again with a healthy, fully differentiated epidermis.

As with burns, physical disruption of the stratum corneum opens the skin in proportion to the extent of damage. Cuts and abrasions are associated with high permeability at and around such injuries. Eruption of the skin in disease has a similar effect to the extent that the stratum corneum's integrity is lost. The skin over eczematous lesions should be regarded as highly permeable. Not all skin diseases raise permeability, however. The states of permeability of ichthyosiform, psoriatic, and lichenified skin have not been well characterized but in all likelihood are low for most drugs. It has been proven difficult to get potent corticosteroids through psoriatic plaque, for instance, and occlusive wrapping is often called for.

Percutaneous Absorption—the Process

The process of percutaneous absorption can be described as follows. When a drug system is applied topically, the drug diffuses passively out of its carrier or vehicle and, depending on where the molecules are placed down, it partitions into either the stratum corneum or the sebum-filled ducts of the pilosebaceous glands. Inward diffusive movement continues from these locations to the viable epidermal and dermal points of entry. In this way a concentration gradient is established across the skin up to the outer reaches of the skin's microcirculation where the drug is swept away by the capillary flow and rapidly distributed throughout the body. The volume of the epidermis and dermis beneath a 100 cm^2 area of application, roughly the size of the back of the hand, is approximately 2 cm^3. The total aqueous volume of a 75 kg ($\sim$165 lb) person is about 50,000 cm^3, yielding a systemic-to-local dilution factor well in excess of 10,000. Consequently, systemic drug levels are usually low and inconsequential. Thus selectively high epidermal concentrations of some drugs can be obtained. However, if massive areas of the body ($\geq$20% of the body surface) are covered with a topical therapeutic, systemic accumulation can be appreciable. For instance, corticosteroids have produced serious systemic toxicities on occasion when they have been applied over large areas of the body [37]. Moreover, as has already been pointed out, if the stratum corneum is not intact, many chemicals can gain systemic entrance at alarming rates. Together these factors may place a patient at grave risk and should always be taken into account when topical drugs are put in use. The pharmacist should therefore carefully measure how topical systems are to be applied and be on the alert for untoward systemic responses when body coverages are unavoidably extensive.

The events governing percutaneous absorption following application of a drug in a thin vehicle film are illustrated in Fig. 2. The important processes of dissolution and diffusion within the vehicle are cataloged. These will be discussed later. Two principal absorption routes are indicated in the figure: (a) the transepidermal route, which involves diffusion directly across the stratum corneum, and (b) the transfollicular route, where diffusion is through the follicular pore. Much has been written concerning the relative importance of these two pathways. Claims that one or the other of the routes is the sole absorption pathway are groundless since percutaneous absorption is a spontaneous, passive diffusional process that takes the path of least resistance. Therefore, depending on the drug in question and the condition of the skin, either or both routes can be important. There are temporal dependencies to the relative importance of the routes as well. Corticosteroids breach the stratum corneum so slowly that clinical responses to them, which are prompt, are reasoned to be due to follicular diffusion [4].

Sight should not be lost of the fact that the chemical barrier of the skin actually consists of all skin tissues between the surface and the systemic entry point. While it is true that the stratum corneum is a source of high diffusional resistance to most compounds and thus the skin's foremost barrier layer, exceptional situations exist where it is not the only or even the major resistance to be encountered. For example, extremely hydrophobic chemicals have as much or more trouble passing across the viable tissues lying immediately beneath the stratum corneum and above the circulatory bed because such drugs have little capacity to partition into these tissues. Backing for the latter assertion comes from extensive clinical experience as well as from physical modeling of percutaneous absorption. Consider that ointments can be used safely over open wounds because their hydrocarbon constituents are not transported significantly across even denuded skin. Similarly, the skin is considerably more impermeable to octanol and higher alkanols than is the stratum corneum alone because of the presence of the viable tissue layer beneath.

Copyright © 2002 Marcel Dekker, Inc.

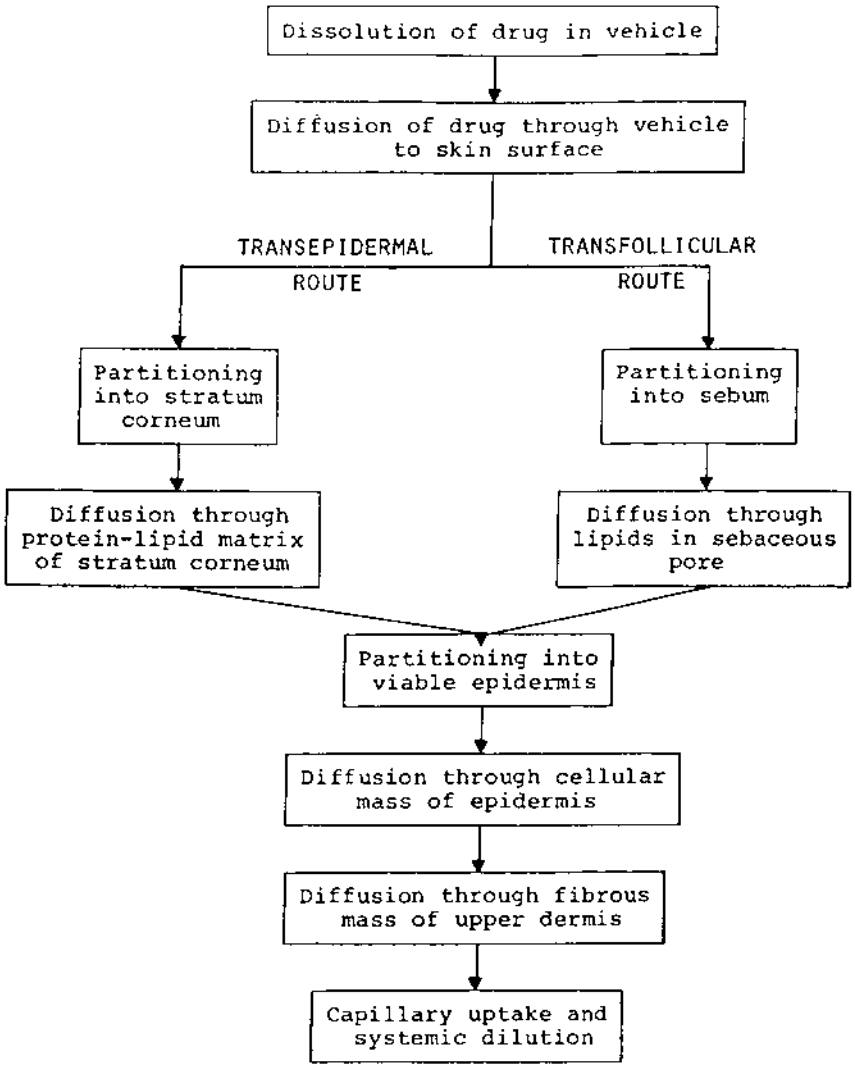


Fig. 2 Events governing percutaneous absorption.

Model of the Skin Barrier

The percutaneous absorption picture can be qualitatively clarified by considering Fig. 3, where the schematic skin cross section is placed side by side with a simple model for percutaneous absorption patterned after an electrical circuit. In the case of absorption across a membrane, the current or flux is in terms of matter or molecules rather than electrons, and the driving force is a concentration gradient (technically, a chemical potential gradient) rather than a voltage drop [38]. Each layer of a membrane acts as a diffusional resistor. The resistance of a layer is proportional to its thickness (h), inversely proportional to the diffusive mobility of a substance within it as reflected in a diffusion coefficient (D), inversely proportional to the capacity of the layer to solubilize the substance relative to all other layers as expressed in a partition coefficient (K), and inversely proportional to the fractional area of the membrane occupied by the diffusion route (f) if there is more than one route in question [39]. In general, an individual resistance in a set may be represented by:

$$R_i = \frac{h_i}{f_i \cdot D_i \cdot K_i} = \frac{[\text{thickness}]}{[\text{fractional area}][\text{diffusion coefficient}][\text{partition coefficient}]} \quad (1)$$

Copyright © 2002 Marcel Dekker, Inc.

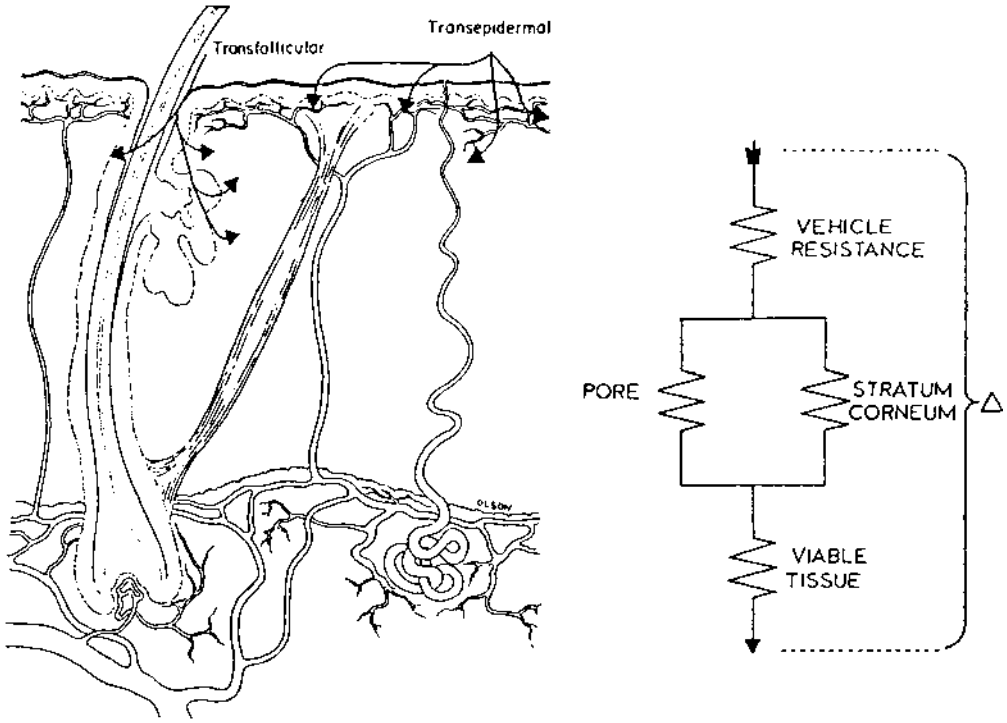


Fig. 3 Skin cross section beside a simple model.

The overall phenomenon of percutaneous absorption is describable upon recognizing that the resistances of phases in series (phases encountered serially) are additive and that diffusional currents (fluxes) through routes in parallel (differing routes through a given phase) are additive. Such considerations applied to skin allow one to explain, in semi-quantitative terms, why percutaneous absorption through intact skin is slow for most chemicals and drugs and why disruption of the horny covering of the skin profoundly increases permeability of the ordinary run of solutes.

First, consider the transepidermal route. The fractional area of this route is virtually 1.0, meaning the route constitutes the bulk of the area available for transport. Molecules passing through this route encounter the stratum corneum and then the viable tissues located above the capillary bed. As a practical matter, the total stratum corneum is considered a singular diffusional resistance. Because the histologically definable layers of the viable tissues are also physicochemically indistinct, the set of strata represented by viable epidermis and dermis is handled comparably and treated as a second diffusional resistance in series.

Estimated diffusion coefficients in the stratum corneum are up to 10,000 times smaller than found anywhere else in the skin, reflecting in part the considerable density of this tissue. If diffusion is almost exclusively through the intercellular lipid regime within the horny tissue as most experts believe, then the estimates, which range between 1×10^{-13} cm^2/s to a low of 1×10^{-9} cm^2/s, have to be tempered with the knowledge that path tortuosity (nonlinearity) and excluded volume were not taken into account. In any case, such low values still speak to the high resistance of the horny tissue [4], particularly given that the thickness of the stratum corneum is only about 1×10^{-3} cm (10 μm). The parameter exhibiting the most variability relative to the stratum corneum's diffusion resistance is the partition coefficient, K_{sc}. The partition coefficient can take values several log orders less than one with highly polar molecules, e.g., glucose, or values several log orders greater than one for hydrophobic molecules, e.g., β-estradiol. The wedge of living tissue lying between the stratum corneum and the capillaries is on the order of 100 μm thick (1×10^{-2} cm). Permeation of this element of the barrier is facile and without great molecular selectivity. Where measured, diffusion coefficients through this cellular mass have proven to be no less than one tenth of the magnitude of those found for the same compounds passing through water [40].

Copyright © 2002 Marcel Dekker, Inc.

The follicular route can be analyzed similarly. The fractional area available for penetration by this route is on the order of one one-thousandth [4], clearly a restricting factor. Here partitioning is into sebum and the distance that has to be traveled through sebaceous medium filling the follicular duct can only be guessed at, but it has to be greater than the thickness of the stratum corneum. Fifty μm seems a reasonable estimate, the actual value almost certainly being within a factor of two of this value. Diffusion coefficients in the quasi-liquid sebum can be guessed to be from a hundred to a thousand times greater than found for the stratum corneum [4,10]. Since sebum is lipoidal, partition coefficients associated with passage through this route must exhibit wide-ranging values even as found for the stratum corneum. A thickness of viable tissue perhaps comparable to that found along the transepidermal route also lies within this pathway to the microcirculation.

Net chemical penetration of the skin is simply the sum of the accumulations by each of the mentioned routes and by other routes, for instance eccrine glands, where these contribute. The latter tiny glands are ubiquitously distributed over the body but are generally discounted in importance due to the limited fractional area they occupy and their unfavorable physiological states, either empty or profusely sweating.

All the salient features described here can be incorporated into a quantitative framework that takes into account stratification within the tissues and the parallel pathways [39]. It is instructive to consider a simple model based on these descriptions that embodies only transepidermal and transfollicular barrier elements. We can assume that each distinct tissue acts as a homogeneous phase, a gross distortion of reality but an assumption that nevertheless leads to a useful conceptual description. The resistance by the transepidermal route would be:

$$R_{\text{TransEpidermal}} = R_{\text{stratum corneum}} + R_{\text{viable tissues transepidermally}}$$

or

$$R_{\text{TE}} = R_{\text{sc}} + R_{\text{vt-TE}} \tag{2}$$

Similarly, the transfollicular resistance would be:

$$R_{\text{TF}} = R_{\text{seb}} + R_{\text{vt-TF}} \tag{3}$$

Since these routes are in parallel, the total resistance on combining them is:

$$R_{\text{total}} = \frac{1}{\frac{1}{R_{\text{sc}}+R_{\text{vt-TE}}} + \frac{1}{R_{\text{seb}}+R_{\text{vt-TF}}}} \tag{4}$$

The mass transfer coefficient (permeability coefficient) of a route is the reciprocal of the resistance of that route, and thus we can amend Eq. (4) to read:

$$R_{\text{total}} = \frac{1}{P_{\text{TE}} + P_{\text{TF}}} \tag{5}$$

Similarly, the overall permeability coefficient is the reciprocal of the total resistance and thus P_{total} is:

$$P_{\text{total}} = P_{\text{TE}} + P_{\text{TF}} \tag{6}$$

For reasons that need not be elaborated here, water is invariably the solvent medium used to experimentally access permeability coefficients. Accordingly, water is assumed to be the vehicle used to apply a drug to the skin. This choice of vehicle effectively sets the partition coefficients between the aqueous tissues and the vehicle roughly to unity. Choosing water as the vehicle of consideration does not in any way invalidate insights that can be drawn from the model as long as saturated solutions, which operate at the thermodynamic activity of the solid drug, are brought into the analysis. Barring specific solvent-induced changes in the physical chemistry of the tissue, from the thermodynamic perspective all saturated solutions should deliver drug at the same rate. Substituting fractional areas, thicknesses, partition coefficients, and diffusion coefficients for all the phases while recognizing that partition coefficients between water and the viable tissues functionally assume values of unity leads to the following expression for the overall permeability coefficient:

$$P_{\text{total}} = f_{\text{TE}}\left(\frac{D_{\text{sc}}K_{\text{sc/w}}D_{\text{vt}}}{D_{\text{sc}}K_{\text{sc/w}}h_{\text{vt-TE}} + D_{\text{vt}}h_{\text{sc}}}\right) + f_{\text{TF}}\left(\frac{D_{\text{seb}}K_{\text{seb/w}}D_{\text{vt}}}{D_{\text{seb}}K_{\text{seb/w}}h_{\text{vt-TF}} + D_{\text{vt}}h_{\text{seb}}}\right) \tag{7}$$

Finally, a general expression describing the steady state flux across a membrane, $\partial M/\partial t$ can be written as:

$$\frac{\partial M}{dt} = A \cdot P_{\text{total}} \cdot \Delta C \tag{8}$$

This equation teaches us that the total stead-state flux (total rate of permeation across a membrane in the steady state of permeation), $\partial M/dt$, is proportional to the involved area (A) and the concentration differential expressed across the membrane, ΔC. In an experiment, flux is the experimentally measured parameter while A and ΔC are fixed in value when setting up an experiment. The value of the permeability coefficient, P_{total}, is what is calculated upon completion of an experiment using Eq. (8). The permeability coefficient, besides having the specific attributes ascribed to it, is

Copyright © 2002 Marcel Dekker, Inc.

the number that converts the combined area and concentration proportionalities of flux to an equality (an equation containing an equal sign). The permeability coefficient takes units of distance over time (cm/h) and can be regarded as the average velocity of molecules penetrating a membrane irrespective of the complexity of the membrane. Its magnitude depends on the properties of the vehicle, membrane, and permeant. Moreover, when the membrane consists of several phases, the permeability coefficient is also dependent on the juxtaposition of the phases. We can now write for the skin:

$$\frac{\partial M}{dt} = A\left\{f_{\mathrm{TE}}\left(\frac{D_{\mathrm{sc}}K_{\mathrm{sc/w}}D_{\mathrm{vt}}}{D_{\mathrm{sc}}K_{\mathrm{sc/w}}h_{\mathrm{vt\text{-}TE}} + D_{\mathrm{vt}}h_{\mathrm{sc}}}\right) + f_{\mathrm{TF}}\left(\frac{D_{\mathrm{seb}}K_{\mathrm{seb/w}}D_{\mathrm{vt}}}{D_{\mathrm{seb}}K_{\mathrm{seb/w}}h_{\mathrm{vt\text{-}TF}} + D_{\mathrm{vt}}h_{\mathrm{seb}}}\right)\right\}\Delta C \quad (9)$$

where A is the involved area of the skin and the term ΔC is the permeant's concentration differential across the skin. In clinical situations ΔC is usually well approximated by the actual concentration in the topical vehicle because dilution by way of systemic absorption of the permeant is so great. In many real instances ΔC reflects saturation of a drug in its vehicle.

Equation (9) defines the steady state in flux in terms of physically meaningful parameters. In other words, it is an anatomically based mathematical representation (model) of the skin barrier. The upper collection of terms in the greater parentheses defines the role of the transepidermal route. The second group of terms below in the greater parentheses characterizes the transfollicular pathway. To breathe life into the model, values for the parameters of Eq. (9) that are representative have to be plugged into the equation. Some parameter estimates are given in Table 8. The listed fractional areas, diffusion coefficients, and strata thicknesses reported in Table 8 are based on the best information available. All the values are approximations; some are no more than best guesses. Regardless, the impressions drawn from the model when they are plugged into it are consistent with known percutaneous absorption behavior.

To illustrate the above point, take the set of largest values given for the diffusion coefficients found in Table 8, that is, 10^{-9} cm^2/s for the stratum corneum, $10^{-7}\,\mathrm{cm}^2/\mathrm{s}$ for sebum, and $10^{-6}\,\mathrm{cm}^2/\mathrm{s}$ for the viable tissue, and convert them to cm^2/h. Conversion of these from reciprocal seconds to reciprocal hours eventually leads to permeability coefficients that are more easily compared with literature values (P in units of cm/h) When these values are substituted into Eq. (7) along with the estimated thicknesses of the respective tissues (cm) and the fractional areas of the routes, the steady-state flux through the transepidermal route is projected to be

Table 8 Representative Parameters to Probe Model

	Diffusion coefficient, D (cm^2/s)[a] [4,8]
Stratum corneum	10^{-9}–10^{-13}
Water	$\simeq 10^{-9}$
n-Alkanols (hydrated tissue)	$\simeq 10^{-9}$
n-A1kanols (dry tissue)	$\simeq 10^{-10}$
Small nonelectrolytes	10^{-9}– 10^{-10}
Progesterone	$\simeq 10^{-11}$
Cortisone	$\simeq 10^{-12}$
Hydrocortisone	$\simeq 10^{-13}$
Follicular pore (sebum)	10^{-7}– 10^{-9}
Viable tissue	$\simeq 10^{-6}$
	Tissue thickness, h (μm) [2,8,11]
Stratum comeum	
Dry (normal state)	~10
Hydrated (as by occlusion) state	20–30
Pore diffusional length	Approximately two to five times greater than the stratum corneum thickness
Viable tissue stratum	150–2000[b] (200)
	Fractional area of the routes, F [4,8]
Transepidermal	~1
Transfollicular	$\sim 10^{-3}$
Transeccrine	$<10^{-4}$[c]
	Tissue/vehicle partition coefficient, K
Stratum corneum	From < 1 to $\gg 1$[d]
Sebum	From $\ll 1$ to $\gg 1$[d]
Viable tissue	
Aqueous vehicle	~1
Nonaqueous vehicle	From $\ll 1$ to $\gg 1$[d]

[a]These diffusivities are estimates obtained by in vitro experiment (stratum corneum) or by comparison with small tissues in which diffusivities have been measured (all others). They do not account for regional variations across the body surface, so on both counts must be considered highly approximate.
[b]Highly approximate and variable, depending on blood flow patterns.
[c]This is sufficiently small to discount transeccrine diffusion contributions in the general treatment.
[d]All depend on the physiocochemical nature of the drug and vehicle as well as the physicochemical nature of the respective tissues.

Copyright © 2002 Marcel Dekker, Inc.

30 times greater than through the follicular pores, everything else being equal.* The suggestion is strong here that high sebum diffusivity by itself is insufficient to offset the small fractional area of the follicular route. Clearly, partitioning tendencies favoring one or the other of these routes will magnify or shrink this multiple. Unfortunately, there is no a priori way to estimate the relative magnitudes of the operative partition coefficients.

Some scientists, including this author, believe that the stratum corneum harbors a minor polar (aqueous pore) pathway† mostly because of evidence that suggests that the stratum corneum offers higher fluxes to polar solutes such as methanol, ethanol, propylene glycol, glycerol, and glucose than one would otherwise expect. In this regard, it is also relevant that ions diffuse through the stratum corneum with deceptive ease considering their solution attributes. Sebum is not well characterized as a diffusion medium, but it is generally taken to be an oily composite. If the latter portrayal is apt, then sebum would have a thermodynamically limited ability to dissolve polar compounds. On these admittedly flimsy grounds, an argument can be mounted that the transepidermal route dominates the transfollicular route with respect to the permeation of small, polar nonelectrolytes and ions [4,8,10].

It is thought that both the stratum corneum and sebum are, to first good approximations, lipoidal routes. Consequently, drug substances of diminishing polarity should partition out of water into the key transport regimes within these routes to increasing extents. Homologs formed by extending the length of an alkyl chain provide a means for testing this hypothesis due to the fact that oil/water (o/w) partition coefficients of alkyl homologs grow exponentially. One can therefore probe the fundamental physical behaviors of lipid membranes as long as (in the permeability domain where) permeability coefficients are directly related to o/w partitioning. The slope of log(partition coefficient) against alkyl chain length plot indicates the sensitivity of partitioning between the phases in question to the addition of a methylene (CH_2) group. The value of $d[\log(K_{o/w})]/dn$, the slope, is referred to as the π-value for the partitioning system in question. The π-value for the partitioning of homologs between octanol and water is very close to 0.5. Thus, regardless of the homologous series in question, octanol/water partition coefficients, increase by a factor of about 10 for every two methylene units added to an alkyl chain. With a π-value of greater than 0.6, hexane/water partitioning evidences an even greater lipoidal sensitivity. Based on permeation of *n*-alkanols through human skin (and the permeability partitioning relationship), human stratum corneum appears have a π-value slightly less than 0.3 [4]. The low partitioning sensitivity to the addition of a methylene group suggests that the stratum corneum's lipoidal phase is considerably more polar than the reference organic solvents. If the substance of the sebaceous gland were more hydrophobic than the polar lipids found between the cell platelets of the stratum corneum, as is thought, then its methylene partitioning sensitivity would be greater than the stratum corneum's, though almost certainly less than seen with the organic solvents. This means, all else being equal, that partition coefficients would increase more rapidly by this route. It also suggests that the region of direct partitioning dependency of permeability by this route would be narrower than found for the transdermal pathway [41,42]. It doesn't appear that these combined effects should lead to a sizable shift in the relative importance of the routes as judged through modeling. This is because by the time $K_{seb/w}$ takes a value of one, passage through the transfollicular route is already approaching the point where the resistance of the viable tissue within the route is reaching a magnitude comparable to that offered by the sebum plug. Consequently, further increases in $K_{seb/w}$ would not lead to increased permeation via this pathway. On the other hand, according to the model, the value of $K_{sc/w}$ has to approach 50 before the viable tissue becomes the rate-controlling element in the transepidermal route. Perhaps by coincidence, 50 is the published partitioning value for octanol, which actually does appear to be at the threshold of viable tissue control of its permeation [4].

Model interpretation takes a different bent when minimum values for the respective diffusion coefficients are incorporated in the steady-state model, i.e., $10^{-13}\,cm^2/s$ for the stratum corneum and $10^{-9}\,cm^2/s$ for the follicular shunt route. Inserting these values, everything else held constant, suggests there should be a substantial upgrading of the importance of the transfollicular contribution. Data with steroids seem to indicate, however, that the transepidermal route retains a dominant position in the steady state even in this case.

In the main, to this point, the permeation model represented by Eq. (9) can be considered esoteric. The value of the model to practicing pharmacists stems substantially from what the model teaches us about barrier damage. To measure the result of extreme barrier damage, one only needs to use zero for

*Everything else equal here means $K_{sc/water}$ would be equal to $K_{seb/water}$.

†This supposition is hotly debated in scientific circles.

Copyright © 2002 Marcel Dekker, Inc.

the value of h_{sc}, an operation equivalent to removing the stratum corneum. Under this circumstance, the permeability coefficient via the expansive transepidermal route takes the simplified value, $D_{vt}/h_{vt\text{-}TE}$. $D_{vt}/h_{vt\text{-}TE}$ can be as much or more that 1000-fold larger for polar compounds than found when the stratum corneum is present. However, for compounds at the extreme of nonpolarity, negligible increases in permeability are projected because $K_{sc/w}$, which appears in reciprocal in the statement of the stratum corneum's resistance, is so large that the stratum corneum is not of much consequence even when it is intact. It follows then that the level of increased permeability resulting from skin damage depends on exactly how lipophilic a compound is. The bottom line, however, is that many compounds that are safe to apply to normal skin experience up to a 1000-fold increases in their skin permeability and become very dangerous when in contact with skin. This is especially true when large areas of contact, A in the model equation, are involved. There are many cases of lethal systemic poisonings resulting from contact of chemicals with barrier-compromised skin in the literature. Pharmacists must understand and take cognizance of the kind of danger outlined here when counseling patients on the use of topical medications.

So far only steady-state conditions for permeation have been considered. But in all phenomena involving the mass transport of substances across membranes, one also has to consider the time it takes for gradients to be set up across the membranes by molecules moving randomly within the substance of the membranes. This early period of permeation is referred to as the nonstationary state period. The duration of this period is characterized by a lag time, as illustrated in Fig. 4*. The nonstationary-state period may be of some importance in the instance of skin permeation [43]. Where independent parallel pathways exist, as they seem to do through the stratum corneum, compounds may gain access to underlying tissues far more readily by one pathway than by the other. We can consider this point in terms of the lag time, t_L, that one expects to obtain for a compound diffusing through a simple isotropic field of a membrane:

$$t_L = \frac{h^2}{6 \cdot D} \tag{10}$$

*The lag time is the intercept that results by extrapolating the steady-state line of a plot of cumulative amount of drug penetrated versus time to the time axis.

where h identifies the membrane's thickness and D is the compound's diffusion coefficient. The equation teaches us that the lag times for parallel routes should differ to the extents that h and D are different for the routes. We can crudely estimate the respective lag times for the transepidermal and transfollicular routes from the diffusion coefficient and thickness estimates tabulated in Table 8. Taking this approach, lag times for the transepidermal route are projected to range from minutes for small nonelectrolytes to multiple days for molecules of the size and properties of the corticosteroids. On the other hand, the lag times for breaching the transfollicular route should range from seconds to minutes. Though relatively few molecules reach the living tissues via the transfollicular route due to the limited fractional area of this pathway, it appears that the first molecules to reach the viable epidermis actually get there via this path. This comparison has led scientists to suggest that the early clinical responses seen with drugs as large as steroids are the result of follicularly transported molecules. However, if the lag time for passage directly through the stratum corneum is also very brief, the follicular shunt diffusion is far less likely to be clinically meaningful. These nonstationary-state considerations do not apply at all when the horny barrier is impaired. Under this circumstance transepidermal lag times can be so brief that they cannot be easily measured.

Equations (8) and (9) point to the fact that the amount of drug delivered through the skin is proportional to the area of application of a topical dosage form or a transdermal patch. Nitroglycerine and other patches are available in different sizes (areas) to take advantage of this proportionality in dosing. Pharmacists have to be cognizant of this area dependency for another important reason, this being that the systemic accumulations, and thus systemic toxicities, of topically applied drugs are proportional to application area. Indeed, excessive area of application is the most frequent cause of the untoward systemic actions seen with topically applied drugs. Tiny infants have been fatally poisoned following too liberal applications of borate-containing talcum powders. As a result, borates are no longer used as lubricants in powder formulations. Babies have also been poisoned upon bathing them with a hexachlorophene-containing soap. Bathing, of course, is virtually synonymous with whole body coverage. Hexachlorophene is no longer used as an antiseptic soap. Diaper dermatitis tends to be expressed over large areas. Anti-inflammatory corticosteroids have occasionally been used too liberally to remedy this condition, again eventuating in serious

Copyright © 2002 Marcel Dekker, Inc.

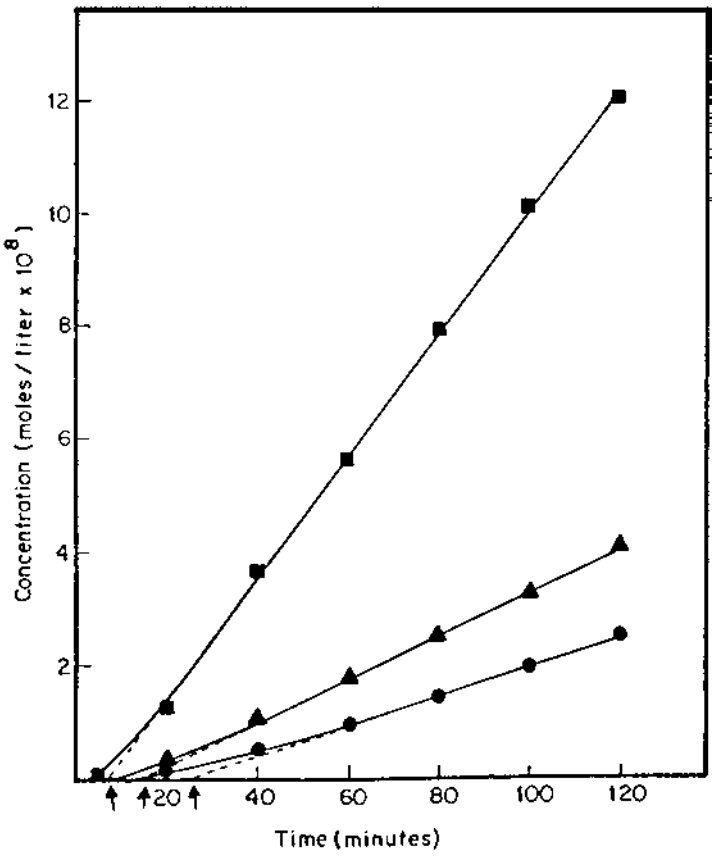


Fig. 4 Generalized permeation profile. From left to right the data are for n-butanol permeating hairless mouse skin at 20°, 25°, and 30°C respectively. Increasing temperature raises the flux (slope) and shortens the lag time.

toxicity. And systemic toxicities have at times accompanied the application of salicylic acid, a keratolytic agent or substance that breaks down keratin, to large psoriatic lesions. There are documented cases of abuse of transdermal patches, including a case in which a patient, confused about the manner of usage, succumbed to fentanyl by wearing four patches simultaneously. A pharmacist must realize that in the instance of topical therapy, area is dose!

Phenomenological Considerations in Percutaneous Delivery

While the model helps us understand the chemical structure dependencies of skin permeability, it isn't all that useful for calculating permeability coefficients because of the many iffy assumptions it contains. A different tack has to be taken to gain a sense of the limits, especially the upper limit, of cutaneous drug delivery. There is no lower limit. Even proteins penetrate intact skin to some extent. Some idea of the upper limit can be gained upon examining the rates of delivery of nitroglycerine and like facile penetrants of the skin. Nitroglycerine, a liquid at room temperature, is a lipophilic nonelectrolyte of 227 dalton molecular weight. These physical properties make this drug about as good a skin penetrant as can be found. It is formulated in transdermal systems virtually as a neat liquid and thus near its upper attainable thermodynamic activity. It diffuses through the skin at between 0.02 and 0.04 mg/cm^2/h from the transdermal reservoirs it is placed in. These rates equate with a delivery of 0.5–1.0 mg/cm^2/day [24]. Consequently, 20 cm^2 patches provide a daily delivery of between 10 and 20 mg of the drug [24]. Nicotine, another low molecular weight, substantially hydrophobic, liquid compound at 25°C, permeates human skin from its transdermal delivery systems with comparable ease. Selegiline is yet another drug with properties like these that is a facile skin penetrant. Putting all of this together, it appears that a percutaneous delivery rate of about 1 mg/cm^2/day is near the upper achievable limit for "skin delivery."

Copyright © 2002 Marcel Dekker, Inc.

Now it is true that water diffuses out through the skin far faster than any of these compounds diffuse in the reverse direction. To be specific, we typically lose about 0.5 mg/cm^2/h of water through the skin into a dry external atmosphere, a flux that exceeds those of the referenced compounds by more than an order of magnitude. However, water is a very unique molecule both with respect to its size and its interactions within the horny layer. It appears that its diffusion is not constrained to the intercellular domain of the stratum corneum, but, rather, some water almost certainly works its way into and up through keratin amassed in the intracellular space. A major fraction of insensible perspiration exits the skin this way. By way of contrast, permeation of nitroglycerine and the other drugs mentioned is thought to mainly take place through the intercellular domain of the stratum corneum.

If all nonpolar drugs were as skin permeable as nitroglycerine, many more transdermal delivery systems might exist. Why are there so few transdermally delivered drugs? The crystallinity of compounds is one important answer, for, in a general way, the level of crystallinity sets a limit on solubility and thus the activity gradient that can be expressed across the skin. Superimposing crystallinity on nitroglycerine illustrates this point. By thermodynamic derivation, the activity of a crystalline compound relative to the activity of its pure liquid state (its supercooled liquid state) is given by:

$$\ln a_2 = \frac{\Delta H_f(T_f - T)}{RT_fT} \tag{11}$$

where a_2 is the activity of a solid, ΔH_f is its heat of fusion, T_f is its melting point (temperature of fusion), T is the experimental temperature (nominally 25°C), and R is the gas constant (1.987 in cal/mol/degree units). Upon hypothetically superimposing a melting point of 100°C and a heat of fusion of 5000 cal/mol (about 20,000 J/mol) on nitroglycerine, the activity of the compound is seen to drop about fivefold. It would now only be possible to deliver 200 μg/cm^2/day of the drug or only about 4 mg of it from a 20 cm^2 patch. If it were that nitroglycerine melted at 200°C, all else being equal, its activity would be only 4% of that of its neat liquid. At 40 μg/cm^2/day, the dose achievable using a 20 cm^2 patch would now be roughly 1 mg. Since most drugs are quite crystalline, the delivery of 1 mg of drug per day through a 20 cm^2 area (roughly the area of a Ritz cracker) can be feat. This is one of the main reasons only exceptionally potent drugs are taken seriously for transdermal delivery. The dose limit results from the fact that, irrespective of solvent, drug solubilities are lowered over what they would otherwise be by crystallinity. Referring back to the model, the impact of lowered solubility is experienced in the term ΔC. This is so because, barring supersaturation, the largest value ΔC can take turns out to be the solubility of the compound. In other words, a drug's solubility ordinarily sets the upper limit on a drug's driving force and therefore on its achievable dose. It is particularly important to give serious weight to the crystallinity factor when selecting compounds for topical or transdermal purposes from the many that may be available. Everything else being equal, low melting compounds are far easier to deliver in therapeutically adequate amounts.

The solubility of a drug in any specified particular solvent is also determined by its interactions within the solvent. The activity of a dissolved solute is related to its concentration through an activity coefficient, i.e.:

$$a_2 = X_2 \cdot \gamma_2 \tag{12}$$

where X_2 is the mole fraction concentration of the solute in a saturated solution. Mole fraction concentration is specified in lieu of molar or percentage concentration because it fits better in theoretical developments that lead up to Raoult's law. γ_2 is the drug's activity coefficient on the mole fraction scale. The activity coefficient is simply a number that is needed turn the proportionality that exists between activity and concentration into an equality. Combining equations, one can now write

$$\ln X_2 = -\frac{\Delta H_f(T_f - T)}{RT_fT} - \ln \gamma_2 \tag{13}$$

The right-hand side of Eq. (13) teaches us that the solubility of a drug in a specified solvent depends on the drug's melting parameters in the same way activity does. However, the drug's solubility also depends on its activity coefficient in the solvent in question as indicated by the second term on the right side of Eq. (13) If the solute's solution phase interactions are strong (energetic), the activity coefficient will take a value less than one. Consequently, the solubility of the compound in this circumstance exceeds that of an ideal solution (where $\gamma_2 = 1.0$ and $a_2 = X_2$. Relatively weak interactions between the solute and solvent, on the other hand, result in activity coefficients greater than unity. In this case, solubilities are lower than ideal. The bottom line of all of this is that a high level of crystallinity is associated

Copyright © 2002 Marcel Dekker, Inc.

with low activity and vice versa. This factor is set apart from solute-solvent interactions. It is differing solute-solvent interactions that cause a compound to express different solubilities in different solvents. However and importantly, the thermodynamic activity of a drug is identical across all its saturated solutions as long as its crystallinity is independent of the solvents it is placed in. This does not necessarily mean that the delivery of a drug has to be the same from every one of its saturated solutions regardless of solvent, for kinetic (as opposed to thermodynamic) factors can exert their influences. Consider, for example, the situation where dissolution of drug particles suspended in a particular vehicle cannot keep up with release of the drug into the skin. The thermodynamic activity of the drug will necessarily drop below that of a saturated solution even when suspended drug is still present. And when skin is the membrane, the differing abilities of various media to alter the physicochemical attributes of the conduit phases of the stratum corneum and thereby enhance permeation have to be taken into account.

Crystallinity aside, the two physical attributes of a drug that most control its skin permeability are its physical size and its lipophilicity [44,45]. When all extant human permeability coefficients (at the time over 90 compounds) were subjected to multiple linear regression by Potts and Guy using the following semiempirical equation [45]:

$$\log 10(P) = \log 10\left(\frac{D^{\circ}}{h}\right) + \alpha \cdot \log 10(K_{\text{oct/w}}) - \beta \cdot (\text{MW}) \qquad (14)$$

they found that almost 70% of the variation exhibited by human skin permeability coefficients is explained by differences in their molecular sizes and partitioning coefficients. D° in equation 14 is the hypothetical diffusivity of a molecule having zero molecular volume (the molecular size influence appears in the molecular weight term). As before, h is meant to represent the length of the diffusion pathway. $K_{\text{oct/w}}$ is the octanol/water partition coefficient of each compound. The octanol/water partition coefficient is a general measure of the relative lipophilicity of each compound and also something of a surrogate for a compound's stratum corneum/water partition coefficient, $K_{\text{sc/w}}$. α is a proportionality factor relating $K_{\text{sc/w}}$ to $K_{\text{oct/w}}$, while β is a constant arising from the dependency of diffusion coefficients on molecular size. The computer-determined expression that best fit all the permeability coefficient data is [44]:

$$\log 10(P) = -6.3 + 0.71(K_{\text{oct/w}}) - 0.0061 \cdot (\text{MW}) \qquad (15)$$

Equation (l5) ostensibly allows one to estimate permeability coefficients in units of cm/s. However, as with any parameter calculated from statistically drawn relationships, such estimates have to be "taken with a grain of salt," because the absolute error of estimation for a single compound can be large.

V. UNIQUE PHYSICOCHEMICAL SYSTEMS USED TOPICALLY

As one scans the products at the drug counter, one finds an enormous variety of formulation types available for topical therapy or for cosmetic purposes. Solutions are commonly found. They come in packages that allow them to be rubbed on, sprayed on by aerosol and atomizers, painted on, rolled on, swabbed on by premoistened pledgets, and dabbed on from applicators. Assorted medicated soaps are available for a range of purposes. Emulsions for the skin are found in the form of shampoos and as medicated lotions. Powders to soothe and lubricate are placed in sprinkling cans, while others containing drugs are formulated into aerosols to be sprayed on the skin. There are numerous fluid suspensions to be used as makeup or for therapeutic purposes. Clear and opaque gels are also to be found in both cosmetic and therapeutic spheres, as are assorted semisolid creams, ointments, and pastes. The physical natures of these latter systems range from soft semisolids that are squeezed out of tubes to hardened systems suitable for application in stick form. There are therapeutic and cosmetic oils for the bath. The list of products and formulation types is nearly endless.

Of all these formulations, it is the diverse semisolids that stand out as being uniquely topical. Semisolid systems fulfill a special topical need as they cling to the surface of the skin to which they are applied, generally until being washed off or worn off. In contrast, fluid systems have poor substantivity and readily streak and run off the desired area. Similarly, powders have poor staying properties. Importantly, the fundamental physicochemical characteristics of solutions, liquid emulsions and suspensions, and powders are independent of their route of application, and are discussed adequately elsewhere in this text and need not be reconsidered. This is not to say the compositions of such systems cannot be uniquely topical, for there are chemicals that can be safely applied to the

Copyright © 2002 Marcel Dekker, Inc.

skin which are unsafe to use systemically. It is necessary to elaborate upon the properties of semisolids.

A. General Behavior of Semisolids

The term semisolid infers a unique rheological character. Like solids, such systems retain their shape until acted upon by an outside force, whereupon, unlike solids, they are easily deformed. Thus, a finger drawn through a semisolid mass leaves a track that does not fill up when the action is complete. Rather, the deformation made is for all practical purposes permanent, an outcome physically characterized by saying semisolids deform plastically. Their overall rheological properties allow them to be spread over the skin to form films that cling tenaciously.

To be semisolid, a system must have a three-dimensional structure that is sufficient to impart solid-like character to the undistributed system that is easily broken down and realigned under an applied force. The semisolid systems used pharmaceutically include ointments and solidified w/o emulsion variants thereof, pastes, o/w creams with solidified internal phases, o/w creams with fluid internal phases, gels, and rigid foams. The natures of the underlying structures differ remarkably across all these systems, but all share the property that their structures are easily broken down, rearranged, and reformed. Only to the extent that one understands the structural sources of these systems does one understand them at all.

B. Ointments

Unless expressly stated otherwise, ointments are hydrocarbon-based semisolids containing dissolved or suspended drugs. They comprise fluid hydrocarbons, C_{16} to perhaps C_{30} straight chain and branched, entrapped in a fine crystalline matrix of yet higher molecular weight hydrocarbons. The high molecular weight fraction precipitates out substantially at room temperature, forming interlocking crystallites [46]. The extent and specific nature of this structure determine the stiffness of the ointment. It follows directly from this that hydrocarbon-based ointments liquefy upon heating when the crystallites melt. Moreover, when cooled very slowly, they assume a fluidity much greater than when rapidly cooled because slow cooling leads to fewer and larger crystallites and therefore less total structure. Ordinary white and yellow petrolatum are examples of such systems.

Several alternative means of forming hydrocarbon ointments illustrate their structural properties. Ointments can be made by incorporating high melting waxes into mineral oil (liquid petrolatum) at high temperature. Upon cooling, interlocking wax crystallites form and the system sets up. Polyethylene too can gel mineral oil if dissolved into this vehicle at high temperature and the solution is then force cooled [47]. A network of polyethylene crystallites provides the requisite solidifying matrix. This polyethylene gelled system is more fluid on the molecular level than are the semisolid petroleum distillates while at the same time macroscopically behaving as an ointment. Consequently, diffusion of drugs through this vehicle is more facile and drug release is somewhat greater than in petrolatum-based systems [48]. Plastibase (Squibb) is the commercially available base of polyethylene gelled mineral oil. It is useful for the extemporaneous preparation of ointments by cold incorporation of drugs. Pharmacists should not melt down this base to incorporate drugs because its gelled state cannot be restored without special processing equipment.

If a material other than a hydrocarbon is used as the base material of an ointment-like system, the ointment bears the name of its principal ingredient. There are silicone ointments that contain polydimethylsiloxane oil in large proportion. These reportedly act as excellent water barriers and superior emollients. Some are actually used to protect skin from the undesirable effects of long immersion in water.

Ointments of the specific kinds mentioned above are taken to be good vehicles to apply to dry lesions but not to moist ones. All are also greasy and stain clothes. The principal ingredients forming the systems, hydrocarbons and silicone oils, are generally poor solvents for most drugs, seemingly setting a low limit on the drug-delivery capabilities of the systems. This solubility disadvantage can be offset somewhat if hydrocarbon miscible solvents are blended into the systems to raise solvency. Alternatively, they can be made over into emulsions to increase their abilities to dissolve drugs. Along these lines, absorption bases are conventional ointments that contain w/o emulsifiers in appreciable quantity. A w/o emulsion is formed when an aqueous medium, perhaps containing the drug in solution, is worked into the base. Such emulsions are still ointments, as structurally defined, for it is the external phase of the formed emulsion that imparts the structure, and this retains its ointment-like character. It is important to note that the term "absorption base" refers to a water-incorporation capacity and infers nothing about bioavailability. This is not to say that it is not better to have water-soluble drugs emulsified than as suspended solids in such systems from a bioavailabi1ity standpoint. In this regard, for optimum

Copyright © 2002 Marcel Dekker, Inc.

results, the internal, presumably aqueous phase should be close to saturated. Diverse additives are used to emulsify water into these systems, including cholesterol, lanolin (which contains cholesterol and cholesterol esters and other emulsifiers), semi-synthetic lanolin derivatives, and assorted ionic and nonionic surfactants singularly or in combination.

Polyethylene glycol *ointment* is a water-soluble system that contains fluid, short-chain polyoxyethylene polymers (polyethylene glycols) in a crystalline network of high-melting, long-chain polyoxyethylene polymers (Carbowaxes, Union Carbide). The structure formed is totally analogous to that of the standard ointment. In one variation, this system functions well as a suppository base. Liquid polyethylene glycols are fully miscible with water, and many drugs that are insoluble in petroleum vehicles readily dissolve in the polar matrix of this base. In fact, with some drugs, delivery (bioavailability) can be compromised by an excessive capacity of the base to dissolve substances, resulting in poor vehicle-into-skin partitioning. Since polyethylene glycols are highly water soluble, bases formed from them literally dissolve off the skin when placed under a stream of running water.

C. Pastes

Pastes are basically ointments into which a high percentage of insoluble particulate solids have been added—as much or more than 50% by weight in some instances. This extraordinary amount of particulate matter stiffens the systems through direct interactions of the dispersed particulates and by adsorbing the liquid hydrocarbon fraction within the vehicle onto the particle surfaces. Insoluble ingredients such as starch, zinc oxide, calcium carbonate, and talc are used as the dispersed phase. Pastes make particularly good protective barriers when placed on the skin for, in addition to forming an unbroken film, the solids they contain can absorb and thereby neutralize certain noxious chemicals before they ever reach the skin. This explains why they are used to ameliorate diaper rash, for when spread over the baby's bottom they absorb irritants (ammonia, others?) formed by bacterial action on urine. Like ointments, pastes form an unbroken, relatively water-impermeable film on the skin surface and thus are emollients; unlike ointments, the film is opaque and therefore an effective sun block. Thus, skiers apply pastes around the nose and lips to gain dual protection. Pastes are actually less greasy than ointments because of the adsorption of the fluid hydrocarbon fraction onto the particulates.

D. Creams

Creams are semisolid emulsion systems having a creamy appearance as the result of reflection of light from their emulsified phases. This contrasts them with simple ointments, which are translucent. Little agreement exists among professionals as to what constitutes a cream, and thus the term has been applied both to absorption bases containing emulsified water (w/o emulsions) and to semisolid o/w systems, which are physicochemically totally different, strictly because of their similar creamy appearances. Logically, classification of these systems should be based on their physical natures, in which case absorption bases would be ointments and the term "cream" could be reserved exclusively for semisolid o/w systems, which in all instances derive their structures from their emulsifiers and internal phases.

The classical o/w cream is vanishing cream that contains only 15% stearic acid or its equivalent as the internal phase. Vanishing cream and its variants are first prepared as ordinary liquid emulsions at high temperature; the structure that gives them their semisolid character forms as the formed emulsions cool. Both the aqueous and stearic acid phases are heated above the point where the waxy components liquefy and then are emulsified. Sufficient emulsifier is either formed in situ or added in to create a substantial micellar phase to exist in equilibrium with the liquefied internal phase of the hot emulsion. In the instance of the classical vanishing cream, about 20% of the stearic acid it contains is neutralized with strong alkali to form the surfactant. Portions of the waxy alcohols and/or undissociated waxy acids are solubilized within such micelles. As these systems are then cooled, their emulsion droplets solidify and the micellar structures linking all together take on a liquid crystalline character [49]. The latter three-dimensional matrix has been referred to as *frozen micelles* and is what actually solidifies such creams. The compositions and amounts of both the internal phase and emulsifiers determine the extent and qualities of the structure. Creams like this are more or less stiff, depending on the level of micellar solubilization and the melting properties of the internal waxy component [49]. Within the family of such creams, the internal phase ranges in composition from about 12 to 40% by weight.

Stiff o/w emulsions can also result from droplet interactions of the internal phase, but this requires emulsifying such a huge amount of internal phase that the droplets exceed close spherical packing. In this state the emulsified particles are squashed together,

Copyright © 2002 Marcel Dekker, Inc.

losing their spherical shape, producing large interfacial areas of contact at the sites where the droplets come into contact. A fragile structure is obtained, somewhere between that of a highly viscid liquid and a true semisolid. This cream type is far less common than systems built around frozen micellar structures.

A semisolid cream of the o/w type containing a solidified, liquid crystalline internal phase is an elegant topical system preferred by many for general purposes. Such o/w systems are readily diluted with water and thus easily rinsed off the skin and are generally nonstaining. Upon application, volatile components of the cream, which may comprise as much as 80% of the total system, evaporate, and the thin application shrinks down into an even thinner layer. Stearic acid creams are particularly interesting in this regard. The small amount of internal phase they contain causes them to evaporate down to near nothingness. The dry, nontacky, translucent nature of the stearic acid crystals left on the skin contributes to the sense of their lack of discernibility. Most hand lotions and creams and foundation creams used to make face powders adherent to the face are variants of the vanishing cream formula.

Through evaporation, the drug in a cream is concentrated in its forming film, a process that can be orchestrated to program drug delivery. If no thought is given to the consequences attending drying out of the formulation, however, the drug is just as likely to precipitate out, in which instance drug delivery comes to an abrupt stop. One must therefore ensure that the formed film has some capacity to dissolve its drug. To this end, low-volatility, water-miscible solvents such as propylene glycol are added to many cream formulations. When ingredients like water and alcohol evaporate, the film left upon applying such creams becomes a rich concentrate of drug, internal phase, and its less volatile external phase components. One strives to add just enough co-solvent to keep the drug solubilized in the equilibrium film but also near saturation. It should be kept in mind that, unless the internal phase liquefies at body temperature, the waxy constituents cannot act as a solvent for the drug and thus do not lend the film much capacity for delivery.

The typical cream, a soft, emulsified mass of solidified particles in an aqueous, micelle-rich medium, does not form a water-impermeable (occlusive) film on the skin. Nevertheless, creams contain lipids and other moisturizers that replace substances lost from the skin in the course of everyday living. Creams thus make good emollients because, by replenishing lipids and in some instances also polar, hygroscopic substances, they restore the skin's ability to hold onto its own moisture.

The oleaginous phases of creams differ compositionally from hydrocarbon ointments. Many, but not all, creams are patterned after vanishing cream and contain considerable stearic acid. In lieu of some or all of the stearic acid, creams sometime contain long-chain waxy alcohols (cetyl, C_{16}; stearyl, C_{18}), long-chain esters (myristates, C_{14}; palmitates, C_{16}; stearates, C_{18}), other long-chain acids (palmatic acid), vegetable and animal oils, and assorted other waxes of both animal and mineral origin.

Properly designed o/w creams are elegant drug-delivery systems, pleasing in both appearance and feel post application. They are nongreasy and are easily rinsed off of the skin. They are good for most topical purposes and are considered particularly suited for application to oozing wounds.

E. Gels (Jellies)

Gels are semisolid systems in which a liquid phase is trapped within an interlocking, three-dimensional polymeric matrix of a natural or synthetic gum. A high degree of physical or chemical cross-linking of the polymer is involved. It only takes 0.5–2.0% of the most commonly used gelants to set up the systems. Some of these systems are as transparent as water, an aesthetically pleasing state. Others are turbid because the polymer is present in colloidal aggregates that disperse light. Clarity of the latter ranges from slightly hazy to a whitish translucence not unlike that observed with petrolatum.

Agarose gels admirably illustrate the properties and to an extent the structural characteristics of most gels. Agarose solutions are water-thin when warm but solidify near room temperature to form systems that are soft to rubbery, depending on the source and concentration of the agarose. A three-dimensional structure arises from the entwining of the ends of polymer strands into double helices. Kinks in the polymer mark the terminal points of these windings. Because individual polymer strands branch to form multiple endings, a three-dimensional array of physically cross-linked polymer strands is formed. The process of physical cross-linking is actually a crystallization phenomenon tying polymeric endings together, fixing the strands in place, yielding a stable, yet pliant structure [50]. Less extensive structure than that found in agar growth media results in a spreadable semisolid suitable for medical application. The structure should persist to temperatures exceeding body temperature for the gel-

Copyright © 2002 Marcel Dekker, Inc.

led systems to be the most useful. It is important to note that gelation is never a result of mere physical entanglement of polymer strands, otherwise the systems would be highly viscid. The polymers used to prepare pharmaceutical gels include natural gums such as tragacanth, pectin, carrageen, agar, and alginic acid and synthetic and semi-synthetic materials such as methylcellulose, hydroxyethylcellulose, carboxymethylcellulose, and carboxypolymethylene (carboxy vinal polymers sold under the name Carbopol, B. F. Goodrich).

Gels or jellies are used pharmaceutically as lubricants and also as carriers for spermacidal agents to be used intravaginally with diaphragms as an adjunctive means of contraception. Since the fluid phase of a gel does not have to be strictly water, gels offer a wide range of uses; by blending solvents, it is possible to form films that exhibit a range of evaporation rates, solvency, and other release-determining attributes [51]. Gel products containing anti-inflammatory steroids are used to treat inflammations of the scalp because this is an area of the body where creams and ointments are too greasy for patient acceptance.

F. Rigid Foams

Foams are systems in which air or some other gas is emulsified in a liquid phase to the point of stiffening. As spreadable topical systems go, medicated foams tend towards the fluid side, but like some shaving creams, they can be stiffer and approximate a true semisolid. Like the second type of o/w emulsion that only borders on semisolidity, these derive structure from an internal phase, bubbles of an entrapped gas so voluminous that it exceeds close spherical packing. Consequently, the bubbles interact with their neighbors over areas rather than points of contact. The interactions are sufficient in many cases to provide a resistance to deformation and something approaching semisolid character. Whipped cream is a common example of this type of system. Here air is literally beaten into the fluid cream until it becomes stiff. Aerosol shaving creams and certain medicated quick-breaking antiseptic foams are examples of the foams currently found in cosmetic and therapeutic practice. These are supplied in pressurized cans that have special valves capable of emulsifying a gas into the extruded preparations.

G. Common Constituents of Dermatological Preparations

So many materials are used as pharmaceutical necessities and as vehicles in topical systems that they defy thorough analysis. The pharmacist should nevertheless make some effort to learn about the more common constituents and their principal functions. The compositions of formulations as presented on product labels are the main source of such information. The compositions presented in Table 9 offer a glimpse into the compositional natures of semisolids.

Due to the large number of materials that are used in topical preparations and the diversity of their physical properties, the formulation of topical dosage forms tends to be something of an art perfected through experience. Only by making myriad recipes does one eventually gain insight into the materials and their use in designing new formulations. Such insight allows the experienced formulator to manipulate the properties of existing formulations to gain a desired characteristic. Often one finds good recipes to use as starting points for formulations in the trade literature. Two factors have to be kept in mind when borrowing the compositions of such trade formulations. First, the trade recipes (recipes supplied with advertising material touting specific components) are often inadequately tested in terms of long-term stability. Second, the dominant features used in judging the merits of trade-promoted formulas tend to be their initial appearances and overall elegance. Little to no attention can be paid to the drug-delivery attributes of the prototypical systems when they are first prepared in the suppliers' laboratories because the drug-delivery attributes are so compound specific. Thus it is left up to the pharmacist (industrial research pharmacist) to make adjustments in the formulas that are consistent with good delivery of specific drugs. Each drug requires unique adjustments in accord with its singular physicochemical properties.

H. General Methods of Preparation of Topical Systems

Irrespective of whether the scale of preparation is large or small, ointments, pastes, and creams tend to be produced by one or the other of two general methods. Some are made at high temperature by blending liquid and melted solid components together and then dispersing all other ingredients within the hot, oily melt. Alternatively, drug and/or adjuvants can be dispersed or dissolved within one of the phases or a fraction of one of the phases of an emulsion prior to forming the emulsion. The drug can be mixed into a freshly formed, still molten emulsion while it is still warm. Finally, a drug can be incorporated into an already solidified base via cold incorporation. As pointed out earlier, the

Copyright © 2002 Marcel Dekker, Inc.

Table 9 Prototype Formulations

I. Ointment (white ointment, USP)	
White petrolatum	95% (w/v)
White wax	5%
Melt the white wax and add the petrolatum; continue heating until a liquid melt is formed. Congeal with stirring. Heating should be gentle to avoid charring (steam is preferred), and *air incorporation* by too vigorous stirring is to be avoided.	
II. Absorption ointment (hydrophilic petrolatum, USP)	
White petrolatum	86% (w/w)
Stearyl alcohol	3%
White wax	8%
Cholesterol	3%
Melt the stearyl alcohol, white wax, and cholesterol (steam bath). Add the petrolatum and continue heating until a liquid melt is formed. Cool with stirring until congealed.	
III. Water-washable ointment (hydrophilic ointment, USP)	
White petrolatum	25% (w/w)
Stearyl alcohol	25%
Propylene glycol	12%
Sodium lauryl sulfate	1%
Methylparaben	0.025%
Propylparaben	0.015%
Purified water	37%
Melt the stearyl alcohol and white petrolatum (steam bath) and warm to about 75°C. Heat the water to 75°C and add the sodium lauryl sulfate, propylene glycol, methylparaben, and propylparaben. Add the aqueous phase and stir until congealed.	
IV. Water-soluble ointment (polyethylene glycol ointment, USP 14)	
Polyethylene glycol 4000 (Carbowax 4000)	50%
Polyethylene glycol 400	50%
Melt the PG 4000 and add the liquid PG 400. Cool with stirring until congealed.	
V. Cream base, w/o (rose water ointment, NF 14)	
Oleaginous phase	
Spermaceti	12.5%
White wax	12.0%
Almond oil	55.58%
Aqueous phase	
Sodium borate	0.5%
Stronger rose water, NF	2.5%
Purified water, USP	16.5%
Aromatic	
Rose oil, NF	0.02%
Melt the spermaceti and white wax on a steam bath. Add the almond oil and continue heating to 70°C. Dissolve the sodium borate in the purified water and stronger rose Water, warmed to 75°C. Gradually add the aqueous phase to the oil phase with stirring. Cool to 45°C with stirring and incorporate the aromatic (rose oil).	
Note: This is a typical cold cream formulation. The cooling effect comes from the slow evaporation of water from the applied films. The aromatic is added at as low a temperature as possible to prevent its loss by volatilization during manufacture.	
VI. Cream base, o/w (general prototype)	
Oleagenous phase	
Stearyl alcohol	15%
Beeswax	8%
Sorbitan monooleate	1.25%
Aqueous phase	
Sorbitol solution, 70% USP	7.5%
Polysorbate 80	3.75%

Copyright © 2002 Marcel Dekker, Inc.

Table 9 (*Continued*)

Methylparaben	0.025%
Propylparaben	0.015%
Purified water, q.s. ad	100%
Heat the oil phase and water phase to 70°C. Add the oil phase slowly to the aqueous phase with stirring to form a crude emulsion. Cool to about 55°C and homogenize. Cool with agitation until congealed.	
VII. Cream base, o/w (vanishing cream)	
Oleagenous phase	
Stearic acid	13%
Stearyl alcohol	1%
Cetyl alcohol	1%
Aqueous phase	
Glycerin	10%
Methylparaben	0.1%
Propylparaben	0.05%
Potassium hydroxide	0.9%
Purified water, q.s. ad	100%
Heat the oil phase and water phase to about 65°C. Add the oil phase slowly to the aqueous phase with stirring to form a crude emulsion. Cool to about 50°C and homogenize. Cool with agitation until congealed.	
Note: In this classic preparation, the stearic acid reacts with the alkaline borate to form the emulsifying stearate soap.	
VIII. Paste (zinc oxide paste, USP)	
Zinc oxide	25%
Starch	25%
Calamine	5%
White petrolatum, q.s. ad	100%
Titrate the calamine with the zinc oxide and starch and incorporate uniformly in the petrolatum by levigation in a mortar or on a glass slab with a spatula. Mineral oil should *not* be used as a levigating agent, since it would soften the product. A portion of the petrolatum can be melted and used as a levigating agent is so desired.	
IX. Gel (lubricating jelly)	
Methocel 90 H.C. 4000	0.8%
Carbopol 934	0.24%
Propylene glycol	16.7%
Methylparaben	10.015%
Sodium hydroxide, q.s. ad	pH 7
Purified water, q.s. ad	100%
Disperse the Methocel in 40ml of hot (80°–90°C) water. Chill overnight in a refrigerator to effect solution. Disperse the Carbopol 934 in 20ml of water. Adjust the pH of the dispersion to 7.0 by adding sufficient 1% sodium hydroxide solution (about 12 ml is required per 100 ml) and bring the volume to 40 ml with purified Water. Dissolve the methylparaben in the propylene glycol. Mix the Methocel, Carbopol 934, and propylene glycol fractions using caution to avoid the incorporation of air.	

first of these methods is commonly used to make o/w creams of the vanishing type. The fusion method is used to prepare many ointments as well. Cold incorporation comes into play in large-scale manufacture when the systems in preparation contain heat-labile drugs. In this instance the drug is first crudely worked into an ointment or cream base using by serial dilution and then distributed uniformly with the aid of a roller mill. Cold incorporation is also necessary when a base is destroyed by heat, as happens with Plastibase (Squibb).

In the fusion method for ointments, mineral oil, petrolatum, waxes, and other ingredients that belong in the formulation are heated together to somewhere

Copyright © 2002 Marcel Dekker, Inc.

between 60 and 80°C, depending on the components, and mixed to a uniform composition while in the fluidized state. Cooling is then effected using some sort of heat exchanger. To prevent decomposition, drugs and certain delicate adjuvants are added sometime during the cooling process. If insoluble solids need to be dispersed, the system is put through a milling process (colloid mill, homogenizer, ultrasonic mixer, etc.) to disperse them fully. A hand homogenizer works well at the prescription counter for small-volume, extemporaneously prepared systems. Systems in preparation are always cooled with mild stirring until they are close to solidification. The rate of cooling is important, for rapid cooling, as mentioned, imparts a finer, more rigid structure. Stirring should be set to minimize vortexing and thereby prevent air incorporation into the solidifying system. Representative formulations with more system-specific, detailed directions are given in Table 9 for ointments and the other semisolid systems of note.

The fusion method for preparing creams is a bit more complex. In this instance the aqueous and oil phases are heated separately to somewhere between 60 and 80°C. As a general rule the oil phase is heated to 5°C above the melting point of the highest-melting waxy ingredient and the water phase is heated to 5°C above the temperature of the oil phase, the latter to prevent premature solidification during the emulsification process. Water-soluble ingredients are dissolved in the heated aqueous phase and oil-soluble ingredients are dissolved in the oily melt, but only as long as they are heat stable and not too volatile. If an o/w system is to be made, the emulsifiers are added to the aqueous phase and the emulsion is formed by slow addition of the oil phase. In the industry the crude emulsion is then passed through a high-shear mixer to form a finely divided emulsion state. Following this, the emulsion is cooled with gentle stirring until congealed, again taking care not to whip air into the formulation. Typically, the emulsions solidify between 40 and 50°C. If a w/o emulsion is to be made, the addition steps are usually reversed. Therefore and generally, the discontinuous phase is added to the continuous, external phase containing the emulsifier. However, methods vary here, and for a particular formula the reverse order of addition may work best. Any means that reliably leads to a good emulsion is obviously acceptable.

A solid can be incorporated directly into an already congealed system in several ways. This is accomplished on a small scale by levigating the solid with a small portion of the total base it is to be suspended in to obtain a paste-like mass. The drug is worked into the base on a glass plate with the aid of a spatula or is triturated in using a mortar and pestle. After the initial mix is smooth, a portion of the vehicle roughly equal in bulk to that of the pasty mass is added and blended in. This latter procedure is repeated several times (geometric dilution) until the drug is uniformly dispersed throughout its total vehicle. In large-scale manufacture, solids are crudely dispersed into the base using a blender and then roller mills, in which a film of the formulation is passed from one roller to another and so on, in each passage with kneading and mixing, are used to obtain fine dispersion. As outlined when discussing absorption bases, the drug may also be dissolved in water to form a solution to be levigated into an ointment base or cream. Such addition softens creams even to the point of converting them to thick lotions. The chosen vehicle, of course, must have an inherent capacity to emulsify or otherwise take up the solution. Aromatic materials such as essential oils, perfume oils, camphor, and menthol, which volatilize if added when the base is hot, are incorporated into these semisolids while they are still being mixed but near the temperature where a particular system starts to congeal. Volatile materials are often introduced into the formulation as hydroalcoholic solutions.

The preparation of gels can also involve high-temperature processing. It is easier to disperse methylcellulose in hot than in cold water, for instance. The polymer then goes into solution and thickens or sets up as the temperature is lowered. Adding the hot methylcel1ulose dispersion to ice water gets one quickly to the final equilibrium state. Tragacanth gels, on the other hand must be prepared at room temperature due to the extreme heat lability of this natural gum. A little alcohol or propylene glycol can be mixed into this gum before adding water to it to facilitate wetting of the gum prior to its dispersion. By way of contrast, Carbopol-containing systems are gelled by neutralizing the medium they have been dispersed in with alkali. Neutralization induces carboxyl groups found on the polymer backbone to ionize, instantaneously drawing these polymers into solution. Organic solvents can be gelled with Carbopol polymers as well, in such instances using soluble amines for the neutralization step.

Several prototype gel formulations are given in Table 9 to illustrate general compositional requirements and manufacturing methods. The design of specific systems tailored to meet predetermined, demanding performance criteria, particularly with respect to bioavailability, generally requires modification of published formulations or a totally original approach.

Copyright © 2002 Marcel Dekker, Inc.

VI. PERFORMANCE OF TOPICAL THERAPEUTIC SYSTEMS

Topical preparations, like all other dosage forms, must be formulated, manufactured, and packaged in a manner that assures that they meet general standards of bioavailability, physical (physical system) stability, chemical (ingredient) stability, freedom from contamination, and elegance. Like all other pharmaceuticals, these factors must remain essentially invariant over the stated shelf life of the product and they must be reproducible from batch to batch.

A. Bioavailability

Chemical Structure, Delivery, Clinical Response

Much has already been said concerning the chemical structural dependencies of skin permeation. However, the goal of all treatment is successful therapy, not delivery per se, and consequently the intrinsic activities of the drugs must also be taken into account when selecting compounds for dermatological and transdermal development. The pharmacological response depends on delivering sufficient drug of a given activity to the target zone. Clearly, the more potent a compound is, the less of it that needs to be delivered. Since topical delivery is difficult at best, potency often dictates which of the compounds found within a family of drugs should be developed, for the highly potent analog, reasonably formulated, offers the best chance of obtaining clinically sufficient delivery. Conversely, marginally potent analogs, even when expertly formulated, often fail because of inadequate delivery. An excellent example of this principle is found with the narcotic analgesics. Because of its extraordinary potency, fentanyl, with a daily palliative requirement of 1 mg, and not morphine, which requires between 60 and 120 mg to alleviate pain over the course of a day, has made it way into transdermal use. The fact that fentanyl is also physicochemically more suited to transdermal delivery than is morphine does not controvert this axiom.

Unlike mass transport across membranes, which relates to chemical structure in predictable ways, the potencies of drugs as seen in pharmacological, pharmacodynamic, or other tests are highly structurally specific within a class of drugs and are without commonality across classes. A drug's activity involves a complex merging of these separate structural influences, with bioavailability always one of the concerns. Such concern is minimal when a truly superficial effect is involved, however. For example, the most potent antiseptic as measured in the test tube is likely to have near the highest topical potency as well. The intrinsic activities of compounds may be poor indicators of relative topical potentials when deep skin penetration is required, however, because the structural features benefiting the biological response are often distinct from than those that favor permeation. Thus, tissue permeability can be an important and sometimes a dominant factor in the clinical structure-activity profile.

We have seen that the determinants of skin permeation are the activity (concentration) of a drug in its vehicle, the drug's distribution coefficients between the vehicle and the skin and across all phases of the skin, and the drug's diffusion coefficients within the skin strata. Congeners, if comparably sized, exhibit little variance in their diffusion coefficients. However, the structural differences seen within congeneric families profoundly affect the solubility, partitioning, and in transit binding tendencies of the family members in addition to determining their binding with receptors. Drug delivery and resulting clinical effectiveness are a function of the former phenomena [52]. For example, the 21-ester of hydrocortisone is more hydrophobic than its parent, its ether/water partition coefficient being about 18 times that of hydrocortisone [53]. Given the strong parallels in partitioning behaviors that exist across partitioning systems, it stands to reason that similar order-of-magnitude increases exist with respect to the acetate's stratum corneum/water and sebum/-water partition coefficients. At the same time, acetylating hydrocortisone at the 21-position increases the melting point 12°C. Consequently, not only does derivatization drop the aqueous solubility precipitously, but it depresses solubility in all other solvents as well [54]. While the increase in partition coefficient raises the permeability coefficient relative to hydrocortisone, this impact is more than offset by reduced solubility, and far less of the acetate derivative can be delivered through the skin from respective saturated solutions [55]. However, as the alky1 chain length of the ester is methodically extended (C_3, C_4, . . .,C_7), the growing bulkiness of the alkyl group increasingly interferes with crystalline packing. Consequently, melting points fall incrementally from the 224°C peak of the acetate ester to 111°C when a chain length of seven, the heptanoate ester, is reached [54]. An especially sharp drop of 69°C is experienced between chain lengths five and six. Because of declining crystallinity beyond the chain length of two, solubilities of the esters in organic solvents rise markedly. Moreover, aqueous solubilities, though methodically depressed by increasing hydrophobicity, remain many times higher than they otherwise would

Copyright © 2002 Marcel Dekker, Inc.

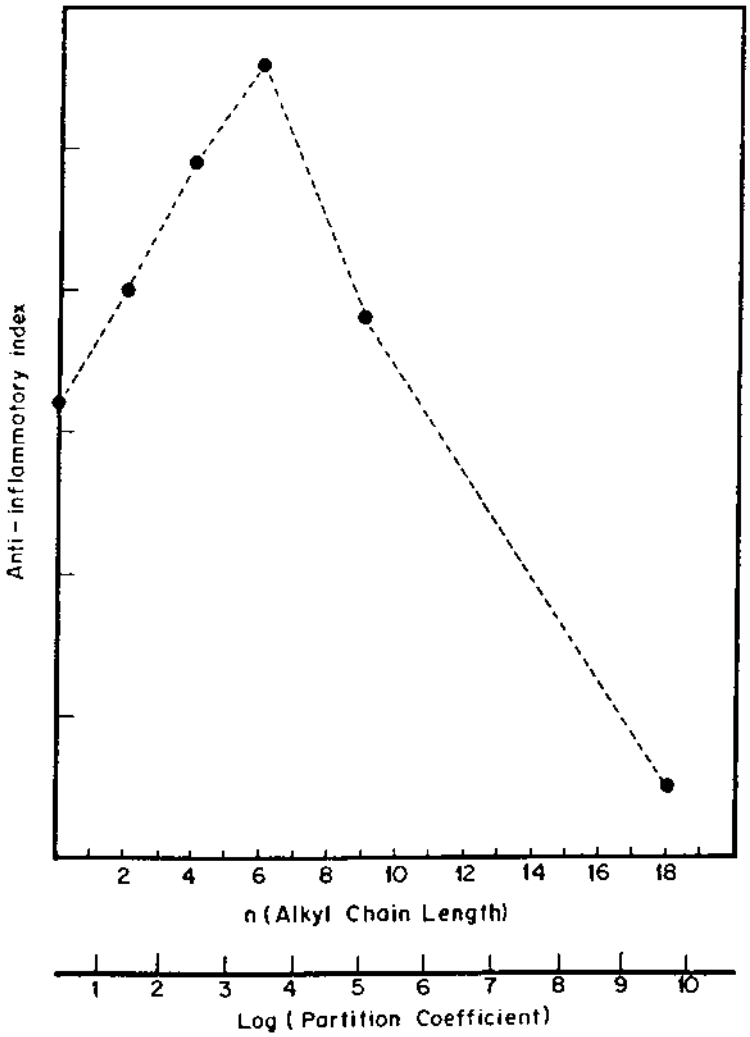


Fig. 5 Ability of hydrocortisone esters to suppress inflammation.

be. The net effect of these concerted forces is that the hexanoate and heptanoate esters of hydrocortisone are well over an order of magnitude more skin permeable than hydrocortisone when they are administered as saturated solutions [54,55].

Armed with the above insight, we can examine the pharmacological ramifications of esterifying hydrocortisone. In Fig. 5 the ability of hydrocortisone esters to suppress inflammation induced by tetrahydrofurfural alcohol, which acts simultaneously as irritant and vehicle, is shown as a function of the alkyl chain length of the esters [56]. An optimum chain in effect is seen at an alkyl chain length of six (hexanoate), with substantially longer and shorter esters being measurably less effective. The behavior is exactly what would be predicted from partitioning and solubility considerations. That this is not an isolated behavioral pattern with corticosteroids can be seen in Fig. 6, where vasoconstriction data of McKenzie and Atkinson for three betamethasone ester families, 21-esters, 17,21-*ortho*-esters, and 17-esters, are shown as functions of the ether/water partition coefficients of the compounds [57]. Vasoconstriction, blanching of the skin under the site of steroid application, is a proven index of a steroid's combined potency and ability to permeate through skin. Maxima are apparent in the

Copyright © 2002 Marcel Dekker, Inc.

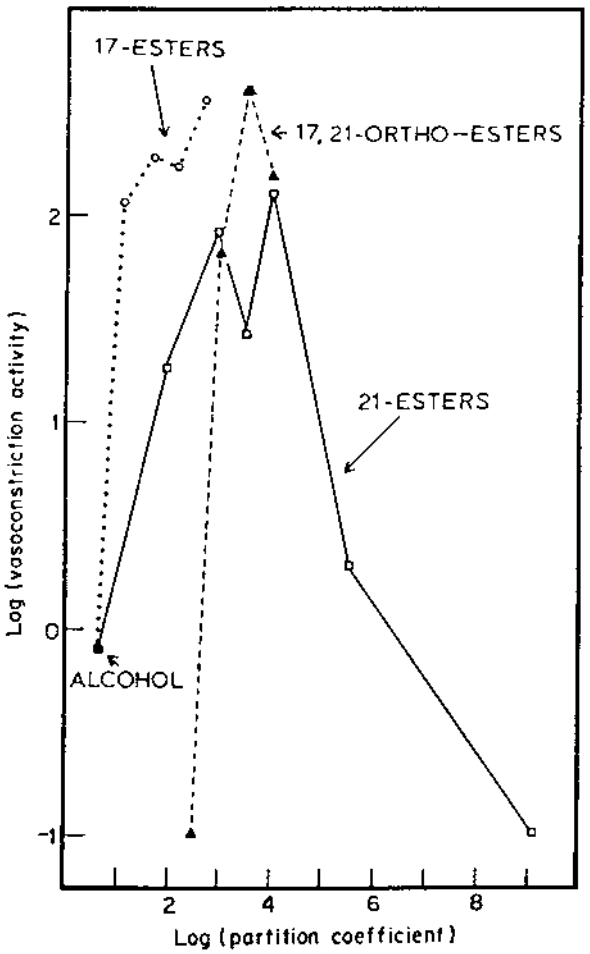


Fig. 6 Vasoconstriction data of betamethasone family.

data for the first two of these series, and the indications are that the 17-ester series is also peaking. Both maxima lie between ether/water partition coefficients of 1000 and 10,000 [53], as, interestingly and probably significantly, does the optimum ether/water partition coefficient of the hydrocortisone esters. The differing shapes and heights of the curves are not readily quantitatively explained but reflect differences in intrinsic vasomotor activities of each ester type. The coincidence of the maxima on the partitioning scale, on the other hand, seemingly relates to an optimum lipophilicity for delivery.

The decline in activity at the longer chain lengths (Fig. 6) also has a plausible explanation. Two factors are reasoned to be operative here: declining solubilities coupled with changes in the absorption mechanism associated with stratification of the barrier. Aqueous solubilities of homologs exponentially decline as alkyl chain length is extended [58,59]. While melting points are negatively impacted and depressed early in the series, the crystalline structure eventually accommodates the alkyl chain, whereupon further increases in chain length reverse the trend, reducing all solubilities [59]. Through all of this, o/w partition coefficients, which are unaffected by crystallinity, increase exponentially, which has the effect of exponentially increasing the ability of the stratum corneum and sebum phases to transport the steroids relative to that of the viable tissue layer. In other words, the resistance of the stratum corneum drops precipitously without a commensurate drop in the resistance of the viable tissue layer. As a result, the latter takes control of the permeation process [10,37,60]. Once this change in mechanism is manifest, the permeation of homologs from their saturated solutions mirrors the downward trend in aqueous solubilities, with further increases in chain length (hydrophobicity) marked with exponential declines in steady-state fluxes. The homologs quickly become inactive [60]. In effect, the homologs become so insoluble in water that they are thermodynamically restrained from partitioning into the viable tissues upon breaching the stratum corneum (or pore lipids). These features are in total agreement with expectations drawn from the earlier presented skin permeability model.

Clearly the physicochemical properties of a drug are a decisive factor in its overall activity. Where possible, molecular structures should be optimized to obtain best clinical performance. Rarely does an oral drug have physicochemical features suitable for topical or transdermal therapy, and it can take a great deal of systematic research to identify where the best balance of activity and permeability lies. Experience with corticosteroids suggests that as much as a 100-fold improvement in clinical activity may be attainable through molecular design, for today's most potent topical corticosteroids are more active than hydrocortisone by a factor at least this large.

Vehicle Properties and Percutaneous Absorption

The role solubility plays relative to maximal flux across membranes is clear from the preceding paragraphs. To kinetically reach the skin's surface, an appreciable fraction of a drug must also be in solution in the vehicle designed around it. Otherwise, diffusion of the drug through the vehicle to its interface with the skin may not completely compensate for drug lost through partitioning into the skin, kinetically dropping the drug's activity within this critical juncture of

Copyright © 2002 Marcel Dekker, Inc.

formulation and skin below saturation, lowering the drug's release into the surface tissues. Taken to an extreme, low vehicle solubility sets up a situation in which drug dissolution within and diffusion through the vehicle becomes delivery rate controlling [61,62]. In instances where a drug's pharmacological activity depends on getting as much drug as possible into the tissues, this is a problem. The outcome is similar when the drug is formulated in a highly unsaturated state in the first place. Again, it will not partition into the skin to the fullest possible extent, resulting in less than maximal bioavailability. Assuming maximal delivery is the goal, the optimum between these extremes is achieved by adjusting the solvency of the vehicle so that all or most of the drug is in solution but at the same time the vehicle is saturated or close thereto. This has the effect of balancing the kinetic and thermodynamic factors. It is for this reason that solvents like propylene glycol are added to topical formulations. Slowly evaporating propylene glycol provides a chemical environment in which drugs dissolve or remain dissolved, facilitating delivery. Therefore, one frequently finds propylene glycol (5–15%) in topical corticosteroid creams and other formulations.

These principles, which have been clinically validated, establish the critical role the vehicle plays in a drug's activity. While pharmacists do not have the wherewithal to actually test products at the dispensing counter, they nevertheless should be aware of these principles in order to select and dispense products from manufacturers who can demonstrate that such formulation factors have been given due consideration. These delivery dependencies also stand to caution the pharmacist. Extemporaneous mixing of commercial products, for example, one containing a steroid and another an antibiotic, and the diluting of products with homemade vehicle are suspect practices because the compositional changes associated with such blending are likely to adversely affect the delivery attributes of otherwise carefully designed systems [63,64].

There is another way vehicles can influence percutaneous absorption, which is by altering the physicochemical properties of the stratum corneum. In the main modification of the barrier results in increased skin permeability, but a buttressing effect is also achievable with substances having the capacity to solidify the horny structure. To repeat a point, simply hydrating the stratum corneum promotes absorption. This may be accomplished by covering the skin with a water-impermeable bandage or other wrapping (an occlusive dressing). The blockage of evaporative water loss leads to hydration of the stratum corneum, softening it and increasing the diffusive mobility of chemicals through it. The occlusive covering also prevents evaporation of volatile vehicle components, compositionally stabilizing a spread film, maintaining its solvency for the drug. It is estimated that occlusive hydration increases percutaneous absorption from 5-to 10-fold, enlargements that are often clinically significant [65]. The technique has been used with corticosteroids in refractory dermatoses such as psoriasis.

The following interesting phenomena associated with the occlusion of corticosteroids are enlightening. When applied under an occlusive dressing corticosteroids induce vasoconstriction at lower concentrations than when applied in the open. When the dressing is removed, vasoconstriction subsides in a few hours. However, as many as several days later blanching can sometimes be restored simply by rewrapping the area of application [66]. This suggests that steroid molecules somehow bottled up in the stratum corneum are released when occlusion is reestablished. It appears that as the stratum corneum dehydrates and returns to its normal state, substances such as the corticosteroids that may be present are entrapped within one of its physical domains, freezing them in place until either the stratum corneum is sloughed or until occlusive hydration is reinstated. The phenomenon is referred to as the skin's *reservoir effect*. The application of drugs dissolved in volatile solvents like acetone and ethanol also creates reservoirs in the stratum corneum for, as the solvents evaporate, the concentrating drug is driven into the skin's surface. Certain solvents also momentarily increase the solvency of the stratum corneum [67].

A few water-miscible organic solvents are taken up by the stratum corneum in amounts that soften its liquid crystalline, lipoidal domain [67], particularly when applied in concentrated form. If used very liberally (under laboratory conditions), these so-called skin penetration enhancers even elute interstitial lipids and denature keratin [68,69]. Under admittedly artificial research conditions, the increases in percutaneous absorption resulting from their actions can be dramatic. Dimethylsulfoxide (DMSO), dimethylacetamide (DMA), and diethyltoluamide (DEET) are key examples. DMSO has long been touted as a skin penetration enhancer. Experimental studies indicate that it reversibly denatures keratin, opening up the protein matrix, facilitating permeation [68,69]. DMSO also extracts lipids from the skin when applied liberally. Even when used sparingly, neat DMSO is imbibed to a degree by the stratum corneum, increasing the ability of the tissue to dissolve substances of all kinds [70].

Copyright © 2002 Marcel Dekker, Inc.

This favors the absorption of drugs by allowing more drug to dissolve in the tissue, steepening diffusion gradients that are expressed across the tissue. Moreover, opposing flows of DMSO (inward) and water (outward) are set up when concentrated DMSO is placed over the skin that delaminates the stratum corneum, apparently with a pooling of solvent between the separated layers. Of all the possible actions of DMSO, the latter two seem to be the most important because extraction of skin lipids and denaturation of keratin require far more DMSO than found in a topical application of ordinary thickness (20–30 μm). These collective factors, in the extreme, effectively chemically remove the stratum corneum as a contributing part of the skin barrier [33]. In reality, however, the limited amounts of enhancer actually applied limit enhancement. Moreover, much of the enhancement capacity is lost if the solvents are in any way diluted. Thus, the use of neat organic solvents as skin penetration enhancers is an infrequent practice.

It has long been known that certain surfactants (e.g., sodium lauryl sulfate) are skin irritants, even in relatively dilute states, in part because they impair the barrier function of the stratum corneum, facilitating their own absorption. Concern about irritation precludes serious consideration of agents like this as enhancers. However, certain weaker amphiphilic substances (e.g., methyl oleate, glyceryl monolaurate, proplyene glycol monolaurate), some of which have long been used as ingredients in cosmetics if not therapeutic systems, are showing they have an unrealized potential as enhancers. Of special importance, surface-active (amphiphilic) substances are effective in the small amounts that can actually be applied to skin in spread films. Amphiphilic molecules penetrate into and blend with the stratum corneum's own lipids, which themselves are polar, amphiphilic substances. Thermoanalytical and spectroscopic evidence indicates that, in doing so, they relax the ordered structure of the stratum corneum's natural lipids, facilitating diffusion through existing channels and perhaps freeing up new channels [71–73]. Moreover, the reduction in liquid crystallinity invariably increases the capacity of the stratum corneum to dissolve substances, further magnifying the effect. It appears that relatively short alkyl chains (C_{10}, C_{12}) and a relatively weak polar end groups favor enhancement.

These emerging structural requirements of enhancement have launched a quest for new, even more powerful enhancers. Several potent amphiphilic compounds have surfaced from this pursuit that, at their worst, are only mildly tissue provocative, key examples being *N*-dodecylazacycloheptan-2-one (Azone®) and methyl decyl sulfoxide. Each of these compounds contains a short alkyl chain (Azone = C_{12}; decyl methyl sulfoxide = C_{10}) attached to a highly water-interactive but nonionic head group. Neither compound is at all water soluble, indicating that neither has the amphiphilic balance to form its own micelles. Of the two compounds, Azone has been the most scrutinized. It promotes the absorption of polar solutes at surprisingly low percentage concentrations. Its effects on animal skins have been especially profound; up to several hundred-fold improvements in the in vitro permeation rates of highly polar cyclic nucleosides through hairless mouse skin have been reported [74], for example. Azone does not appear to be comparably effective on human skin, but those actions it has are effected at low concentration [75]. As with any agent of the kind, its actions are dependent on formulation and how this affects the thermodynamic activity of the enhancer in the delivery system. Concern over toxicity and the availability of alternative substances with established safety pedigrees have become impediments to the introduction enhancers having new, totally unfamiliar chemical structures.

The effects of skin penetration enhancers on the stratum corneum may or may not be lasting, depending on the degree of chemical alteration of the stratum corneum that is experienced. Irreversibility is a perceived problem to the extent that the skin is left vulnerable to the absorption of other chemicals that come in contact with the conditioned area for as long at the area remains highly permeable. The fear is that such vulnerability will stay high until the greater part of the stratum corneum is renewed through mitosis, which minimally takes several days. Moreover, the enhancing solvents are themselves absorbed to some degree, another source of toxicological concern. DMSO, for instance, is known to increase intraocular pressure. DMA has been associated with liver damage. Azone may irritate, but its real liability is that its chemical structure is totally novel and without toxicological precedent. While worry over toxicity may be out of proportion with actual degrees of exposure attending the ordinary circumstances of clinical use of dermatological products, concern is nevertheless warranted given the occasional use of products over expansive areas.

Transdermal Delivery—Attributes of Transdermal Delivery Systems

We are learning more and more that the conditions of use of topical delivery systems have a profound

Copyright © 2002 Marcel Dekker, Inc.

influence on their performance. In this regard, transdermal systems, specifically the adhesive patches that are used to treat systemic disease, and dermatological products are subject to very different operating environments and conditions [76]. Transdermal delivery is aimed at achieving systemically active levels of a drug. A level of percutaneous absorption that leads to appreciable systemic drug accumulation is absolutely essential. Ideally one would like to avoid any build-up of a drug within the local tissues, but build-up is nevertheless unavoidable, for the drug is driven through a relatively small diffusional area of the skin defined by the contact area (absorption window) of the application. Consequently, high accumulations of drug in the viable tissues underlying the patch are preordained by the nature of the delivery process. Irritation and sensitization can be associated with such high levels, and therefore careful testing is done to rule out these complications before a transdermal delivery system gets far along in development.

Table 10 outlines general expectations associated with transdermal delivery. The water-impermeable backing materials of present and presumably most future transdermal systems cause make them occlusive. There is good reason for this. Foremost is again the fact that occlusion facilitates drug delivery. In labs where these systems are designed it has been learned that occlusion is often essential to achieving adequate rates of delivery. Furthermore, transdermal drugs such as nitroglycerin and nicotine are themselves relatively volatile compounds. While they can be packaged in a fashion that prevents drug loss, a backing material that is substantially impermeable is also needed to prevent these compounds from evaporating off into space after placing patches containing them on the skin. Impermeable polymer or foil backings also block the diffusive transport of body water to the atmosphere by way of the patch. Insensible perspiration at the site of the patch is thus held in check, but not without creating a substantially moist environment at the interface of the patch with the skin. Consequently, if given enough time, organisms already in the skin can colonize within this interface.

Table 10 Norms of Operation of Transdermal Patches

Occluded applications
Composition relatively invariant in use
System size (area) predetermined
Specific site prescribed for application
Application technique highly reproducible
Delivery is sustained
Generally operate at unit drug activity, at least operate at steady activity
Delivery is zero-order
Serum levels related to product efficacy
Bioequivalency based on pharmacokinetic (blood level) endpoint
Unavoidable local tissue levels consequential only to system toxicity
Individual dose interruptable
Whole system removed when spent
Delivery efficiency is low (only a fraction of drug content is delivered)

Other than possibly for the insensible perspiration they absorb, transdermal patches tend to operate as thermodynamically static systems, meaning as compositionally fixed systems, from the moment they are applied until their removal. Marketed ethanol-driven estradiol and fentanyl patches are exceptions because they meter out ethanol and drive it into the stratum corneum to propel the absorption process. Compositional steadfastness is still the rule, however, and it is this feature that bestows the zero-order delivery attribute on the ordinary transdermal patch. Drug is present within the patches in reservoir amounts whether or not the reservoir compartment is easily distinguished, for there must be enough drug to sustain delivery over the full course of patch wear.

In some prototypes, e.g., the nitroglycerin transdermal systems, huge excesses of drug are placed in the patch to assure that the drug's activity remains essentially level. Only a small fraction of the drug, well under 50% of the patch's total content, is actually delivered during the prescribed time the patch is to be worn. In situations where the drug is prohibitively expensive or prone to abuse (e.g., fentanyl), efficiencies have to be raised to the maximum physically achievable, and the fractional delivery of formulated drug has been made to approach or exceed 50%. The inherent stability of the delivery environment of patches results partly from their main materials of construction are polymers, fabricating laminates, and adhesives, all of which tend to be chemically robust. Solubilizing solvents (e.g., ethanol) and skin penetration enhancers (e.g., propylene glycol monolaurate) may also be present and their absorption into the skin may change patch compositions, but even here the processes are carefully orchestrated to gain a stable delivery environment over the long term.

A transdermal patch is a self-contained system that is applied as it is packaged, with its only

Copyright © 2002 Marcel Dekker, Inc.

manipulation being removal of the release liner to expose and ready its adhesive surface. The size of a patch, meaning its area of contact with the skin, is determined before it is made. All of this area or only an inner portion of it may actually be involved in drug delivery, but, either way, the area is fixed. Since absorption is proportional to area, to meet the differing drug requirements of individual patients, patches of different sizes are generally made available. The application site is also a constant of therapy in that a specific site or sites are recommended for use (not always for scientifically supportable reasons, e.g., nitroglycerine patches are worn over the heart!). Users tend to follow such dictates. Beyond this, the manner of application is also highly reproducible. Thus there is as tight a control over absorption area and application site variables here as can be found in all of therapy. The only variability not customarily controlled for is that associated with the skin's permeability itself, but even here some systems have been made to operate with high delivery precision by incorporating a rate-controlling membrane into them. Altogether the manner of use of the systems is highly reproduced from one application to the next.

Measures of function of transdermal systems distinguish them among the systems we use topically. Since systemic actions are sought, blood levels of the drug in question must reach and remain within therapeutic bounds. More often than not the requisite blood level is known from a drug's use by other routes of administration. Thus, a clear systemic target level usually exists, and an absolute rate of delivery commensurate with reaching this is a built-in feature of the patch. The requisite delivery rate can be estimated in several ways even before an attempt is made to design a transdermal system. For instance, once an upper size limit is set for the patch, the total daily oral requirement of the drug can be used to calculate the minimal delivery rate in $mg/cm^2/h$. This is done after making an appropriate downward adjustment in the transdermal dose to account for oral drug losses attributable to first-pass metabolism. Alternatively, the rate can be estimated from an established blood level and a known rate of systemic clearance (as the product of these). Comparing performances of transdermal delivery systems is a straightforward matter too. Bioequivalency of different systems built around a specific drug are easily measured in terms of the blood levels they produce. And if therapy isn't going well, one can bring delivery to a reasonably abrupt halt by simply removing a patch.

Topical Delivery—Attributes of Topical Delivery Systems

Topical delivery systems fill an important niche in therapy. Despite the fact that often less than 1% and almost never more than 15% of the drug in a dermatological application is systemically absorbed (systemically recoverable), topical delivery nevertheless allows one to achieve drug levels in local tissues far in excess of those that can be achieved by other means of administration. At the same time, systemic toxicities of the drug are rarely encountered with topical administration, with the exceptions occurring when dermatological formulations are used liberally over extensive areas. Because only small amounts of a drug are ordinarily applied topically, in most instances the amounts absorbed are so limited that one has trouble even measuring them. Thus, albeit imprecise, topical therapy actually represents a brute force form of drug targeting and has been discussed in this context. The principal drug-delivery systems for this purpose are of course ointments, creams, and gels, with miscellaneous other powder, liquid, and semisolid vehicles sometimes being employed. The *norms* of topical delivery, which are in striking contrast to the *norms* of transdermal delivery, are outlined in Table 11.

We tend to think of them being much the same, but the functioning of semisolid dermatological products stands in stark contrast with that of transdermal delivery systems. To begin with, most topical applications

Table 11 Norms of Operation of Dermatological Formulations

Open application
Experience profound compositional shifts in use
May experience phases changes in use
Site in the disease's location
Operate at variable drug activity
Highly nonstationary state kinetics
Application technique and amount are highly individualized
Applications short-acting
Local tissue levels tied to efficacy
Used on diseased, damaged skin
No easy bioequivalency endpoint
Systemic absorption absolutely undesirable, but some unavoidable
Therapy interruptable by washing off application
System removal inadvertent—wear and tear
Delivery efficiency is low (only a small fraction of drug is delivered)

Copyright © 2002 Marcel Dekker, Inc.

are left open to the atmosphere. Amounts applied per unit area depend on the individual making the application. Of singular importance with respect to system function, extraordinary physicochemical changes accompany the evaporative concentration of these formulations, possibly including the precipitation of the drug or other substances that were comfortably in solution at the moment of their application. Evaporative concentration can also upset the oil-to-water balance of emulsions, destabilizing them, at times causing them to break or invert. In a matter of hours, if not just minutes, a surface film or dry residue having a totally different delivery faculty than the bulk formulation may be all that is left of the application. Such precipitous changes, if out of control, can bring drug delivery to an abrupt halt.

The amounts of ointments and creams people apply are highly individualized. So are the techniques of application. Some patients vigorously rub semisolid formulations into the skin, while others just spread films until they are more or less uniform over the desired area. While pharmacokinetic assessments of a system's delivery attributes is ordinarily done using normal skin (in vitro) or on healthy volunteers (in vivo), the site of its clinical deployment is usually anything but normal. Rather, it is determined by the skin condition to be treated. Clearly, the manufacturer is without control over how a disease is expressed in a particular patient. For many diseases, disease manifestation can be anywhere on the body. Moreover, from individual to individual it varies in intensity and vastness. Thus, more area may be involved in one case than in another, and the barrier function of the skin may be more or less intact in any instance. This creates a set of imponderables with respect to delivery, efficacy, and safety.

The removal of the dermatological application is rarely deliberate. Rather, some substance is usually transferred to clothing, etc., some is absorbed, some evaporates, and some is inadvertently removed by activities such as bathing. Of course, applications can be deliberately washed from the skin if one wishes to terminate therapy. Partly because of their temporal, inhabitancy, local applications tend to be short acting relative to transdermal delivery systems. Other factors are the finite doses that are actually administered and the often rapid evaporative concentration of such films to compositions that cease supporting dissolution of the drug and its diffusion to the skin's surface. Consider that the application of a topical product to the skin in a representative, 20 μm thick layer places only 20 μg of drug over each cm^2 of skin when the drug is formulated at the relatively high concentration of 1%. Roughly 10 mg of stratum corneum covers each cm^2 of skin, enough for 20 μg of drug to effectively get lost in. Consequently, only a fraction of the drug that does enter the skin actually reaches the live tissues. Such finite doses do not sustain delivery, and thus delivery wanes after several hours irrespective of the wearability of the application or processes attending its evaporative shrinkage. Since all these attending processes defy quantification, there is precious little existing information to guide one concerning a fitting regimen of application for most topical dosage forms. Rather, dosing regimens evolve historically from collective clinical experience. All in all, topical therapy is an extraordinarily complex operation.

Compositional changes following the application of certain topical systems are unavoidable. Many o/w creams contain as much as 80–85% external phase, usually primarily water. Lotions and gels also contain volatile constituents in large proportion. All rapidly evaporate after their application, and, consequently, the drug-delivery system is the formed, concentrated film that develops on the skin and not the medium as packaged in the tube, jar, or bottle. Ingredients should be chosen to assure that compositional changes as invariably occur interfere as little as possible with delivery and therapy. In this regard, the rate at which the volatile components evaporate to form the equilibrium film can itself be a factor in bioavailability [77]. It has been reported, for instance, that a thinly applied corticosteroid preparation produced greater vasoconstriction than did thicker applications of the same material [78]. Though the total amounts of drug per unit area were greater with the thick films, responses were less in their case because evaporative concentration of the steroid in the applied medium proceeded more slowly. Even without knowing the mechanistic details, we can conclude from this that less steroid was driven into the skin from the thick applications in the course of the test. It has also been demonstrated that vasoconstriction is more pronounced at low concentrations when steroid is applied in volatile ethanol than when applied in propylene glycol [56]. While differences in solvency play a role here, it is also clear that the rapid evaporation of solvents like ethanol drives drug into the skin. Such observations emphasize the importance of distinguishing between the system as packaged and the transitional system following application. Unfortunately, this distinction is not always made and much topical delivery research aimed at assessing the relative abilities formulations have to deliver drug has been performed by placing extra-

Copyright © 2002 Marcel Dekker, Inc.

ordinarily thick layers of formulation over the skin. Such thick applications do not even remotely simulate the clinical release situation, especially when it comes to creams and gels. This area of drug delivery is in need of much research.

Formulators can use the tendency of creams, gels, and other systems to evaporatively concentrate to advantage. Solvents are chosen and blended so that the drug remains soluble in the formed film long after application is made. This can be accomplished by replacing a fraction of the water or other highly volatile solvent found in these systems with solvents of far lower volatility. As previously pointed out, 5–15% propylene glycol is found in many topical corticosteroid creams and lotions just for this reason.

In summary, the way a topical drug is formulated has a great deal to do with its clinical effectiveness, not a surprising conclusion given what is known about the relationships between bioavailability and formulation for other modes of administration. Yet in the area of topical drug performance, antiquated concepts and approaches to system design linger. In days when topical bioavailability was little understood and therefore ignored, formulators concentrated on vehicle elegance and stability. Attempts were made to design vehicles compatible with all types of drugs, so-called universal vehicles. Universal vehicles are still discussed in many standard texts. Today's technology and science clearly indicate that the universal vehicle is akin to a unicorn, beautiful but totally mythical. In the real world, each system must be designed around the drug it contains to optimize the clinical potential of the active ingredient. The duration of action will depend on how long the drug remains appreciably in solution within its spread film. These matters are carefully examined when a drug-delivery system, topical or not, reaches the FDA as an Investigational New Drug (IND), New Drug Application (NDA) or Abbreviated New Drug Application (ANDA) [79].

B. Aspects of Physical and Chemical Stability

Concern for the physical and chemical integrity of topical systems is no different than for other dosage forms. However, there are some unique and germane dimensions to stability associated with semisolid systems. A short list of some of the factors to be evaluated for semisolids is given in Table 12. All factors must be acceptable initially (within prescribed specifications), and all must remain so over the stated lifetime for the product (the product's shelf life).

Table 12 Factors for Evaluation of Semisolids

Stability of the active ingredient(s)
Stability of the adjuvants
Visual appearance
Color
Odor (development of pungent odor or loss of fragrance)
Viscosity, extrudability
Loss of water and other volatile vehicle components
Phase distribution (homogeneity or phase separation, bleeding)
Particle size distribution of dispersed phases
pH
Texture, feel upon application (stiffness, grittiness, greasiness, tackiness)
Particulate contamination
Microbial contamination and sterility (in the unopened container and under conditions of use)
Release and bioavailability

The chemical integrities of drugs, preservatives, and other key adjuvants must be assessed as a function of time to establish a product's useful shelf life from a chemical standpoint. Semisolid systems present two special problems in this regard. First, semisolids are chemically complex to the point that just separating drug and adjuvants from all other components is an analyst's nightmare. Many components interfere with standard assays, and therefore difficult separations are the rule before anything can be analyzed. Also, since semisolids undergo phase changes upon heating, one cannot use high-temperature kinetics for stability prediction. Thus, stability has to be evaluated at the storage temperature of the formulation, and this of course takes a long time. Under these circumstances problematic stability may not be evident until studies have been in progress for a year or more. Be this as it may, stability details are worked out in the laboratories of industry, the pharmacist ordinarily accepting projected shelf lives as fact. Some qualitative indicators of chemical instability that the pharmacist might look for are the development of color (or a change in color and/or its intensity) and the development of an off-odor. Often products yellow or brown with age as the result of oxidative reactions occurring in the base. Discolorations of the kind are commonly seen when natural fats and oils, e.g., lanolins, are used to build the vehicle. Extensive oxidation of natural fatty materials (ranci-

Copyright © 2002 Marcel Dekker, Inc.

dification) is accompanied by development of a disagreeable odor. One may also notice phase and texture changes in a suspect product. Pharmacists should take note of the appearances of the topical products they dispense, removing all those from circulation that exhibit color changes or become fetid. Changes in product pH also indicate chemical decomposition, most probably of a hydrolytic nature, and, if somehow detected, are reason to return a product.

Time-variable rheological behavior of a semisolid may also signal physical and/or chemical change. However, measures such as spreadability and feel upon application are probably unreliable indicators of a changing rheology, and more exacting measurements are necessary. A pharmacist does not ordinarily have the tools at hand to make accurate rheological assessments, but the equipment to do so is generally available and used within the development labs of the industry. Here one may find exquisitely sensitive plate and cone research viscometers that in principle precisely quantify viscosity and also utilitarian rheometers that put viscosities on a relative scale. The latter include extrusion rheometers, which measure the force it takes to extrude a semisolid through a narrow orifice, penetrometers, which characterize viscosity in terms of the penetration of a weighted cone into a semisolid, and Brookfield viscometers with spindle and helipath attachments. The latter measure the force it takes to drive a spindle helically through a semisolid. As used with semisolids, the utilitarian rheometers only provide relative, although quite useful, measures of viscosity. Increases (or decreases) in viscosity by any of these measuring tools indicate changes in the structural elements of the formulation. The gradual transformations in semisolid structure that take place are more often than not impermanent, in which case the systems are restored to their initial condition simply by mixing them. Substantial irreversible rheological changes are a sign of poor physical stability.

Changes in the natures of individual phases of or phase separation within a formulation are reasons to discontinue use of a product. Phase separation may result from emulsion breakage, clearly an acute instability. More often it appears more subtly as bleeding—the formation of visible droplets of an emulsion's internal phase in the continuum of the semisolid. This problem is the result of slow rearrangement and contraction of internal structure. Eventually, here and there, globules of what is often clear liquid internal phase are squeezed out of the matrix. Warm storage temperatures can induce or accelerate structural crenulation such as this; thus, storage of dermatological products in a cool place is prudent. The main concern with a system that has undergone such separation is that a patient will not be applying a medium of uniform composition application after application. Due to unequal distribution between phases (internal partitioning), one phase will invariably have a higher concentration of the drug than the other. Therefore, since semisolid emulsions, unlike liquid emulsions, cannot be returned to an even distribution by shaking, formulations exhibiting separation are functionally suspect and should be removed from circulation.

Pharmacists should also take a dim view of changes in the particle size, size distribution, or particulate nature of semisolid suspensions. They are the consequence of crystal growth, changes in crystalline habit, or the reversion of the crystalline materials to a more stable polymorphic form. Any crystalline alteration can lead to a pronounced reduction in the drug-delivery capabilities and therapeutic utility of a formulation. Thus, products exhibiting such changes are seriously physically unstable and unusable.

A more commonly encountered change in formulations is the evaporative loss of water or other volatile phases from a preparation while it is in storage. This can occur as the result inappropriate packaging or a flaw made in packaging. Some plastic collapsible tubes allow diffusive loss of volatile substances through the container walls. One will find this phenomenon occasionally in cosmetics that are hurried to the marketplace without adequate stability assessment, but rarely in ethical pharmaceuticals, which are time-tested. However, a bad seal may occur in any tube or jar no matter what its contents, with eventual loss of volatile ingredients around the cap or through the crimp. Such evaporative losses cause a formulation to stiffen and become puffy and its application characteristics change noticeably. There is corresponding weight loss. Under this influence the contents of a formulation may shrink and pull away from the container wall. These phenomena are most likely to be seen in creams and gels due the high fractions of volatile components that characterize them. Problems here are exacerbated when products are stored in warm locations.

Gross phase changes are detectable by eye upon close inspection of products. The package of course gets in the way of such analysis, but if a product is truly suspect, it should be closely examined by opening and inspecting the full contents of the container. A jar can be opened and its contents probed with a spatula without ruining the container. Close inspection of the

Copyright © 2002 Marcel Dekker, Inc.

contents of a tube requires destruction of the package, however. The easy way to do this is to cut off the seal along the bottom of the tube with scissors and then make a perpendicular cut up the length of the tube to the edge of the platform to which the cap is anchored. Careful further trimming a quarter of the way around the platform in each direction creates left and right panels that can be pealed back with tweezers to expose the tube's contents. Textural changes such as graininess, bleeding, and other phase irregularities are easily seen on the unfolded, flat surface. Normally it takes a microscope to reveal changes in crystalline size, shape and/or distribution, but palpable grit is a sure sign that a problem exists. Weight loss of a product, which is easily checked at the prescription counter, clearly indicates the loss of volatile ingredients (the weight of a suspect tube can be directly compared to the weight of a fresh tube, etc.). On the rare occasions when deterioration like this is noted by a pharmacist in the course of handling products or is reported to the pharmacist by a knowledgeable patient, the suspect package should be removed from circulation and the manufacturer informed of the action. If a problem seems general rather than isolated (i.e., to a single bad package), the FDA should be notified as well to safeguard the public. This agency will determine if a product has gone bad and a general recall is warranted.

C. Freedom from Contamination

Particulates

Numerous topical preparations contain finely dispersed solids. Pastes, for example, contain as much as or more than 50% solids dispersed in an ointment medium. Powders themselves are used topically. Many dermatological liquids and semisolids contain suspended matter. However present, the particles should be impalpable, i.e., incapable of being individually perceived by touch, so that the formulations don't feel gritty. The palpability of a particle is a function of its hardness, shape, and size. The pharmacist can only manipulate the latter, and thus it is important to prepare or use finely subdivided solids when making topical dosage forms. Individual particles greater than 50 μm in their longest dimension can be individually perceived by touch. The surface of the eye is substantially more sensitive, and a 10 μm particle can be distinguished here. Clearly, the presence of hard, palpable particulates in semisolids makes them abrasive, particularly when applied to disease- or damage-sensitized skin. Severe eye irritation is possible if ophthalmic ointments contain them. One particularly troublesome source of particulate contamination is flashings (tiny metal slivers and shavings) left over from the production tin and aluminum collapsible tubes. These often adhere electrostatically and tenaciously to tubing walls following cutting of the containers down to a particular size. Some escape removal by the washing and rinsing done to cleanse the empty containers. Consequently, a jet of exceedingly high-velocity air is blown into the open end of tubes just prior to their filling to remove all particulates. If this precaution is not taken, tiny metal slivers may be packaged with the product, posing the threat that they will become dislodged and instilled into the eye while the product is in use. For these reasons, the United States Pharmacopoeia/National Formulary (USP/NF) has a particulate test for ophthalmic ointments. In this test the ointments are liquefied in a petri dish at high temperature (85°C) for 2 hours and then solidified by cooling. Particles that have settled to the bottom of the shallow glass container are counted by microscopic scanning at 30 times magnification. The requirements are met is the total number of particles 50 μm or larger in any dimension does not exceed 50 in the 10 tubes tested and if not more than 8 particles are found in a single tube. Products that are put into the distribution channels have to meet this test. Nevertheless, the pharmacist should be on the lookout for particulate problems associated with commercial products. The pharmacist must also take measures to ensure that extemporaneously compounded formulations are free of particulates. Particular attention must be paid to the cleaning of collapsible tubes and other package parts prior to their use.

Microbial Specifications and Sterility

As of the USP XIX, ophthalmic ointments must be prepared and dispensed as sterile products (until opened for use). Presently in the United States, non-ophthalmic topical preparations do not need to be sterile, although they cannot contain pathogens and must have low microbe counts. The reasons ophthalmic sterility requirements were broadened to cover ointments are enlightening. In the mid-1960s there was an outbreak of extremely serious *Pseudomonas* eye infections in the Scandinavian block of countries, in some instances resulting in loss of sight. The source of the contamination was traced to antibiotic-containing ophthalmic ointments made by a regional manufacturer known for its high standards of manufacturing and quality control [80]. Pathogenic *Pseudomonas* organisms were found in both the products and the

Copyright © 2002 Marcel Dekker, Inc.

manufacturing facilities where the ointments were prepared. It was widely believed up until this time that pathogens could not and would not survive and grow in ointments and similar media. The presence of antibiotics in the preparations could only have added to the false sense of security this company had. This incident sent shock waves throughout the pharmaceutical world and spawned revisions in all world compendia. In the United States ophthalmic ointments must be sterile when dispensed. In Europe dermatological products which are to be used over broken skin must also be sterile.

The foregoing incident has special meaning to the dispensing pharmacist. Unopened ophthalmic ointments should be dispensed for each condition and should be given very short shelf-life datings. Patients should be advised to discard unused quantities of old preparations and to return for fresh supplies if and when chronic symptoms reappear. Similar advice and precautions are good practice with dermatological products like ointments that do not ordinarily contain microbial preservatives. Lotions, creams, and topical solutions that contain preservatives tend to remain pathogen-free after their packages have been opened, providing an extra measure of safety.

Preservatives have an important purpose in topical medications. Systems containing them tend to remain aseptic. Even if a few organisms subsist in the presence of the preservatives, these tend to be nonvegetative. Importantly, no pathogenic forms survive to cause problems. Preservatives are necessary for systems that have an aqueous phase, because water offers an environment that is particularly conducive to microbial growth. Therefore, all emulsions and aqueous solutions and suspensions should be preserved. However, choosing a preservative is no easy task, for the physical systems tend to be compositionally complex and polyphasic, affording many possible means for specific preservatives to be inactivated. In mass-produced products the effectiveness of the preservation system of formulations is checked by the USP preservative challenge test.

D. Pharmaceutical Elegance

A number of attributes of topical drug systems that may be classified as cosmetic can make patients more or less willing to use their medications (compliant). These include ease of application, the feel of the preparation once it is on the skin, and the appearance of the applied film. Ideally, the application should be undetectable to the eye and neither tacky nor greasy. Certain items, such as ointments and pastes, are, of course, intrinsically greasy, and suspensions of all types tend to leave an opaque, easily detectable film. Thus the extent to which the cosmetic features can be idealized is dependent on the nature and purpose of the dosage form.

The ease of application and method of application of a formulation depend on the physicochemical attributes of the system involved. Solutions and other highly fluid systems may be swabbed on, sprayed on, or rolled on. A cotton pledget or other applicator is often necessary to obtain an even application. Soft semisolid systems, on the other hand, may be spread evenly and massaged into the skin with the fingers, a procedure technically referred to as inunction. The spreadability is a rheological quality related to the nature and degree of internal structure of the formulation. Formulations such as pastes that are very stiff tend to be hard to apply; their application over broken or irritated skin can be disagreeable. The stiffness of a preparation can be upregulated or downregulated by manipulating the amounts of structure-building components of a vehicle and in some instances by adjusting the phase/volume ratio of semisolid emulsions. Thus, for ointments, increased spreadability can be obtained by decreasing the ratio of the waxy components (waxes and petrolatum) to fluid vehicle components (mineral oil, fixed oils). Greasiness of such preparations goes in the opposite direction. For o/w creams, decreasing the ratio of the internal phase to the external phase tends to make the systems more fluid. Substitution of more liquid oils for some of the high-melting waxy components of creams achieves the same end.

Tackiness and greasiness are determined by the physicochemical properties of the vehicle constituents that comprise the formed film on the skin. A sticky film is extremely uncomfortable, and, generally, considerable effort is directed to minimizing this inelegant feature. Where creams are concerned, waxy ingredients such as stearic acid and cetyl alcohol produce noticeably nontacky films. Stearic acid is the principal internal phase component of vanishing creams, systems that are virtually undetectable visually or by touch after inunction. On the other hand, propylene glycol, which may be added to creams and gels to solubilize a drug, tends to make systems tacky. The synthetic and natural gums used as thickening and suspending agents in gels and lotions tends to increase their tackiness, and, therefore, these materials are used as sparingly as function allows.

Creams tend to be invisible on the skin. The same is true for ointments, although the oiliness of ointments

Copyright © 2002 Marcel Dekker, Inc.

causes them to glisten to an extent. Whatever opacity creams and ointments have is due primarily to the presence of insoluble solids. These often imbue applications with a powdery or even crusty appearance. Dispersed solids are usually functional, as in calamine lotion, zinc sulfide lotion, zinc oxide paste, and so on, and are an implacable feature of these preparations. However, at times insoluble solids are added as tints to match the color of the skin and to impart opacity. Since individual skins vary widely in hue (pigmentation) and texture, tinting to a single color and texture is generally unsuccessful.

Evaluation of the cosmetic elegance of topical preparations can be accomplished scientifically, but it is questionable whether physical experiments on system rheology and the like offer appreciable advantage over the subjective evaluations of the pharmacist, the formulator, or other experienced people. Persons who use cosmetics are particularly adept and helpful as evaluators.

E. Skin Sensitivity—A Specific Toxicological Concern

One further problem of topical formulations associated with many ingredients and of special concern with preservatives is the development of skin sensitivity [81]. The skins of some individuals are particularly susceptible to an allergic conditioning to chemicals known as Type IV contact hypersensitivity. Haptens (chemicals like urushiol found in poison ivy) are absorbed through the skin and, while in the local tissues, chemically react with local proteins. Langerhans cells, the local cells involved in immunological surveillance, identify these now denatured proteins as foreign (nonhost). The Langerhans cells then leave the dermis by way of the lymphatics and enter the draining lymph node, where they complete the sensitization process by passing the allergen message on to resident lymphocytes (antigen presentation). Once sensitized, subsequent contact with the offending chemical (hapten) leads to inflammation and skin eruption. Many of the preservatives used in pharmacy are phenols and comparably reactive substances, compounds that have a high propensity to sensitize susceptible individuals. The pharmacist should be alert to this possibility and prepared to recommend discontinuing therapy and physician referral when allergic outbreak is evident or suspected. Moreover, the pharmacist should be ready to recommend alternative products that do not contain an allergically offending substance once it has been identified, assuming of course that therapeutically suitable alternatives exist.

Allergic incidents are widespread, and, from an allergy standpoint, it is useful that the ingredients of dermatological medications are listed on the package or in the package insert. This allows the pharmacist to screen products for their suitability for individuals with known sensitivities. Over-the-counter medications and cosmetics also contain a qualitative listing of their ingredients. The pharmacist thus has access to critical information he or she needs to safeguard patients relative to their known hypersensitivity.

REFERENCES

1. A. P. Lemberger, in *Handbook of Non-Prescription Drugs*, American Pharmaceutical Association, Washington, DC, 1973, p. 161.
2. G. L. Wilkes, I. A. Brown, and R. H. Wildnauer, CRC Crit. Rev. Bioeng., Aug., 453 (1973).
3. K. F. Rushmer, K. J. K. Buettner, J. M. Short, and G. F. Odland, Science, 154, 343 (1966).
4. R. J. Scheuplein and I. H. Blank, Physiol. Rev., 51, 762 (l971).
5. W. Montagna, *The Structure and Function of Skin*, 3rd ed., Academic Press, New York,1974.
6. P. M. Elias, S. Grayson, M. A. Lampe, M. L. Williams, and B. E. Brown, in *Stratum Corneum* (R. Marks and G. Plewig, eds.), Springer-Verlag, New York, 1983, pp. 53–67.
7. J. W. Tingstad, D. E. Wurster, and T. Higuchi, J. Am. Pharm. Assoc., 47, 187 (1958).
8. R. J. Scheuplein, J. Invest. Dermatol., 47, 334 (1965).
9. H. Baker and A. M. Kligman, Arch. Dermatol., 96, 441 (1967).
10. R. J. Scheuplein, J. Invest. Dermatol., 48, 79 (l967).
11. R. T. Woodburne, *Essentials of Human Anatomy*, Oxford University Press, New York, 1965, p.6.
12. S. Rothman, *Physiology and Biochemistry of the Skin*, University of Chicago Press, Chicago, 1954, p. 26.
13. A. E. McKenzie, Arch. Dermatol., 86, 611 (1962).
14. R. B. Stoughton, Southern Med. J., 55, 1134 (1962).
15. G. Sazbo, Adv. Biol. Skin, 3, 1 (1962).
16. M. Katz and B. J. Poulsen, in *Handbook of Experimental Pharmacology, New Series*, Vol. 28 (B. B. Brodie and J. Gillette, eds.), Springer-Verlag, New York, 1972, p. 103.
17. S. M. Peck and W. R. Russ, Arch. Dermatol. Syphilol., 56, 601 (1947).
18. M. R. Liggins, Patient Care, July 1, 56 (1974).
19. B. Idson, J. Soc. Cosmet. Chem., 24, 1972 (1973).
20. K. Laden, Am. Perfum. Cosmet., 82, 77 (1967).
21. J. N. Robinson, in *Handbook of Non-Prescription Drugs*, American Pharmaceutical Association, Washington, DC, 1973, p. 209.
22. R. B. Stoughton, Clin. Pharmacol. Ther., 16, 869 (1974).

Copyright © 2002 Marcel Dekker, Inc.

23. J. J. Voorhees, E. A. Duell, M. Stawiski, and E. R. Harrell, Clin. Pharmacol. Ther., 16, 919 (1973).
24. W. R. Good, Med. Device Diag. Ind., Feb., 35 (1986).
25. G. W. Cleary, in *Medical Applications of Controlled Release*, Vol. 1 (R. S. Langer and D. L. Wise eds.), CRC Press, Boca Raton, FL, 1983, pp. 203–251.
26. Y. W. Chien, *Novel Drug Delivery Systems*, Marcel Dekker, New York, 1982, pp. 149–217.
27. J. P. Marty, R. H. Guy, and H. I. Maibach, *Percutaneous Absorption*, 2nd ed. (R. L. Bronaugh and H. I. Maibach, eds.), Marcel Dekker, New York, 1989, pp. 511–529.
28. C. R. Behl, N. H. Bellantone, and G. L. Flynn, in *Percutaneous Absorption* (R. L. Bronaugh and H. I. Maibach, eds.), Marcel Dekker, New York, 1985, pp. 183–212.
29. L. B. Fisher, in *Percutaneous Absorption*, 2nd ed. (R. L. Bronaugh and H. I. Maibach, eds.), Marcel Dekker, New York, 1989, pp. 213–222.
30. D. A. Weingand, C. Haygood, J. R. Gaylor, and J. H. Anglin, Jr., in *Current Concepts in Cutaneous Toxicity* (V. A. Drill and P. Lazar, eds.), Academic Press, New York, 1980, pp. 221–235.
31. D. A. Weingand, C. Haygood, and J. R. Gaylor, J. Invest. Dermatol., 62, 563 (1974).
32. H. H. Durrheim, G. L. Flynn, W. I. Higuchi, and C. R. Behl, J. Pharm. Sci., 69, 781 (1980).
33. G. L. Flynn, E. E. Linn, T. Kurihara-Bergstrom, S. K. Govil, and S. Y. E. Hou, in *Transdermal Delivery of Drugs*, Vol. 2, CRC Press, Boca Raton, FL, 1987.
34. C. G. T. Mathias, in *Dermatotoxicology*, 2nd ed. (F. N. Marzulli and H. I. Maibach, eds.), Hemisphere Publishing, New York, 1983, pp. 167–183.
35. M. S. Roberts, A. A. Anderson, J. Swarbrik, and D. E. Moore, J. Pharm. Pharmacol., 30, 486 (1978).
36. H. P. Bader, L. A. Goldsmith, and L. Bonar, J. Invest. Dermatol., 60, 215 (1971).
37. J. A. Keipert, Med. J. Aust., 1, 1021 (1971).
38. G. L. Flynn, in *Percutaneous Absorption: Mechanism—Methodology—Drug Delivery*, 2nd ed. (R. Bronaugh and H. I. Maibach, eds.), Marcel Dekker, New York, 1989, pp. 27–51.
39. G. L. Flynn, S. H. Yalkowsky, and T. J. Roscman, J. Pharm. Sci., 63, 479 (1974).
40. R. J. Scheuplein and R. L. Bronaugh, in *Biochemistry and Physiology of the Skin*, Vol. 2 (L. A. Goldsmith, ed.), Oxford University Press, New York, 1983, pp. 1255–1295.
41. T. Higuchi and S. S. Davis, J. Pharm. Sci., 59, 1376 (1970).
42. S. S. Davis, T. Higuchi, and J. H. Rytting, Adv. Pharm. Sci., 4, 73 (1974).
43. R. J. Scheuplein, I. H. Blank, G. J. Brauner, and D. J. MacFarlane, J. Invest. Dermatol., 52, 63 (1969).
44. G. L. Flynn, in *Principles of Route-to-Route Extrapolation for Risk Assessment* (T. R. Gerrity and C. J. Henry, eds.), Elsevier, New York, 1990, pp. 93–128.
45. R. O. Potts and R. H. Guy, Pharm. Res., 9, 663 (1992).
46. N. Z. Erdi, M. M. Cruz, and O. A. Battista, J. Colloid Interface Sci., 28, 36 (1968).
47. P. Thau and C. Fox, J. Soc. Cosmet. Chem., 16, 359 (1965).
48. S. Foster, D. E. Wurster, T. Higuchi, and L. W. Busse, J. Am. Pharm. Assoc. (Sci. Ed.), 40, 123 (1951).
49. B. W. Barry, J. Pharm. Pharmacol., 21, 533 (1969).
50. S. Amott, A. Fulmer, W. E. Scott, I. C. M. Dea, R. Moorhouse, and D. A. Rees, J. Mol. Biol., 90, 269 (1974).
51. *Physician's Desk Reference*, 28th ed., Medical Economics, Oradell, NJ, 1974, p. 1460.
52. G. L. Flynn and T. J. Roseman, J. Pharm. Sci., 60, 1778 (1971).
53. G. L. Flynn, J. Pharm. Sci., 60, 345 (1971).
54. T. A. Hagen, *Physicochemical Study of Hydrocortisone and Hydrocortisone n-Alkyl-21-Esters*, thesis, University of Michigan, 1979.
55. W. M. Smith, *An Inquiry into the Mechanism of Percutaneous Absorption of Hydrocortisone and its n-Alkyl Esters*, thesis, University of Michigan, 1982.
56. C. A. Schlagel, Adv. Biol. Skin, 12, 339 (1972).
57. A. W. McKenzie and R. M. Atkinson, Arch. Dermatol., 89, 741 (1964).
58. G. Saracco and B. Spaccamella-Marcheti, Ann. Chem., 48, 1357 (1958).
59. S. H. Yalkowsky, G. L. Flynn, and T. G. Slunick, J. Pharm. Sci., 61, 852 (1972).
60. G. L. Flynn and S. H. Yalkowsky, J. Pharm. Sci., 61, 838 (1972).
61. D. E. Wurster, Am. Perfum. Cosmet., 80, 21 (1965).
62. T. Higuchi, J. Soc. Cosmet. Chem., 11, 85 (1960).
63. K. H. Burdick, B. J. Poulsen, and V. A. Place, JAMA, 211, 462 (1970).
64. H. Schaefer and T. E. Redelmeier, *Skin Barrier—Principles of Percutaneous Absorption*, Karger, Basel, 1996, pp. 213–223.
65. A. W. McKenzie and R. B. Stoughton, Arch. Dermatol., 86, 608 (1962).
66. C. F. H. Vickers, Adv. Biol. Skin, 12, 177 (1972).
67. R. B. Stoughton, Arch. Dermatol., 91, 657 (1965).
68. R. J. Scheuplein and L. Ross, J. Soc. Cosmet. Chem., 21, 853 (1970).
69. S. G. Elfbaum and K. Laden, J. Soc. Cosmet. Chem., 19, 163 (1968).
70. R. Jones, Excipient effects on topical drug delivery paper presented at the Industrial Pharmaceutical Technology Section Symposium, 121st Annual Meeting of the American Pharmaceutical Association, Chicago, Abstracts, Vol. 4, p. 24.
71. M. Goodman and B. W. Barry, in *Percutaneous Absorption*, 2nd ed. (R. L. Bronaugh and

Copyright © 2002 Marcel Dekker, Inc.

H. I. Maibach, eds.), Marcel Dekker, New York, 1989, pp. 567–593.
72. W. Abraham and D. T. Downing, in *Prediction of Percutaneous Absorption* (R. C. Scott, R. H. Guy, and J. Hadgraft, eds.), IBC Technical Services, London, 1990, pp. 110–122.
73. B. D. Anderson, W. I. Higuchi, and P. Raykar, Pharm. Res., 5, 566 (1988).
74. W. I. Higuchi, W. W. Shannon, J. L. Fox, G. L. Flynn, W. F. H. Ho, R. Vaidyanathan, and D. C. Baker, in *Recent Advances in Drug Delivery Systems* (J. M. Anderson and S. W. Kim, eds.), Plenum Press, New York, 1984, pp. 1–7.
75. S. Y. E. Hou and G. L. Flynn, Pharm. Res., 3(Suppl.), 525 (1986).
76. G. L. Flynn and N. D. Weiner, in *Dermal and Transdermal Delivery—New Insights and Perspectives* (R. Gurny, ed.), APV, Stuttgart, 1993, pp. 33–65.
77. M. Goldman, B. J. Poulsen, and T. Higuchi, J. Pharm. Sci., 58, 1098 (1969)
78. B. J. Poulsen, E. Young, V. Coquilla, and M. Katz, J. Pharm. Sci., 57, 928 (1968).
79. V. P. Shah, T. M. Ludden, S. V. Dighe, J. P. Skelly and R. L. Williams, in *Topical Drug Bioavailability, Bioequivalence, and Penetration* (V. P. Shah and H. I. Maibach, eds.), Plenum Press, New York, 1993, pp. 414–424.
80. L. O. Killings, O. Ringerts, and L. Silverstolpe, Acta Pharm. Suec., 3, 219 (1966).
81. A. A. Fisher, E. Pascher, and N. B. Kanot, Arch. Dermatol., 104, 286 (1971).

Copyright © 2002 Marcel Dekker, Inc.

Chapter 9

Disperse Systems

Wandee Im-Emsap and Juergen Siepmann

Freie Universität Berlin, Berlin, Germany

Ornlaksana Paeratakul

Srinakharinwirot University, Nakhonnayok, Thailand

I. INTRODUCTION

A disperse system is defined as a heterogenous, two-phase system in which the internal (dispersed, discontinuous) phase is distributed or dispersed within the continuous (external) phase or vehicle. Various pharmaceutical systems are included in this definition, the internal and external phases being gases, liquids, or solids. Disperse systems are also important in other fields of application, e.g., processing and manufacturing of household and industrial products such as cosmetics, foods, and paints.

This chapter describes the basic principles involved in the development of disperse systems. Emphasis is laid on systems that are of particular pharmaceutical interest, namely, suspensions, emulsions, and colloids. Theoretical concepts, preparation techniques, and methods used to characterize and stabilize disperse systems are presented. The term "particle" is used in its broadest sense, including gases, liquids, solids, molecules, and aggregates. The reader may find it useful to read this chapter in conjuction with Chapters 8, 12, and 14, since they include some of the most important applications of disperse systems as pharmaceutical dosage forms [1].

A. Classification

Disperse systems can be classified in various ways. Classification based on the physical state of the two constituent phases is presented in Table 1. The dispersed phase and the dispersion medium can be either solids, liquids, or gases. Pharmaceutically most important are suspensions, emulsions, and aerosols. (Suspensions and emulsions are described in detail in Secs. IV and V; pharmaceutical aerosols are treated in Chapter 14.) A suspension is a solid/liquid dispersion, e.g., a solid drug that is dispersed within a liquid that is a poor solvent for the drug. An emulsion is a liquid/liquid dispersion in which the two phases are either completely immiscible or saturated with each other. In the case of aerosols, either a liquid (e.g., drug solution) or a solid (e.g., fine drug particles) is dispersed within a gaseous phase. There is no disperse system in which both phases are gases.

Another classification scheme is based on the size of the dispersed particles within the dispersion medium (Table 2). The particles of the dispersed phase may vary considerably in size, from large particles visible to the naked eye, down to particles in the colloidal size range, and particles of atomic

Copyright © 2002 Marcel Dekker, Inc.

Table 1 Classification Scheme of Disperse Systems on the Basis of the Physical State of the Dispersed Phase and the Dispersion Medium

Dispersed phase	Dispersion medium		
	Solid	Liquid	Gas
Solid	Solid suspension (zinc oxide paste, toothpaste)	Suspension (tetracycline oral suspension USP, bentonite magma NF)	Solid aerosol (epinephrine bitartrate inhalation aerosol USP, smoke)
Liquid	Solid emulsion (hydrophilic petrolatum USP, butter)	Emulsion (mineral oil emulsion USP, Milk)	Liquid aerosol (nasal sprays, fog)
Gas	Solid foam Foamed plastics Pumice	Foam Effervescent salts in water Carbonated beverages	None

Pharmaceutical and other examples are given in parentheses.

and molecular dimensions. Generally, three classes are distinguished: molecular, colloidal, and coarse dispersions. Molecular dispersions are homogeneous in character and form true solutions. Colloidal dispersions are intermediate in size between true solutions and coarse dispersions. The term "colloidal" is usually applied to systems in which the particle size of the dispersed phase is in the range of 1–1000 nm and where the dispersion medium is a liquid. Nanoparticles distributed within a liquid and aqueous polymer dispersions for pharmaceutical coating applications are examples of colloidal systems. (Colloidal dispersions are described in more detail in Sec. VI.) Dispersions containing larger dispersed phases, usually 10–50 μm in size, are referred to as "coarse dispersions," which include most pharmaceutical suspensions and emulsions.

The defined size ranges and limits are somewhat arbitrary since there are no specific boundaries between the categories. The transition of size ranges, either from molecular dispersions to colloids or from colloids to coarse dispersions, is very gradual. For example, an emulsion may exhibit colloidal properties, and yet the average droplet size may be larger than 1 μm. This is due to the fact that most disperse systems are heterogeneous with respect to their particle size [1–2].

Table 2 Classification Scheme of Disperse Systems on the Basis of the Particle Size of the Dispersed Phase

Category	Range of particle size	Characteristics of system	Examples
Molecular dispersion	< 1.0 nm	Particles invisible by electron microscopy; pass through semipermeable membranes; generally rapid diffusion	Oxygen molecules, potassium and chloride ions dissolved in water
Colloidal dispersion	1.0 nm–1.0 μm	Particles not resolved by ordinary microscope but visible by electron microscopy; pass through filter paper but not semipermeable membranes; generally slow diffusion	Colloidal silver sols, surfactant micelles in an aqueous phase, aqueous latices and pseudolatices
Coarse dispersion	> 1.0 μm	Particles visible by ordinary microscopy; do not pass through normal filter paper or semipermeable membranes	Pharmaceutical emulsions and suspensions

Copyright © 2002 Marcel Dekker, Inc.

In colloid science, colloidal systems are commonly classified as being lyophilic or lyophobic, based on the interaction between the dispersed phase and the dispersion medium. In lyophilic dispersions, there is a considerable affinity between the two constituent phases (e.g., hydrophilic polymers in water, polystyrene in benzene). The more restrictive terms hydrophilic and oleophilic can be used when the external phase is water and a nonpolar liquid, respectively. In contrast, in lyophobic systems there is little attraction between the two phases (e.g., aqueous dispersions of sulfur). If the dispersion medium is water, the term hydrophobic can be used. Resulting from the high affinity between the dispersed phase and the dispersion medium, lyophilic systems often form spontaneously and are considered as being thermodynamically stable. On the other hand, lyophobic systems generally do not form spontaneously and are intrinsically unstable.

The number of the constituent phases of a disperse system can be higher than two. Many commercial multiphase pharmaceutical products cannot be categorized easily and should be classified as complex disperse systems. Examples include various types of multiple emulsions and suspensions in which solid particles are dispersed within an emulsion base. These complexities influence the physicochemical properties of the system, which, in turn, determine the overall characteristics of the dosage forms with which the formulators are concerned.

Disperse systems can also be classified on the basis of their aggregation behavior as molecular or micellar (association) systems. Molecular dispersions are composed of single macromolecules distributed uniformly within the medium, e.g., protein and polymer solutions. In micellar systems, the units of the dispersed phase consist of several molecules, which arrange themselves to form aggregates, such as surfactant micelles in aqueous solutions.

B. Pharmaceutical Applications

Disperse systems have found a wide variety of applications in pharmacy [2]. Liquid dispersions, such as emulsions and suspensions, have the advantage of being easily swallowed and flexibly dosed compared to solid dosage forms. This is particularly important in the case of infants, children, and elderly patients (patient compliance). In addition, the small particle size of the drug present in disperse systems results in a large specific surface area. This leads to a higher rate of drug dissolution and possibly a superior bioavailability compared to solid dosage forms containing larger drug particles [3–7]. This can be of major importance in the case of poorly soluble drugs.

A suspension dosage form is often selected if the drug is insoluble in aqueous vehicles at the dosage required and/or when the attempts to solubilize the drug through the use of cosolvents, surfactants, and other solubilizing agents would compromise the stability or the safety of the product or, in the case of oral administration, its organoleptic properties [8–11]. The bitter or unpleasant tastes of dissolved drug molecules can often be improved by the selection of an insoluble form of the drug.

Colloidal systems have been used extensively in pharmacy, and their applications have grown rapidly over the last decades. Colloids have been employed in nuclear medicine as diagnostic and therapeutic aids (e.g., colloidal ^{198}Au, ^{99m}Tc, and sulfur); as adjuvants enhancing the immune effect of various agents (e.g., toxins adsorbed onto a colloidal carrier); and as anticancer agents (e.g., colloidal copper) [12]. Certain drugs show improved therapeutic effects when formulated in a colloidal state [13]. Colloidal silver chloride, silver iodide, and silver protein are effective germicides and do not cause irritation that is characteristic of ionic silver salts. More advanced applications of colloids include a variety of drug-delivery systems, the development of polypeptide chemotherapeutic agents, and the utilization of colloidal systems as pharmaceutical excipients, product components, vehicles and carriers. Colloidal drug-delivery systems, administered topically, orally, parenterally, or by inhalation, have been prepared for the purposes of drug targeting, controlled release, and/or protection of the drugs [14–18]. Naturally occurring plant macromolecules are capable of existing in the colloidal state and have been used for medical purposes, for example, hydroxyethyl starch (HES) has been used widely as plasma extender. Aqueous polymer latexes and pseudolatexes based on cellulosic and acrylic polymers have replaced organic solvent–based systems in a wide range of pharmaceutical coatings and controlled release technologies [19–22].

Patient acceptance is probably the most important reason why emulsions are popular oral and topical dosage forms. Oils and drugs having an objectionable taste or texture can be made more palatable for oral administration by formulating into emulsions [23–28]. As a result, mineral oil–based laxatives, oil-soluble vitamins, vegetable oils, high-fat nutritive preparations for enteral feeding, and certain drugs such as valproic acid, are formulated frequently in an oil-in-water (O/W) emulsion form. With topically applied

Copyright © 2002 Marcel Dekker, Inc.

emulsions, the formulation scientist can control the viscosity, appearance, and degree of greasiness of cosmetic and dermatological products. O/W emulsions are most useful as water-washable bases, whereas water-in-oil (W/O) emulsions are used widely for the treatment of dry skin and emollient applications to provide an occlusive effect. Semisolid preparations, such as ointments and creams, represent the dispersions of liquids in solids that are used topically. Emulsions are also employed in many other clinical applications as radiopaque emulsions, parenteral emulsions, and in blood replacement therapy [29–36].

Solid dispersions are disperse systems where both phases, the dispersed phase as well as the dispersion medium, are in the solid state. The drug can, for example, be present in the form of very fine particles or molecules distributed within an inert water-soluble carrier. This type of system is frequently used to enhance the dissolution rate and oral absorption of poorly water-soluble drugs [2,37]. Several techniques have been used to prepare such solid dispersions. The drug may be dissolved in the molten carrier. After cooling, a mixture of drug and carrier or a solid solution of the drug in the carrier results. Another method is to dissolve the drug and the carrier in a suitable organic solvent, followed by evaporation of the solvent and subsequent co-precipitation of the drug and carrier.

A solid or a liquid aerosol can be defined as a solid/gas or liquid/gas disperse system [38,39]. The main purpose of this dosage form is to deliver the drug to body surfaces or cavities, such as nasal passages or the respiratory tract. Pharmaceutical dosage forms include aerosol sprays, inhalations, and insufflations. Another type of the dispersion involving gas is the foam product, in which air or a propellant is emulsified within a liquid phase. Such preparations containing drugs (e.g., antibiotics or steroids) are often used topically. Foams containing spermicides are used for topical application within the vagina. Others have also been used widely for cosmetic purposes.

From a technical point of view, disperse systems are often involved in various steps of pharmaceutical manufacturing processes. The principles of colloid science govern many practices and operations in industrial pharmacy, such as particle size reduction (e.g., milling of pharmaceuticals), coating of pharmaceutical solids/solid dosage forms, microencapsulation, solubilization, and complexation of drugs. Solutions of water-soluble polymers act as viscosity-imparting agents, thus permitting the improvement of the physical stability of emulsions and suspensions. Organic polymer solutions or aqueous polymer dispersions are used as coating materials for solid dosage forms to achieve controlled drug release, to improve chemical stability, to provide taste masking and/or the identification of the product. Dispersions of surfactants provide a means of enhancing the solubility of a drug via micellar solubilization.

II. FUNDAMENTAL PROPERTIES

A. Particle Properties

The most significant characteristics of a dispersion are the size and shape of the dispersed particles. Both properties depend largely on the chemical and physical nature of the dispersed phase and on the method used to prepare the dispersion. The mean particle diameter as well as the particle size distribution of the dispersed phase have a profound effect on the properties of the dosage forms, such as product appearance, settling rate, drug solubility, resuspendability, and stability. Clinically important, the particle size can affect the drug release from the dosage forms that are administered orally, parenterally, rectally, and topically. Thus, these parameters have to be taken into account when formulating pharmaceutical products with good physical stability and reproducible bioavailability. This section discusses the basic principles of micromeritics, the science and technology of small particles. (Further information on the techniques and equipment currently used in the determination of particle size can be found in Sec. VIII.A.)

Particle Shape

Insoluble particles of drugs or pharmaceutical excipients obtained from various sources and manufacturing processes are seldom uniform spheres even after size reduction and classification. For example, precipitation and mechanical comminution generally produce randomly shaped particles unless the solids possess pronounced crystal habits or the solids being ground possess strongly developed cleavage planes. Primary particles may exist in various shapes, ranging from simple to irregular geometries. Their aggregation behavior produces an even greater variety of shapes and structures. The terminology and definitions of different particle shapes have been described previously [40]. Most common shapes found with pharmaceutical solids are spheres, cylinders, rods, needles, and various crystalline shapes.

Emulsification processes produce spherical droplets of the internal phase to minimize the interfacial area

Copyright © 2002 Marcel Dekker, Inc.

between the two phases. For example, the particles of true polymer latexes are spherical because they are typically prepared by emulsion polymerization, and the polymerization of solubilized monomer is initiated inside the spherical, swollen surfactant micelles [21,41]. Polymeric particles of pseudolatexes prepared from preformed polymers by either solvent evaporation or inverse emulsification processes are also spherical or near-spherical in shape for analogous reasons. Various pharmaceutical excipients exhibit characteristic particle morphologies determined by their molecular structure and the arrangement of the molecules. Some clay particles have plate-like structures possessing straight edges and hexagonal angles, e.g., bentonite and kaolin, while other clays may have lath-shaped or rod-shaped particles.

Characterization of the particle shape is generally described by the deviation from sphericity, as in the case of ellipsoids where the ratio of the two radii is the measure of deviation. The surface and volume are important properties affected by the overall shape of a particle. A more complicated relationship for particle characterization was described by Heywood, who introduced shape coefficients such as surface and volume coefficients and elongation and flatness ratios [42].

The information on particle shape is particularly important for the understanding of the behavior of suspensions during storage. The particle shape of the suspended particles (suspensoids) may have an impact on the packing of sediment (e.g., packing density and settling characteristics) and thus the product's resuspendability and stability. Packing density is defined as the weight-to-volume ratio of the sediment at equilibrium. Goodarznia and Sutherland [43] reported that the deviations from spherical shape and/or size uniformity can affect the packing density of suspensions containing cubes and spheroids. A wide particle size distribution often results in a high-density suspension, whereas widely differing particle shapes (e.g., plates, needles, filaments, and prisms) often produce low-density slurries. Symmetrical barrel-shaped particles of calcium carbonate were found to produce stable suspensions without caking upon storage, while asymmetrical needle-shaped particles formed a tenacious sediment cake, which could not be easily redispersed [44].

The viscosity of colloidal dispersions is affected by the shape of the dispersed phases. Sphero-colloids form dispersions of relatively low viscosity, while systems containing linear particles are generally more viscous. The relationship of particle shape and viscosity reflects the degree of solvation of the particles. In a good solvent, a colloidal particle unrolls and exposes its maximum surface area due to an extensive interaction between the dispersed phase and the dispersion medium. In contrast, in a poor solvent, the particle tends to coil up to assume a spherical shape, and the viscosity drops accordingly. Properties such as flow, sedimentation, and osmotic pressure are also affected by the changes in the particle shape of colloids.

Particle Size and Size Distribution

The particle size and size distribution of the dispersed phase represent a very important part of the knowledge required for a thorough understanding of any disperse system. The particle size can significantly affect the absorption behavior of a drug [6,7,45]. Certain types of dosage forms require specific size ranges, for example, suspension aerosols delivering drugs into the respiratory tract should contain particles in the order of 1–5 μm and no particles larger than 10 μm.

The size of a spherical particle is readily expressed in terms of its diameter. With asymmetrical particles, an equivalent spherical diameter is used to relate the size of the particle to the diameter of a perfect sphere having the same surface area (surface diameter, d_s), the same volume (volume diameter, d_v), or the same observed area in its most stable plane (projected diameter, d_p) [46]. The size may also be expressed using the Stokes' diameter, d_{st}, which describes an equivalent sphere undergoing sedimentation at the same rate as the sample particle. Obviously, the type of diameter reflects the method and equipment employed in determining the particle size. Since any collection of particles is usually polydisperse (as opposed to a monodisperse sample in which particles are fairly uniform in size), it is necessary to know not only the mean size of the particles, but also the particle size distribution.

The particle size data can be presented by graphical and digital methods. When the number or weight of particles lying within a certain size range is plotted against the size range or mean particle size, a bar graph (histogram) or a frequency distribution curve is obtained [13]. Alternatively, the cumulative percentage over or under a particular size can be plotted against the particle size. This results in a typical sigmoidal curve called a cumulative frequency plot. From these data, the mean particle size, standard deviation, and the extent of polydispersity may be determined.

Various theoretical distribution functions have been proposed, such as normal or Gaussian distribution and the log-normal distribution. The simplest case is

Copyright © 2002 Marcel Dekker, Inc.

described by the normal distribution equation with a specific mean value and standard deviation. However, for most pharmaceutical disperse systems, the normal equation is usually not adequate and the log-normal equation is commonly applied. For example, the log-normal distributions are often used to describe the particle size distributions of ground drug powders [40,47].

Special attention must be paid to the interpretation of particle size data presented in terms of either weight or number of particles. Particle weight data may be more useful in sedimentation studies, whereas number data are of particular value in surface-related phenomena such as dissolution. Values on the basis of number can be collected by a counting technique such as microscopy, while values based on weight are usually obtained by sedimentation or sieving methods. Conversion of the estimates from a number distribution to a weight distribution, or vice versa, is also possible using adequate mathematical approaches, e.g., the Hatch-Choate equations.

B. Surface Properties and Interfacial Phenomena

One of the most obvious properties of a disperse system is the vast interfacial area that exists between the dispersed phase and the dispersion medium [48–50]. When considering the surface and interfacial properties of the dispersed particles, two factors must be taken into account: the first relates to an increase in the surface free energy as the particle size is reduced and the specific surface increased; the second deals with the presence of an electrical charge on the particle surface. This section covers the basic theoretical concepts related to interfacial phenomena and the characteristics of colloids that are fundamental to an understanding of the behavior of any disperse systems having larger dispersed phases.

The Interface

An interface is defined as a boundary between two phases. The solid/liquid and the liquid/liquid interfaces are of primary interest in suspensions and emulsion, respectively. Other types of interfaces such as liquid/gas (foams) or solid/gas interfaces also p1ay a major role in certain pharmaceutical dosage forms, e.g., aerosols.

A large surface area of the dispersed particles is associated with a high surface free energy that renders the system thermodynamically unstable. The surface free energy, ΔG, can be calculated from the total surface area, ΔA, as follows:

$$\Delta G = \gamma_{SL} \cdot \Delta A \qquad \text{or} \qquad \Delta G = \gamma_{LL} \cdot \Delta A \tag{1}$$

where γ_{SL} and γ_{LL} are the interfacial tension between the solid particles and the liquid medium and the liquid/liquid interfacial tension, respectively.

Very small dispersed particles are highly energetic. In order to approach a stable state, they tend to regroup themselves in order to reduce the surface free energy of the system. An equilibrium will be reached when $\Delta G = 0$. This condition may be accomplished either by a reduction of the interfacial tension or by a decrease of the total surface area.

Particle Interactions

The interactions between similar particles, dissimilar particles, and the dispersion medium constitute a complex but essential part of dispersion technology. Such interparticle interactions include both attractive and repulsive forces. These forces depend upon the nature, size, and orientation of the species, as well as on the distance of separation between and among the particles of the dispersed phase and the dispersion medium, respectively. The balance between these forces determines the overall characteristics of the system.

The particles in a disperse system with a liquid or gas being the dispersion medium are thermally mobile and occasionally collide as a result of the Brownian motion. As the particles approach one another, both attractive and repulsive forces are operative. If the attractive forces prevail, agglomerates result indicating an instability of the system. If repulsive forces dominate, a homogeneously dispersed or stable dispersion remains.

Various types of attractive interaction are operating: (a) dipole-dipole or Keesom orientation forces, (b) dipole-induced dipole or Debye induction forces, (c) induced dipole–induced dipole or London dispersion forces, and (d) electrostatic forces between charged particles. The strongest forces are the electrostatic interactions between charged particles. These forces, either attractive or repulsive, are effective over a relatively long range and are dependent on the ionic charge and size of the particle. Forces that are weaker and effective over a shorter distance include other types of electrostatic forces such as ion-dipole, ion-induced dipole, and (a)–(c). The group of these last three forces are often referred to as van der Waals forces, which result from the interaction of electromagnetic dipoles within the particles. They produce a short-range type of interaction, varying inversely with r^6, where r is the interparticle distance. Even though they are relatively weak, van der Waals forces coupled with hydrogen

Copyright © 2002 Marcel Dekker, Inc.

bond interactions are significant factors describing the behavior of most nonionic compounds in liquids and other dispersion media. These various types of attractive interparticulate forces can lead to the instabilities of disperse systems described in Sec. II.C.

Electrical Properties of the Interfaces

Most insoluble materials, either solids or liquids, develop a surface charge when dispersed within an aqueous medium. These surface charges of particles may arise from several mechanisms. For example, the ionization of functional groups present at the surface of the particle, such as carboxylic acid or amine groups, can be involved. The charge is then dependent on the extent of ionization and is a function of the pH of the dispersion medium. Furthermore, surface charges can be developed due to the adsorption or desorption of protons. A variety of colloidal systems (e.g., polymers, metal oxides) fall into this group. For example, the surface hydroxyl groups of aluminum hydroxide gel can adsorb protons and become positively charged or may donate protons to become negatively charged. An important property of such systems is the point of zero charge (PZC), which represents the pH at which the net surface charge is zero. More general, surface charges of dispersed particles may be caused by preferential adsorption of specific ions onto the surface. Surface charges can also be introduced by ionic surfactants. For example, oil globules in an O/W emulsion exhibit a surface charge if anionic or cationic surface-active agents are used. The surfactant molecules are oriented at the oil/water interface so that the charged hydrophilic groups are directed towards the water phase. Ion deficiencies in the crystal lattice or interior of the particles may also cause surface charges. Many mineral clays exhibit a negative surface charge because of isomorphic substitution, e.g., an Al^{3+} ion occupies a site that is usually occupied by a Si^{4+} ion, resulting in a deficit of charge. Similarly, antacid magaldrate exhibits a positive surface charge: Mg^{2+} ions in the crystal lattice are partially substituted by Al^{3+} ions.

When a charged particle is dispersed within a dispersion medium containing dissolved cations and anions, the surface charges of the particle interact with the dissolved ions in solution. For a detailed description of the occurring phenomena, the reader is referred to the literature [13]. Roughly, the electric distribution at the interface is equivalent to a double layer of charge: a first layer being tightly bound, and a second layer that is more diffusive. The zeta (ζ) potential is defined as the potential at the shear plane (interface separating layers that move with the particle from layers that do not move with the particle) and can be regarded as the "apparent" or "active" charge of the particle.

Derjaguin, Landau, Verwey, and Overbeek developed a theory giving insight into the energy of interaction between suspended particles [51–53]. This theory is thus often referred to as the DLVO theory. It relates the stability of a disperse system to the electrolyte content in the continuous phase and provides an insight into the factors responsible for controlling the rate at which particles in disperse systems come into contact or aggregate. The process of aggregation subsequently accelerates particle sedimentation and affects redispersibility of the disperse systems. In general, the DLVO theory is applicable to pharmaceutical dosage forms [54–56] such as colloids, suspensions, and O/W emulsions. In the case of W/O systems, it must be applied with extreme caution. Figure 1 [13] shows typical potential energy curves for the interaction of two charged particles as a function of interparticle distances: the attractive energy curve (V_A), the repulsive energy curve (V_R), and the net or total potential energy curve (V_T). The total energy of interaction,

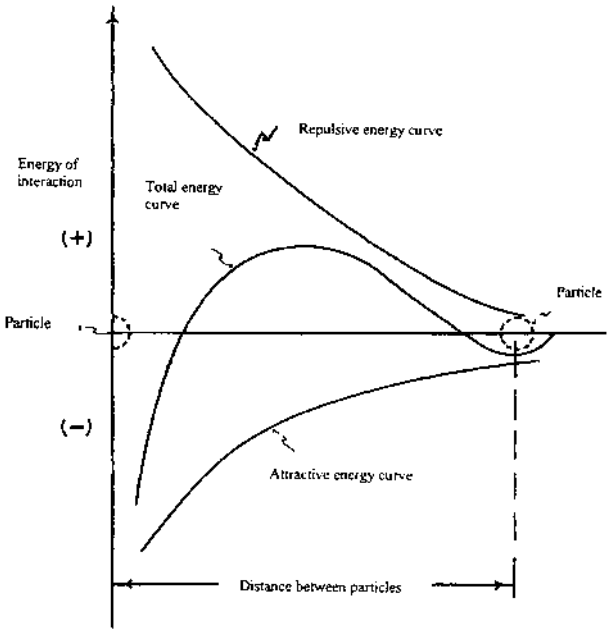


Fig. 1 Illustration of the DLVO theory: interaction of two charged particles as a function of the interparticle distance (attractive energy curve, V_A; repulsive energy curve, V_R; and net or total potential energy curve, V_T).

Copyright © 2002 Marcel Dekker, Inc.

V_T, between two particles is defined as the sum of the attraction and repulsion energies:

$$V_T = V_A + V_R \tag{2}$$

From these curves, it is evident that the attractive potential is predominant at short distances, and the net interaction is attraction in the deep potential energy minimum, V_p. At greater distances, the electrostatic repulsion energy falls off more rapidly with increasing separation distance than the van der Waals attraction energy, and the net interaction is attraction in the shallow secondary minimum, V_s. At intermediate distances, the electrostatic repulsion predominates and the net interaction is repulsion with the maximum potential, $V_{\max}$. The stability of a disperse system is indicated by the height of the maximum in the potential energy curve, $V_{\max}$. This potential energy barrier must be surmounted if the particles are to approach each other sufficiently closely to fall into the deep primary energy minimum of irreversible coagulation. To gain information about the stability of a disperse system, it is important to compare the total energy of the particles with the kinetic energy of the particles, kT, where k is the Boltzmann constant and T is the absolute temperature. The value of $V_{\max}$ necessary to prevent the irreversible contact of particles is considered to be about 10–20 kT, which correponds to a zeta potential of approximately 50 mV [41]. If the secondary minimum is deep enough (about 5 kT or greater), the particles might associate to form a cluster that is loose and reversible, called floc or floccule. Such flocculation in the secondary minimum, however, does not occur with small colloidal particles since the energy minimum is of the same order as the thermal energy of the particles, and the aggregation is easily reveresed by Brownian motion.

According to the DLVO theory, disperse systems become unstable whenever their kinetic energy is sufficient to overcome the potential energy barrier, $V_{\max}$. Thus, the instability of disperse systems increases when decreasing the height of this energy barrier and when increasing the kinetic energy of the particles. The reduction of $V_{\max}$ can result from the addition of substances that (a) neutralize the surface particle charge or cause the loss of the hydration layer, (b) compress the double layer, and/or (c) cause adsorbed species (e.g., surfactants) to desorb from the particle surface. The primary factor determining the thickness of the electric double layer is the potential energy drop-off. The potential gradient strongly depends on the concentration and charge of any electrolyte present in the dispersion medium. In water at room temperature, the thickness of the electric double layer ranges from 1 to 1000 Å, depending on the concentration of ions in the bulk phase. Increasing electrolyte concentrations lead to compressed double layers (decreasing double layer thicknesses). Thus, the particles can approach each other more closely, and the attractive forces become more important. The concentration of foreign electrolytes required to cause flocculation decreases as the valence of the coagulating ion increases. For example, fewer Al^{3+} ions are required to flocculate a suspension than Na^+ ions. This has been observed in many instances and is known as the Schulze-Hardy rule [13]. It has been shown that the quantity of electrolyte required to cause flocculation decreases by a factor of 10 when a monovalent electrolyte, e.g., Na^+, is replaced by a divalent electrolyte, e.g., Mg^{2+}. Furthermore, the efficiency of electrolytes in precipitating a dispersion depends on how extensively the electrolytes are hydrated. The Hofmeister or lyotropic series arranges ions in the order of increasing hydration and increasing efficiency in causing the instability. For example, the series for monovalent cations is $Cs^+ < Rb^+ < NH_4^+ < Na^+ < Li^+$ and for divalent ions is $Ba^{2+} < Sr^{2+} < Ca^{2+} < Mg^{2+}$.

Wetting

The wetting process is a primary concern, particularly in the preparation of a liquid disperse system in which the internal phase is a solid (suspension). A solid material to be suspended must first be separated into single particles, and the particles must be individually wetted by the dispersion medium to achieve a homogeneous distribution of the internal phase. Upon wetting, the air at the solid surface is replaced by the liquid medium. Obviously, the tendency of a solid to be wetted by a liquid is primarily determined by the interaction between the three phases. Experimentally, the degree of wetting of a powder can be evaluated by observing the contact angle, θ, which is defined by the boundaries of the solid surace and the tangent to the curvature of the liquid drop. The contact angle results from an equilibrium involing three interfacial tensions: those acting at the liquid/gas, solid/liquid, and solid/gas interfaces. A contact angle of 0° indicates an extensive interaction between the solid and the liquid phase. Thus, the solid is completely wetted by the liquid. Partial wetting occurs when the contact angles are between 0 and 90°. In contrast, contact angles greater than 90°are classified as nonwetting situations in which the liquid cannot spread over the solid surface

Copyright © 2002 Marcel Dekker, Inc.

spontaneously. If the angle is close to 180°, the solid substance is considered to be unwettable by the liquid.

Hydrophilic substances are readily wetted by water or other polar liquids due to the good interfacial interaction resulting in small contact angles. Once dispersed, they may significantly increase the viscosity of the liquid system. On the other hand, hydrophobic substances repel water but can easily be wetted by nonpolar liquids. They usually do not alter the viscocity of aqueous systems. In pharmaceutical dosage forms, aqueous vehicles or hydroalcoholic mixtures are often used. Thus, proper wetting of hydrophobic drugs is a necessary first step when preparing suspensions. Hydrophobic materials are often extremely difficult to disperse because of poor wetting or presence of entrained air pockets, minute quantities of grease, and other contaminants. The powder may just float on the surface of the liquid despite its higher density. Fine powders are particularly susceptible to this effect, and they may fail to become wetted even when mechanically forced below the surface of the suspending medium. To overcome this problem, to improve the wetting characteristics of hydrophobic drug powders, often anionic or nonionic surfactants are used. These substances decrease the solid/liquid interfacial tension, thus facilitating the wetting process. The mechanism of surfactant action involves the preferential adsorption of the hydrophobic part of the surfactant onto the hydrophobic surface of the particle. The polar part of the surfactant is directed towards the aqueous medium.

Different experimental methods have been developed to provide a measure for the efficiency of a wetting agent. For example, narrow lyophobic troughs can be used, in which one end holds the powder while a solution of a wetting agent is placed in the other end. The rate of penetration is observed and measured as the degree of wettability. Another technique involves measuring the relative ability of solutions of different wetting agents to carry a drug powder through a gauze when the solutions are dropped onto the gauze supporting a powder. Good wetting agents are able to carry greater amounts of powder through the gauze than poor wetting agents. Also for nonaqueous systems, different methods have been developed to compare the efficiency of wetting agents (e.g., for certain lanolin derivatives). In the paint industry two techniques are often used that could also be applied to pharmaceutical systems: the so-called wet and flow point methods. In the first case, the amount of vehicle required to wet a known amount of powder is measured, whereas in the second case, the amount of vehicle needed to prepare a pourable system is measured.

Adsorption

Adsorption is the tendency of atoms, molecules, ions, etc. to locate at a particular surface/interface in a concentration different from the concentrations in the surrounding bulk media. The adsorption process occurs as a result of the unequal distribution of the forces at the interfaces. The adsorbing species may be gases, solvents, or solutes, and the interfaces may be solid/solid, solid/liquid, solid/gas, liquid/liquid, or liquid/gas interfaces. Adsorption is termed "positive" if the concentration of adsorbed species at the interface is greater than that in the bulk and "negative" when the opposite is true. According to the nature of interaction, adsorption can be divided into two main classes: (a) physical adsorption or physisorption, where the forces and processes are reversible, nonspecific, and of relatively low energy, and (b) chemisorption, where the forces and processes of adsorption are specific, irreversible, and of higher energy.

Of special interest in liquid dispersions are the surface-active agents that tend to accumulate at air/liquid, liquid/liquid, and/or solid/liquid interfaces. Surfactants can arrange themselves to form a coherent film surrounding the dispersed droplets (in emulsions) or suspended particles (in suspensions). This process is an oriented physical adsorption. Adsorption at the interface tends to increase with increasing thermodynamic activity of the surfactant in solution until a complete monolayer is formed at the interface or until the active sites are saturated with surfactant molecules. Also, a multilayer of adsorbed surfactant molecules may occur, resulting in more complex adsorption isotherms.

Adsorption of species onto a particle surface is responsible for many resultant properties of the system. The adsorption of surfactants alters the properties at the interface and promotes wetting of the dispersed phase in suspensions. The reduction of the interfacial tension can effectively decrease the resulting surface free energy and hence the tendency for coalescence or aggregation. Thus, the formation and stabilization of the disperse systems can be significantly facilitated. An additional effect promoting stability is the presence of a surface charge, which causes electrostatic repulsion between adjacent particles. The adsorption of ionic surfactants generally increases the charge density on the surface of the dispersed particles, thus improving the stability of the dispersion.

Protective colloids or polymeric materials can be adsorbed onto the surface of the dispersed phase. Polymer adsorption can be accomplished simply by

Copyright © 2002 Marcel Dekker, Inc.

adding a solution of adsorbable polymeric species into a slurry of the dispersed particles. Then, adequate time must be provided for the system to equilibrate and complete the interaction between the adsorbent and the adsorbate. The mode of adsorption strongly depends on the number of sites and functional groups of the polymer chains available for the interaction with the particle surface. Most nonelectrolyte polymers promote steric stabilization, generally the result of either entropic stabilization or osmotic repulsion [41]. Entropic stabilization arises when two opposing adsorbed polymer layers of adjacent particles overlap, resulting in compression and interpenetration of their chain segments. The restriction of the movements of the polymer chains in the overlap region leads to a negative entropy change. Thus, the reverse process of disentanglement of the two adsorbed layers occurs and is energetically more favorable. This mechanism predominates when the concentration of polymer in the adsorbed layer is low. The osmotic repulsion is operative in cases where more polymer segments are involved and become crowded in the overlap regions. The increase in polymer concentration within these areas causes a local increase in osmotic pressure and results in an influx of water. This influx pushes the particles apart.

C. Instabilities and Stabilization Concepts

Content uniformity and long-term stability of a pharmaceutical product are required for a consistent and accurate dosing. Aggregation of dispersed particles and resulting instabilities such as flocculation, sedimentation (in suspensions), or creaming and coalescence (in emulsions) often represent major problems in formulating pharmaceutical disperse systems.

Instabilities

As discussed above, an instability of a dispersion may result from the tendency of the system to reduce its surface free energy, ΔG. Unfortunately, the terminology used in the literature to describe the occurring phenomena is not uniform. Flocculation is generally understood as a process in which particles are allowed to come together and form loosely bound clusters having an open structure. Unlike coalescence, the total surface area is not reduced during the flocculation process. Deflocculation is the opposite, i.e., breakdown of clusters into individual particles. Some authors differentiate between the terms aggregates and agglomerates: in aggregates the particles are more strongly bound than in flocculates, thus these systems are more difficult to be redispersed. The term agglomerates is then used as a general expression covering both flocculates and aggregates. Other authors use the terms aggregates and agglomerates as synonyms. In the case of coalescence, the total surface area of the particles is reduced. (For a more detailed description of the instabilities of emulsions and suspensions, the reader is referred to Secs. IV and V, respectively)

Although disperse systems are thermodynamically unstable, certain systems can remain "stable" over a prolonged period of time. Thermodynamically driven changes to a lower energy state may be reversible or irreversible. The type and kinetics of these changes determine the usefulness of a product, as indicated, for example, by its shelf life. A disperse system remains stable as long as the repulsive forces are sufficiently strong to outweigh the van der Waals and/or other attractive forces. These repulsive forces are generally acquired through one or both of the following mechanisms: (a) electrostatic repulsion, which arises from the presence of an ionic charge on the surface of the dispersed particles; (b) steric repulsion, which arises from the presence of uncharged molecules on the surface of the particles.

Stabilization by Electrostatic Repulsion

The electrostatic stabilization of disperse systems can be described by the DLVO theory (see above). The electrostatic repulsive forces can prevent the dispersed particles having surface charges of the same sign from approaching each other, thus stabilizing the dispersion against interparticle attraction or coagulation. As long as the height of the potential energy barrier, V_{max}, exceeds the kinetic energy, the approaching particles do not come sufficiently close to each other to establish important attractive forces (van der Waals forces), but move away from each other due to steric or electrostatic effects. A net positive potential energy of about $20\,kT$ is usually sufficient to keep the particles apart, rendering the dispersion permanently stable. At $T = 298\,\text{K}$, this equals 1×10^{-12} erg or 1×10^{-5} J. Correspondingly, moderate physical stability is achieved when the zeta potential is between ± 30 and $\pm 60\,\text{mV}$, and good to excellent physical stability is achieved when the zeta potential is between ± 60 and $\pm 100\,\text{mV}$. However, when the size of the dispersed particles exceeds $1\,\mu\text{m}$ and the density is greater than $1.0\,\text{g/cm}^3$, the effect of the zeta potential becomes less significant.

Ionic solids with surface layers containing the ionic species in near proper stoichiometric balance and most

Copyright © 2002 Marcel Dekker, Inc.

water-insoluble organic compounds have relatively low surface charge densities. They may adsorb ionic, equally charged surfactants from solutions, which increases their surface charge densities and the magnitude of their ζ potentials, resulting in an enhanced electrostatic stabilizing effect. In suspensions, the addition of ions that are adsorbed onto the surface of the dispersed particles generally creates strong repulsion forces between suspended particles and stabilizes the system. In contrast, the addition of water-miscible solvents (e.g., alcohols, glycerin, or propylene glycol) to aqueous dispersions can lower the dielectric constant of the medium. This results in a reduction of the thickness of the electric double layer and the magnitude of the potential energy barrier. Thus, the addition of these solvents tends to cause instability or makes the system more sensitive to coagulation.

Stabilization by Steric Repulsion

When a strongly hydrated hydrophilic polymer is adsorbed onto the surface of a hydrophobic solid particle surrounded by an aqueous medium, the affinity of the polymer for water can exceed the attractive forces between the suspended particles. For example, a hydrophilic polymer such as gelatin can increase the strength of the protective hydration layer formed around the dispersed particles. Water-soluble polymers whose adsorption stabilizes dispersions and protects them against coagulation are also called protective colloids. The hydrated, protective shell becomes an integral part of the particles' surface. Adsorption of nonionic polymers (e.g., gums or water-soluble cellulosic derivatives) or surfactants (e.g., polysorbate 80) of sufficient chain length can create steric hindrance between adjacent particles. A polymer can be adsorbed in the forms of loops, tails, and trains. Parts of the long-chain molecules are adsorbed onto the surface of the particles (e.g., trains), whereas other parts extend into the bulk liquid (e.g., tails) [57]. Some molecules are adsorbed onto the solid surfaces in the form of loops projecting into the aqueous phase rather than lying flat against the solid substrate. The stabilizing efficiency of the adsorbed macromolecules depends on the presence of sufficiently long tails or loops. The adsorbed species usually form multi- rather than monomolecular films at the interface.

Two mechanisms of steric stabilization can be distinguished: entropic stabilization and osmotic repulsion. Entropic stabilization arises when two opposing adsorbed polymer layers of adjacent particles overlap, resulting in compression and interpenetration of their chain segments. The restriction of the movements of the polymer chains in the overlap region leads to a negative entropy change. Thus, the reverse process of disentanglement of the two adsorbed layers occurs and is energetically more favorable. This mechanism predominates when the concentration of polymer in the adsorbed layer is low. The osmotic repulsion mechanism is operative in cases where more polymer segments are involved and become crowded in the overlap regions. The increase in polymer concentration within these areas causes a local increase in osmotic pressure and results in an influx of water. This influx pushes the particles apart.

Steric stabilization is particularly useful and widely applied during emulsification processes. One major advantage compared to electrostatic stabilization is the relative insensitivity towards added electrolytes. An auxiliary effect promoting dispersion stability is the significant viscosity increase of the dispersion medium.

D. Rheological Properties

Rheology is the study of flow and deformation of materials under the influence of external forces. It involves the viscosity characteristics of powders, liquids, and semisolids. Rheological studies are also important in the industrial manufacture and applications of plastic materials, lubricating materials, coatings, inks, adhesives, and food products. Flow properties of pharmaceutical disperse systems can be of particular importance, especially for topical products. Such systems often exhibit rather complex rheological properties, and pharmaceutical scientists have conducted fundamental investigations in this area [58–64].

Newton's Law and Newtonian Flow

To characterize rheological behavior of materials, some basic terms need to be defined. Consider a liquid material that is subjected to a shearing force as illustrated in Fig. 2. The liquid is assumed to consist of a series of parallel layers with the surface area A, the bottom layer being fixed. When a force is applied on the top layer, the top plane moves at a constant velocity, whereas each lower layer moves with a velocity directly proportional to its distance from the stationary bottom layer. The velocity gradient (dv/dr, the difference in velocity, dv, between the top and bottom planes of liquid separated by the distance, dr) is also called the rate of shear, G:

$$G = \frac{dv}{dr} \tag{3}$$

Copyright © 2002 Marcel Dekker, Inc.

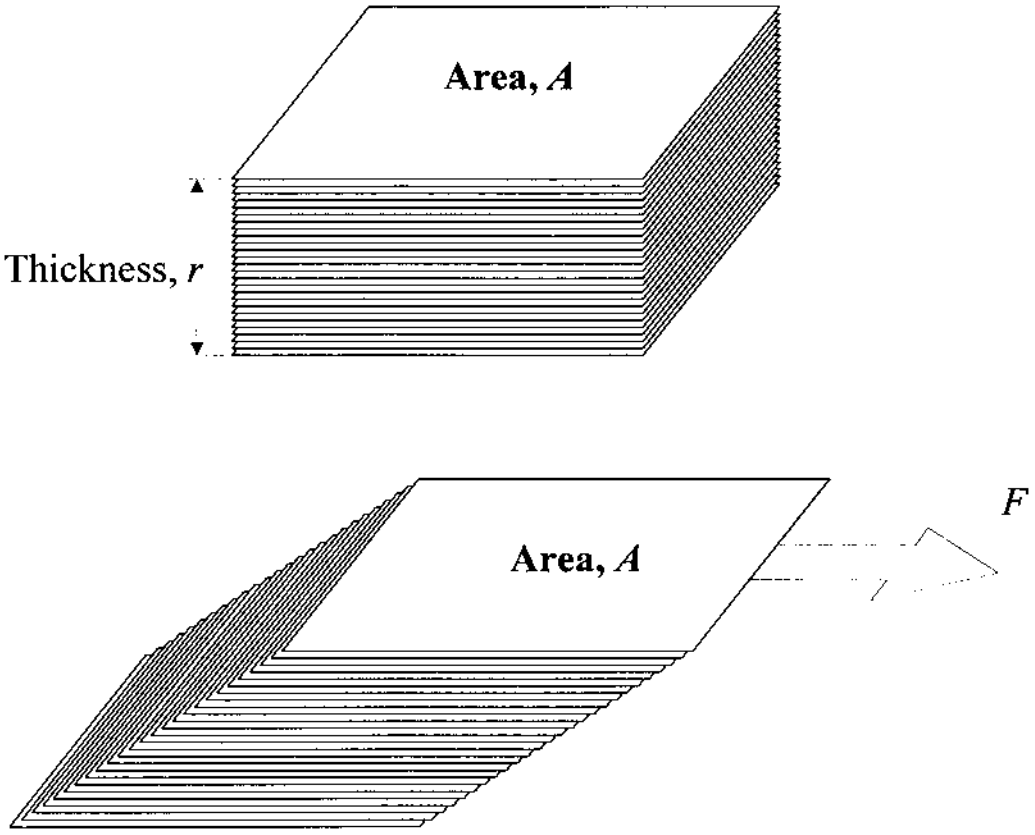


Fig. 2 Model of a liquid material that is subjected to a shearing force. The liquid is assumed to consist of a series of parallel layers, the bottom layer being fixed.

The force per unit area, F'/A, required to cause flow is called the shearing stress (F):

$$F = \frac{F'}{A} \tag{4}$$

The rate of shear, G, is directly proportional to the shearing stress, F, and the relationship can be expressed as:

$$\frac{F'}{A} = \eta \frac{dv}{dr} \quad \text{or} \quad \eta = \frac{F}{G} \tag{5}$$

where η is the coefficient of viscosity, or viscosity.

A plot of F vs. G yields a rheogram or a flow curve. Flow curves are usually plotted on a log-log scale to include the many decades of shear rate and the measured shear stress or viscosity. The higher the viscosity of a liquid, the greater the shearing stress required to produce a certain rate of shear. Dividing the shear stress by the shear rate at each point results in a viscosity curve (or a viscosity profile), which describes the relationship between the viscosity and shear rate. The unit of viscosity is poise, which is the shearing force required to produce a velocity of 1 cm/s between two parallel planes of liquid, each 1 cm^2 in area and separated by a distance of 1 cm. The most often used unit is centipoise, or cp (equivalent to 0.01 poise). Another term, fluidity, ϕ, is defined as the reciprocal of viscosity:

$$\phi = \frac{1}{\eta} \tag{6}$$

In addition, the U.S. Pharmacopoeia includes the explanation of kinematic viscosity [units in stoke (s) and centistoke (cs)], defined as the absolute viscosity divided by the density of the liquid at a definite temperature, ρ:

$$\text{Kinematic viscosity} = \frac{\eta}{\rho} \tag{7}$$

For simple Newtonian fluids (e.g., pure water), a plot of rate of shear against shearing stress gives a

Copyright © 2002 Marcel Dekker, Inc.

straight line, thus the slope (η) is constant (Fig. 3a, curve A) [65]. In other words, the viscosity is constant, depending on neither the shear rate nor time. However, pharmaceutical disperse systems rarely exhibit simple Newtonian flow, and their viscosity is not constant but changes as a function of shear rate and/or time. The rheological properties of such systems cannot be defined simply in terms of one value. These non-Newtonian phenomena are either time-independent or time-dependent. In the first case, the systems can be classified as pseudoplastic, plastic, or dilatant, in the second case as thixotropic or rheopective.

Pseudoplastic Flow

Many fluids show a decrease in viscosity with increasing shear rate. This behavior is referred to as "shear thinning," which means that the resistance of the material to flow decreases and the energy required to sustain flow at high shear rates is reduced. These materials are called "pseudoplastic" (Fig. 3a and b, curves B). At rest the material forms a network structure, which may be an agglomerate of many molecules attracted to each other or an entangled network of polymer chains. Under shear this structure is broken down, resulting in a shear

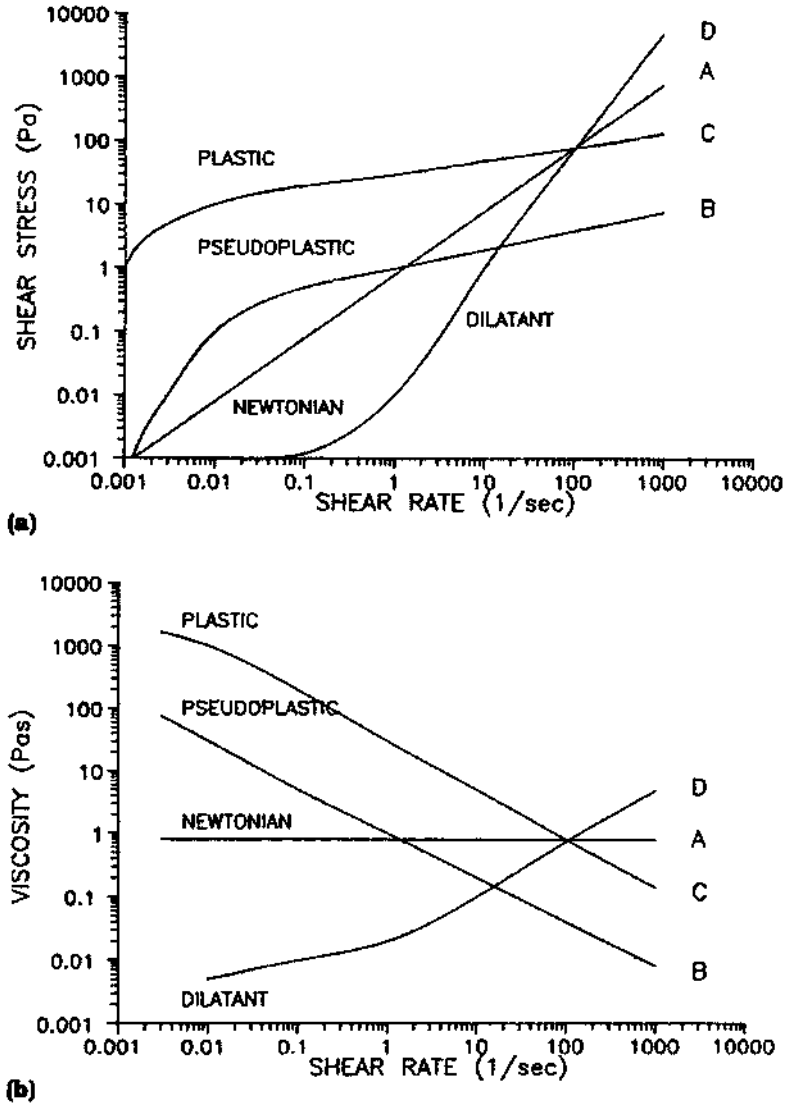


Fig. 3 Newtonian and non-Newtonian behaviours as a function of shear rate: (a) flow profile; (b) viscosity profile. (From Ref. 65.)

Copyright © 2002 Marcel Dekker, Inc.

thinning behavior. With linear polymers, it is speculated that the axes of such molecules become more oriented in the direction of flow as shearing stress is increased. Pseudoplastic flow can be found in emulsions, suspensions, as well as various pharmaceutical thickening agents, e.g., polymer solutions.

Plastic Flow

Plastic fluids are Newtonian or pseudoplastic liquids that exhibit a "yield value" (Fig. 3a and b, curves C). At rest they behave like a solid due to their interparticle association. The external force has to overcome these attractive forces between the particles and disrupt the structure. Beyond this point, the material changes its behavior from that of a solid to that of a liquid. The viscosity can then either be a constant (ideal Bingham liquid) or a function of the shear rate. In the latter case, the viscosity can initially decrease and then become a constant (real Bingham liquid) or continuously decrease, as in the case of a pseudoplastic liquid (Casson liquid). Plastic flow is often observed in flocculated suspensions.

Dilatant Flow

In some cases, the viscosity increases with increasing shear rate. The system appears to become more structured and more viscous with increasing shear stress. This flow behavior is called "dilatant" (Fig. 3a and b, curves D). Examples include pastes containing plasticizers, ionic polymers, suspensions with high solids content, highly pigmented systems, and quicksand.

Thixotropy and Rheopexy

Thixotropy is a phenomenon that occurs frequently in dispersed systems. It is defined as a reversible, time-dependent decrease in viscosity at a constant shear rate. Generally, a dispersion that shows an isothermal gel-sol-gel transformation is a thixotropic material. The mechanism of thixotropy is the breakdown and reforming of the gel structure.

An example for a rheogram of a thixotropic material is given in Fig. 4 [13]. Here, the rate of shear, G, is plotted as a function of the shearing stress, F. Due to the time dependence of the viscosity, a hysteresis loop results: the up curve (early times) is not identical with the down curve (late times). The up curve shows the rate of shear when increasing the shearing stress; the down curve shows the rate of shear when subsequently decreasing the shearing stress. As the shear rate decreases, the structure rebuilds and viscosity is restored to its original value. If the recovery is fast, as in many water-based systems, the down curve is superimposed on the up curve. If the rearrangement is slow, as in

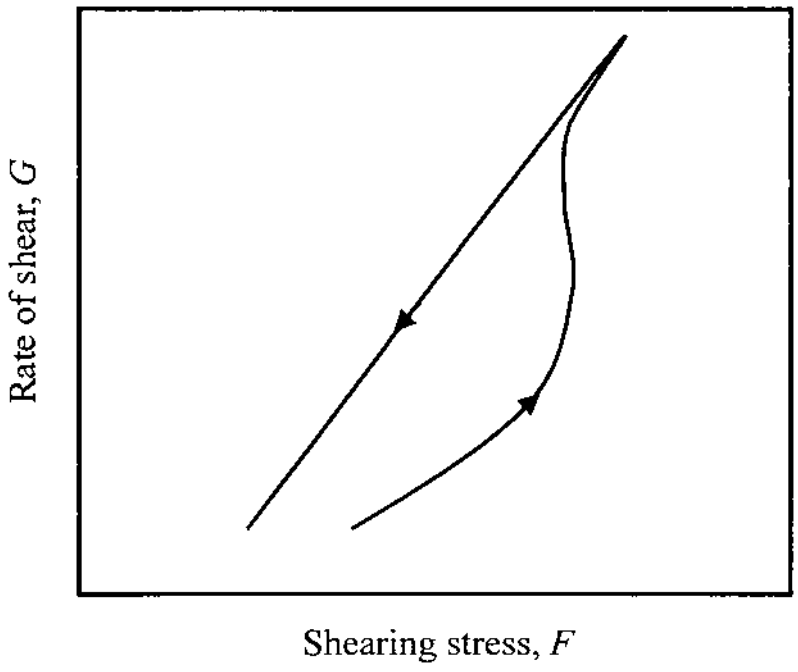


Fig. 4 Flow curves of a system exhibiting thixotropy.

Copyright © 2002 Marcel Dekker, Inc.

many organic solvent–based systems, it may take longer for the fluid to regain its initial properties after shearing, and therefore the down curve will be above the up curve. The hysteresis area between the up curve and the down curve defines the energy required to break down the network structure of the material; its dimensions are energy/volume.

If, in contrast, the viscosity *in*creases with time while the material is being sheared and recovers its original viscosity when allowed to rest, the material is called "rheopective." In this case the down curve is positioned below the up curve.

III. FORMULATION ADDITIVES

A. Surfactants

According to the colloid scientist Winsor, surfactants are defined as compounds "which possess in the same molecule distinct regions of hydrophilic and lipophilic character." For example, in the oleate ion there is an alkyl chain that is basically hydrophobic (lipophilic tail) and a COO^- headgroup that is hydrophilic (lipophobic). Being amphiphilic in nature, surfactants have the ability to modify the interface between various phases [66]. Their effects on the interface are the result of their ability to orient themselves in accordance with the polarities of the two opposing phases. The polar part can be expected to be oriented towards the more polar (hydrophilic, aqueous) phase, whereas the nonpolar tails should direct towards the nonpolar (lipophilic, oil) phase.

Surfactants are useful in formulating a wide variety of disperse systems. They are required not only during manufacture but also for maintaining an acceptable physical stability of these thermodynamically unstable systems. Besides the stabilizing efficiency, the criteria influencing the selection of surfactants for pharmaceutical or cosmetic products include safety, odor, color, and purity.

When the variation of any colligative property of a surfactant in aqueous solution is examined, two types of behavior are apparent. At low concentrations, properties approximate those to be expected from ideal behavior. However, at a concentration value that is characteristic for a given surfactant system (critical micelle concentration, CMC), an abrupt deviation from such behavior is observed. At concentrations above the CMC, molecular aggregates called micelles are formed. By increasing the concentration of the surfactant, depending on the chemical and physical nature of the molecule, structural changes to a more orderly state than micellar structure in the molecular assembly of amphiphiles occur. These are, generally, hexagonal, cubic, and lamellar liquid crystalline structures; their occurrence relates mostly to the concentration of the surfactant in the system. Liquid crystals have been known for more than a century and have been formulated as parts of many creams, lotions, and cosmetic products.

Surfactants can be classified according to their functionality and their natures of hydrophilic and hydrophobic portions. A typical functional scheme was developed in the CTFA (Cosmetic Ingredient Handbook) by creating six functional categories for surfactants: cleansing agents, emulsifying agents, foam boosters, hydrotropes, solubilizing agent, and suspending agents. Another means for classification is based on the nature of the hydrophobic portions of surfactants. Such a classification would create groups based on the presence of hydrophobes derived from paraffinic, olefinic, aromatic, cycloaliphatic, or heterocyclic hydrophobes. This type of classification is of particular interest in comparing surfactants' characteristics with regard to their physiological effects related to the origin of the lipophilic constituents.

The most useful and widely accepted classification is based on the nature of the hydrophilic head groups. This classification system has universal acceptance and has been found to be practical throughout the surfactant industry. This approach creates four large groups of surfactant structures: anionics, cationics, amphoterics (zwitterionics), and nonionics. Anionic surfactants carry negative charges and include a large group of surfactants used in pharmaceutical products, such as soaps, sulfates, and sulfonates. Soaps can be prepared in situ by a reaction between a fatty acid and an alkali. Cationic surfactants, carrying positively charged head groups, are of special pharmaceutical interest since they often possess antimicrobial activity. Examples include quaternary ammonium compounds and cetyltrimethylammonium bromide. Cationic surfactants should not be used in the same formulation with anionic surfactants, since they will interact. Amphoteric surfactants possess both positive and negative charges in their structures and include substances like lecithin and *N*-dodecyl-*N*,*N*-dimethylglycine. Nonionic surfactants are not dissociated and, thus, are not charged. Unlike the anionic and cationic types, they are not susceptible to pH changes and to the presence of electrolytes. *N*-Alkyl polyoxyethylene surfactants, of general formula $CH_3(CH_2)_n(OCH_2CH_2)_mOH$, where n is often between 10 and 18 and m between 6 and 60, are good examples of nonionic surfactants.

Copyright © 2002 Marcel Dekker, Inc.

Particularly useful is the physical classification of surfactants based on the hydrophile-lipophile balance (HLB) system [67,68] established by Griffin [69,70]. More than 50 years ago he introduced an empirical scale of HLB values for a variety of nonionic surfactants. Griffin's original concept defined HLB as the percentage (by weight) of the hydrophile divided by 5 to yield more manageable values:

$$\text{HLB} = \frac{\text{wt\% hydrophile}}{5} \tag{8}$$

Griffin studied primarily ethylene oxide (EO) adducts and routinely substituted "% EO" for "% hydrophile." Since that time, the HLB system has become very popular, especially to characterize emulsifying agents. (The reader is referred to sec. V.B. for a more detailed discussion of the use of the HLB system for the identification of adequate emulsifiers and combinations thereof.)

Finely divided solid particles that are wetted to some degree by both oil and water can also act as emulsifying agents. This results from the fact that they can form a particulate film around dispersed droplets, preventing coalescence. Powders that are wetted preferentially by water form O/W emulsions, whereas those more easily wetted by oil form W/O emulsions. The compounds most frequently used in pharmacy are colloidal clays, such as bentonite (aluminum silicate) and veegum (magnesium aluminum silicate). These compounds tend to be adsorbed at the interface and also increase the viscosity of the aqueous phase. They are frequently used in conjunction with a surfactant for external purposes, such as lotions or creams.

The common concentration of a surfactant used in a formulation varies from 0.05 to 0.5% and depends on the surfactant type and the solids content of the dispersion. In practice, very often combinations of surfactants rather than single agents are used to prepare and stabilize disperse systems. The combination of a more hydrophilic surfactant with a more hydrophobic surfactant leads to the formation of a complex film at the interface. A good example for such a surfactant pair is the Tween-Span system of Atlas-ICI [71].

B. Protective Colloids and Viscosity-Imparting Agents

Protective colloids can be divided into synthetic and natural materials. Table 3 classifies the pharmaceutical gums, thickeners, and other hydrophilic polymers according to their origins [72]. Protective colloids of natural origin, such as gelatin, acacia, and tragacanth, have been used for years as emulsifying agents [73,74]. Gelatin and serum albumin are preferred protective colloids for stabilizing parenteral suspensions because of their biocompatibility. These two polymers, as well as casein, dextrin, and plant gums, can be metabolized in the human body. Dispersing a natural gum within water yields a thick, viscous liquid called a mucilage. Mucilages are used primarily to aid in suspending insoluble substances within liquids because their colloidal character and viscosity can help to prevent sedimentation. However, the drawbacks of mucilages include their sensitivity to pH change, addition of electrolytes, and/or heat, which adversely affect viscosity. Mucilages of vegetable gums are prone to microbial decomposition and show appreciable decrease in viscosity upon storage.

Besides these naturally derived protective colloids, several synthetic, water-soluble substances are widely used at appropriate concentrations as mucilage substitutes, suspending agents, emulsifying agents, and/or viscosity-imparting agents. Synthetic cellulose derivatives include sodium carboxymethylcellulose, hydroxypropylcellulose, hydroxypropyl methyl cellulose, and methyl cellulose. The resulting viscosity depends on the concentration and specification (e.g., molecular weight) of the polymers. Many of these substances are used as protective colloids at low concentrations ($<0.1\%$) and as viscosity builders at relatively high concentrations ($>0.1\%$). Since these agents do not reduce the surface and interfacial tension significantly, it is often advantageous to use them in combination with surfactants. Unlike many natural polymers, cellulose derivatives and most synthetic protective colloids are not biotransformed. When administered orally, they are not absorbed, but excreted intact.

When polymeric substances and hydrophilic colloids are used as suspending agents, it is recommended that appropriate tests be performed to show that the agent does not interfere with the drug and does not modify the resulting therapeutic effect. With regard to the raw material selection, protective colloids may be chosen according to their abilities to aid in stabilizing a dispersion as well as other factors, such as cost, toxicity, and resistance to chemical and/or microbial attack. Although many natural substances are nontoxic and relatively inexpensive, they often show considerable batch-to-batch variations and a relatively high risk for microbial contamination.

C. pH-Controlling Agents

A properly formulated disperse system should exhibit an acceptable physical stability over a wide range of

Copyright © 2002 Marcel Dekker, Inc.

Table 3 Classification of Protective Colloids (gums, thickeners, and polymers) According to Their Origins

Classification		Origin	Products
Natural	Plant	Tree and shrub exudates	Karaya gum
			Tragacanth gum
			Gum acacia
		Seed extracts	Guar gum
			Locust bean gum
			Psyllium seed
			Quince seed
		Seaweed extracts	Carrageenan
			Alginates
			Agar
		Tree extracts	Larch gum
		Fruit extracts	Pectins
		Grains and roots	Starches
	Microbial	Exocellular polysaccharides	Xanthan gum
			Dextran
	Animal	Milk protein	Casein
		Skin and bones	Gelatin
			Keratin
		Insect secretion	Shellac
Modified natural	Plant	Wood pulp and cotton	Cellulose derivatives: methylcellulose, sodium carboxy-methylcellulose
		Seed extracts	Guar derivatives
Synthetic		Petroleum based	Acrylic acid polymers
			Polyacrylamides
			Alkylene/alkylene oxide polymers
Inorganic		Clays	Smectite hydrophilic and organoclays
		Amorphous silicon dioxide	Hydrated silica
			Fumed silica

Source: Ref. 72.

pH values. If a specific pH is required, the system can be maintained at the desired pH value using an adequate buffer. This is especially important for drugs that possess ionizable acidic or basic groups in which the pH of the vehicle often influences the stability and/or solubility. These buffering salts or electrolytes must be used with extreme caution since small changes in electrolyte concentration can alter the surface charge of the dispersed phase and, thus, affect the overall stability of the system. Osmotic active salts (e.g., sodium chloride) and/or stabilizers (e.g., disodium edetate) may be replaced by organic nonelectrolytes, such as dextrose, mannitol, or sorbitol, in order to avoid possible destabilization. Adjustments in osmolarity or tonicity are generally required when preparing ophthalmic or injectable disperse systems.

For suspensions primarily stabilized by a polymeric material, it is important to carefully consider the optimal pH value of the product since certain polymer properties, especially the rheological behavior, can strongly depend on the pH of the system. For example, the viscosity of hydrophilic colloids, such as xanthan gums and colloidal microcrystalline cellulose, is known to be somewhat pH- dependent. Most disperse systems are stable over a pH range of 4–10 but may flocculate under extreme pH conditions. Therefore, each dispersion should be examined for pH stability over an adequate storage period. Any

Copyright © 2002 Marcel Dekker, Inc.

changes in the pH values could be indicative of a potential instability problem.

D. Preservatives

Preservation against microbial growth is an important aspect of disperse systems, not only with respect to the primary microbiological contamination, but also in terms of the physical and chemical integrity of the system [75]. Aqueous liquid products are prone to microbial contamination because water in combination with excipients derived from natural sources (e.g., polypeptides, carbohydrates) serves as an excellent medium for the growth of microorganisms. Besides the pathogenic hazards, microbial contamination can result in discoloration or cracking (separation into two bulk phases) of the products. The evolution of carbon dioxide gas may result in the explosion of containers. Colloidal dispersions and systems prepared by a controlled flocculation may become unstable and deflocculate due to a decrease in the zeta potential in the absence of adequate preservation.

The types of microorganisms found in various products are *Pseudomonas* species, including *Pseudomonas aeruginosa*, *Salmonella*, species, *Staphylococcus aureus*, and *Escherichia coli*. The USP and other pharmacopoeias recommend certain classes of products to be tested for specified microbial contaminants, e.g., natural plant, animal, and some mineral products for the absence of *Salmonella* species, suspensions for the absence of *E. coli*, and topically administered products for the absence of *P. aeruginosa* and *S. aureus*. Emulsions are especially susceptible to contamination by fungi and yeasts. Consumer use may also result in the introduction of microorganisms. For aqueous-based products, it is therefore mandatory to include a preservative in the formulation in order to provide further assurance that the product retains its pharmaceutically acceptable characteristics until it is used by the patient.

Substances that have been used as preservatives for disperse systems include chlorocresol, chlorobutanol, benzoates, phenylmercuric nitrate, parabens, and others [76,77]. The use of cationic antimicrobial agents such as quaternary ammonium compounds (e.g., benzalkonium chloride) is contraindicated in many cases because they may be inactivated by other formulation components and/or they may alter the charge of the dispersed phase. Clay suspensions and gels should be adequately preserved with nonionic antimicrobial preservatives. The use of preservatives is generally limited to products that are not intended for parenteral use. Intravenous injectable preparations should be prepared according to stringent sterile conditions from pyrogen-free raw materials and normally be terminally sterilized by either autoclaving or aseptic filtration.

In emulsions, partitioning of the incorporated preservative can occur between the aqueous and the oil phase. A lipophilic preservative may pass into the oil phase so that a significant portion is removed from the aqueous phase. Since it is the latter in which microorganisms tend to grow, the use of water-soluble preservatives can be more effective, especially for O/W emulsions. For most emulsion systems, the esters of *p*-hydroxybenzoic acid (parabens) appear to be the most satisfactory. Since microorganisms can also reside within the oil phase, it is further recommended that a pair of preservatives having different oil and water solubilities be used in order to ensure appropriate concentrations in both phases.

The major criteria for the selection of an appropriate preservative include efficiency against a wide spectrum of microorganisms, stability (shelf life), toxicity, sensitizing effects, compatibility with other ingredients in the dosage form, and taste and odor. Other specific factors to be considered may include the application site (e.g., external, ophthalmic, parenteral), pH of the dispersion medium, solubility in the vehicle, partitioning into an oil phase (in emulsions), and adsorption onto the solid phase (in suspensions) or onto the packaging materials [78]. Finally, the establishment of microbiological standards for raw materials and strict adherence to the Good Manufacturing Practice (GMP) protocols during industrial-scale production can efficiently reduce the severity of the contamination problem.

E. Antioxidants

Many pharmaceutical products undergo oxidative deterioration upon storage because the therapeutic ingredients or adjuvants oxidize in the presence of atmospheric oxygen. Vitamins, essential oils, and almost all fats and oils can be oxidized readily. The decomposition can be particularly significant in disperse systems, such as emulsions, because of the large area of interfacial contact and because the manufacturing process may introduce air into the product. Many drugs commonly incorporated into emulsions are subjected to autooxidation and subsequent decomposition. Traces of oxidation products are undesirable because they are generally easily noticed by their smell or taste. The term autooxidation is used when the ingredient(s) in the product react(s) with oxygen without drastic external interference. Series of autooxidative

Copyright © 2002 Marcel Dekker, Inc.

reactions involve the initiation step or the formation of a free radical, the propagation step where the free radical is regenerated and reacts with more oxygen, followed by the termination step as the free radicals react with each other resulting in the inactive products.

Antioxidants are classified into three main groups. The first group, comprising the true antioxidants, probably inhibits oxidation by reacting with free radicals, thus blocking the chain reaction. Examples include tocopherols, alkyl gallates, butylated hydroxyanisole (BHA), and butylated hydroxytoluene (BHT). They are commonly used at concentrations ranging from 0.001 to 0.1%. Antioxidants of the second group, comprising the reducing agents, have a lower redox potential than the drug or other substances that they should protect and are therefore more readily oxidized. They may also react with free radicals. Examples are ascorbic and isoascorbic acids and the potassium or sodium salts of sulfurous acid. The latter group, such as sodium metabisulfite, has been reported to produce sensitivity reactions. The third group comprises the antioxidant synergists, including sequestering and chelating agents, which possess little antioxidant effect themselves but enhance the action of a true antioxidant by reacting with heavy metal ions that catalyze oxidation. The synergist class includes citric acid, tartaric acid, disodium edetate, and lecithin.

The selection of an appropriate antioxidant depends on factors such as stability, toxicity, efficiency, odor, taste, compatibility with other ingredients, and distribution phenomena between the two phases. Antioxidants that give protection primarily in the aqueous phase include sodium metabisulfite, ascorbic acid, thioglycerol, and cysteine hydrochloride. Oil-soluble antioxidants include lecithin, propyl gallate, ascorbyl palmitate, and butylated hydroxytoluene. Vitamin E has also been used, but its virtue as a natural antioxidant has been the subject of some controversy.

IV. SUSPENSIONS

A suspension is defined as a disperse system in which the dispersed phase is in the solid state, whereas the dispersion medium is in the liquid state. For example, the internal phase (suspensoid, suspended phase) can be uniformly distributed throughout the suspending medium (suspending vehicle) in which it exhibits a minimum degree of solubility. The internal phase consists of a homogeneous or heterogeneous distribution of solid particles having a specific range of size. Coarse suspensions contain suspended particles, which are larger than about 1 μm in size with the practical upper limit of about 75 μm. The particle diameter of most good pharmaceutical suspensions lies between 1 and 50 μm. When one or more of the ingredients constituting the internal phase are pharmaceutically useful and/or pharmacologically active, the system is called a pharmaceutical suspension.

Suspensions have several advantages as a dosage form. They allow the development of a liquid product containing an appropriate quantity of the active ingredient in a reasonable volume. Resistance to hydrolysis and oxidation is generally good in suspensions when compared with that in respective aqueous solutions. In such a case, a suspension would insure chemical stability while permitting liquid therapy. Suspensions can also be used for taste-masking purposes. A disagreeable taste of certain drugs when given in solution form can be overcome when administered as undissolved particles in an oral suspension. A number of chemical derivatives of many poor-tasting drugs, such as chloramphenicol, have been specifically developed for their insolubility in a desired vehicle for the purpose of preparing a palatable liquid dosage form.

On the other hand, there are a number of disadvantages of suspensions that should be noted. The content uniformity and dosage accuracy of suspensions are unlikely to compare favorably with that obtainable by the use of tablets, capsules, or solutions. Sedimentation and compaction of sediment, so-called caking, frequently cause problems that are not always easy to solve. In addition, the product is liquid and relatively bulky compared to solid dosage forms. These properties are disadvantageous for both pharmacists and patients. However, the above-mentioned advantages can, under certain circumstances, outweigh the disadvantages of this type of dosage forms.

A. Sedimentation and Stokes' Law

The control of sedimentation is required to ensure a sufficient and uniform dosage. Sedimentation behavior of a disperse system depends largely on the motion of the particles which may be thermally or gravitationally induced. If a suspended particle is sufficiently small in size, the thermal forces will dominate the gravitational forces and the particle will follow a random motion owing to molecular bombardment, called "Brownian motion." The distance moved or displacement, D_i, is given by:

$$D_i^2 = \frac{RTt}{3N\pi\eta r} \tag{9}$$

Copyright © 2002 Marcel Dekker, Inc.

where R is the universal gas constant, T is the absolute temperature, t is the time, N is the Avogadro's number, η is the viscosity, and r is the particle radius. Note that the displacement, D_i, decreases with increasing radius of the particle, r. For any given system, one can define a "no sedimentation diameter" (NSD), below which value the Brownian motion will be sufficient to keep particles from sedimentation. The value of NSD obviously depends on the density and viscosity of the respective system.

Increasing the radius of the suspended particles, Brownian motion becomes less important and sedimentation becomes more dominant. These larger particles therefore settle gradually under gravitational forces. The basic equation describing the sedimentation of spherical, monodisperse particles in a suspension is Stokes' law. It states that the velocity of sedimentation, v, can be calculated as follows:

$$v = \frac{2gr^2(\rho_1 - \rho_2)}{9\eta} = \frac{gd^2(\rho_1 - \rho_2)}{18\eta} \tag{10}$$

where r is the particle radius, d is the particle diameter, ρ_1 and ρ_2 are the densities of the particle and dispersion medium, respectively, g is the acceleration caused by gravity, and η is the viscosity of the medium.

The settling velocity is proportional to the second power of the particle radius or particle diameter. It is apparent that agglomerates and flocculates settle more rapidly than individual particles. Since both gravity and buoyancy are operating simultaneously on the particle, either upward movement or downward movement results. Suspensions generally undergo sedimentation upon storage while emulsions may exhibit either upward creaming (O/W type) or downward creaming (W/O type). The determining factor is the difference in the densities of the internal and the external phases.

Stokes' law is rigorously applicable only for the ideal situation in which uniform and perfectly spherical particles in a very dilute suspension settle without turbulence, interparticle collisions, and without chemical/physical attraction or affinity for the dispersion medium [79]. Obviously, the equation does not apply precisely to common pharmaceutical suspensions in which the above-mentioned assumptions are most often not completely fulfilled. However, the basic concept of the equation does provide a valid indication of the many important factors controlling the rate of particle sedimentation and, therefore, a guideline for possible adjustments that can be made to a suspension formulation.

Since the sedimentation rate increases with increasing particle size, it is apparent that particle size reduction is beneficial to the stability of suspensions. However, one should avoid reducing the particle to an extreme degree of fineness since very fine particles have the tendency to form compact cakes upon storage. The rate of sedimentation may also be appreciably reduced by increasing the viscosity of the continuous phase. But this can only be achieved within a practical limit. From Stokes' law, it can also be seen that if the density difference of both phases is eliminated, sedimentation can be entirely prevented. However, it is seldom if ever possible to increase vehicle density above about 1.3. In addition, a product having a too high viscosity is not desirable because of its poor pourability and difficulty of redispersion. A wide variety of suspending agents are available. Some excipients, such as cellulose derivatives, have a pronounced effect on the viscosity, but hardly any on density. Others, such as sorbitol, can modify both density and viscosity.

Generally, the physical stability of a pharmaceutical suspension can be appropriately adjusted by an alteration in the dispersed phase rather than by significant modifications in the dispersion medium. These adjustments are mainly concerned with particle size, uniformity of particle size, and separation of the particles so that they are not likely to become larger or to form a solid cake upon standing.

B. Flocculation and Deflocculation Phenomena

The zeta potential is a measurable indication of the apparent particle charge in the dispersion medium. When its value is relatively high, the repulsive forces usually exceed the attractive forces. Accordingly, the particles are individually dispersed and are said to be "deflocculated." Thus, each particle settles separately, and the rate of sedimentation is relatively small. The settling particles have plenty of time to pack tightly by falling over one another to form an impacted bed. The sedimentation volume of such a system is low, and the sediment is difficult to redisperse. The supernatant remains cloudy even when settling is apparent.

Controlled flocculation is the intentional formation of loose agglomerates of particles held together by comparatively weak bonding forces. This can be achieved by the addition of a preferentially adsorbed ion whose charge is opposite in sign to that of the zeta potential–determining ions. Thus, the apparent active charge of the particles is progressively lowered. At certain concentration of the added ion, the forces of repulsion are sufficiently small that the forces of attraction

Copyright © 2002 Marcel Dekker, Inc.

start to predominate. Under these conditions the particles may approach each other closely and form loose agglomerates, termed "flocs," "flocculates," or "floccules" ("flocculated" system). When compared to the deflocculated particles, the flocs settle rapidly and form a higher sedimentation volume. The loose structure permits the flocculates to break up easily and distribute uniformly with only a small amount of agitation. Controlled flocculated systems usually develop a clear supernatant solution above the loose sediment. Thus, they might look less uniform upon standing, even though they provide easy redispersion and better dose uniformity than many other types of suspensions.

Flocculating agents can be simple electrolytes that are capable of reducing the zeta potential of suspended charged particles. Examples include small concentrations (0.01–1%) of monovalent ions (e.g., sodium chloride, potassium chloride) and di- or trivalent ions (e.g., calcium salts, alums, sulfates, citrates or phosphates) [80–83]. These salts are often used jointly in the formulations as pH buffers and flocculating agents. Controlled flocculation of suspensions can also be achieved by the addition of polymeric colloids or alteration of the pH of the preparation.

A caking diagram, described by Martin [81], shows the flocculation of a bismuth subnitrate suspension by means of the flocculating agent monobasic potassium phosphate. The addition of the negatively charged phosphate ions to the suspended drug particles causes the positive zeta potential to decrease. Upon further phosphate addition, the zeta potential falls to zero and then becomes negative. Microscopic examination reveals that maximum flocculation occurs and persists until the zeta potential becomes sufficiently negative for deflocculation. The onset of flocculation coincides with the maximum sedimentation volume determined.

The controlled flocculation method may be used in conjunction with the addition of a polymeric material to form a structured vehicle. After the formation of the flocs, an aqueous solution of polymeric material, usually negatively charged, such as carboxymethylcellulose or carbopol, is added. The concentration employed depends on the consistency desired for the suspension, which also relates to the size and density of the dispersed phase. Care must be taken to ensure the absence of any incompatibility between the flocculating agent and the polymer used for the formation of the structured vehicle.

When formulating a deflocculated system, a number of materials may be used as dispersion aids. Deflocculating agents include polymerized organic salts of sulfonic acid of both types—alkyl-aryl and aryl-alkyl—that can alter the surface charge of the particles by physical adsorption. Unlike surfactants, these agents do not lower surface and interfacial tension appreciably; hence, they have little or no tendency to create foam or wet particles. Most deflocculants are not generally considered safe for internal use; the only acceptable dispersant for parenteral products is lecithin.

C. Crystal Growth and Polymorphism

The growth in time of unprotected slightly soluble drug solids and subsequent changes in their particle size distribution in suspension have been reviewed by several researchers [84–89]. The crystal growth in disperse systems may be attributed to one or more of the following mechanisms: Ostwald ripening, temperature changes, and polymorphic transformation.

Ostwald ripening is the growth of large particles at the expense of smaller ones as a result of a difference in the solubility of the particles of varying sizes. The surface free energy of small particles is greater than that of large particles. Therefore, small particles can be appreciably more soluble than larger ones. A concentration gradient results, with higher drug concentrations in the area of small particles and lower drug concentrations in the area of large particles. Thus, according to Fick's law of diffusion, the dissolved particle molecules diffuse from the environment of the smaller particles to the environment of the larger particles. An increase in the concentration in the area of the large particles leads to crystallization and particle growth. Thus, small particles become smaller, whereas large particles become larger upon storage. Small fluctuations in temperature can accelerate this effect. Small particles dissolve to a greater extent when the temperature increases and then recrystallize on the surface of existing larger particles as the temperature drops. The suspension becomes coarser, and the mean particle size spectrum shifts to higher values. This effect can be expressed by the following relationship:

$$\log \frac{S}{S_o} = \frac{k}{2.303 \cdot r} \tag{11}$$

where S is the initial solubility rate of small particles; S_o is the solubility rate of large particles at equilibrium, r is the particle radius in cm, and k is a constant that includes surface tension, temperature, molar volume, and thermodynamic terms ($k = 1.21 \times 10^{-6}$). Thus, it can be seen that the increase in solubility rate of a 0.2 μm particle is 13%, 1% for a 2 μm particle, and negligible for particles of 20 μm and larger.

Copyright © 2002 Marcel Dekker, Inc.

Crystal growth due to temperature fluctuations during storage is of importance especially when the suspensions are subjected to temperature cycling of 20°C or more. These effects depend on the magnitude of temperature change, the time interval, and the effect of temperature on the drug's solubility and subsequent recrystallization process. At each crystal contact point there exists a thin layer of supersaturated solution that facilitates crystal growth. The type of inherent crystal form is determined by the factors governing the rate of crystal growth. The degree of supersaturation is dependent on the rate of cooling, the extent and degree of agitation, and the size and number of nuclei available for the particle nucleation. Other factors include pH, solvent effects, and the impurities present.

As crystal growth generally increases with an increase in particle solubility, the excipients that tend to increase the particle solubility should be kept to a minimum. Many pharmaceutical gums can absorb onto the surfaces of drug crystals and, thus, can be used to inhibit the crystal growth. In a dilute suspension, crystal growth increases with the degree of agitation because the mass transfer in the bulk fluid is increased (conventional transport). If sedimentation occurs, the local increase in particle concentration decreases the mean free diffusion path of the solute molecules and may thus promote the particle growth.

Polymorphism refers to the different internal crystal structures of a chemically identical compound. Drugs may undergo a change from one metastable polymorphic form to a more stable polymorphic form. Also, the crystal habit might change due to the degree of solvation or hydration. The formation of distinct new crystalline entities during storage is possible. For example, an originally anhydrous drug in a suspension may rapidly or slowly form a hydrate. These various forms may exhibit different solubilities, melting points, and x-ray diffraction patterns. In the preparation of suspensions using precipitation methods, the solvent and the rate of cooling are important factors determining the type of polymorph(s) obtained.

Various drugs are known to exist in different polymorphic forms (e.g., cortisone and prednisolone). The rate of conversion from a metastable into the stable form is an important criteria to be considered with respect to the shelf life of a pharmaceutical product. Polymorphic changes have also been observed during the manufacture of steroid suspensions. When steroid powders are subjected to dry heat sterilization, subsequent rehydration of anhydrous steroid in the presence of an aqueous vehicle results in the formation of large, needle-like crystals. A similar effect may be produced by subject finished suspensions to moist heat sterilization in an autoclave.

Higuchi showed that crystal growth may also arise when the more energetic amorphous or glassy forms of a drug exhibit significantly greater initial solubility in water than their corresponding crystalline forms [84]. In addition, size reduction by crushing and grinding can produce particles whose different surfaces exhibit high or low dissolution rates. This effect can be correlated to differences in the free surface energy introduced during comminution.

To prevent crystal growth and possible changes in particle size distribution, one or more of the following procedures and techniques may be employed [9,89]: (a) selection of particles with a narrow size range; (b) selection of a more stable crystalline form of the drug; (c) avoidance of the use of high-energy milling during particle size reduction; (d) incorporation of a wetting agent (e.g., surfactant) and/or a protective colloid (e.g., cellulose derivatives forming film barriers around the particles); (e) increase of the viscosity of the vehicle to retard particle dissolution and subsequent crystal growth; and (f) avoidance of temperature extremes during storage.

D. Pharmaceutical Suspensions

In the preparation of physically stable pharmaceutical suspensions, a number of formulation components can be incorporated to maintain the solid particles in the dispersed state. These substances can be classified as (a) components of the suspending system, including wetting agents, dispersants or deflocculating agents, flocculating agents, and thickeners, and (b) components of the suspending vehicle (external phase), including pH-control agents and buffers, osmotic agents, coloring/flavoring agents, preservatives, and liquid vehicles. The components of each category are individually selected for their use in the preparation of orally, topically, or parenterally administered suspensions.

Orally administered suspensions containing a wide class of active ingredients (e.g., antibiotics, antacids, radiopaque agents) are of major commercial importance. The solids content of an oral suspension may vary considerably. For example, antibiotic preparations may contain 125–500 mg solid drug per 5 mL or a teaspoonful dose, while a drop concentrate may provide the same amount of drug in only 1–2 mL. Antacid or radiopaque suspensions also contain relatively high amounts of suspended material for oral administration. The suspending vehicle can, for example, be a syrup, sorbitol solution, or gum-thickened water with added

Copyright © 2002 Marcel Dekker, Inc.

artificial sweeteners. Taste and mouthfeel are important considerations when formulating oral suspensions.

Many antibiotic drugs are unstable in the presence of an aqueous vehicle and, therefore, are frequently supplied as dry powder mixtures for reconstitution at the time of dispensing. Generally, this type of product is either a powder mixture or a completely/partially granulated product, which upon dilution and agitation with a specified quantity of vehicle (e.g., water) results in the formation of a suspension suitable for administration [90]. The preparation is typically designated in the USP by a title of the form "For Oral Suspension," whereas the ready-to-use suspension preparations are simply designated as "Oral Suspension." The dry mix products often contain drugs, colorants, flavorants, sweeteners (e.g., sucrose or sodium saccharin), stabilizing agents (e.g., citric acid, sodium citrate), suspending agents (e.g., guar gum, xanthan gum, methylcellulose), and preservatives (e.g., parabens, sodium benzoate).

The formulation of oral sustained-release suspensions has resulted in only limited success due to the difficulty in maintaining the stability of sustained-release particles when present in liquid systems. Formulation techniques, such as coated beads, drug-impregnated wax matrix, microencapsulation, and ion-exchange resins, have been used for this purpose [9l–93]. The combination of an ion-exchange resin complex with particle coating has resulted in a commercial product, the so-called Pennkinetic system. By this technique, ionic drugs are complexed with ion-exchange resins, and the drug-resin complex particles are then treated with an impregnating polymer such as polyethylene glycol 4000 followed by coating with a sustaining polymer, such as ethyl cellulose [94,95]. In liquid suspension (the dispersion medium being free of ions that could replace drug ions in the resin complex), the drug remains absorbed to the resin. However, upon swallowing ions from the gastrointestinal liquid can penetrate the particles and replace the drug ions, which subsequently diffuse out of the system (at a controlled, slow rate). Drug release from these systems depends on the type of drug-resin complex, on the ionic environment (e.g., pH and electrolyte concentration within the GI tract), as well as on properties of the resin. Most ion-exchange resins currently employed in sustained-release products contain sulfonic acid groups that exchange cationic drugs possessing amine functionality. An example is hydrocodone polistirex [Tussionex® Pennkinetic Extended-Release Suspension (Pennwalt)].

Topical suspensions are intended to be applied externally. Shake lotion and calamine lotion are good examples of historical products in this class. Because safety and toxicity are dealt with in terms of dermatological acceptability, many useful new suspending agents have been introduced in topical formulations. The protective action and cosmetic properties of topical lotions usually require the use of high concentrations of disperse phase.

In contrast, parenteral suspensions have relatively low solids contents, usually between 0.5 and 5%, with the exception of insoluble forms of penicillin in which concentrations of the antibiotic may exceed 30%. These sterile preparations are designed for intramuscular, intradermal, intralesional, intraarticular, or subcutaneous injection. Syringeability is an important factor to be taken into consideration with injectable dosage forms. The viscosity of a parenteral suspension should be sufficiently low to facilitate injection. Common suspending vehicles include preserved isotonic saline solution or a parenterally acceptable vegetable oil. Ophthalmic and optic suspensions that are instilled into the eye/ear must also be prepared in a sterile manner. The vehicles are essentially isotonic and aqueous in composition. The reader should refer to Chapter 12 for further discussion on parenteral products.

E. Methods of Evaluating Suspensions

Suspensions are generally evaluated with respect to their particle size, electrokinetic properties (zeta potential), and rheological characteristics. A detailed discussion on the methods/techniques and relevant instrumentation is given in Sec. VII. A number of evaluating methods done specifically with suspension dosage forms, such as sedimentation volume, redispersibility, and specific gravity measurements, will be treated in this section.

The sedimentation volume of a pharmaceutical suspension can be evaluated using simple, inexpensive, graduated, cylindrical graduates (100–1000 mL). It is defined as the ratio of the equilibrium volume of sediment, V_u, to the total volume of the suspension, V_o.

$$F = \frac{V_u}{V_o} \tag{12}$$

The value of F ranges between 0 and l and increases as the volume of suspension that appears occupied by the sediment increases. For example, if 100 mL of a well-shaken test formulation is placed in a graduate cylinder and the final height of the sediment is at the 20 mL line, then F is 0.2. It is normally found that the greater the value of F, the more stable the product.

Copyright © 2002 Marcel Dekker, Inc.

When $F = 1$, no sediment is apparent and caking is absent, and the suspension is considered esthetically pleasing. This method of evaluation is quite useful in determining the physical stability of suspensions. It can be used to determine the settling rates of flocculated and deflocculated suspensions by making periodic measurement of sedimentation height. Tingstad [96] indicated that a flocculated suspension that settles to a level that is 90% of the initial suspension height ($F = 0.9$) and no further is probably satisfactory.

The degree of flocculation, β, is defined as the ratio of the sedimentation volume of the flocculated suspension, F, to the sedimentation volume of the suspension when deflocculated, F_∞. It is expressed as:

$$\beta = \frac{F}{F_\infty} \tag{13}$$

The degree of flocculation, therefore, is an expression of the increased sediment volume resulting from flocculation. For example, if β has a value of 5.0, this means that the volume of sediment in the flocculated system is five times that in the deflocculated state. As the value of β approaches unity, the degree of flocculation decreases.

Great care must be exercised when using graduated cylinders because decreases in the diameter of small containers can produce a "wall effect," which often affects the settling rate or ultimate sedimentation volume of flocculated suspensions. Such small containers have a tendency to hold up the suspensions due to adhesive forces acting between the container's inner surface and the suspended particles.

If a pharmaceutical suspension produces a sediment upon storage, it is essential that it should be readily dispersible so that uniformity of dose is assured. The amount of shaking required to achieve this end should be minimal. Various redispersibility tests have been described. For example, the test suspension is placed in a 100 mL graduated cylinder, which, after storage and sedimentation, is rotated through 360° at 20 rpm. The endpoint is taken when the inside of the base of the graduated cylinder is clear of sediment. The ultimate test of redispersibility is the uniformity of suspended drug dosage delivered from a product, from the first to the last volumetric dose out of the bottle, under one or more standard shaking conditions.

V. EMULSIONS

An emulsion is a two-phase system consisting of at least two immiscible liquids (or two liquids that are saturated with each other), one of which is dispersed as globules (internal or dispersed phase) within the other liquid phase (external or continuous phase), generally stabilized by an emulsifying agent [81,97]. Emulsions have been widely used in many areas of application: in industries [98–100], agricultures [101,102], food technologies [103–109], pharmaceutics [110–114], and cosmetics [115–119]. Very frequently emulsions are used in cosmetic products as topical vehicles for dermal application since they have high patient/consumer acceptance. Pharmaceutical emulsions are currently used internally for the administration of nutrients, drugs, and diagnostic agents. Emulsions discussed in this chapter include macroemulsions, multiple emulsions, microemulsions, liposomes, and a special emulsion type: in situ forming microparticle (ISM) systems.

A. Types of Emulsions

Macroemulsions

Based on the nature of the dispersed phase and the dispersion medium, two types of macroemulsions can be distinguished. If the continuous phase is an aqueous solution and the dispersed phase an oil, the system is called an oil-in-water emulsion. Such an O/W emulsion is generally formed if the aqueous phase constitutes more than 45% of the total weight and a hydrophilic emulsifier is used. Conversely, when the aqueous phase is dispersed within the oil, the system is called a water-in-oil emulsion. W/O emulsions are generally formed when the aqueous phase constitutes less than 45% of the total weight and a lipophilic emulsifier is used. Generally, O/W emulsions are more popular than W/O emulsions in the pharmaceutical field, especially when they are designed for oral administration. In the cosmetic industries, lotions or creams are of either the O/W or the W/O type, depending on their applications. O/W emulsions are most useful as water-washable drug bases and for general cosmetic purposes, while W/O emulsions are employed more widely for the treatment of dry skin and emollient applications. It is important for a pharmacist to know the type of emulsion, since it can significantly affect its properties and performance. There are several methods of determining the emulsion type, as summarized in Table 4 [120].

Multiple Emulsions

Multiple emulsions are more complex systems. If a simple emulsion is further dispersed within another continuous phase, these systems are called

Copyright © 2002 Marcel Dekker, Inc.

Table 4 Methods to Determine Type of Macroemulsion (W/O or O/W)

Test	Method/Observation
Phase dilution test	Place two emulsion droplets on a glass slide and add a droplet of one component to each emulsion droplet, stir, and observe under a microscope. This test is based on the principle that an emulsion can only be diluted with the liquid that constitutes the continuous phase.
Dye solubility test	A colored dye soluble in only one component is added to the emulsion. If the color spreads throughout the whole emulsion, the phase in which the dye is soluble is the continuous phase.
Conductivity test	Immerse a pair of electrodes connected to an external electric source in the emulsion. If the external phase is water, a current passes through the emulsion. If the oil is the continuous phase, the emulsion fails to carry the current.

multiple emulsions. Figure 5 presents schematically the preparation of a W/O/W emulsion [122,123]. In step 1 an aqueous phase is added to an oily phase, containing a lipophilic surfactant. Upon mixing, a W/O emulsion is formed. In the second step, this W/O emulsion is then poured into a second aqueous solution, containing a hydrophilic surfactant. Upon mixing, the multiple W/O/W emulsion is formed.

Water-in-oil-in-water (W/O/W) and oil-in-water-in-oil systems (O/W/O) are the two basic types of multiple emulsions. Recently multiple emulsions have been developed with the purpose of prolonging the release of incorporated drugs. This technology has gained particular interest for the microencapsulation of peptides/proteins and hydrophilic drugs [124]. To increase the encapsulation efficiency of hydrophilic compounds W/O/W, O/W/O, W/O/O, and W/O/O/O emulsions have been used [125]. Cosmetically interesting is the application of multiple emulsions as supermoisturizing product bases, combining the benefits of O/W and W/O emulsions.

Microemulsion

In 1959, J. H. Schulman introduced the term "microemulsion" for transparent-solutions of a model four-component system [126]. Basically, microemulsions consist of water, an oily component, surfactant, and co-surfactant. A three phase diagram illustrating the area of existence of microemulsions is presented in Fig. 6 [24]. The phase equilibria, structures, applications, and chemical reactions of microemulsion have been reviewed by Sjöblom et al. [127]. In contrast to macroemulsions, microemulsions are optically transparent, isotropic, and thermodynamically stable [128, 129]. Microemulsions have been subject of various investigations because their unique properties allow a wide range of potential practical applications [130–132]. However, there is still substantial controversy concerning the exact nature of these systems and the appropriateness of the terminology. Terms as transparent emulsions, micellar solutions, solubilized systems, and swollen micelles have all been applied to the same or similar systems. Nonetheless, emulsions and microemulsions may be differentiated on the basis of particle size: microemulsions contain particles in the nanometer size range (typically 10–100 nm), whereas conventional emulsions (or macroemulsions, coarse emulsions) contain particles in the micrometer range.

Liposomes

Liposomes are vesicular lipid systems with diameter ranging between 50 nm and a few μm. They are composed of membrane-like lipid layers surrounding aqueous compartments. The lipid layers consist of phospholipids, making liposomes biocompatible and biodegradable. The phospholipids have a hydrophilic head and a lipophilic tail. Different preparation methods for liposomes have been described by Vemuri et al. [133], leading to different types of vesicular structures (Fig. 7) [134]. Nowadays, liposomes are widely used for drug delivery and drug targeting [135–137]. For a detailed description of this type of drug delivery system, the reader is referred to Chapters 15 and 16.

In Situ Forming Microparticle Systems

ISM systems were first introduced by Bodmeier and coworkers [138–140]. They are included in this chapter as special disperse systems because these

Copyright © 2002 Marcel Dekker, Inc.

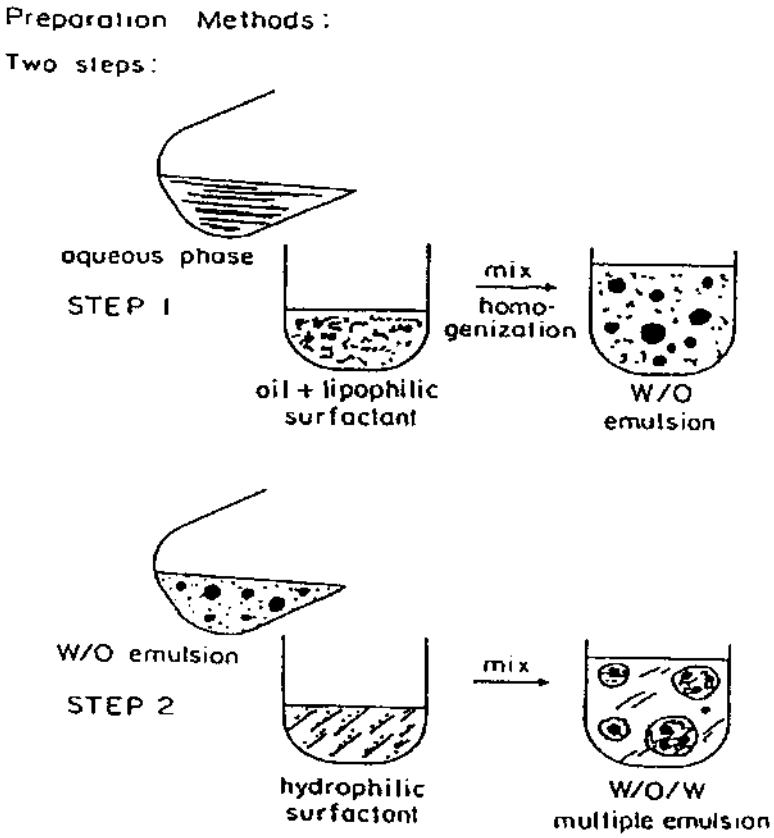


Fig. 5 Schematic of a two-step procedure to produce a W/O/W emulsion. (From Ref. 28.)

formulations are based on O/O or O/W emulsions. Being liquids, the ISM systems can easily be injected intramuscularly or subcutaneously and subsequently form microparticles within the human body upon contact with physiological fluids. The ISM system is composed of a drug and biodegradable polymer, which are co-dissolved in a water-miscible, biocompatible solvent. This solution is emulsified into an external, either oily or aqueous phase containing an emulsion stabilizer to form O/O or O/W emulsions. Upon contact with aqueous physiological fluids, the polymer solvent dissipates, leading to polymer solidification and hardening of the emulsion droplets. Thus, the microparticle formation occurs *in situ*. The resulting microparticles are regular in shape and reproducible. Furthermore, drug release can be prolonged and controlled during desired periods of time. The major advantages of this new technology over implants and the classical solvent evaporation method used to prepare microparticles are the ease of preparation and the avoidance of surgical insertion or removal of empty remnants.

B. Formulation and Preparation Techniques for Emulsions

The preparation techniques for emulsions can be divided into laboratory-scale productions and large-scale productions. Details on the latter are given in the manufacturing and equipment part of this chapter (see Sec. IX). Each method requires that energy be introduced into the system, either by trituration, homogenization, agitation or heat. The production of satisfactory stable emulsions requires adequate formulations and preparation techniques. It has been suggested that the formulator first determine the physical and chemical characteristics of the drug, which include the structure formula, melting point, solubilities in different media, stability, dose, and specific chemical incompatibilities. This information is essential for the selection of the appropriate type of emulsion. Then the required emulsifying agent(s) and its (their) concentration(s) should be identified. The choice of materials to be used largely depends on the purpose for which the emulsion is designed.

Copyright © 2002 Marcel Dekker, Inc.

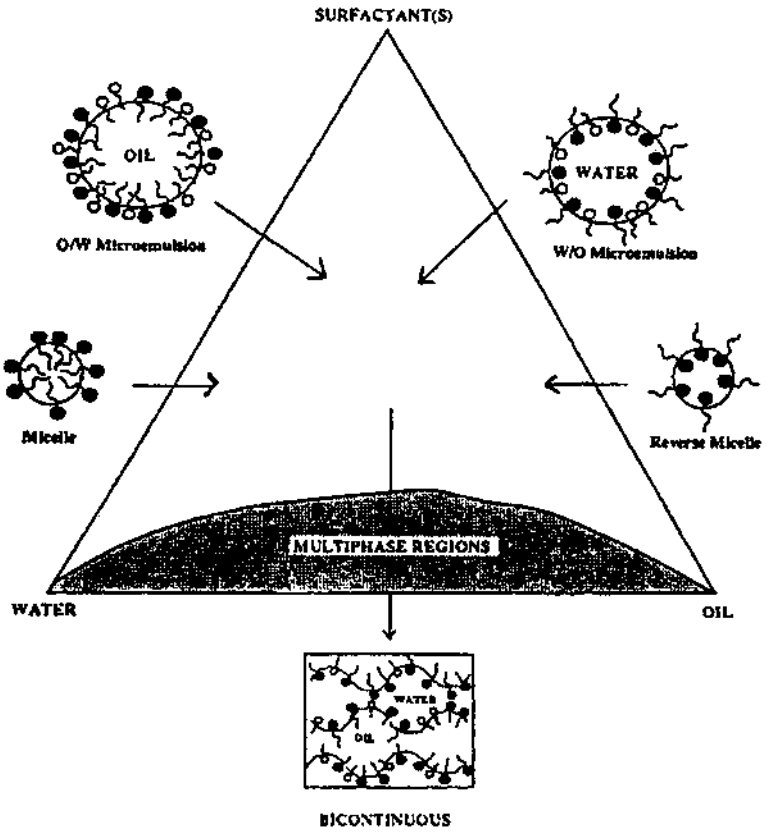


Fig. 6 Three-phase diagram illustrating the area of existence of microemulsions. (From Ref. 24.)

Phase Volume Ratio

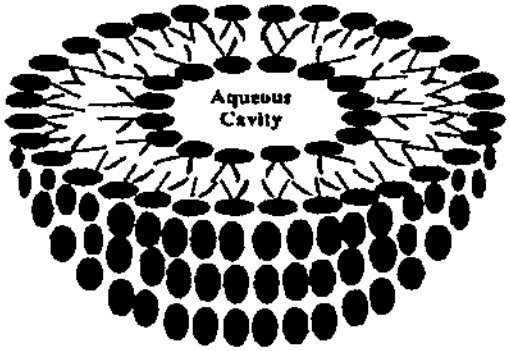
The ratio of volume of disperse phase to volume of the dispersion medium greatly influences the characteristics of an emulsion. The choice of the phase volume ratio depends on a number of factors, including the requested consistency. It is generally difficult to formulate emulsions containing less than about 25% of disperse phase due to their susceptibility to severe creaming or sedimentation problems. However, a combination of proper emulsifying agents and suitable processing technology makes it possible to prepare emulsions with only 10% disperse phase without considerable stability problems. Conversely, products containing a high percentage of disperse phase (more than about 70%) are likely to exhibit phase inversion (the disperse phase becomes the dispersion medium).

Emulsifying Agents

Emulsions are thermodynamically unstable systems. However, using appropriate emulsifying agents (emulsifiers) to decrease the interfacial tension, the stability of these systems can be significantly increased. A satisfactory emulsifier should have a reasonable balance between its hydrophilic and hydrophobic groups, produce a stable emulsion (both initially and during storage), be stable itself, be chemically inert, be nontoxic and cause no irritation upon application, be odorless, tasteless, and colorless, and be inexpensive. Emulsifying agents can be classified into the following three groups:

1. Surfactants that are adsorbed at oil/water interfaces to form monomolecular films and reduce interfacial tension.
2. Natural macromolecular materials, which form multimolecular films around the disperse droplets of O/W emulsions. They are frequently called auxiliary emulsifying agents and have the desirable effect of increasing the viscosity of the dispersion medium. However, they often suffer from the disadvantages of being subject to hydrolysis and sensitive to variations in pH.

Copyright © 2002 Marcel Dekker, Inc.



Liposome model

DENOMINATION	Large unilamellar vesicles LUV	Multi- or Oligolamellar vesicles MLV/OLV	Small unilamellar vesicles SUV
DIAMETER (nm)	80-1000	100-4000	20-80

Fig. 7 Different types of vesicular structures of liposomes. (From Ref. 134.)

3. Very finely disperse solids, which are adsorbed at the liquid/liquid interfaces, forming films of particles around the disperse globules. Certain powders can very effectively stabilize against coalescence. The solid's particle size must be very small compared with the emulsion droplet size and must exhibit an appropriate angle of contact at the three-phase (oil/water/solid) boundary [141].

The selection of a suitable emulsifying agent and its appropriate concentration are matters of experience and of trial and error. It is not necessary to use emulsifier amounts above the required quantities to produce complete interfacial films, unless an increase in the viscosity of the dispersion medium is intended. Reducing the interfacial tension makes emulsification easy but does not by itself prevent coalescence of the particles and resultant phase separation. Frequently, combinations of two or more emulsifying agents are used [2] to (a) adequately reduce the interfacial tension, (b) produce a sufficiently rigid interfacial film, and (c) achieve the most suitable viscosity of the external phase. Care must be taken to ensure the compatibility between the different emulsifiers. For example, charged emulsifying agents of opposite sign are likely to interact and coagulate when combined.

As described in Sec. III, surfactants can be classified using the hydrophile-lipophile balance (HLB) system [142], established by Griffin (see Sec. III.A) The HLB system provides a scale of surfactant hydrophilicity (HLB values range from 0 to 20, 20 corresponding to the highest possible hydrophilicity) that simplifies emulsifier selection and blending. In Table 5 the HLB values of some pharmaceutically relevant surfactants are listed. Surfactants with a low HLB value (<6) tend to provide stable W/O emulsions; those with a high HLB value (>8) tend to stabilize O/W emulsions. Table 6 gives an overview of the classical use of surfactants in the preparation of pharmaceutical dosage forms depending on their HLB value.

The HLB values required for the emulsification of commonly used oils and waxes in pharmaceutical applications are given in Table 7 [143]. To obtain a

Copyright © 2002 Marcel Dekker, Inc.

Table 5 HLB Values of Surfactants Commonly Used in Pharmaceutical Products

Surfactant	HLB
Nonionic	
Sorbitan monolaurate (Span 20)	8.6
Sorbitan monopalmitate (Span 40)	6.7
Sorbitan monostearate (Span 60)	4.7
Sorbitan monooleate (Span 80)	4.3
Sorbitan trioleatem (Span 85)	1.8
Polyoxyethylene (20) sorbitan monolaurate (Tween 20)	16.7
Polyoxyethylene (20) sorbitan monopalmitate (Tween 40)	15.6
Polyoxyethylene (20) sorbitan monostearate (Tween 60)	14.9
Polyoxyethylene (20) sorbitan monooleate (Tween 80)	15.0
Polyoxyethylerie-polyoxypropylene block copolymers (Pluronics or Poloxamer)	29.0
Amphoteric	
Lecithin (from egg yolk or soybean)	7–10
Hydrogenated lecithin	7–10
Anionic	
Sodium doecyl sulfate	40.0
Bile salts	20–25

desired HLB value, lipophilic and hydrophilic surfactants can also be mixed. The following equation allows the calculation of the resulting HLB value of a mixture of two surfactants (A and B):

$$\mathrm{HLB(blend)} = f \times \mathrm{HLB(A)} + (1 - f) \times \mathrm{HLB(B)} \tag{14}$$

where f is the fraction of surfactant A in the blend. For example, a mixture of 30% Span 80 (HLB = 4.3) and 70% Tween 80 (HLB = 15) has an overall HLB value of $\mathrm{HLB} = (0.3 \times 4.3) + (0.7 \times 15) = 11.8$.

Table 6 Relationship Between the Use of Surfactants in the Preparation of Pharmaceutical Dosage Forms, Their Dispersibility in Water and Their HLB Values

HLB	Dispersibility in water	Suitable application
1–4	Nil	
3–6	Poor	W/O emulsifier
6–8	Milky dispersion on agitation	Wetting agent
8–10	Stable milky dispersion	Wetting agent; O/W emulsifier
10–13	Translucent to clear dispersion	O/W emulsifier
>13	Clear solution	O/W emulsifier; solubilizing agent

Viscosity-Imparting Agents

The viscosity of an emulsion can be of crucial importance for its stability, especially the viscosity of the external phase. A high viscosity reduces creaming and also lessens the tendency of particles to coalescence and produce phase separation. Examples of the widely used viscosity-imparting agents are alginates, bentonite, carboxymethylcellulose, polyvinyl pyrrolidone, hydroxypropylcellulose, and carbomer.

Emulsification Techniques

Techniques of emulsification of pharmaceutical products have been reviewed by Block [27]. The location of the emulsifier, the method of incorporation of the phases, the rates of addition, the temperature of each phase, and the rate of cooling after mixing of the phases considerably affect the droplet size distribution, viscosity, and stability of the final emulsion. Roughly four emulsification methods can be distinguished:

1. Addition of the internal phase to the external phase, while subjecting the system to shear or fracture.
2. Phase inversion technique: the external phase is added to the internal phase. For example, if an O/W emulsion is to be prepared, the aqueous phase is added to the oil phase. First a W/O emulsion is formed. At the inversion point, the

Copyright © 2002 Marcel Dekker, Inc.

Table 7 Required HLB Values for the Emulsification of Oils and Waxes Commonly Used in Pharmaceutical Applications

Compound	HLB	Compound	HLB
O/W emulsions			
Isostearic acid	15–16	Ethyl benzoate	13
Linoleic acid	16	Isopropyl myristate	12
Oleic acid	17	Isopropyl palmitate	12
Ricinoleic acid	16	Kerosene	12
Cetyl alcohol	16	Lanolin, anhydrous	12
Stearyl alcohol	15–16	Mineral oil, aromatic	12
Tridecyl alcohol	14	Mineral oil, paraffinic	10
Beeswax	9	Mink oil	9
Carbon tetrachloride	16	Paraffin wax	10
Carnaubawax	15	Patrolatum	7–8
Caster oil	14	Pine oil	16
Coco butter	6	Rapeseed oil	7
Corn oil	8	Safflower oil	7
Cottonseed oil	6	Soyabean oil	6
W/O emulsion			
Mineral oil	6	Stearyl alcohol	7

Source: Adapted from Ref. 143.

addition of more water results in the inversion of the emulsion system and the formation of an O/W emulsion. The phase inversion technique allows the formation of small droplets with minimal mechanical action and attendant heat. A classical example is the so-called *dry gum* technique, which is a phase inversion technique when hydrophilic colloids are part of the formulation. First the dry hydrocolloid (e.g., acacia, tragacanth, methyl cellulose) is dispersed within the oil phase. Then water is added until the phase inversion occurs to form a O/W emulsion.

3. Mixing both phases after warming each. This method is frequently used in the preparation of ointments and creams.
4. Alternate addition of the two phases to the emulsifying agent. In this method, the water and oil are added alternatively, in small portions, to the emulsifier. This technique is especially suitable for the preparation of food emulsions.

C. Emulsion Stability

Types of Instability

Four major phenomena are associated with the physical instability of emulsions: flocculation, creaming, coalescence, and breaking (Fig. 8) [144]. *Flocculation* is best defined as the association of particles within an emulsion to form large aggregates, which can easily be redispersed upon shaking (see also sec. IV.B). Flocculation is generally regarded as a precursor to the irreversible process of coalescence. It differs from coalescence primarily in that the interfacial film and individual droplets remain intact. The reversibility of flocculation depends mainly on the strength of the interaction between the droplets and on the phase volume ratio. The relationship of droplet deformation, surfactant transfer, and interfacial rheology for emulsion stability has been discussed by Ivanov et al. [145]. *Creaming* occurs when the disperse droplets or floccules separate from the disperse medium under the influence of gravitational force. Generally, a creamed emulsion can be restored to its original state by gentle agitation. According to Stokes' law, the direction in which the droplet moves depends on the relative values of the densities of the two phases. Furthermore, larger droplets cream more rapidly than smaller droplets. Reducing the droplet sizes and thickening the continuous phase can minimize the rate of creaming. *Coalescence* is a much more serious type of instability. It occurs when the mechanical or electrical barrier is insufficient to prevent the formation of progressively larger droplets (Fig. 9) [144], which can finally lead to *breaking* (complete phase separation). Coalescence can be avoided by the formation of thick interfacial films consisting of macromolecules or particulate solids.

Copyright © 2002 Marcel Dekker, Inc.

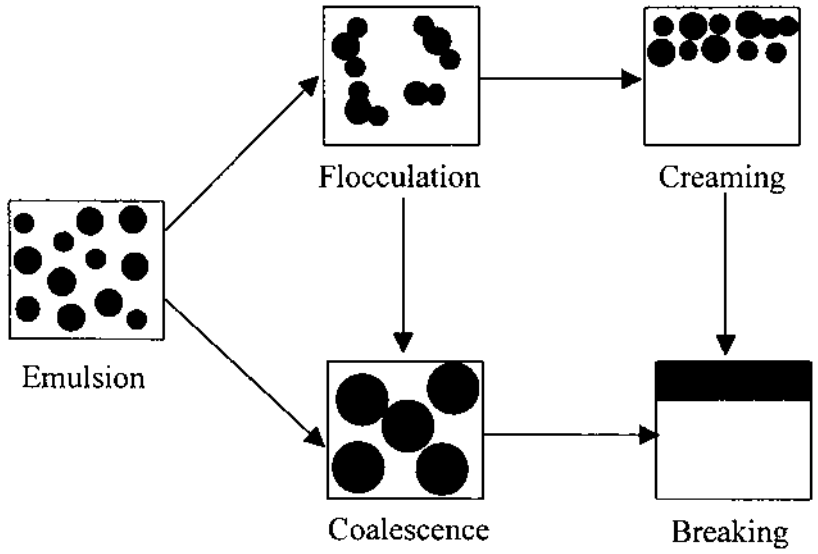


Fig. 8 Schematic illustration of different types of instability of emulsions.

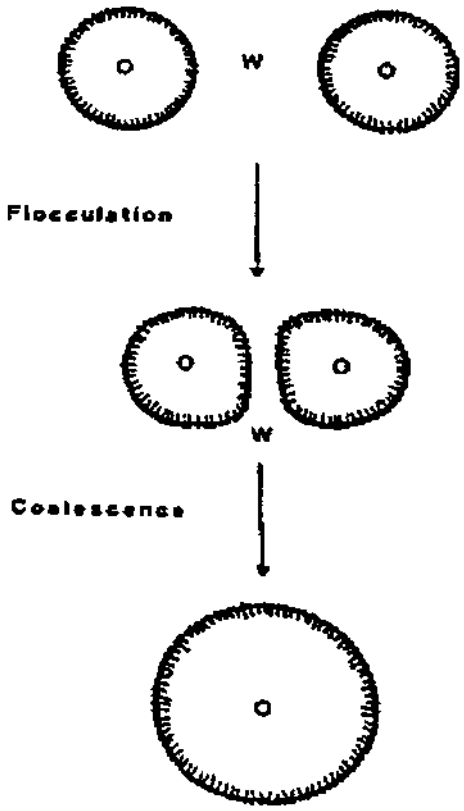


Fig. 9 Schematic presentation of flocculation and coalescence of emulsion droplets. (From Ref. 144.)

Stress Tests

The primary objective when studying the stability of emulsions is to predict the shelf life of these systems under normal storage conditions. The problem of stability assessment studies under normal storage conditions is that they last long periods of time. To shorten the time, many types of stress tests and stress conditions are used to provide a basis for the prediction of the stability of an emulsion. These methods can provide valuable information, but one should always be aware of the risk that changes occurring under stress conditions must not necessarily occur under normal market conditions. Steps should be taken so that the accelerated conditions do not introduce new and unanticipated mechanisms of instability, especially those bearing little relationship to what happens long-term. Stress conditions normally employed for evaluating the stability of emulsions include thermal stress and gravitational stress.

Thermal Stress. The Arrhenius equation states that a 10°C increase in the temperature doubles the rate of most chemical reactions. However, this approach is generally only useful to predict a product's shelf life if the instability of the emulsion is due to a chemical degradation process. Furthermore, this degradation must be identical in mechanism but different in rate at the investigated temperatures. Thus, the instability of

Copyright © 2002 Marcel Dekker, Inc.

emulsion systems seldom obeys the Arrhenius equation. Emulsion systems are often more complex; their instability might include phenomena such as (a) temperature-dependent solubilities and phase distributions of emulsifiers, (b) degradation reactions occurring only at higher temperatures, (c) temperature-induced phase changes, resulting in composition changes and altered rheological behavior, and (d) structural deformations and reformations, which may markedly vary with time and temperature [144]. It is considered reasonable to use the time for destabilization at 40°C multiplied by 4 to give an estimate shelf life at room temperature [134]. In addition, low temperature may also cause the precipitation of the emulsifiers and thus accelerate emulsion instability [142].

Gravitational Stress. Gravitational stress such as centrifugation may allow phase separation to occur quickly. Although this technique has been considered in a casual manner in the evaluation of emulsion stability, prediction of emulsion shelf life from this method was not much investigated until the use of ultracentrifuges [141]. Becher [120] indicates that centrifugation at 3750 rpm in a 10 cm radius centrifuge for a period of 5 hours is equivalent to the effect of gravity for about one year. It has been reported that the logarithm of the root mean cube diameter of particles after being centrifuged at 20,000 rpm produced a linear relationship to the centrifuging time [142]. In general, centrifugation is a useful tool to predict the emulsion shelf life, but the technique may not be suitable for very viscous or semisolid products.

Evaluation of Emulsion Stability

The most useful parameters commonly measured to assess the effect of stress conditions on emulsions include phase separation, rheological property determination, electrical property measurements, and particle size analysis.

Phase Separation. An approximate estimation of phase separation may be obtained visually. In general, creaming, flocculation, and coalescence have occurred before phase separation is visible, thus sometimes making quantitative evaluations more difficult. Accelerating the separation by centrifugation followed by appropriate analysis of the specimens may be useful to quantitatively determine the phase separation. Details on mechanisms of creaming and phase separation as well as some advances in the monitoring techniques of emulsion stability have been reviewed by Robins [146].

Rheological Property Determination. The rheology of an emulsion is often an important factor in determining its stability. Any variation in droplet size distribution, degree of flocculation, or phase separation frequently results in viscosity changes. Since most emulsions are non-Newtonian, the cone-plate type device should be used to determine their viscosity rather than the capillary viscometer.

Electrical Property Measurements. The surface charge and zeta potential (measure of the apparent, active charge) of emulsified droplets is often an important indicator for the stability of the system because electrostatic repulsion can significantly contribute to the avoidance of flocculation and coalescence (see also Sec. II.C). Recently, a new method to determine the creaming state of an emulsion has been established [147, 148]. The conductivity of the emulsion at different heights of the storage cell and at different times is measured, and the values are converted into volume fraction of the disperse phase. The creaming profile of O/W emulsions was described based on a kinetic model to extrapolate for the long-term stability prediction [149, 150].

Particle Size Measurement. The best way to evaluate an emulsion's stability is probably to measure its particle size distribution. A number of methods are available for droplet size determination (see Sec. VIII.A). Optical microscopy, although a time-consuming technique, is a direct way of measuring droplets larger than 1 μm. Nowadays, laser light-scattering, diffraction, and transmission methods are becoming popular for routine determination of particle size [151, 152].

VI. COLLOIDS AND COLLOIDAL DISPERSIONS

A colloid is defined as a system consisting of discrete particles in the size range of 1 nm to 1 μm, distributed within a continuous phase [153]. On the basis of the interaction of particles, molecules, or ions of the disperse phase with molecules of the dispersion medium-, colloidal systems can be classified as being lyophilic or lyophobic. In lyophilic systems, the disperse phase molecules are dissolved within the continuous phase and in the colloidal size range or spontaneously form aggregates in the colloidal size range (association systems). In lyophobic systems, the disperse phase is very poorly soluble or insoluble in the continuous phase. During the last several decades, the use of colloids in

Copyright © 2002 Marcel Dekker, Inc.

the pharmaceutical sciences has gained considerable attention. These systems have been investigated primarily for site-specific drug delivery, for controlled drug delivery, and also for the enhancement of the dissolution rate/bioavailability of poorly water-soluble drugs. Applications of colloidal materials in the preparation of pharmaceutical dosage forms are presented in Table 8 [12]. In this section two types of colloidal systems, namely drug-free aqueous colloidal polymeric dispersions (latices and pseudolatices) and drug-loaded nanoparticles, wi1l be discussed.

A. Latices and Pseudolatices

An aqueous colloidal polymeric dispersion by definition is a two-phase system comprised of a disperse phase and a dispersion medium. The disperse phase consists of spherical polymer particles, usually with an average diameter of 200–300 nm. According to their method of preparation, aqueous colloidal polymer dispersions can be divided into two categories: (true) latices and pseudolatices. True latices are prepared by controlled polymerization of emulsified monomer droplets in aqueous solutions, whereas pseudolatices are prepared starting from already polymerized macromolecules using different emulsification techniques.

To prepare (*true*) latices, the monomers are emulsified in water with stirring and addition of emulsifiers, which help stabilize the monomer droplets. The molecular weight of the polymer molecules in the resultant latex can be controlled by the concentration and decomposition rate of added polymerization initiators. The residual monomer content can be reduced by optimizing the polymerization conditions and can also later be eliminated by steam distillation.

For the preparation of *pseudolatices*, the following three methods are commonly used [41]:

1. Direct emulsification: A solution of the polymer within a volatile, water-immiscible organic solvent (or mixture of solvents), or a polymer melt is emulsified within a surfactant-containing aqueous phase. If used, the organic solvent is then removed by steam distillation to obtain the pseudolatices.
2. Inverse emulsification: A solution of the polymer within a volatile, water-immiscible organic solvent (or mixture of solvents) or a polymer melt is compounded with a long-chain fatty acid (e.g., oleic acid) using conventional rubber-mixing equipment and mixed slowly with a dilute aqueous phase to give a W/O emulsion, which then inverts to an O/W emulsion upon further addition of the aqueous phase. If used, the organic solvent is then removed by steam distillation to obtain the pseudolatices [154, 155].
3. Self-emulsification: The polymer molecules are modified chemically by the introduction of basic (e.g., amino) or acidic (e.g., carboxyl) groups in such concentration and location that the polymer undergoes self-emulsification without surfactant after dispersion in an acidic or basic solution.

Latices and pseudolatices are widely used for the coating of solid pharmaceutical dosage forms such as tablets, capsules, and pellets. The major advantage compared to conventional coating techniques is the avoidance of organic solvents.

B. Nanoparticles

As indicated by the name, nanoparticles are colloidal particles in the size range of 1 to 1000 nm. The term nanoparticles encompasses both nanocapsules and nanospheres. Nanocapsules have a core-shell structure (reservoir system), while nanospheres represent matrix systems. Analogously to the above-described latices and pseudolatices (which are spherical, drug-free, aqueous, polymeric, nanoparticulate dispersions with a narrow size distribution), the techniques used to prepare polymeric nanoparticles can be classified in two groups. Nanoparticles are prepared either from preformed polymers or through various polymerization reactions of lipophilic or hydrophilic monomers. The choice of a particular preparation method and a suitable polymer depends on the physicochemical properties of the drug, the desired release characteristics, the therapeutic goal, the route of administration, the biodegradability/biocompatibility of the carrier material and regulatory considerations. From a technological point of view, the successful selection of a preparation method is determined by the ability to achieve high drug loadings, high encapsulation efficiencies, high product yields, and the potential for an easy scale-up.

Nanoparticles Based on Water-Insoluble Polymers

To prepare nanoparticles based on water-insoluble polymers, a solution of the polymer in an organic solvent can be emulsified within an aqueous phase. This O/W emulsion is then homogenized under high shear with an appropriate homogenization equipment

Copyright © 2002 Marcel Dekker, Inc.

Table 8 Examples of Colloidal Materials Used in the Preparation of Pharmaceutical Dosage Forms

Colloidal materials	Uses
Acacia	Emulsifier, microcapsule wall material, suspending agent
Ailgnic acid	Microcapsule wall material, viscosity-imparting agent, disintegration, binder
Bentonite	Stabilizer, gelling agent, suspending agent, viscosity-imparting agent
Calcium carbonate	Diluent
Carbomer	Emulsifier, gelling agent, suspending agent
Carboxymethyl cellulose	Emulsifier, gelling agent, suspending agent, disintegrant, binder
Cellulose acetate phthalate	Enteric-coating agent
Cellulose	Binder, disintegrant, filler
Dextrin	Stabilizer, viscosity-imparting agent, adhesive, binder, plasticizer, sugar coating agent
Ethylcellulose	Viscosity-imparting agent, microcapsule wall material, binder
Gelatin	Microcapsule wall material, suspending agent, binder, coating agent
Guar gum	Viscosity-imparting agent, stabilizer, binder
Hydroxypropylcellulose	Viscosity-imparting agent, microcapsule wall material, granulating agent
Methylcellulose	Emulsifier, viscosity-imparting agent, gelling agent, suspending agent, binder, disintegrant, sugar coating agent
Poloxamer 188	Emulsifier, stabilizing agent, gelling agent base material for suppositories, coating agent
Polyethylene glycols	Gelling agent, base material for suppositories, coating agent, lubricant, plasticizer
Polymethacrylate	Film coating material, granulating binder
Silicon dioxide	Lubricant for capsules and tablets, viscosity-imparting agent, drying agent for powder
Starch	Binder, disintegrant
Tragacanth	Emulsifier, viscosity-imparting agent, suspending agent

Source: Adapted from Ref. 12.

(e.g., microfluidizer, sonication) prior to the precipitation of the polymer to further reduce the particle size of the internal organic phase into the colloidal size range [156–159]. The high pressure emulsification–solvent evaporation method is limited to water-insoluble drugs. Unless specific binding exists, water-soluble drugs are difficult to encapsulate because of partitioning into the external aqueous phase during emulsification. The particle size of the solidified nanoparticles is determined by the size of the emulsified polymer–drug–solvent droplets and hence depends on the homogenization equipment. Nanoparticles have been prepared with conventional laboratory homogenizers [12], by ultrasonication [157–159], and by microfluidization [159]. With the microfluidizer, the particle size decreases with increasing operating pressure and increasing number of cycles [159]. The particle size of the nanoparticles is also affected by the type of surfactant and surfactant concentration, the viscosity of the organic polymer solution, and the phase ratio of the internal to the aqueous phase [159, 160].

A similar technique, the so-called spontaneous emulsification solvent diffusion method, is derived from the solvent injection method to prepare liposomes [161]. Kawashima et al. [162] used a mixed-solvent system of methylene chloride and acetone to prepare PLGA nanoparticles. The addition of the water-miscible solvent acetone results in nanoparticles in the submicrometer range; this is not possible with only the water-immiscible organic solvent. The addition of acetone decreases the interfacial tension between the organic and the aqueous phase and, in addition, results in the perturbation of the droplet interface because of the rapid diffusion of acetone into the aqueous phase.

Nanocapsules of biodegradable polymers, such as PLA and PLA copolymers or poly (ε-caprolactone), have been prepared by an ''interfacial polymer deposition mechanism'' [163–166]. An additional component, a water-immiscible oil, is added to the drug-polymer-solvent mixture. A solution of the polymer, the drug, and a water-immiscible oil in a water-miscible solvent such as acetone is added to an external

Copyright © 2002 Marcel Dekker, Inc.

aqueous phase. This technique results in nanocapsules with a core-shell structure. The drug must have a high solubility in the oil-solvent mixture in order to obtain nanoparticles with high drug loadings.

An interesting modification of the solvent evaporation method is the "salting-out procedure" [167–172]. The nanoparticles are prepared by adding an aqueous phase saturated with an electrolyte or nonelectrolyte to a solution of the polymer and drug in a water-miscible organic solvent under agitation until an O/W emulsion forms. Saturating the aqueous phase reduces the miscibility of acetone and water by a salting-out process and allows the formation of an O/W emulsion from the otherwise miscible phases. The major advantages of the salting-out method are the avoidance of chlorinated solvents and surfactants commonly used with the conventional solvent evaporation method.

Nanoparticles Based on Hydrophilic Polymers

Nanoparticles based on hydrophilic polymers (e.g., albumin, chitosan, gelatin, or carbohydrates) can be prepared by W/O emulsification techniques. This process has been developed for the preparation of albumin microspheres [173–176]. However, the use of high-shear homogenization equipment or u1trasonication has allowed the formation of emulsions in the nanometer size range. Basically, an aqueous polymer solution is emulsified into an external, water-immiscible phase, such as an oil or organic solvent, followed by homogenization. Upon water removal, the polymer droplets solidify. Insoluble nanoparticles can be obtained by further hardening/insolubilizing the polymer through chemical cross-linking with aldehydes or other cross-linking agents or through denaturation at elevated temperatures. In order to obtain high encapsulation efficiencies, the drug must be insoluble in the external phase. Gelatin nanoparticles have been prepared by emulsifying a concentrated gelatin solution (30%) into hydrogenated castor oil containing proper emulsifying agents above the gelation temperature of the gelatin solution [177]. The W/O emulsion is then cooled in order to gel the aqueous gelatin droplets. Chitosan is one of the few naturally derived polysaccharides carrying basic functional groups. Since it is only soluble in acidic media, chitosan micro-nanoparticles are prepared by emulsifying aqueous solutions of chitosan in acetic acid into an external oil phase [178]. The cross-linking agent can be added to the aqueous chitosan solution prior to the emulsification. The polysaccharides chitosan and sodium alginate form gels with counterions such as tripolyphosphate or calcium chloride. Nanoparticles can be prepared by emulsifying an aqueous polysaccharide solution into the oil phase followed by emulsification of an aqueous solution of the counterion solution.

The drawbacks of the W/O emulsification method include the use of large amounts of oils as the external phase, which must be removed by washing with organic solvents, heat stability problems of drugs, possible interactions of the cross-linking agent with the drug, and, as with all nanoparticles prepared by emulsification techniques, a fairly broad particle size distribution.

Gelatin and albumin nanoparticles have been prepared through desolvation of the dissolved macromolecules by either salts (e.g., sodium sulfate or ammonium sulfate) or ethanol [179–182]. This is, in principle, similar to a simple coacervation method. The particles can then be insolubilized through cross-linking with an optimum amount of aldehydes. These phase separation methods avoid the use of oils as the external phase.

Alginate nanoparticles have been prepared by the addition of an optimized amount of calcium chloride to a sodium alginate solution [183]. This so-called "microgel" can be crosslinked with the oppositely charged polymer, poly-*l*-lysine, to form nanospheres. This technique uses the well-known gelification phenomenon of sodium alginate with calcium ions, but at much lower concentrations than normally used for gel formation. Again, since only one phase is used, high drug loadings are difficult to obtain unless specific binding of the drug to the anionic polymeric carrier occurs.

VII. PARENTERAL DISPERSE SYSTEMS

During the past 30 years, there have been significant developments of parenteral disperse formulations. The use of parenteral emulsions can overcome the problems of low aqueous solubility and water hydrolysis of many drugs [184, 185]. Such formulations can avoid the use of conventional co-solvent systems and the undesirable effects caused by precipitation of drugs at the injection site. Recent developments of parenteral disperse formulations have the potential to provide sustained release and targeting of drugs [186–189].

A. Parenteral Emulsions

Like all parenteral products, parenteral emulsions are required to be sterile, isotonic, nonpyrogenic, nontoxic,

Copyright © 2002 Marcel Dekker, Inc.

and stable. Unique to parenteral emulsions are the requirements for droplet size and surface charge [190], since the parameters have a direct effect on both toxicity and stability.

Particle Size

The British Pharmacopoeia (1980) states that the diameter of emulsion globules for intravenous administration may not exceed 5 μm. Emulsion globules larger than 4–6 μm are known to increase the incidence of emboli and may cause changes in blood pressure [32, 190]. However, lipid particles larger than 7.5 μm have been reported to deform and pass through the pulmonary vasculature without difficulty [191]. In general, emulsions that contain globules 200–500 nm in size tend to be most physically stable. Smaller emulsions are utilized by the body more rapidly than larger ones [190].

Surface Charge

A reduction in the electrical charge is known to increase the flocculation and coalescence rates. Sufficient high zeta potential (> -30 mV) ensures a stable emulsion by causing repulsion of adjacent droplets. The selection of suitable surfactants can help to optimize droplet surface charges and thus enhance emulsion stability. Lipid particles with either positive or negative surface charges are more stable and are cleared from the bloodstream more rapidly than those with neutral charge [192, 193].

Ingredients

Oils. Among many investigated oils, soy oil, middle-chain triglycerides (MCT), safflower oil, and cottonseed oils are favoured for use as the oil phase due to their low incidence of toxic reactions. Purity is regarded as an important criterion for parenteral use. Undesirable contaminants such as hydrogenated oil, saturated fatty materials, pigments, or oxidative decomposition products should be minimized.

Emulsifiers. Natural lecithin is one of the most widely used emulsifiers because it is metabolized in the body. However, type I allergic reaction to soybean lecithin emulsified in lipid solutions has been observed [195]. Among the synthetic emulsifying agents, block copolymers of polyoxyethylene-polyoxypropylene (poloxamer) have attracted increasing interest for parenteral emulsions. Other examples of emulsifiers commonly found in parenteral formulations are given in Table 9 [190].

Table 9 Most Commonly Used Emulsifiers in Parenteral Emulsions

Emulsifiers	Range
Egg lecithin	1–3% w/w
Soyabean lecithin	1–3% w/w
Phosphatidylcholine	Approximately 50%
Glycerol/propylene glycol	30–70% w/w individually
Glyceryl fatty acid esters	3% w/w
Pluronic F68, 88, 108	1.5–10%
Poloxamer 401 (Pluronic L121)	5%
Polysorbate 80	4%

Source: Adapted from Ref. 190.

Additives. Additives are used to adjust the tonicity and pH of emulsions and to increase emulsion stability. Glycerol is a commonly used isotonic agent for parenteral emulsions. The pH of the emulsion should be also adjusted close to a physiological pH of 7–8 by adding aqueous solutions of NaOH or HCl. Tocopherols and oleic acid may be added as antioxidants and stabilizer, respectively. Some commercially available fat emulsions and their compositions are listed in Table l0 [144].

Sterilization

Autoclaving can be used for the sterilization of parenteral emulsions, provided the required sterilization conditions do not lead to significant product degradation. Alternatively, parenteral emulsions can be

Table 10 Compositions of Some Commercially Available Fat Emulsions

Trade name (company)	Oil phase	Emulsifier	Other components
Intralipid	Soybean oil	Lecithin	Glycerol
(Kabi-Vitrum)	10–20%	1.2%	2.5%
Lipofundins	Soyabean oil	Soyabean lecithin	Xylitol
(Braun)	10–20%	0.75–1.2%	5%
Lipofundins	Cottonseed oil	Soyabean lecithin	Sorbitol
(Braun)	10%	0.75%	5%
Liposyn	Safflower oil	Lecithin	Glycerol
(Abbot)	10–20%	1.2%	2.5%
Travemulsion	Soybean protein	Lecithin	Glycerol
(Travenol)	10–20%	1.2%	2.5%

Source: Ref. 144.

Copyright © 2002 Marcel Dekker, Inc.

sterilized by filtration if the emulsion globules can pass through the filter. For thermolabile products and/or products in which sterile filtration is almost impossible, terminal gamma irradiation may be useful. However, restrictions in the use of gamma sterilization differ from country to country, and in some cases changes in the product might occur. Most desirable in terms of stability, is aseptic processing in clean rooms, but there is also a high risk of contamination in the final product.

B. Parenteral Suspensions

Parenteral suspensions are heterogeneous systems consisting of a solid phase that is disperse within a liquid phase. The requirements for and limitations involved in preparing parenteral suspensions include microbiological control, ingredient allowance, and mechanical flow properties [196–198]. Excipients used in parenteral suspensions should be nonpyrogenic, nontoxic, and chemically inert. Flocculating and wetting agents are important ingredients to maintain product stability and good mechanical properties. Water for injection is the most widely used solvent in parenteral formulations but is sometimes not applicable to lipophilic compounds. Nonaqueous solvents can be used to promote solubility and stability. Examples of nonaqueous solvents used in parenteral products include benzyl benzoate, glycofurol, ethyl lactate, polyethylene glycols, glycerin, and ethanol [199]. The basic sterilization methods for solid materials used in parenteral suspensions are recrystallization, spray-drying, lyophilization, ethylene oxide gas, dry heat, and gamma irradiation.

Due to particle sizes in the micrometer range, parenteral suspensions are generally limited to either subcutaneous or intramuscular routes of administration. However, ultrafine suspensions can be approached by high-pressure homogenization [200]. The particle size obtained from this technique is in the 100–500 nm range, thus intravenous administration is possible [201]. General information on parenteral formulations is given in Chapter 12.

VIII. CHARACTERIZATION OF DISPERSE SYSTEMS

A. Particle Size Analysis

Particle size is a major characteristic of a disperse system. Particle size measurements provide the evaluation of possible particle aggregation or crystal growth, and particle size distribution data are often employed as fundamental quality control standards for pharmaceutical disperse systems. A broad particle size distribution is known to promote uneven settling in suspensions and coalescence in emulsions.

Various techniques and equipment are available for the measurement of particle size, shape, and volume. These include for microscopy, sieve analysis, sedimentation methods, photon correlation spectroscopy, and the Coulter counter or other electrical sensing devices. The specific surface area of original drug powders can also be assessed using gas adsorption or gas permeability techniques. It should be noted that most particle size measurements are not truly direct. Because the type of equipment used yields different equivalent spherical diameter, which are based on totally different principles, the particle size obtained from one method may or may not be compared with those obtained from other methods.

Microscopy

Microscopic or optical methods range in complexity and cost from a simple laboratory microscope to refined pieces of automated equipment. The microscope provides a direct observation of the sample and allows the evaluation of particle diameters. An ordinary microscope generally measures the particle size in the range of 0.2 μm to about 100 μm. With this method, a diluted or undiluted sample, e.g., an emulsion or suspension, is mounted on a slide or ruled cell and placed on a mechanical stage. The microscope eyepiece is fitted with a micrometer or a graticule by which the linear dimension of the particles may be estimated. The field can be photographed or projected onto a screen to ease the process of particle measurements. Electronic scanners and video recording equipment have been developed to remove the necessity of measuring particles by visual observation.

To obtain a value for the dimensions of an irregular particle, several measurement approaches can be used: Martin's diameter (defined as the length of a line that bisects the particle image), Feret's diameter (or end-to-end measurement, defined as the distance between two tangents on opposite sides of the particle parallel to some fixed direction), and the projected area diameter (defined as the diameter of a circle having the same area as that of the particle observed perpendicular to the surface on which the particle rests). With any technique, a sufficiently large number of particles is required in order to obtain a statistically valid conclusion. This is best accomplished by using a

Copyright © 2002 Marcel Dekker, Inc.

sampling procedure such as quartering or a device that riffles the samples.

One advantage of the optical method is to obtain information about the particle shape and presence of particle aggregations, e.g., degree and character of aggregations. A disadvantage is that the diameter is obtained from only two (of the three) dimensions of the particle, e.g., length and breadth; no information on the thickness or depth is available. The large number of particles that must be counted (300–500) in order to obtain a good estimation of the distribution makes the method rather slow and tedious.

Electron microscopes are high-resolution equipment used to visualize the samples having particle sizes below the limit of resolution of an optical microscope. The radiation source is a beam of high-energy electrons having wavelengths in the region of 0.1 Å. With current instrumentation, this results in smallest distances by which two objects are separated and remain distinguishable of approximately 5 Å. Scanning electron microscopy (SEM) finds wide application in the field of pharmaceutical technology followed by transmission electron microscopy (TEM) [202].

Sieve Analysis

Sieve analysis, based on either vibratory or suction principle, uses a series of standard sieves calibrated by the National Bureau of Standards. The method is generally used for screening coarse particles down to a material as fine as 44 μm (No. 325 sieve). Sieves produced by photoetching and electroforming techniques are now available with apertures from 90 down to 5 μm.

A typical testing procedure involves several steps. First, the selected number and size of sieves are stacked upon one another, with the largest openings (inversely related to mesh per inch) being at the top of the stack, and beneath that a pan to collect the particles finer than the smallest sieve. The known amount of powder to be analyzed is then placed on the top sieve and the set is vibrated in a mechanical device for a predetermined time period. The results are obtained by weighing the amount of material retained on each sieve and on the collecting pan. The suction method uses one sieve at a time and examines the amount retained on the screen. In both methods the data are expressed as frequency or cumulative frequency plots, respectively.

Sieving errors can arise from a number of variables, including sieve loading and duration and intensity of agitation [203]. Sieving can also cause attrition and, thus, size reduction of granular pharmaceutical materials. Standard sieving protocols must be employed in order to ensure reproducible particle size distributions between different batches of raw materials.

Sedimentation Method

This method is based on Stokes' law (see Sec. IV.A.). Practical techniques include simple sedimentation methods such as the Anderson pipette and the Cahn sedimentation balance. The Andreason pipette involves measuring the percentage of solids that settle with time in a graduated sedimentation vessel. Samples are withdrawn from the bottom of the vessel using a pipette, and the amount of solids is determined by drying and weighing. The Cahn sedimentation balance electronically records the amount of disperse powder settling out as a function of time. One advantage of this method is that the mass of particles is recorded continuously. Both methods are appropriate when the particle size is generally larger than 5 μm. In case particles smaller than 5 μm are to be evaluated, a centrifugal method is often employed to speed up the data collection by overcoming the effects of convection and diffusion. The method involves a photosedimentation device in which the light transmittance is measured through a cell filled with a dispersion of interest. An ultracentrifuge can also be used in such cases in order to increase the sedimentation rate of colloidal particles.

Coulter Counter and Electrical Sensing Devices

Counters, such as the Coulter Counter, determine the number of particles in a known volume of an electrolyte solution. The device employs the electrolyte displacement method and measures the equivalent spherical volume diameter, d_v. This type of equipment is used primarily to obtain the particle size distribution of the sample. It measures the change in an electrical sensing zone that occurs when a particle passes through an orifice positioned between two electrodes. The dispersion is drawn through a small orifice that has an electrode on either side. The disperse particles interfere with the current flow, causing the resistance to change. The resistance changes are related to the particle volumes and, when amplified as voltage pulses, may be counted. The instrument is capable of counting particles at a rate of approximately 4000 per second, and so both gross counts and particle size distributions are obtained in a relatively short period of time. Electronic pulse counters are also useful in studying

Copyright © 2002 Marcel Dekker, Inc.

particle growth, dissolution, and particulate contamination. Examples are the Coulter Counter and the Electrozone Counter. Coulter Electronics also manufactures a submicrometer particle sizing instrument, the Coulter Model N4, for analyzing particles in the size range of 0.003–0.3 μm. The instrument also provides molecular weights and diffusion coefficients.

Light Scattering

The quasi-elastic light-scattering technique (QELS), also called photon correlation spectroscopy (PCS), is based on the principle of light scattering and has been employed to determine the mean particle diameter and size distribution (polydispersity). Light scattering depends on the Faraday-Tyndall effect and is used widely in the determination of molecular weight, size, and shape of colloidal particles. Scattering may be described in terms of the turbidity, τ, the fractional decrease in intensity due to scattering as the incident light passes through a solution. When the particle is asymmetrical, the intensity of scattered light varies with the angle of observation, permitting an estimation of the shape and size of the particle. Light scattering has been used to study proteins, synthetic polymers, association colloids, and lyophobic sols.

Racey et al. [204, 205] applied QELS that uses laser light to determine diffusion coefficients and particle sizes as the Stokes' diameters. It consists of a laser light source, a temperature-controlled sample compartment, and a photomultiplier to detect the scattered light at a certain angle (usually 90°). QELS has been used in the examination of heparin aggregates in commercial preparations. The method is applicable for measuring the particle size ranging from 5 nm to approximately 3 μm. In this size range, the particles exhibit Brownian motion as a result of collisions with the molecules of surrounding liquid medium, which causes fluctuation in the intensity of scattered light. PCS, however, cannot characterize systems having broadly distributed particles. This problem has been overcome by combining sedimentation field-flow fractionation and QELS, which presents a detailed record of the particle size at each size interval.

Hydrodynamic Chromatography

This method is used particularly for colloids. A colloidal dispersion is forced through a long column packed with nonporous beads with an approximate radius of 10 μm. Particles of different particle size travel with different speeds around the beads and are thus collected in size fractions.

B. Determination of Electrical Properties

The movement of a charged particle with respect to an adjacent liquid phase is the basic principle underlying four electrokinetic phenomena: electrophoresis, electroosmosis, sedimentation potential, and streaming potential.

Electrophoresis involves the movement of a charged particle through a liquid under the influence of an applied potential difference. A sample is placed in an electrophoresis cell, usually a horizontal tube of circular cross section, fitted with two electrodes. When a known potential is applied across the electrodes, the particles migrate to the oppositely charged electrode. The direct current voltage applied needs to be adjusted to obtain a particle velocity that is neither too fast nor too slow to allow for errors in measurement and Brownian motion, respectively. It is also important that the measurement is taken reasonably quickly in order to avoid sedimentation in the cell. Prior to each measurement, the apparatus should be calibrated with particles of known zeta potential, such as rabbit erythrocytes.

The velocity of particle migration, v, across the field is a function of the surface charge or zeta potential and is observed visually by means of an ultramicroscope equipped with a calibrated eyepiece and a scale. The movement is measured by timing the individual particles over a certain distance, and the results of approximately 10–15 timing measurements are then averaged. From the measured particle velocity, the electrophoretic mobility (defined as v/E, where E is the potential gradient) can be calculated.

Electroosmosis is essentially opposite in principle to that electrophoresis. In this case, the particles are rendered immobile by forming a capillary or a porous plug. The liquid now moves through this plug or membrane, across which a potential is applied. The rate of flow of liquid through the plug is then determined under standard conditions. Electroosmosis provides another method for obtaining the zeta potential. Streaming potential differs from electroosmosis in that the potential is created by forcing a liquid to flow through a plug or bed of particles. Sedimentation potential measures the potential generated when the particles undergo sedimentation.

The determination of the zeta potential of particles in a disperse system provides useful information concerning the sign and magnitude of the charge and its effect on the stability of the system (see Sec. II.B) [56, 206–208]. It can be of value in the development of pharmaceutical suspensions, particularly if the

Copyright © 2002 Marcel Dekker, Inc.

controlled aggregation approach is used. The use of the zeta potential to evaluate the electrophoretic properties of other dosage forms such as liposomes and microcapsules has also been reported [209, 210].

A number of semi-automated or fully automated instruments are currently used for studying the electrokinetic properties. These include the electrophoretic mass transport analyzer (Zeta Potential Analyzer, Micrometrics Instrument Corp.), the streaming current detector (Hydroscan, Leeds and Northrup Co.), the electrokinetic sonic amplitude device (Matec Instruments, Inc.), and the instruments that determine zeta potential by measuring the electrophoretic mobility of the disperse particles in a charged electric field (Zeta Reader, Komline-Sanderson; Zeta Meter, Laser Zee Meter, Pen Kem Co.). The first three devices are suitable for determining the average zeta potential of coarse suspensions in systems having high solids contents, whereas the latter three are more useful for the measurement of colloidal particles in systems with low solids contents.

C. Rheological Measurements

Rheological evaluations can be simple or complex, depending on the nature of the product. They feature important quality control of all disperse systems. For example, the adequacy of hydration and quality control of the gums used as viscosity-imparting agents in the products is best confirmed by a rheological test. These natural or synthetic gums are polymers of varying molecular weights from lot to lot, which directly affect the product's viscosity. In addition, the degree of dispersion, particle size, and size distribution influence the viscosity and dispersion consistency. Also, formulation additives, such as sweeteners, surfactants, and flocculating agents, influence the rheological properties of the dispersions. Viscosity characteristics of a suspension may be altered not only by the vehicle used but also by the solids content. Increasing amounts of solid particles lead to increased viscosities of the disperse system. The viscosity of lyophobic dispersions is not much higher than that of the original liquid because of minimal interaction between the internal and external phases. The viscosity of lyophilic systems, especially polymer solutions, is much higher because of the interaction between the two phases. As a result of this interaction, dispersions possessing both Newtonian and non-Newtonian properties are formed. Highly viscous or rigid gels can be formed at higher polymer concentrations.

A wide variety of viscometers (ranging from simple to sophisticated) has been designed to evaluate the rheological properties of various disperse systems. It is necessary to select one suitable for the viscosity ranges encountered in a given application. The instrument should provide the required rheological data over the desired range of shear, time under shear, time and temperature. Viscometers are classified as those that operate at a single shear rate and those that allow more than one rate of shear to be examined. The simplest qualitative measurement of viscosity may be obtained by bubble, cup, falling-ball, falling-rod, and capillary viscometers. These simple devices provide a one-point measurement and are widely used for Newtonian fluids and dilute solutions. The glass capillary Cannon-Fenske, Ubbelohde, and Ostwald viscometers are the most popular instruments based on this method.

Most rheometers in use to date are based on rotating the samples and measuring its response to the applied stress by a variety of sensors. The advantages of rotational viscometers are that the shear rate can be varied over a wide range, and that continuous measurements at a given shear rate or shear stress can be made for extended periods of time, allowing the evaluation of time-dependency and shear-dependency properties. Viscometers of this type are routinely found in many pharmaceutical laboratories. Simple rotational viscometers are the Stormer viscometer, the ICI viscometer, and the Brookfield viscometer. A Brookfield viscometer with the helipath attachment is a valuable piece of rheological equipment for measuring the settling behavior and structure of pharmaceutical suspensions. The instrument consists of a slowly rotating T-bar spindle, which is lowered into an undisturbed sample. The dial reading of the viscometer measures the resistance to flow that the spindle encounters from the structure at various levels in the sediment. Taking rheograms at various time intervals, under standard conditions, gives a description of the suspension and its physical stability. The technique is most useful for viscous suspensions high in solids that develop sufficient shear stress for measurement. The instrument is also excellent for characterizing flocculated systems. To simulate the situation illustrated in Fig. 2, coaxial-cylinder viscometers have been designed in which the plates are bent to form cylindrical surfaces, one of which is stationary and the other is rotating. There are two types of concentric cylinder viscometers, namely, the Searle system and the Couette system.

In the Couette type, the material is contained in an annular gap between the inner cylindrical bob or spindle

Copyright © 2002 Marcel Dekker, Inc.

and the outer, concentric cylindrical cup. In the Stormer viscometer or the more advanced Haake Rotovisco, the cup is stationary. Cone-and-plate and parallel-plate rheometers are designed to handle small-quantity samples, from about 0.2 to 5 mL. The plate may be stationary while the cone is rotating, or vice versa. The angle between the cone and the plate has to be extremely small, usually 0.1–3°. Since the sample size is very small, this method is very sensitive to sample drying and temperature changes. Another disadvantage of this method is the loss of material due to slinging from the gap and sample heating, especially at high shear rates.

D. Temperature and Gravitational Stress Tests

Disperse systems can be subjected to cyclic temperature testing, e.g., to conditions of repeated freezing and thawing (e.g., -50°C to $+40$ °C in 24 h) exposing them to elevated temperatures (>40 °C) for short periods of storage to test for the physical stability. The value of such a stress test procedure may be questionable since the exposure to elevated temperatures often results in a drastic change in the drug's solubility. Significant amounts of drug may go into solution and reprecipitate upon subsequent cooling. Most suspension systems contain surfactants and protective colloids to prevent crystal growth, and, thus, inducing crystal growth during stress testing may be of limited value. The use of stressful aging tests, however, has one advantage. If a given suspension is able to withstand exposure to extreme temperatures, it is safe to assume that the product will have good physical stability during prolonged storage at ambient temperature. On the other hand, failure of any preparations to meet such stringent testing procedures should not be considered a bar to further testing, since many marketed pharmaceutical suspensions would have been rejected from further consideration on this basis alone.

The use of stress tests to study the stability of emulsions has been described in Sec. V.C.

IX. MANUFACTURING AND EQUIPMENT

The preparation of satisfactory disperse systems consists of three main steps: preparing the internal phase in the proper size range, dispersing the internal phase in the dispersion medium, and, finally, stabilizing the resultant product. These three steps may be done sequentially, but in many cases (e.g., emulsions), they are usually done simultaneously.

Suspensions can be prepared by either dispersing finely divided powders in an appropriate vehicle or causing precipitation within the vehicle. For most pharmaceutical preparations, the proper size of the internal phase is obtained by a mechanical breakdown of the solid material. One of the most rapid, convenient, and inexpensive methods of producing fine drug powders about 10–50 μm in size is micropulverization. Micropulverizers are high-speed attrition or impact mills, which are efficient in reducing powders to the size acceptable for most oral and topical suspensions. For finer particles, fluid energy grinding, sometimes referred to as jet-milling or micronizing, is quite effective. By this process, the shearing action of high-velocity compressed air streams on the particles in a confined chamber produces the desired ultra-fine or micronized particles. The particles are accelerated to high velocities by the sonic and supersonic velocity of turbulent air streams and collide with one another, resulting in fragmentation and size reduction. These micronized drugs are often employed in parenteral and ophthalmic preparations. Particles of extremely small dimension may also be produced by spray-drying techniques [211]. This method results in a porous, free flowing, easily wetted, monodisperse microcrystalline powder. Other equipment for particle size reduction include colloid mills, ball mills, ultrasonic generators, or homogenizers.

The precipitation (condensation, aggregation) method is more complex and has been used to create particles of suitable size by permitting the atoms or molecules of the material to gather together. In this process, the material is dissolved in a suitable solvent, then a miscible nonsolvent is added and a precipitate is formed. For example, fine sulfur particles can be obtained when water (nonsolvent) is added into an alcoholic solution of sulfur. Suspensions can be prepared by controlled crystallization wherein a supersaturated solution is cooled rapidly with constant agitation to obtain a large number of small, uniform drug crystals. Microcrystals can also be conveniently produced by bubbling liquid nitrogen through a saturated solution of solute prior to solvent freezing. Other methods for obtaining fine particles include phase change such as sublimation and pH adjustment. Shock cool methods can offer a possible solution to the electrostatic charge problems associated with dry particle milling. Chemical reactions, such as double decomposition, have also been used for certain pharmaceutical preparations, e.g., White lotion.

The second step is to disperse the particles within the dispersion medium. In emulsions, this step is accomplished at the same time that particle reduction

Copyright © 2002 Marcel Dekker, Inc.

occurs (Sec. V.B. gives more details on the commonly applied techniques to obtain pharmaceutically relevant emulsions). Surfactants are usually added in order to promote both particle size reduction and uniform dispersion. Suspensions may be prepared by dispersing the particles resulting from the mechanically breakdown process, along with an addition of a surfactant to aid the wetting process. Alternatively, both size reduction and dispersion processes can take place simultaneously in a manner similar to that for emulsions. Homogenization is normally required when various other additives are incorporated. On a laboratory scale, an ultrasonic generator may be used, whereas a conventional colloid mill is more suitable for an industrial-scale production. The most important techniques to prepare nanoparticles are described in Sec. VI.B.

For the production of emulsions on a large scale, the oil and water phases are often heated separately in large tanks. When waxes are present, both phases must be heated at a temperature above the highest melting point of any one component. One phase is then pumped into the tank containing the second phase with constant agitation throughout the time of addition. After cooling, the product is once again homogenized and then packaged. Semisolid preparations, such as creams and ointments, are prepared using appropriate mixing machinery. Often the dispersion step is carried out with colloid mills or homogenizers at elevated temperatures to reduce the viscosity of waxy constituents. On a small scale, a mortar and pestle may be used for communition, and dispersion may be done either sequentially or simultaneously (see Sec. V.B.). Once a satisfactory R&D laboratory-scale formulation has been developed, the problem of pilot- and manufacturing-scale production must be solved. In this scale-up step, formulators may encounter problems of mixing and dispersion not found in a small-scale production [212, 213].

REFERENCES

1. SE Tabibi, CT Rhodes. Disperse systems. In: GS Banker, CT Rhodes, eds. *Modern Pharmaceutics*, 3rd ed. New York: Marcel Dekker, 1996, pp 299–331.
2. JG Nairn. Disperse systems. In: J Swarbrick, JC Boylan, eds. *Encyclopedia of Pharmaceutical Technology*, Vol. 4. New York: Marcel Dekker, 1992, pp 107–l20.
3. RB Pates, IJR Patel, MC Roggle, VP Shah, VK Prasad, A Selen, PG Welling. Bioavailability of hydrochlorothiazide from tablets and suspensions. J Pharm Sci 73:359–365, 1984.
4. PJ Stout, SA Howard, JW Mauger. Dissolution of pharmaceutical suspensions. In: J Swarbrick, JC Boylan, eds. *Encyclopedia of Pharmaceutical Technology*, Vol. 4. New York: Marcel Dekker, 1992, pp 169–192.
5. MD Donovan, DR Flanagan. Bioavailability of disperse dosage form. In: HA Lieberman, MM Rieger, GS Banker, eds. *Pharmaceutical Dosage Forms: Disperse Systems*, Vol. 1. 2nd ed. New York: Marcel Dekker, 1996, pp 315–376.
6. JW Mauger, SA Howard, K Amin. Dissolution profiles for finely divided drug suspensions. J Pharm Sci 72(2):190–193, 1983.
7. K Hirano, H Yamada. Studies on the absorption of practically water-insoluble drugs following injection VI: subcutaneous absorption from aqueous suspensions in rats. J Pharm Sci 7l(5):500–505, 1982.
8. RA Nash. Pharmaceutical suspensions. In: HA Lieberman, MM Rieger, GS Banker, eds. *Pharmaceutical Dosage Forms: Disperse Systems*, Vol. 1. New York: Marcel Dekker, 1988, pp l51–198.
9. NK Patel, L Kennon, RS Levinson. Pharmaceutical suspensions. In: L Lachman, HA Lieberman, JL Kanig, eds. *The Theory and Practice of Industrial Pharmacy*, 3rd ed. Philadelphia: Lea & Febiger, 1986, pp 479–501.
10. J Swarbrick. Coarse dispersions. In: AR Gennaro, ed. *Remington: The Science and Practice of Pharmacy*, 19th ed. Easton, Pennsylvania: Mack Publishing Co., 1995, pp 278–291.
11. MJ Falkiewicz. Theory of suspensions. In: HA Lieberman, MM Rieger, GS Banker, eds. *Pharmaceutical Dosage Forms: Disperse Systems*, Vol. 1. New York: Marcel Dekker, l998, pp 13–48.
12. DJ Burgess. Colloids and colloid drug delivery systems. In: J Swarbrick, JC Boylan, eds. *Encyclopedia of Pharmaceutical Technology*, Vol. 3. New York: Marcel Dekker, 1992, pp 3l–64.
13. A Martin. *Physical Pharmacy*, 4th ed. Philadelphia: Lea & Febiger, 1993, pp 393–422.
14. MJ Alonso. Nanoparticulate drug carrier technology. In: S Cohen, H Bernstein, eds. *Microparticulate Systems for the Delivery of Proteins and Vaccines*. New York: Marcel Dekker, 1996, pp 203–242.
15. R Bodmeier, P Maincent. Polymeric dispersions as drug carriers. In: HA Lieberman, MM Rieger, GS Banker, eds. *Pharmaceutical Dosage Forms: Disperse Systems*, Vol. 3. 2nd ed. New York: Marcel Dekker, 1998, pp 87–128.
16. F Jaeghere, E Doelker, R Gurny. Nanoparticles. In: E Mathiowitz, ed. *Encyclopdia of Controlled Drug Delivery*, Vol. 2. New York: John Wiley & Sons, 1999, pp 641–664.
17. JS Lucks, BW Müller, RH Müller. Polymeric and emulsion carriers—interaction with antiflocculants and ionic surfactants. Int J Pharm 63:183–188, 1990.
18. RH Müller, BHL Böhm. Nanosuspensions. In: RH Müller, S Benita, BHL Böhm, eds. *Emulsions and*

Copyright © 2002 Marcel Dekker, Inc.

Nanosuspensions for the Formulation of Poorly Soluble Drugs. Stuttgart: Medpharm Scientific Publishers, 1998, pp 149–l74.
19. TA Wheatley, CR Steuernagel. Latex emulsions for controlled drug delivery. In: JW McGinity, ed. *Aqueous Polymeric Coatings for Pharmaceutical Dosage Forms*,
2nd ed. New York: Marcel Dekker, 1997, pp 1–54.
20. MR Harris, I Ghebre-Sellassie. Aqueous polymeric coating for modified release oral dosage forms. In: JW McGinity, ed. *Aqueous Polymeric Coatings for Pharmaceutical Dosage forms*, 2nd ed. New York: Marcel Dekker, 1997, pp 81–100.
21. KOR Lehmann. Chemistry and application properties of polymethacrylate coating systems. In: JW McGinity, ed. *Aqueous Polymeric Coatings for Pharmaceutical Dosage forms*, 2nd ed. New York: Marcel Dekker, 1997, pp 101–176.
22. J Wang, I Ghebre-Sellassie. Aqueous polymer dispersions as film formers. In: HA Lieberman, MM Rieger, GS Banker, eds. Pharmaceutical Dosage Forms: Disperse Systems, Vol. 3. 2nd ed. New York: Marcel Dekker, 1998, pp 129–161.
23. B Idson. Pharmaceutical emulsions. In: HA Lieberman, MM Rieger, GS Banker, eds. *Pharmaceutical Dosage Forms: Disperse Systems*, Vol. 1. New York: Marcel Dekker, 1988, pp 199–244.
24. GM Eccleston. Emulsions. In: J Swarbrick, JC Boylan, eds. *Encyclopedia of Pharmaceutical Technology*, Vol. 5. New York: Marcel Dekker, 1992, pp 137–188.
25. BA Mulley. Medicinal emulsions. In: KJ Lissant, ed. *Emulsions and Emulsion Technology, Part I*. New York: Marcel Dekker, 1974, pp 291–350.
26. MM Rieger. Emulsions. In: L Lachman, HA Lieberman, JL Kanig, eds. *The Theory and Practice of Industrial Pharmacy*, 3rd ed. Philadelphia: Lea & Febiger, 1986, pp 502–532.
27. LH Block. Pharmaceutical emulsions and microemulsions. In: HA Lieberman, MM Rieger, GS Banker, eds. *Pharmaceutical Dosage Forms: Disperse Systems*, Vol. 2. New York: Marcel Dekker, 1996, pp 47–110.
28. N Garti. Double emulsions-scope, limitations and new achievements. Colloids Surfaces A: Physicochem Eng Aspects 123–124:233–246, 1997.
29. SH Klang, M Parnas, S Benita. Emulsions as drug carriers – possibilities, limitations and future perspectives. In: RH Müller, S. Benita, BHL Böhm, eds. *Emulsions and Nanosuspensions for the Formulation of Poorly Soluble Drugs*. Stuttgart: Medpharm Scientific Publishers, 1998, pp 31–78.
30. B Lundberg. Preparation of drug-carrier emulsions stabilized with phosphatidylcholine-surfactant mixtures. J Pharm Sci 83(1):72–75, 1994.
31. DP Bluhm, RS Summers, MMJ Lowes, HH Durrheim. Lipid emulsion content and vitamin A stability in TPN admixtures. Int J Pharm 68:277–280, 1991.
32. JB Boyett, CW Davis. Injectable emulsions and suspensions. In: HA Lieberman, HA, MM Rieger, GS Banker, eds. *Pharmaceutical Dosage Forms: Disperse Systems*, Vol. 2, New York: Marcel Dekker, 1989, pp 379–416.
33. PJ Breen, DT Wasan, Y-H Kim, AD Nikolov, CS Shetty. Emulsions and emulsion stability. In: J Sjöblom, ed. *Emulsions and Emulsion Stability*. New York: Marcel Dekker, 1996, pp 23–286.
34. MJ Parnham. Safety and tolerability of intravenously administered phospholipids and emulsions. In: RH Müller, S Benita, BHL Böhm, eds. *Emulsions and Nanosuspensions for the Formulation of Poorly Soluble Drugs*. Stuttgart: Medpharm Scientific Publishers, 1998, pp 131–140.
35. M Rosoff. Specialized pharmaceutical emulsions. In: HA Lieberman, MM Rieger, GS Banker, eds. *Pharmaceutical Dosage Forms: Disperse Systems*, Vol. 3, 2nd ed. New York: Marcel Dekker, 1998, pp 1–42.
36. K Westesen, T Wehler. Physicochemical characterization of a model intravenous oil-in-water emulsion. J Pharm Sci 81(8):777–786, 1992.
37. C Leuner, J Dressman. Improving drug solubility for oral delivery using solid dispersions. Eur J Pharm Biopharm 50(1):47–60, 2000.
38. JJ Sciarra. Aerosols. In: AR Gennaro, ed. *Remington: The Science and Practice of Pharmacy*, 19th ed. Easton, Pennsylvania: Mack Publishing Co., 1995, pp 1676–1692.
39. MJ Clarke, MJ Tobyn, TN Staniforth. Physicochemical factors governing the performance of nedocromil sodium as a dry powder aerosol. J Pharm Sci 89(9):11601–1169, 2000.
40. T. Allen. Particle size measurement. In: JC Williams, ed. *The Powder Technology Series*. London: Chapman and Hall, 1975.
41. JW Vanderhoff, MS El-Aasser. Theory of colloids. In: HA Lieberman, MM Rieger, GS Banker, eds. *Pharmaceutical Dosage Forms: Disperse Systems*, Vol. 1, 2nd ed. New York: Marcel Dekker, 1996, pp 91–152.
42. H Heywood. Symposium on particle size analysis. Inst Chem Eng Supp 25:14–24, 1947.
43. I Goodarznia, DN Sutherland. Floc simulation: Effects of particle size and shape. Chem Eng Sci 30:407–412, 1975.
44. A Heyd, D Dhabhar. Particle shape effect on caking of coarse granulated antacid suspensions. Drug Cosmet Ind 125:42–45. 1979.
45. M Bisrat, C Nyström. Physicochemical aspects of drug release. VIII. The relation between particle size and surface specific dissolution rate in agitated suspensions. Int J Pharm 47:223–231, 1988.
46. A Martin. *Physical Pharmacy*, 4th ed. Philadelphia: Lea & Febiger, 1993, pp 423–452.
47. DP McNamara, ML Vieira, JR Crison. Dissolution of pharmaceuticals in simple and complex systems. In: GL Amidon, PI Lee, EM Topp, eds. *Transport Processes in*

Copyright © 2002 Marcel Dekker, Inc.

Pharmaceuticl Systems. New York: Marcel Dekker, 2000.
48. PC Hiemenz. *Principles of Colloid and Surface chemistry,* 2nd ed. New York: Marcel Dekker, 1986.
49. T Handa, Y Maitani, K Miyazima, M Nakagaki. Interfacial phenomena. In: J Swarbrick, JC Boylan, eds. *Encyclopedia of Pharmaceutical Technology*, Vol. 8, New York: Marcel Dekker, 1993, pp 131–174.
50. G Zografi. Interfacial phenomena. In: AR Gennaro, ed. *Remington: The Science and Practice of Pharmacy*, 19th ed. Easton, Pennsylvania: Mack Publishing Co., 1995, pp 241–251.
51. EJW Verwey, JThG Overbeek. *Theory of the Stability of Lyophobic Colloids*. Amsterdam: Elsevier, 1948.
52. B Derjaguin, L Landau. Acta Physica Chim, USSR 14:663, 1941.
53. B Derjaguin, L Landau. J Exp Theor Physics, USSR 11:802, 1941.
54. BA Matthews, CT Rhodes. Use of the Derjaguin, Landau, Verwey and Overbeek theory to interpret pharmaceutical suspension stability. J Pharm Sci 59:521–525, 1970.
55. JB Kayes. Pharmaceutical suspensions: relation between zeta potential, sedimentation volume and suspension stability. J Pharm Pharmacol 29:199–204, 1977.
56. A Delgado, V Gallardo, A Parrera, F González-Caballero. A study of the electrokinetic and stability properties of nitrofurantoin suspensions. II: Flocculation and redispersion properties as compared with theoretical interaction energy curves. J Pharm Sci 79:709–718, 1990.
57. ThF Tadros. Polymeric surfactants:, stabilization of emulsions and dispersions. In: E Desmond Goddard, JV Gruber, eds. *Principles of Polymer Science and Technology in Cosmetics and Personal Care*. New York: Marcel Dekker, 1999, pp 73–112.
58. RJ Meyer, L Cohen. Rheology of natural and synthetic hydrophilic polymer solutions and related to suspending ability. J Soc Cosmet Chem 10:1–11, 1959.
59. P Sherman. The flow properties of emulsions. J Pharm Pharmacol 16:1–25, 1964.
60. JT Carstensen. *Theory of Pharmaceutical Systems*, Vol. 2, *Heterogeneous Systems*. New York: Academic Press, 1973, pp 1–89.
61. DE Deem. Rheology of dispersed systems. In: HA Lieberman, MM Rieger, GS Banker, eds. *Pharmaceutical Dosage Forms: Disperse Systems*, Vol. 1, New York: Marcel Dekker, 1988, pp 367–514.
62. MA Ramadan, R Tawashi. Effect of surface geometry and morphic features on the flow characteristics of microsphere suspensions. J Pharm Sci 79:929–933, 1990.
63. PE Miner. Emulsion rheology: creams and lotions. In: D Laba, ed. *Rheological Properties of Cosmetics and Toiletries*. New York: Marcel Deklcer, 1993, pp 313–370.
64. CD Vaughan. Predicting stability in rheologically modified systems. In: D Laba, ed. *Rheological Properties of Cosmetics and Toiletries*. New York: Marcel Dekker, 1993, pp 371–401.
65. HN Naé. Introduction to rheology. In: D Laba, ed. *Rheological Properties of Cosmetics and Toiletries*. New York: Marcel Dekker, 1993, pp 9–33.
66. MM Rieger. Surfactants. In: HA Lieberman, MM Rieger, GS Banker, eds. *Pharmaceutical Dosage Forms: Disperse Systems*, Vol. 1, 2nd ed. New York: Marcel Dekker, 1996, pp 211–286.
67. RG Laughlin. HLB, from a thermodynamic perspective. J Soc Cosmet Chem 32:371–378, 1981.
68. H Schott. Solubility parameter and hydrophilic lipophilic balance of nonionic surfactants. J Pharm Sci 73:790–792, 1984.
69. WC Griffin. Classification of surface-active agents by "HLB". J Soc Cosmet Chem 1:311–326, 1949.
70. WC Griffin. Emulsions. In: RE Kirk, DF Othmer, eds. *Encyclopedia of Chemical Technology*, Vol. 5, New York: Interscience, 1950, pp 692–718.
71. Atlas Booklet. A Guide to Formulation of Industrial Emulsions with Atlas Surfactants, Wilmington: Atlas Powder Co., 1953.
72. MT Clarke. Rheological additives. In: D Laba, ed. *Rheological Properties of Cosmetics and Toiletries*. New York: Marcel Dekker, 1993, pp 55–152.
73. JL Zatz, JJ Berry, DA Alderman. Viscosity-imparting agents in disperse systems. In: HA Lieberman, MM Rieger, GS Banker, eds. *Pharmaceutical Dosage Forms: Disperse Systems*, Vol. 2, New York: Marcel Dekker, 1989, pp 171–204.
74. JA Ranucci, IB Silverstein. Polymeric pharmaceutical excipients. In: HA Lieberman, MM Rieger, GS Banker, eds. *Pharmaceutical Dosage Forms: Disperse Systems*, Vol. 3, 2nd ed. New York: Marcel Dekkter, 1998, pp 243–289.
75. CB Anger, D Rupp, P Lo, H Takuri. Preservation of dispersed systems. In: HA Lieberman, MM Rieger, GS Banker, eds. *Pharmaceutical Dosage Forms: Disperse Systems*, Vol. 1, 2nd ed. New York: Marcel Dekker, 1996, pp 377–435.
76. TE Haag, DF Loncrini. Esters of para-hydroxybenzoic acid. In: JJ Kabara, ed. *Cosmetic and Drug Preservation*. New York: Marcel Dekker, 1984, pp 63–77.
77. WB Hugo. Phenols as preservatives for pharmaceutical and cosmetic products. In: JJ Kabara, ed. *Cosmetic and Drug Preservation*. New York: Marcel Dekker, 1984, pp 109–113.
78. TJ McCarthy. Formulated factors affecting the activity of preservatives. In: JJ Kabara, ed. *Cosmetic and Drug Preservation*. New York: Marcel Dekker, 1984, pp 359–388.
79. KS Alexander, J Azizi, D Dollimore, V Upaala. Interpretation of the hindered settling of calcium carbonate

Copyright © 2002 Marcel Dekker, Inc.

suspensions in terms of permeability. J Pharm Sci 79:401–406, 1990.
80. RDC Jones, BA Matthews, CT Rhodes. Physical stability of sulfaguanidine suspensions. J Pharm Sci 59:518–520, 1970.
81. A Martin. *Physical Pharmacy*, 4th ed. Philadelphia: Lea & Febiger, 1993, pp 477–511.
82. JL Zatz, L Schnitzer, P Sarpotdar. Flocculation of sulfamerazine suspensions by a cationic polymer. J Pharm Sci 68(12):1491–1494, 1979.
83. GC Jeffrey, RH Ottewill. Reversible aggregation Part reversible flocculation monitored by turbidity measurements. Colloids Polymer Sci 266:173–179, 1988.
84. T Higuchi. Some physical chemical aspects of suspension formulation. J Am Pharm Assoc, Sci Ed 47:657–660, 1958.
85. KJ Frederick. Performance and problems of pharmaceutical suspensions. J Pharm Sci 50:531–535, 1961.
86. BA Matthews. The use of the Coulter counter in emulsion and suspension studies. Can J Pharm Sci 6:29–34, 1971.
87. JE Carless. Dissolution and crystal growth in aqueous suspension of cortisone acetate. J Pharm Pharmacol 10:630–639, 1968.
88. SA Young, G Buckton. Particle growth in aqueous suspensions: the influence of surface energy and polarity. Int J Pharm 60:235–241, 1990.
89. NB Shah, BB Sheth. Effect of polymers on dissolution from drug suspensions. J Pharm Sci 65(11):1618–1623, 1976.
90. CM Ofner III, RL Schnaare, JB Schwartz. Reconstitutable suspensions. In: HA Lieberman, MM Rieger, GS Banker, eds. *Pharmaceutical Dosage Forms: Disoerse Systems*, Vol. 2, New York: Marcel Dekker, 1989, pp 317–334.
91. R Bodmeier, O Paeratakul. Suspensions and dispersible dosage forms of multiparticulates. In: I Ghebre-Sellassie, ed. *Multiparticulate Oral Drug Delivery*. New York: Marcel Dekker, 1994, pp 143–157.
92. Y Kawashima, T Iwamoto, T Niwa, H Takeuchi. Preparation and characterization of a new controlled release ibuprofen suspension for improving suspendability. Int J Pharm 75:25–36, 1991.
93. R Sjöqvist, C Graffner, I Ekman, W Sinclair, JP Woods. In vivo validation of the release rate and palatability of remoxipride-modified release suspension. Pharm Res 10(7):1020–1026, 1993.
94. Y Raghunathan (to Pennwalt Corporation). U.S. Patent 4,221,778 (1980).
95. Y Raghunathan, L Amsel, O Hinsvark, W Bryant. Sustained-release drug delivery system I: coated ion-exchange resin system for phenylpropanolamine and other drugs. J Pharm Sci 70:379–384, 1981.
96. JE Tingstad Physical stability testing of pharmaceuticals. J Pharm Sci 53(8):955–962, 1964.
97. TF Tadros, B Vincent. Emulsion stability. In: P Becher, ed. *Encyclopedia of Emulsion Technology*. Vol. 1, New York: Dekker, 1983, pp 129–285.
98. J Mikula, VA Munoz. Characterization of emulsions and suspensions in the petroleum industry using cryo-SEM and CLSM. Colloids Surfaces A: Physicochem Eng Aspects, 174(1–2):23–36, 2000.
99. DBR Kumar, MR Reddy, VN Mulay, N Krishnamurti. Acrylic co-polymer emulsion binders for green machining of ceramics. Eur Polym J 36(7):1503–1510, 2000.
100. M Chappat. Some applications of emulsions. Colloids Surfaces A: Physicochem Eng Aspects 9:57–77, 1994.
101. Z Ju, Y Duan, Z Ju. Plant oil emulsion modifies internal atmosphere, delays fruit ripening and inhibits internal browning in Chinese pears. Postharvest Biol and Tech 20(3):243–250, 2000.
102. Y Sela, S Magdassi, N Garti. Release of markers from the inner water phase of W/O/W emulsions stabilized by silicone based polymeric surfactants. J Control Release 33(1):1–12, 1995.
103. D Rousseau. Fat crystals and emulsion stability—a review. Food Res Inter 33(1):3–14, 2000.
104. K Heinzelmann, K Franke. Using freezing and drying techniques of emulsions for the microencapsulation of fish oil to improve oxidation stability. Colloids Surfaces B: Biointerfaces 12(3–6):223–229, 1999.
105. P Paquin. Technological properties of high pressure homogenizers: the effect of fat globules, milk proteins and polysaccharides. Int Dairy J 9(3–6):329–335, 1999.
106. SM Joscelyne, G Traegardh. Food emulsions using membrane emulsification: conditions for producing small droplets. J Food Eng 39(1):59–64, 1999.
107. HD Goff. Colloidal Aspects of Ice Cream-A Review, Int Dairy J 7(6–7):363–373, 1997.
108. DJ McClements. Advances in the application of ultrasound in food analysis and processing. Trends Food Sci Tech 6(9):293–299, 1995.
109. D Peressini, A Sensidoni, B de Cindio. Rheological Characterization of Traditional and light mayonnaises. J Food Eng 35(4):409–417, 1998.
110. A M. Lipid emulsions in parenteral nutrition. Ann Nutr Metab 43(1):1–13, 1999.
111. S Ding. Recent developments in ophthalmic drug delivery. Pharm Sci Tech Today 1(8):328–335, 1998.
112. CA Gogos, F Kalfarentzos. Total parenteral nutrition and immune system activity: A review. Nutrition 11(4):339–344, 1995.
113. H Okochi, M Nakano. Preparation and evaluation of W/O/W type emulsions containing vancomycin. Adv Drug Del Rev 45(1):5–26, 2000.
114. F Nielloud, G Marti-Mestres, JP Laget, C Fernandez, H Maillols. Emulsion formulations:study of the influence of parameters with experimental designs. Drug Dev Ind Pharm 22(2):159–166, 1996.

Copyright © 2002 Marcel Dekker, Inc.

115. M Gallarate, ME Carlotti, M Trotta, S Bovo. On the stability of ascorbic acid in emulsified systems for topical and cosmetic use. Int J Pharm 188(2):233–241, 1999.
116. C Gallegos, JM Franco. Rheology of food, cosmetics and pharmaceuticals. Curr Opin Colloid Interf Sci 4(4):288–293, 1999.
117. D Miller, EM Wiener, A Turowski, C Thunig, H Hoffmann. O/W emulsions for cosmetic products stabilized by alkyl phosphates rheology and storage tests. Colloids Surfaces A: Physicochem Eng Aspects 152(1–2):155–160, 1999.
118. R Clark. Cosmetic emulsions. In: HW Hibbott, ed. *Handbook of Cosmetic Science*. New York: Pergamon Press, 1963, pp 175–204.
119. BW Barry. *Dermatological Formulations*. New York: Marcel Dekker, 1983, pp 296–350.
120. P Becher. Testing of emulsion properties. In: *Emulsions Theory and Practice*, 2nd ed. New York: Reinhold Publ. Co., 1957, pp 381–429.
121. P Couvreur, MJ Blanco-Prieto, F Puisieux, B Roques, E Fattal. Multiple emulsion technology for the design of microspheres containing peptides and olegopeptides. Adv Drug Del Rev 28:85–96, 1997.
122. N Garti, A Aserin. Double emulsions stabilized by macromolecular surfactants. Advances in colloid and interfaces science 65:37–69, 1996.
123. N Garti. Double emulsions-scope, limitations and new achievements. Colloids Surfaces A: Physicochem Eng Aspects 123–124:233–246, 1997.
124. J Herrmann, R Bodmeier. Biodegradable somatostatin acetate containing microspheres prepared by various aqueous and non-aqueous solvent evaporation method. Eur J Pharm Biopharm 45(1):75–82, 1998.
125. PB O'Donnell, JW McGinity. Preparation of microspheres by the solvent evaporation technique. Adv Drug Del Rev 28(1):25–42, 1997.
126. FB Rosevear. Liquid crystals: The mesomorphic phases of surfactant compositions. J Soc Cosmet Chem 19:581, 1968.
127. J Sjöblom, R Lindberg, SE Friberg. Microemulsions-phase equilibria characterization, structures, applications and chemical reactions, Adv Colloid Interf Sci 65:125–287, 1996.
128. HN Bhargava, A Narurkar, LM Lieb. Using microemulsions for drug delivery. Pharm Tech 11:46–54, 1987.
129. H Wennerström, O Söderman, U Olsson, B Lindman. Macroemulsions versus microemulsions. Colloids Surfaces A: Physicochem Eng Aspects 123–124:13–26; 1997.
130. WA Ritschel. Microemulsions for improved peptide absorption from the gastrointestinal tract. Methods Find Exp Clin Pharmacol 13:205–220, 1991.
131. A Haße, S Keipert. Development and characterization of microemulsions for ocular application. Eur J Pharm Biopharm 43(2): 179–183, 1997.
132. MB Charro, GI Vilas, TB Méndez, MAL Q, JP Marty, RH Guy. Delivery of a hydrophilic solute through the skin from novel microemulsion systems. Eur J Pharm Biopharm 43(1):37–42, 1997.
133. S Vemuri, CT Rhodes. Preparation and characterization of liposomes as therapeutic delivery systems: a review. Pharm Acta Helv 70(2):95–111, 1995.
134. A Bochot, P Couvreur, E Fattal. Intravitreal administration of antisense oligonucleotides: potential of liposomal delivery. Prog Retin Eye Res 19(2):131–147, 2000.
135. C Nastruzzi, R Cortesi, E Esposito, R Gambari, M Borgatti, N Bianchi, G Feriotto, C Mischiati. Liposomes as carriers for DNA PNA hybrids. J Control Rel 68(2):237–249, 2000.
136. C Welz, W Neuhuber, H Schreier, R Repp, W Rascher, A Fahr. Nuclear gene targeting using negatively charged liposomes. Int J Pharm 196(2):251–252, 2000.
137. C Mader, S Küpcü, UB Sleytr, M Sára. S-layer-coated liposomes as a versatile system for entrapping and binding target molecules. Biochem Biophys Acta Biomembranes 1463(1):142–150, 2000.
138. R Bodmeier. Multiphase system. Int Patent Application WO 98/55100 (1998).
139. W Im-Emsap, GA Brazeau, JW Simpkins, R Bodmeier. Sustained drug delivery of 17-β estradiol from injectable biodegradble in situ forming microparticles (ISM) system. AAPS PharmSci Supplement 2(4), AAPS Annual Meeting Abstracts, 2000.
140. W Im-Emsap, R Bodmeier. In vitro drug release from in situ forming microparticle (ISM)-systems with dispersed drug. AAPS PharmSci Supplement 2(4), AAPS Annual Meeting Abstracts, 2000.
141. SE Friberg, LS Quencer, ML Hilton. Theory of emulsions. In: HA Lieberman, MM Rieger, GS Banker, eds. *Pharmaceutical Dosage Forms: Disperse Systems*, 2nd ed. Vol. 1, New York: Marcel Dekker, 1996, pp 53–90.
142. ER Garrett. Stability oil-in-water emulsions. J Pharm Sci 54(11):1557–1570, 1965.
143. K Shinoda, H Kunieda. Phase properties of emulsions: PIT and HLB. In: P Becher, ed. *Encyclopedia of Emulsion Technology*, Vol. 1, New York: Dekker, 1983, pp. 337–368.
144. N Garti, A Aserin. Pharmaceutical emulsions double emulsions and microemulsions. In: S Benita, ed. *Microencapsulation-Methods and Industrial Applications*. New York: Marcel Dekker, 1996, pp 411–534.
145. IB Ivanov, KB Danov, PA Kralchevsky. Flocculation and coalescence of micron-size emulsion droplets. Colloids Surfaces A: Physicochem Eng Aspects 152:161–168, 1999.
146. MM Robins. Emulsions-creaming phenomena. Curr Opin Colloid Interf Sci 5:265–272, 2000.

Copyright © 2002 Marcel Dekker, Inc.

147. M Bury. J Gerhards, W Erni. Monitoring sedimentation processes by conductivity measurements. Int J Pharm 76:207–217, 1991.
148. M Bury, J Gerhards, W Erni, A Stamm. Application of a new method based on conductivity measurements to determine the creaming stability of O/W emulsions. Int Pharm 124(2):183–194, 1995.
149. KI Al-Malah , MOJ Azzam, RM Omari. Emulsifying properties of BSA in different vegetable oil emulsions using conductivity technique. Food Hydrocolloids 14:485–490, 2000.
150. F Babick, F Hinze, S Ripperger. Dependence of ultrasonic attenuation on the material properties. Colloids Surfaces A: Physicochem Eng Aspects 172:33–46, 2000.
151. WD Bachalo. Method for measuring the size and velocity of spheres by dual-beam light-scatter interferometry. Apply Opt 19:73–83, 1980.
152. DL Black, MQ McQuay, MP Bonin. Laser-based techniques for particle-size measurement: a review of sizing methods and their industrial applications. Prog Energy Combustion Sci 22(3):267–306, 1996.
153. E Dikinson, G Stainsby. *Colloidals in Food*. New York: Applied Sci Publ., 1982.
154. RK Chang, C Hsiao, JR Robinson. A review of aqueous coating techniques and preliminary data on release from a theophylline product. Pharm Technol 11(3):56–68, 1987.
155. KL Moore. Physicochemical properties of Opadry®, Coateric®, and Surelease®. In: JW Mcginity, ed. *Aqueous Polymeric Coatings for Pharmaceutical Dosage Forms*. New York: Marcel Dekker, 1989, pp 303–3l5.
156. R Gurny, NA Peppas, DD Harrington, GS Banker. Development of biodegradable and injectable latices for controlled release of potent drugs. Drug Dev Ind Pharm 7:1–25, 1981.
157. HJ Krause, A Schwarz, P Rohdewald. Polylactic acid nanoparticles, a colloidal drug delivery system for lipophilic drugs. Int J Pharm 27: 145–155, 1985.
158. F Koosha, RH Müller, SS Davis, MC Davies. The surface chemical structure of poly (β-hydroxybutyrate) microparticles produced by solvent evaporation process. J Control Rel 9:149, 1989.
159. R Bodmeier, H Chen. Indomethacin polymeric nanosuspensions prepared by microfluidization. J Control Rel 12:223–233, 1990.
160. PD Scholes, AGA Coombes, L Illum, SS Davis, M Vert, MC Davies. The preparation of sub-200 nm poly (lactide-co-glycolide) microspheres for site-specific drug delivery. J Control Rel 25:145–153, 1993.
161. S Batzri, ED Korn. Single bilayer liposomes prepared without sonication. Biochem Biophys Acta 443:629–634, 1973.
162. T Niwa, H Takeuchi, T Hino, N Kunou, Y Kawashima. Preparations of biodegradable nanospheres of water-soluble and insoluble drugs with d,l-lactide/glycolide copolymer by a novel spontaneous emulsification solvent diffusion method, and the drug release behavior. J Control Rel 25:89–98, 1993.
163. H Fessi, F Puisieux, JP Devissaguet. Procede de preparation de systemes colloidaux dispersibles d'une substance, sous forme de nanocapsules. European patent 274, 961 (1987).
164. N Ammoury, H Fessi, JP Devissaguet, M Allix, M Plotkine, RG Boulu. Effect on cerebral blood flow of orally administered indomethacin-loaded poly (isobutylcyanoacrylate) and poly(d,l-lactide) nanocapsules. J Pharm Pharmacol 42:558–561, 1990.
165. N Ammoury, H Fessi, JP Devissaguet, M Dubrasquet, S Benita. Jejunal absorption, pharmacological activity, and pharmacokinetic evaluation of indomethacin-loaded poly(d,l-lactide) and poly(-isobutyl-cyanoacrylate) nanocapsules in rats. Pharm Res 8:101–105, 1991.
166. N Ammoury, H Fessi, JP Devissaguet, F Puisieux, S Benita. In vitro release kinetic pattern of indomethacin from poly (d,l-lactide) nanocapsules. J Pharm Sci 79(9):763–767, 1990.
167. C Bindschaedler, R Gurny, E Doelker. Process for preparing a powder of water insoluble polymer which can be redispersed in a liquid phase, the resulting powder and utilization thereof. Swiss Patent 1497/88 (1988).
168. H Ibrahim, C Bindschaedler, E Doelker, P Buri, R Gurny. 1991. Concept and development of opthalmic pseudo-latices triggered by pH. Int J Pharm 77:211–219, 1991.
169. H Ibrahim, C Bindschaedler, E Doelker, P Buri, R Gurny. Aqueous nanodispersions prepared by a salting-out process. Int J Pharm 87:239–246, 1992.
170. E Allemann, R Gurny, E Doelker. Preparation of aqueous polymeric nanodispersions by a reversible salting-out process, influence of process parameters on particle size. Int J Pharm 87:247–253, 1992.
171. E Allemann, JC Leroux, R Gurny, E Doelker. In *vitro* extended-release properties of drug loaded poly (dl-lactic acid) nanoparticles produced by a salting-out procedure. Pharm Res 10(12):1732–1737, 1993.
172. E Allemann, E Doelker, R Gurny. Drug loaded poly (lactic acid) nanoparticles produced by a reversible, salting-out process: purification of an injectable dosage form. Eur J Pharm Biopharm 39:13–18, 1992.
173. B Ekman, I Sjöholm. Incorporation of macromolecules in microparticles: preparation and characteristics. Biochem 15:5115, 1976.
174. AFJr Yapel. Albumin microspheres: heat and chemical stabilization. In: KJ Widder, R Green, eds. *Methods in Enzymology:* Part A, *Drug and Enzyme Targeting*. Orlando: Academic Press, 1985, pp 3–18.

Copyright © 2002 Marcel Dekker, Inc.

175. WE Longo, EP Goldberg. Hydrophilic albumin microspheres. In: KJ Widder, R Green, eds. *Methods in Enzymology:* Part A, *Drug and Enzyme Targeting*. Orlando: Academic Press, 1985, pp 18–26.
176. E Tomlinson, JJ Burger. Incorporation of water-soluble drugs in albumin microspheres. In: KJ Widder, R Green, eds. *Methods in Enzymology:* Part A, *Drug and Enzyme Targeting*. Orlando: Academic Press, 1985, pp 27–43.
177. T Yoshioka, M Hashida, S Muranishi, H Sezaki. Specific delivery of mitomycin C to the liver, spleen, and lung: Nano-and microspherical carriers of gelatin. Int J Pharm 8:131–141, 1981.
178. EE Hassan, RC Parish, JM Gallo. Optimized formulation of magnetic microspheres containing the anticancer agent, oxantrazole. Pharm Res 9:390–397, 1992.
179. JJ Marty, RC Oppenheim, P Speiser. Nanoparticles-a new colloidal drug delivery system. Pharm Acta Helv 53:17–22, 1978.
180. M El-Samaligy, P Rohdewald. Triamcinolone diacetate nanoparticles, a sustained release drug delivery system suitable for parenteral administration. Pharm Acta Helv 57:201, 1982.
181. M El-Samaligy, P Rohdewald. Reconstituted collagen nanoparticles, a novel drug carrier delivery system. J Pharm Pharmacol 35:537–539, 1983.
182. RC Oppenheim. Solid colloidal drug delivery systems: nanoparticles. Int J Pharm 8:217–234, 1981.
183. M Rajanorivony, C Vauthier, G Couarraze, F Puisieux, P Couvreur. Development of a new drug carrier made from alginate. J Pharm Sci 82(9):912–917, 1993.
184. WR Perkins, I Ahmad, X Li, DJ Hirsh, GR Masters, CJ Fecko, JK Lee, S Ali, J Nguyen, J Schupsky, C Herbert, AS Janoff, E Mayhew. Novel therapeutic nano-particles (lipocores): trapping poorly water soluble compounds. Int J Pharm 200(1):27–39, 2000.
185. DF Driscoll, MN Bacon, BR Bistrian. Physicochemical stability of two types of intravenous lipid emulsion as total nutrient admixtures. J Parent Enteral Nutr 24(1):15–22, 2000.
186. HS Yoo, JE Oh, KH Lee, TG Park. Biodegradable nanoparticles containing doxorubicin-PLGA conjugate for sustained release. Pharm Res 16(7): 1114–1118, 1999.
187. RH Müller, K Mäder, S Gohla. Solid lipid nanoparticles (SLN) for controlled drug delivery-A review of the state of the art. Eur J Pharm Biopharm 50(1): 161–177, 2000.
188. S Harnisch, RH Müller. Adsorption kinetics of plasma proteins on oil-in-water emulsions for parenteral nutrition. Eur J Pharm Biopharm 49(1):41–46, 2000.
189. PP Constantinides, KJ Lambert, AK Tustian, B Schneider, S Lalji, W Ma, B Wentzel, D Kessler, D Worah, SC Quay. Formulation development and antitumor activity of a filter-sterilizable emulsion of paclitaxel. Pharm Res 17(2):175–182, 2000.
190. AG Floyd. Top ten considerations in the development of parenteral emulsions. Pharm Sci Tech Today 2(4):134–143, 1999.
191. VS Koster, PFM Kuks, R Langer, H Talsma. Particle size in parenteral fat emulsions, what are the true limitations? Int J Pharm 134:235–238, 1996.
192. G Chansiri, RT Lyons, MV Patel, SL Hem. Effect of surface charge on the stability of oil/water emulsions during steam sterilization. J Pharm Sci 88(4):453–458, 1999.
193. V Labhasetwar, C Song, W Humphrey, R Shebuski, RJ Levy. Arterial uptake of biodegradable nanoparticles: effect of surface modifications. J Pharm Sci 87(10): 1229–1234, 1998.
194. M Jumaa, BW Müller. The effect of oil components and homogenization conditions on the physicochemical properties and stability of parenteral fat emulsions. Int J Pharm 163(1–2):81–89, 1998.
195. Y Kawano, T Noma. Inhibition by lecithin-bound iodine (LBI) of inducible allergen-specific T lymphocytes' responses in allergic diseases. Int J Immunopharmacol 18(4):241–249, 1996.
196. GC Na, HJ Stevens, BO Yuan, N Rajagopalan. Physical stability of ethyl diatrizoate nanocrystalline suspension in steam sterilization. Pharm Res 16(4):569–574, 1999.
197. K Westesen, B Siekmann. Investigation of the gel formation of phospholipid-stabilized solid lipid nanoparticles. Int J Pharm 151(1):35–45, 1997.
198. J Mewis. Flow behaviour of concentrated suspensions: predictions and measurements. Int J Miner Proc 44–45:17–27, 1996.
199. AJ Spiegel, MM Noseworthy. Use of nonaqueous solvents in parenteral products. J Pharm Sci 52(10):917–927, 1963.
200. RH Müller, K Peters. Nanosuspensions for the formulation of poorly soluble drugs I.Preparation by a size-reduction technique. Int J Pharm 160(2):229–237, 1998.
201. K Peters, S Leitzke, SE Diederichs, K Borner, H Hahn, RE Müller, S Ehlers. Preparation of a clofazimine nanosuspension for intravenous use and evaluation of its therapeutic efficacy in murine Mycobacterium avium infection. J Antimicrobial Chemother 45(1):77–83, 2000.
202. PC Schmidt. Secondary electron microscopy in pharmaceutical technology, In: J Swarbrick, JC Boylan, eds. *Encyclopedia of Pharmaceutical Technology*, Vol. 19. New York: Marcel Dekker, 2000, pp 311–356.
203. JW Mullin. Sieving of pharmaceuticals. In: J Swarbrick, JC Boylan, eds. *Encyclopedia of Pharmaceutical Technology*, Vol. 14. New York: Marcel Dekker, 1996, pp 63–86.
204. TJ Racey, P Rochon, DVC Awang, GA Neville. Aggregation of commercial heparin samples in storage. J Pharm Sci 76: 314–318, 1987.

Copyright © 2002 Marcel Dekker, Inc.

205. TJ Racey, P Rochon, F Mori, GA Neville. Examination of a possible role of dermatan sulfate in the aggregation of commercial heparin samples. J Pharm Sci 78:214–218, 1989.
206. JB Kayes. Pharmaceutical suspensions: microelectrophoretic properties. J Pharm Pharmacol 29:163–168, 1977.
207. V Gallardo, L Zurita, A Ontiveros, JDG Durán. Interfacial properties of barium sulfate suspensions. Implications in their stability. J Pharm Sci 89:1134–1142, 2000.
208. K Thode, RH Müller, M Kresse. Two-time window and multiangle photon correlation spectroscopy size and zeta potential analysis—highly sensitive rapid assay for dispersion stability. J Pharm Sci 89:1317–1324, 2000.
209. DJA Crommelin. Influence of lipid composition and ionic strength on the physical stability of liposomes. J Pharm Sci 73:1559–1563, 1984.
210. H Takenaka, Y Kawashima, SY Lin. Electrophoretic properties of sulfamethoxazole microcapsules and gelatin-acacia coacervates. J Pharm Sci 70:302–305, 1981.
211. MJ Killeen. Spray drying and spray congealing of pharmaceuticals. In: J Swarbrick, JC Boylan, eds. *Encyclopedia of Pharmaceutical Technology*, Vol. 14. New York: Marcel Dekker, 1996, pp. 207–222.
212. LH Block. Scale-up of disperse systems: Theoretical and practical aspects. In: HA Lieberman, MM Rieger, GS Banker, eds. *Pharmaceutical Dosage Forms: Disperse Systems*, Vol. 3, 2nd ed. New York: Marcel Dekker, 1998, pp 363–394.
213. RA Nash. Validation of disperse systems. In: HA Lieberman, MM Rieger, GS Banker, eds. *Pharmaceutical Dosage Forms: Disperse Systems*, Vol. 3, 2nd ed. New York: Marcel Dekker, 1998, pp 473–512.

Copyright © 2002 Marcel Dekker, Inc.