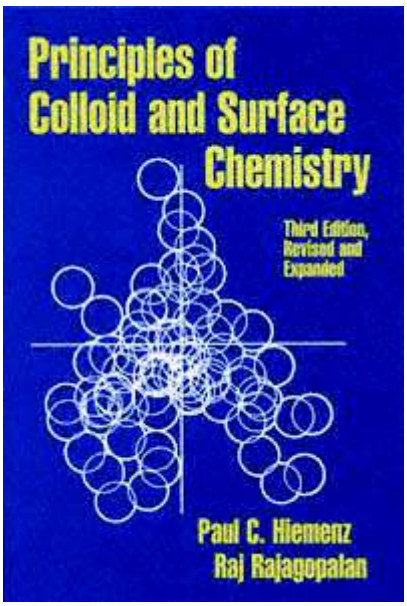

Principles of
Colloid and Surface
Chemistry
Third Edition,
Revised and
Expanded
Paul C. Hiemenz
Raj Rajagopalan

Principles of Colloid and Surface Chemistry

Third Edition, Revised and Expanded

Paul C. Hiemenz
California State Polytechnic University
Pomona, California

Raj Rajagopalan
University of Florida
Gainesville, Florida



MARCEL DEKKER, INC. NEW YORK • BASEL • HONG KONG

Library of Congress Cataloging-in-Publication Data

Hiemenz, Paul C.
Principles of colloid and surface chemistry. — 3rd ed., rev. and expanded / Paul C. Hiemenz, Raj Rajagopalan.
p. cm.
Includes bibliographical references and index.
ISBN 0-8247-9397-8 (hardcover : alk. paper)
1. Colloids. 2. Surface chemistry. I. Rajagopalan, Raj.
II. Title.
QD549.H53 1997
541.3'45—dc21

97-4015
CIP

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

This book is printed on acid-free paper.

Copyright © 1997 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.

Marcel Dekker, Inc.
270 Madison Avenue, New York, New York 10016

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

PRINTED IN THE UNITED STATES OF AMERICA

Preface to the Third Edition

The face of colloid and surface chemistry has changed dramatically in the 10 years since the last edition of this book appeared in print. Advances in instrumentation now make it possible for us to "see"—and, indeed, to *manipulate*—individual atoms on a surface. Molecular engineering of polymers, surfactants, and particles is now within reach for fabricating novel advanced materials and for preparing capsules (liposomes) for drug delivery and gene therapy. Direct measurements of colloidal and surface forces are now commonplace, and colloidal dispersions are now used as models for studying equilibrium and non-equilibrium thermodynamics and flow behavior of *atomic* (as well as *geological*) systems—phenomena that are otherwise not easily accessible to the experimentalist. Colloid and surface chemistry is a truly interdisciplinary subject today, and its content and significance have no bounds.

This richness has its rewards, as has been documented in a number of advanced monographs and a few graduate-level books in recent years. However, it has its drawbacks as well, as those who use colloids and surface chemistry are drawn from increasingly diverse backgrounds. Introductory textbooks that are suitable for such a diverse group of students and professionals remain almost as rare today as they were a decade ago. This need for an introductory book accessible to everyone, regardless of background, even from the perspectives of the traditional applications of colloid and surface chemistry, cannot be overemphasized. For example, in the chemical industries alone, almost two-thirds of the operations or products involve powders and suspensions. A report from Du Pont researchers [Ennis, B., Green, J., and Davies, R., *Chemical Engineering Progress*, **90**(4), 32–43 (1994)] notes that of the roughly 3000 products manufactured by Du Pont, about 60% are marketed as powders, dispersions, or suspensions and that the processing of another 20% involves particles or suspensions. Industries employ engineers and scientists of all types, and the rather tight curriculum requirements in most areas of science and engineering seldom allow even a minimal preparation of university graduates for working with materials in particulate form.

Factors such as the above have shaped our approach to the present edition. The philosophy behind the first two editions was that most of the material in the book should be a natural extension of an introductory physical chemistry course at the undergraduate level and should be accessible to anyone (students as well as practicing engineers and scientists) with such a background. This remains the rule in the present edition as well. In our opinion, this orientation is essential for giving a firm foundation in colloids and surfaces to the diverse audience, which typically includes individuals from almost all branches of engineering and the sciences.

Keeping these factors in mind, we have updated many topics and have introduced new features that draw attention to exciting developments that go beyond the scope of the present volume. Among the major revisions and additions made are the following:

1. A major portion of Chapter 1 has been rewritten. In addition to providing a broad overview of the basic issues of importance in colloids and to introducing much of

what is presented in the rest of the book, Chapter 1 now includes modern developments in instrumentation such as scanning probe microscopy and spectroscopy, and atomic and surface force measurements; a brief introduction to "model" colloids and their uses; and an introduction to fractal dimensions of aggregates and surfaces.

2. A number of vignettes are sprinkled throughout Chapter 1 to provide motivation for and to illustrate the concepts introduced in the chapter and in the rest of the book. These vignettes also serve as vehicles for conveying the excitement of the subject and as introductions to advanced concepts that are beyond the scope of the book. They cover a rather broad array of disciplines ranging from environmental remediation and separation technologies to biological and medical sciences and polymer composites, and they thus reveal the breadth of the present as well as the potential applications. In addition to the vignettes in Chapter 1, each of the other 12 chapters contains a vignette highlighting the material covered in the chapter and establishing the connections to other chapters or advanced developments.
3. Chapter 4 now goes beyond an introduction to Newtonian viscosity and flow of Newtonian dispersions. In addition to a basic introduction to viscosity and the Navier-Stokes equation, the chapter now includes the Krieger-Dougherty relation for concentrated dispersions, elements of non-Newtonian rheological behavior of dispersions, electroviscous effects, and a summary of the definitions and physical significance of dimensionless groups essential for estimating and quantifying rheological properties of dispersions.
4. Chapter 5, on scattering techniques, has been updated to include measurement of fractal dimensions of aggregates, intraparticle and interparticle structure, dynamic light scattering and relation between light scattering and complementary measurements based on scattering of X-ray and neutrons.
5. Chapter 7 now includes an introduction to structural transitions in Langmuir layers and examples of applications of Langmuir and Langmuir–Blodgett films.
6. Chapter 8 has been revised to include a discussion of the critical packing parameter of surfactants and its relation to the structure of resulting surfactant aggregates. This simple geometric basis for the formation of micelles, bilayers, and other structures is intuitively easier to understand for a beginning student.
7. Chapters 9 and 10 of the second edition (dealing with physical adsorption at gas/solid interfaces and microscopy, spectroscopy and diffractometry of metal surfaces, respectively) have been pared down and consolidated into one chapter (new Chapter 9, on adsorption on gas/solid interfaces). The outdated materials or materials extraneous to the main thrust of the book from the old Chapter 10 have been eliminated and only the discussion of low-energy electron diffraction (LEED) has been retained.
8. Chapters 11–13 of the second edition, which discussed van der Waals forces (old Chapter 11), electrical double layers (old Chapter 12), electrokinetic phenomena (old Chapter 13), and colloid stability (old Chapters 11 and 12), have been restructured and new materials on colloid stability and polymer/colloid interactions have been added. For example:
   - In the new version, Chapter 10 focuses exclusively on van der Waals forces and their implications for macroscopic phenomena and properties (e.g., structure of materials and surface tension). It also includes new tables and examples and some additional methods for estimating Hamaker constants from macroscopic properties or concepts such as surface tension, the parameters of the van der Waals equation of state, and the corresponding state principle.
   - Chapters 11 and 12 in the present edition focus exclusively on the theories of electrical double layers and forces due to double-layer interactions (Chapter 11) and electrokinetic phenomena (Chapter 12). Chapter 11 includes expressions for interacting spherical double layers, and both chapters provide additional examples of applications of the concepts covered.

- The new Chapter 13 collects the material on colloid stability previously distributed between old Chapters 11 and 12 and integrates it with new material on stability ratio and slow and fast coagulation, polymer-induced forces, and polymer-induced stabilization and destabilization.

This rearrangement makes the presentation of colloidal forces, colloid stability, and electrokinetic phenomena more logical and pedagogically more appealing.

Numerous cross-references link the materials in the various chapters; however, many of the chapters are self-contained and can stand alone, thus offering the instructor the flexibility to mix and match the topics needed in a particular course. Moreover, we have also introduced a number of other features to make the present edition more useful and convenient as a textbook *and* for self-study. For example:

1. A collection of Review Questions has been added at the end of each chapter. These should be useful not only for class assignments but also as a self-assessment tool for students.
2. The references at the end of each chapter have been grouped into two sets, one containing annotated general references (usually other textbooks, monographs or review articles) and the second with special references of narrower scope. The level of each annotated reference (undergraduate, graduate, or advanced) has also been included in the annotations, as an additional aid to students and instructors.
3. All illustrations have been redrawn, and a number of new illustrations, tables, and examples have been added. (Appendix D presents a list of all the worked-out examples and the corresponding page numbers.)
4. The presentation in each chapter has been divided into a number of sections and subsections so that the headings of the sections and subsections can serve as guideposts for students.

As it is often said, when it comes to textbooks, one never "finishes" writing a book – one just "abandons" it! There is a lot we would like to do further with this book, but these changes must wait for the next edition. We would be pleased indeed to hear comments, suggestions, and recommendations for potential revisions and additions from anyone who uses this book, so that everyone can benefit from the collective wisdom of the community at large.

Paul C. Hiemenz
Raj Rajagopalan

This page intentionally left blank

Acknowledgments

A number of colleagues have been kind enough to send us comments or recent research papers, reviews, or related materials for the third edition. The information we received has shaped our thinking and has influenced the contents and the organization of the present edition, and we would like to recognize the assistance and interest of these individuals here. In particular, we would like to thank Prof. Nicholas Abbott (UC, Davis), Prof. John Anderson (Carnegie Mellon), Dr. Rafat R. Ansari (NASA Lewis, Cleveland, OH), Prof. John C. Berg (Univ. of Washington), Prof. Daniel Blankschtein (MIT), Dr. Krishnan Chari (Kodak, Rochester, NY), Dr. Reg Davies (Du Pont, Wilmington, DE), Prof. Menachem Elimelech (UCLA), Prof. H. Scott Fogler (Michigan), Prof. George Hirasaki (Rice), Dr. Norio Ise (Rengo Corp., Fukui, Japan), Prof. Jacob N. Israelachvili (UC, Santa Barbara), Prof. Charles Knobler (UCLA), Prof. Josip Kratohvil (Clarkson), Prof. J. Adin Mann (Case Western), Prof. Jacob Masliyah (Alberta), Prof. Egon Matijevic (Clarkson), Dr. S. Mehta (ARCO, Plano, TX), Prof. Clarence Miller (Rice), Dr. Ian Morrison (Xerox, Rochester, NY), Prof. Brij M. Moudgil (Florida), Dr. Cherry A. Murray (AT&T, Murray Hill, NJ), Prof. R. Nagarajan (Penn State), Prof. Robert Ofoli (Michigan State), Prof. Kyriakos Papadopoulos (Tulane), Prof. Dennis C. Prieve (Carnegie Mellon), Prof. Clay Radke (UC, Berkeley), Dr. Mihail C. Roco (NSF), Prof. Robert Rowell (Univ. of Massachusetts), Prof. William B. Russel (Princeton), Prof. Robert Schechter (UT, Austin), Prof. Ken Schmitz (Univ. of Missouri, Kansas City), Prof. Dinesh Shah (Florida), Prof. P. Somasundaran (Columbia), Prof. Frank M. Tiller (Houston), Prof. Theo van de Ven (McGill), Prof. Carol Van Oss (SUNY, Buffalo), Prof. Darsh T. Wasan (IIT, Chicago), and Prof. Mark Wiesner (Rice).

Some of RR's colleagues at the University of Houston, especially Profs. Kishore K. Mohanty, Richard Pollard, Jay D. Schieber, Cindy L. Stokes, Frank M. Tiller, and Richard C. Willson, read and offered comments on some of the vignettes used in the book. Their interest and assistance have contributed significantly to the clarity of the final versions that appear in the book.

RR would also like to thank his students and coworkers, who played a major role throughout the preparation of the present edition. Mr. Roger Seow did much of the laborious task of scanning the second edition into the computer and fixing all the errors generated by the usually less-than-perfect optical character recognition programs. His remarkable attention to details and organizational skills made the initial stages of the preparation of the text appear almost effortless. Mr. Nikhil Joshi stepped in subsequently to attend to the endless printing and reprinting of the figures, tables, and text for the final stage. We are greatly indebted to both. Dr. Jean François Guérin, Dr. Yongmei Wang, Ms. Carlisa Harris, Ms. Sheryl L. Horton, Mr. Randy Diermeier, Mr. Jorge Jimenez, Mr. Dang Nhan, Mr. K. Srinivasa Rao, Mr. Antonio Rodriguez, and Mr. Sameer Talsania have been extremely patient despite the fact that their research projects frequently took second place to the preparation of the book.

The Media Center at the University of Houston deserves special recognition for its excellent work in preparing the initial versions of the drawings of most of the illustrations.

Finally, we would like to thank the folks at Marcel Dekker, Inc.—particularly Ted Allen and Eric F. Stannard, Production Editors, and Sandy Beberman, Assistant Vice President—for their infinite patience and, of course, for their characteristically friendly and courteous assistance.

This page intentionally left blank

Preface to the Second Edition

In the preface to the first edition, I stated that this is "primarily a textbook, written with student backgrounds, needs, and objectives in mind." This continues to be true of the second edition, and many of the revisions I have made are attempts to make the book even more useful than its predecessor to student readers. In addition, colloid and surface science continue to flourish. In preparing the second edition I have also attempted to "open up" the coverage to include some of the newer topics from an ever-broadening field.

A number of differences between the first and second editions can be cited which are readily traceable to either one or both of the foregoing considerations.

Two new chapters have been added which explore—via micelles and related structures and metal surfaces under ultra-high vacuum—both "wet" and "dry" facets of colloids and surfaces. Although neither of these areas is new, both are experiencing an upsurge of activity as new instrumentation is developed and new applications are found.

A number of chapters have been overhauled so thoroughly that they bear only minor resemblance to their counterparts in the first edition. The thermodynamics of polymer solutions is introduced in connection with osmometry and the drainage and spatial extension of polymer coils is discussed in connection with viscosity. The treatment of contact angle is expanded so that it is presented on a more equal footing with surface tension in the presentation of liquid surfaces. Steric stabilization as a protective mechanism against flocculation is discussed along with the classical DLVO theory.

Solved problems have been scattered throughout the text. I am convinced that students must work problems to gain mastery of the topics we discuss. Including examples helps bridge the gap between the textbook presentation of theory and student labors over the analysis and mechanics of problem solving.

SI units have been used fairly consistently throughout the book. Since the problems at the ends of the chapters are based on data from the literature and since cgs units were commonly used in the older literature, the problems contain a wider assortment of units. A table of conversions between cgs and SI units is contained in Appendix B.

I am very much aware of the many important topics that the book fails to cover or, worse yet, even mention. However, lines must be drawn somewhere both to keep the book manageable in size and cost and to have it useful as the basis for a course. As it is, I have added a good deal of new material without deleting anywhere near as much of the old. I have tried to select for inclusion topics of fundamental importance which could be developed with some internal coherence and with some continuity from a (prerequisite) physical chemistry course.

A number of users of the first edition communicated with me, pointing out errors and offering suggestions for improvement. I appreciate the feedback of these correspondents, and hope that the revisions I have made at least partially reflect their input.

I want to express my thanks to Carol Truett for expertly drawing the new illustrations and giving a "new look" to the old ones. I also appreciate the assistance of Lisa Scott in the preparation of the manuscript. Thanks, also, to Reuben Martinez for helping me with the proofreading and indexing. Lastly, let me again invite users to call errors and/or obscurities to my attention, and to thank them in advance for doing so.

Paul C. Hiemenz

This page intentionally left blank

Preface to the First Edition

Colloid and surface chemistry occupy a paradoxical position among the topics of physical chemistry. These are areas which have traditionally been considered part of physical chemistry and are currently enjoying more widespread application than ever due to their relevance to environmental and biological problems. At the same time, however, colloid and surface chemistry have virtually disappeared from physical chemistry courses. These topics are largely absent from the contemporary general chemistry course as well. It is possible, therefore, that a student could complete a degree in chemistry without even being able to identify what colloid and surface chemistry are about.

The primary objective of this book is to bridge the gap between today's typical physical chemistry course and the literature of colloid and surface chemistry. The reader is assumed to have completed a course in physical chemistry, but no prior knowledge of the topics under consideration is assumed. The book is, therefore, introductory as far as the topic subjects are concerned, although familiarity with numerous other aspects of physical chemistry is required background.

Since physical chemistry is the point of departure for this presentation, the undergraduate chemistry major is the model reader toward whom the book is addressed. This in no way implies that these are the only students who will study the material contained herein. Students majoring in engineering, biology, physics, materials science, and so on, at both the undergraduate and graduate levels will find aspects of this subject highly useful. The interdisciplinary nature of colloid and surface chemistry is another aspect of these subjects that contributes to their relevance in today's curricula.

This is primarily a textbook, written with student backgrounds, needs, and objectives in mind. There are several ways in which this fact manifests itself in the organization of this book. First, no attempt has been made to review the literature or to describe research frontiers in colloid and surface chemistry. A large literature exists which does these things admirably. Our purpose is to provide the beginner with enough background to make intelligible the journals and monographs which present these topics. References have been limited to monographs, textbooks, and reviews which are especially comprehensive and/or accessible. Second, where derivations are presented, this is done in sufficient detail so that the reader should find them self-explanatory. In areas in which undergraduate chemistry majors have minimal backgrounds or have chronic difficulties—for example, fluid mechanics, classical electromagnetic theory, and electrostatics—the presentations begin at the level of general physics, which may be the student's only prior contact with these topics. Third, an effort has been made to facilitate calculations by paying special attention to dimensional considerations. The cgs-esu system of units has been used throughout, even though this is gradually being phased out of most books. The reason for keeping these units is the stated objective of relating the student's experiences to the existing literature of colloid and surface chemistry. At present, the cgs-esu system is still the common denominator between the two. A fairly detailed list of conversions between cgs and SI units is included in Appendix C. Finally, a few problems are included in

each chapter. These provide an opportunity to apply the concepts of the chapter and indicate the kinds of applications these ideas find.

Not all who use the book will have the time or interest to cover it entirely. In the author's course, about two-thirds of the material is discussed in a one-quarter course. With the same level of coverage, the entire book could be completed in a semester. To cover the amount of material involved, very little time is devoted to derivations except to answer questions. Lecture time is devoted instead to outlining highlights of the material and presenting supplementary examples.

The underlying unity which connects the various topics discussed here is seen most clearly when the book is studied in its entirety and in the order presented. Time limitations and special interests often interfere with this ideal. Those who choose to rearrange the sequence of topics should note the subthemes that unify certain blocks of chapters. Chapters 1 through 5 are primarily concerned with particle characterization, especially with respect to molecular weight; Chapters 6 through 8, with surface tension/free energy and adsorption; and Chapters 9 through 11, with flocculation and the electrical double layer. Subjects of special interest to students of the biological sciences are given in Chapters 2 to 5, 7, and 11.

Colloid chemistry and surface chemistry each span virtually the entire field of chemistry. The former may be visualized as a chemistry whose "atoms" are considerably larger than actual atoms; the latter, as a two-dimensional chemistry. The point is that each encompasses all the usual subdivisions of chemistry: reaction chemistry, analytical chemistry, physical chemistry, and so on. The various subdivisions of physical chemistry are also represented: thermodynamics, structure elucidation, rate processes, and so on. As a consequence, these traditional categories could be used as the basis for organization in a book of this sort. For example, "The Thermodynamics of Surfaces" would be a logical chapter heading according to such a plan of organization. In this book, however, no such chapter exists (although not only chapters but entire volumes on this topic exist elsewhere). The reason goes back to the premise stated earlier: These days most undergraduates know more about thermodynamics than about surfaces, and this is probably true regardless of their thermodynamic literacy/illiteracy! Accordingly, this book discusses surfaces: flat and curved, rigid and mobile, pure substances and solutions, condensed phases and gases. Thermodynamic arguments are presented—along with arguments derived from other sources—in developing an overview of surface chemistry (with the emphasis on "surface"). A more systematic, formal presentation of surface thermodynamics (with the emphasis on "thermodynamics") would be a likely sequel to the study of this book for those who desire still more insight into that aspect of two-dimensional chemistry. Similarly, other topics could be organized differently as well. Only time will tell whether the plan followed in this book succeeds in convincing students that chemistry they have learned in other courses is also applicable to the "in between" dimensions of colloids and the two dimensions of surface chemistry.

The notion that molecules at a surface are in a two-dimensional state of matter is reminiscent of E. A. Abbott's science fiction classic, *Flatland.** Perusal of this book for quotations suitable for Chapters 6, 7, and 8 revealed other parallels also: the color revolt and light scattering, "Attend to Your Configuration" and the shape of polymer molecules, and so on. Eventually, the objective of beginning each chapter with a quote from *Flatland* replaced the requirement that the passage cited have some actual connection with the contents of the chapter. As it ends up, the quotes are merely for fun: Perhaps those who are not captivated by colloids and surfaces will at least enjoy this glimpse of *Flatland.*

Finally, it is a pleasure to acknowledge those whose contributions helped bring this book into existence. I am grateful to Maurits and Marcel Dekker for the confidence they showed and the encouragement they gave throughout the entire project. I wish to thank Phyllis Bartosh, Felecia Granderson, Jennifer Woodruff, and, especially, Mickie McConnell and Lynda Parzick for making my sloppy manuscript presentable. My appreciation also goes to Bob Marvos, George Phillips, and, especially, Dottie Holmquist for their work on the figures,

*E. A. Abbott, *Flatland* (6th ed.), Dover, New York, 1952. Used with permission.

which are such an important part of any textbook. I also wish to thank Michael Goett for helping with proofreading and indexing. Finally, due to the diligence of the class on whom this material was tested in manuscript form, the book has 395 fewer errors than when it started. For the errors that remain, and I hope they are few in number and minor in magnitude, I am responsible. Reports from readers of errors and/or obscurities will be very much appreciated.

Paul C. Hiemenz

This page intentionally left blank

Contents

Preface to the Third Edition *iii*

Acknowledgments *vii*

Preface to the Second Edition *ix*

Preface to the First Edition *xi*

1. Colloid and Surface Chemistry: Scope and Variables **1**

1.1 Introduction 1
1.2 The Importance of the Surface for Small Particles 6
1.3 Classification of Colloids Based on Affinity to Carrier Fluid 10
1.4 Concept of Stability of Colloidal Systems 14
1.5 Some Physical Characteristics of Colloids 19
1.6 Some Classical and Emerging Experimental Tools 38
1.7 An Overview of the Book 56
Review Questions 57
References 57
Problems 59

2. Sedimentation and Diffusion and Their Equilibrium **62**

2.1 Introduction 62
2.2 Sedimentation: Some Basic Considerations 65
2.3 Gravitational Sedimentation 67
2.4 Centrifugal Sedimentation 74
2.5 Diffusion 78
2.6 Brownian Motion and Diffusion 85
2.7 The Random Coil and Random Walk Statistics 94
2.8 Equilibrium Between Sedimentation and Diffusion 98
Review Questions 101
References 101
Problems 102

3. Solution Thermodynamics: Osmotic and Donnan Equilibria **105**

3.1 Introduction 105
3.2 Osmotic Pressure: Thermodynamic Foundations 108
3.3 Osmometry: Some Applications 114
3.4 Statistical Foundations of Solution Thermodynamics 120
3.5 Osmotic Equilibrium in Charged Systems 132
3.6 Some Applications of Osmotic Phenomena 139

Review Questions 140
References 141
Problems 142

4. The Rheology of Dispersions 145

4.1 Introduction 145
4.2 Newton's Law of Viscosity 148
4.3 Concentric-Cylinder and Cone-and-Plate Viscometers 150
4.4 The Poiseuille Equation and Capillary Viscometers 154
4.5 The Equation of Motion: The Navier-Stokes Equation 158
4.6 Einstein's Theory of Viscosity of Dispersions 161
4.7 Deviations from the Einstein Model 168
4.8 Non-Newtonian Behavior 174
4.9 Viscosity of Polymer Solutions 181
Review Questions 188
References 188
Problems 189

5. Static and Dynamic Light Scattering and Other Radiation Scattering 193

5.1 Introduction 193
5.2 Interaction of Radiation with Matter 196
5.3 Scattering by Small Particles: Theory of Rayleigh Scattering 202
5.4 Experimental Aspects of Light Scattering 207
5.5 Extension to Larger Particles and to Intraparticle Interference Effects 213
5.6 Interference Effects and Structure of Particles 223
5.7 Scattering by Large, Absorbing Particles 229
5.8 Dynamic Light Scattering 236
Review Questions 242
References 243
Problems 244

6. Surface Tension and Contact Angle: Application to Pure Substances 248

6.1 Introduction 248
6.2 Surface Tension and Contact Angle: A First Look 251
6.3 Thermodynamics of Surfaces: Surface Tension as Surface Free Energy 255
6.4 Surface Tension: Implications for Curved Interfaces and Capillarity 257
6.5 Effects of Curved Interfaces on Phase Equilibria and Nucleation: The Kelvin Equation 261
6.6 Surface Tension and Contact Angle: Their Relation to Wetting and Spreading Phenomena 265
6.7 Contact Angles: Some Complications 272
6.8 Measuring Surface Tension and Contact Angle: Round Two 276
6.9 Contact of Liquids with Porous Solids and Powders 283
6.10 Molecular Interpretation of Surface Tension 286
Review Questions 291
References 293
Problems 297

7. Adsorption from Solution and Monolayer Formation 297

7.1 Introduction 297
7.2 Insoluble Monolayers: Spreading of Surfactants on Aqueous Surfaces 300

7.3	Experimental Measurement of Film Pressure	304
7.4	Results of Film Balance Studies	308
7.5	Viscous Behavior of Two-Dimensional Phases	318
7.6	Applications of Monolayers and Monolayer Concepts	320
7.7	Adsorption from Solution: Thermodynamics	323
7.8	The Gibbs Equation: Experimental Results	327
7.9	Adsorption on Solid Surfaces	331
7.10	Applications of Adsorption from Solution	338
7.11	Adsorption in the Presence of an Applied Potential	343
	Review Questions	348
	References	348
	Problems	349

8. Colloidal Structures in Surfactant Solutions: Association Colloids 355

8.1	Introduction	355
8.2	Surfactants in Solution: Experimental Observations and Models	357
8.3	Structure of Micelles	362
8.4	Molecular Architecture of Surfactants, Packing Considerations, and Shapes of Micelles	367
8.5	Critical Micelle Concentration and the Thermodynamics of Micellization	370
8.6	Solubilization	375
8.7	Catalysis by Micelles	380
8.8	Reverse Micelles	386
8.9	Emulsions and Microemulsions	389
8.10	Some Applications of Microemulsions	392
8.11	Biological Membranes	395
	Review Questions	398
	References	399
	Problems	399

9. Adsorption at Gas–Solid Interfaces 405

9.1	Introduction	405
9.2	Experimental and Theoretical Treatments of Adsorption: An Overview	409
9.3	Thermodynamics of Adsorption: Phenomenological Perspective	413
9.4	Thermodynamics of Adsorption: A Statistical Perspective	419
9.5	Multilayer Adsorption: The Brunauer-Emmett-Teller Equation	425
9.6	Energetics of Adsorption	433
9.7	Adsorption in Porous Solids	436
9.8	Adsorption on Crystal Surfaces	439
9.9	Metal Surfaces and Heterogeneous Catalysis	451
	Review Questions	455
	References	456
	Problems	456

10. van der Waals Forces 462

10.1	Introduction	462
10.2	van der Waals Forces and Their Importance in Colloid and Surface Chemistry	464
10.3	Molecular Interactions and Power Laws	467
10.4	Molecular Origins and the Macroscopic Implications of van der Waals Forces	471

10.5 van der Waals Forces Between Large Particles and Over Large Distances 479
10.6 Calculating van der Waals Forces Between Macroscopic Bodies 483
10.7 Theories of van der Waals Forces Based on Bulk Properties 486
10.8 Effect of the Medium on the van der Waals Attraction 490
Review Questions 495
References 495
Problems 496

11. The Electrical Double Layer and Double-Layer Interactions 499

11.1 Introduction 499
11.2 Surface Charges and Electrical Double Layer: Background 502
11.3 The Capacitor Model of the Double Layer 504
11.4 The Diffuse Double Layer: The Debye-Hückel Approximation 508
11.5 The Debye-Hückel Approximation: Results 512
11.6 The Electrical Double Layer: Gouy-Chapman Theory 516
11.7 Overlapping Double Layers and Interparticle Repulsion 520
11.8 "Not-Quite-Indifferent" Electrolytes: Stern Adsorption 527
Review Questions 530
References 530
Problems 531

12. Electrophoresis and Other Electrokinetic Phenomena 534

12.1 Introduction 534
12.2 Mobilities of Small Ions and Macroions in Electric Fields: A Comparison 536
12.3 Zeta Potential: Thick Electrical Double Layers 538
12.4 Zeta Potential: Thin Electrical Double Layers 544
12.5 Zeta Potential: General Theory for Spherical Particles 546
12.6 Electroosmosis 550
12.7 Streaming Potential 553
12.8 The Surface of Shear and Viscoelectric Effect 555
12.9 Experimental Aspects of Electrophoresis 559
12.10 Determining the Surface Charge from Electrokinetic Measurements 565
12.11 Applications of Electrokinetic Phenomena 567
Review Questions 570
References 570
Problems 571

13. Electrostatic and Polymer-Induced Colloid Stability 575

13.1 Introduction 575
13.2 Interparticle Forces and the Structure and Stability of Dispersions 578
13.3 The Derjaguin-Landau-Verwey-Overbeek Theory of Colloid Stability 585
13.4 Theory of Coagulation in Dilute Dispersions 592
13.5 Polymer-Colloid Mixtures: A Phenomenological Perspective 604
13.6 Polymer-Induced Forces 610
13.7 Steric Stabilization: A Model for the Initial Encounter 614
Review Questions 619
References 620
Problems 621

Appendix A: Examples of Expansions Encountered in This Book *625*
Reference 625
Appendix B: Units: CGS-SI Interconversions *626*
References 626
Appendix C: Statistics of Discrete and Continuous Distributions of Data *631*

C.1 Introduction 631
C.2 Discrete Distributions 631
C.3 Theoretical Distribution Functions 634
References 637

Appendix D: List of Worked-Out Examples *638*

Index *641*

This page intentionally left blank

1

Colloid and Surface Chemistry:

Scope and Variables

Next . . . come the Nobility, of whom there are several degrees, beginning at Six-Sided Figures, . . . and from thence rising in the number of their sides till they receive the honorable title of Polygonal. . . . Finally when the number of sides becomes so numerous, and the sides themselves so small, that the figure cannot be distinguished from a circle, he is included in . . . the highest class of all.

From Abbott's *Flatland*

1.1 INTRODUCTION

"Yesterday, I couldn't define colloid chemistry; today, I'm doing it." This variation of an old quip could apply to many recent chemistry and engineering graduates on entering employment in the "real world." Two facts underlie this situation. First, colloid and surface science, although traditional parts of physical chemistry, have largely disappeared from introductory physical chemistry courses. Second, in research, technology, and manufacturing, countless problems are encountered that fall squarely within the purview of colloid and surface science. In this section we enumerate some examples that illustrate this statement. The nine "vignettes" included in this chapter also illustrate the importance of colloid and surface science in a broad range of scientific and technological areas.

The paradoxical situation just described means that it is entirely possible for a science or an engineering student to have completed a course in physical chemistry and still not have any clear idea of what colloid and surface science are about. A book like this one is therefore in the curious position of being simultaneously "advanced" and "introductory." Our discussions are often advanced in the sense of building on topics from physical chemistry. At the same time, we have to describe the phenomena under consideration pretty much from scratch since they are largely unfamiliar. In keeping with this, this chapter is concerned primarily with a broad description of the scope of colloid and surface science and the kinds of variables with which they deal. In subsequent chapters different specific phenomena are developed in detail.

1.1a Colloid and Surface Chemistry: Some Definitions

1.1a.1 Definition of Colloids

Our first tasks are to define what we mean by colloid science and how this is related to surface science. For our purposes, any particle that has some linear dimension between 10^{-9} m (10 Å) and 10^{-6} m (1 μm or 1 μ)* is considered a colloid. For us, linear dimensions rather than

*In this book SI (International System of Units) units are used fairly consistently in keeping with current practice. Some quantities are traditionally expressed in hybrid units—for example, the specific area is usually measured in $m^2\ g^{-1}$—and we continue this practice. The older literature uses cgs (centimeter-gram-second) units almost exclusively, so the reader must be cautious in consulting other sources. Appendix B contains a list of conversion factors between SI and cgs units.

particle weights or the number of atoms in a particle will define the colloidal size range; however, other definitions may be encountered elsewhere. It should be emphasized that these limits are rather arbitrary. Smaller particles are considered within other branches of chemistry, and larger ones are considered within sciences other than chemistry. The statement above may be expanded still further. Colloid science is interdisciplinary in many respects; its field of interest overlaps physics, biology, materials science, and several other disciplines. It is the particle dimension—not the chemical composition (organic or inorganic), sources of the sample (e.g., biological or mineralogical), or physical state (e.g., one or two phases)—that consigns it to our attention. With this in mind, it is evident that colloid science is the science of both large molecules and finely subdivided multiphase systems.

1.1a.2 Surfaces and Interfaces

It is in systems of more than one phase that colloid and surface science meet. The word *surface* is thus used in the chemical sense of a phase boundary rather than in a strictly geometrical sense. Geometrically, a surface has area but not thickness. Chemically, however, it is a region in which the properties vary from those of one phase to those of the adjoining phase. This transition occurs over distances of molecular dimensions at least. For us, therefore, a surface has a thickness that we may imagine as shrinking to zero when we desire a purely geometric description. The term *interface* is also used in this context. This term simply highlights the fact that the surface of interest is the dividing region between two phases.

It is self-evident that the more finely subdivided a given weight of material is, the higher the surface area will be for that weight of sample. In the following section we discuss this in considerable detail since it is the basis for combining a discussion of surface and colloid science in a single book.

1.1a.3 Other Concepts and Classification of Colloids

In subsequent sections of this chapter, we discuss further the distinction between macromolecular colloids and multiphase dispersions (Section 1.3), the use of the term *stability* in colloid science (Section 1.4), the size and shape of colloidal particles, the states of aggregation among particles, and the distribution of particle sizes that is typical of virtually all colloidal preparations (Section 1.5). The fact that particles in the colloidal size range are not all identical in size also requires a preliminary discussion of statistics, which is the subject of Section 1.5c and Appendix C.

VIGNETTE I.1 ENVIRONMENTAL SCIENCES: Colloid-Enhanced Transport in Unconsolidated Media

Contaminated bed sediments exist at numerous locations in the United States and around the world. These result mainly from past indiscriminate pollution of our aquatic environments and consist of freshwater and marine bodies including streams, lakes, wetlands, and estuaries. The bed sediments contain many hydrophobic organic compounds and metal ions that in the course of time act as sources of pollutants of the overlying aqueous phase. There are a number of transport pathways by which pollutants are transferred to the aqueous phase from contaminated sediments. One of the lesser known, but potentially important, modes of transport of pollutants from bed sediments is by diffusion and advection of contaminants associated with colloidal-size dissolved macromolecules in pore water. These colloids are measured in the aqueous phase as dissolved organic compounds (DOCs). (These are defined operationally as particles with a diameter smaller than 0.45 micrometer.)

The facilitated transport of compounds by colloids, illustrated schematically in Figure 1.1, is important in several areas and especially in the study of the fate and transport processes of hydrophobic organic compounds and metal ions in the environment. This facilitated transport also has implications in other areas in which colloid diffusion through porous

media is important, such as those processes utilizing porous sorbents to clean up water and *in situ* flushing of subsurface soil with surfactant solutions for both oil recovery and cleanup of soils contaminated with oil. The understanding of facilitated transport of hydrophobic compounds from unconsolidated media such as bed sediments and soils is therefore very important and requires a knowledge of colloid and surface science of pollutant/particle interactions and transport.

In general, colloids are known to have a large capacity to bind hydrophobic compounds and have been reported to transport contaminants to very large distances in groundwater. It has been shown using simple mathematical models that orders of magnitude larger fluxes of contaminants are possible in the presence of colloids as compared to simple molecular diffusion of contaminants in the absence of any association with colloids. It is also known that contaminant binding to colloids is a fast, equilibrium process with partition constants larger than that between the bed sediment and water. In the literature there exists some information on the possible effects of colloids on the transport of pollutants from bed sediments. However, detailed studies on the flux of contaminants by colloids under controlled laboratory conditions are not available.

This vignette, in fact, captures many of the concepts and phenomena we discuss in this book and highlights them in the context of a very practical problem of considerable significance to our environment (see also the legend for Fig. 1.1).

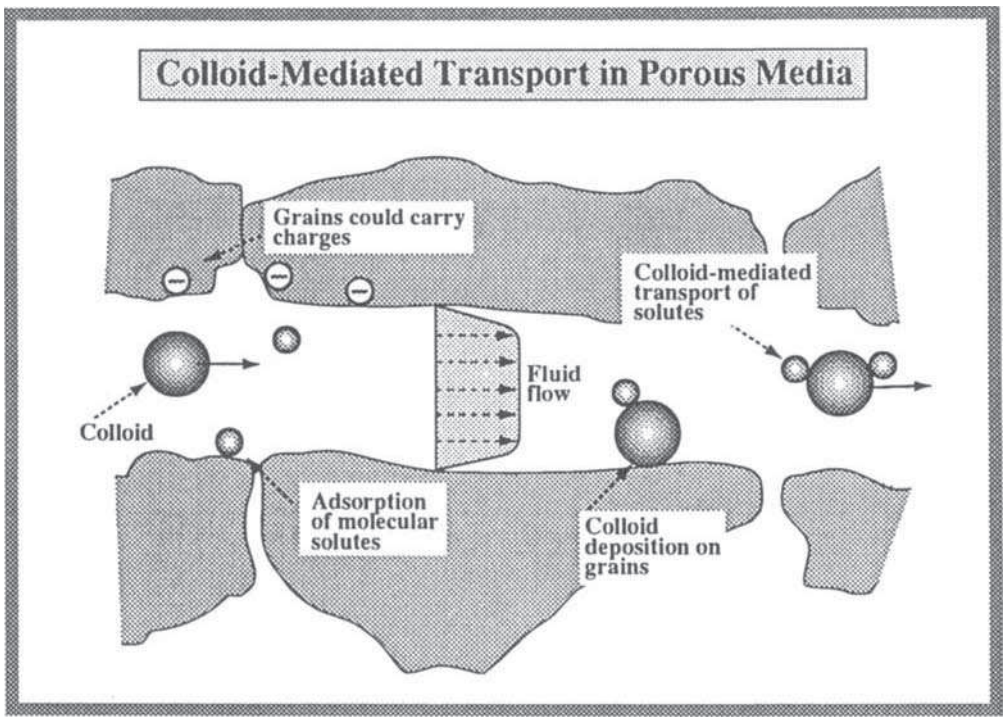


FIG. 1.1 A schematic illustration of colloid-mediated transport in porous media. The sketch illustrates the transport of molecular solutes by colloidal particles. The extent of such transport and its importance are determined by a number of factors, such as the extent of adsorption of molecular solutes on the colloids and on the grains, the deposition and retention of colloids in the pores, the influence of charges on the colloids and on the pore walls, and so on.

1.1b Impact of Colloid and Surface Chemistry in Science, Engineering, and Technology

One of the basic premises underlying the selection of topics included in this book is that areas of similarity between diverse fields should be stressed. This is not to say, of course, that differences are unimportant. Rather, it seems more valuable to point out to the beginner that useful methods and insights are frequently part of the well-established procedures of other disciplines that deal with related phenomena. Too provincial a viewpoint, especially at the beginning, is apt to isolate the student from many potentially valuable sources of information. In the long run, this seems like a greater loss than the loss of time that occurs when one concludes too hastily that a technique that works well in one system should work equally well in another system in which the particles are larger or smaller by several orders of magnitude. Errors of this last sort are generally discovered quickly enough!

1.1b.1 Examples of Applications

Any attempt to enumerate the areas in which surface and colloid chemical concepts find applicability is bound to be incomplete and quite variable with time because of changing technology. Nonetheless, we conclude this section with a partial listing of such applications. If there is any difficulty in doing this, it is because of the abundance rather than scarcity of such examples.

It should be evident from the partial list of examples summarized in Table 1.1 how many materials or phenomena of current scientific or everyday interest touch on colloid and surface science to some extent. Many of these areas, of course, have enormous technological and/or theoretical facets that are totally outside our perspective. Nevertheless, all share a common interest in small particles and/or large molecules.

1.1b.2 Vignettes

In addition to Table 1.1 and the applications discussed in this book, we have included nine "vignettes" in this chapter and one each in the other 12 chapters.* These vignettes are examples selected from physical, biological, and engineering sciences and highlight the broad scope, the rich variety, and the scientific and engineering challenges in colloid and surface science. These vignettes are not meant to be comprehensive and complete; rather, the idea is to expose the student to fascinating topics and ideas that currently occupy researchers and technologists in the chosen areas and to convey a sense of excitement about the subject.

The vignette included in each chapter (with the exception of Chapter 1) highlights an application (in practice or in advanced research) of some of the concepts discussed in that chapter. The vignettes included in Chapter 1, in addition to serving the above purpose and elaborating on the material discussed in the chapter, seek to link the introductory material in this chapter to more detailed, advanced topics treated in other chapters.

VIGNETTE I.2 BIOLOGICAL AND LIFE SCIENCES: Biological Membranes and Cells

Life sciences provide a fascinating array of examples in which colloid and surface science plays a vital (pun intended!) role in maintaining and promoting supramolecular structures and processes that sustain life. A specific example is the phospholipid "bilayers" that form the "walls" of biological cells and separate the interior of the cells from the rest of the environment (see Fig. 1.2; see also Chapter 8, Section 11). These bilayers arise from self-assembly of component molecules, each of which consists of a hydrophilic "head" group

*For the purpose of cross-referencing, the vignettes in this chapter are numbered sequentially as Vignette I.1, I.2, and so on, whereas those in the other chapters carry only the chapter number as a Roman numeral (e.g., Vignette X for the vignette in Chapter 10).

and hydrophobic "tails." The nature of the forces that create such layers (or membranes) and maintain their structure and functionality is a major topic of research today and is pursued by cell biologists, physiologists, biophysicists, biochemists, and engineers, among others. However, a bilayer by itself is insufficient to create and maintain a "living" cell. The cell membrane is actually a mosaic of a number of functional units that include protein molecules embedded in the membrane. The protein molecules themselves contain hydrophobic and hydrophilic components and provide the pathways for life-sustaining ions and polar molecules to move across the cell membrane. The arrangements of the hydrophilic and hydrophobic parts of a protein molecule are far more complex than those in a simpler structure such as a bilayer. The configuration of the hydrophilic and hydrophobic parts of a protein and how a protein is embedded in a bilayer are important for the molecular events that contribute to the functions of a biological cell (see Fig. 1.2 for some additional details). It therefore comes as no surprise that colloid and surface science plays a central role in such phenomena in life sciences (see Alberts et al. 1989; Tanford 1989; Bergethon and Simons 1990; Goodsell 1993; Morowitz 1992*).

In Figure 1.2, the hydrophobic parts of the protein molecules form the outer surface of the cylinders shown and hold the molecules in place, away from water, in the hydrocarbon part of the bilayer. The inner surfaces of the cylinders are hydrophilic and allow the transfer of ions and polar molecules from one side of the bilayer to the other. These pathways remain closed until an appropriate internal or external stimulus triggers them to open to allow transport across the membrane.

In addition to illustrating the importance of colloid and surface science in biological and life sciences, this vignette draws our attention to the importance of phenomena peculiar to surfactant systems. We discuss in Chapters 7 and 8 the behavior of surfactants in solutions and their tendency to self-assemble when dissolved in water or in water-oil mixtures.

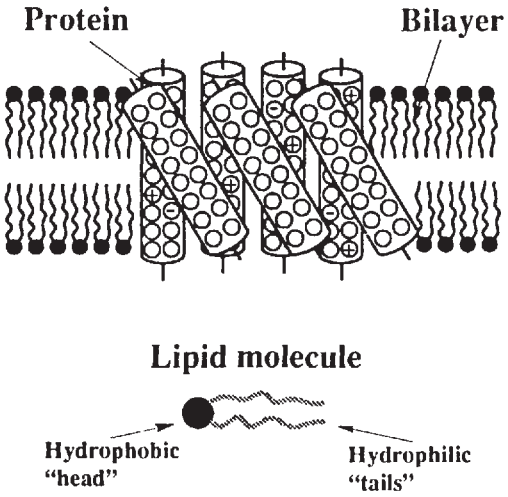


FIG. 1.2 A simplified sketch of a hypothetical protein molecule embedded in a bilayer (a biological membrane). The bilayer shown is a two-dimensional cross section of a membrane. The bundle of cylinders shown represents the "helices" of a protein. The cylinders are part of the same protein and are joined together by other segments (not shown) of the protein protruding out of the bilayer on either side.

*Morowitz presents an interesting (and controversial) theory of the beginning of cellular life; the theory is based on the spontaneous condensation of amphiphilic molecules to form vesicles ("protocells").

TABLE 1.1 Some Examples of Disciplines and Topics for which Colloids and Colloidal Phenomena Are Important

Discipline	Examples
Analytical chemistry	Adsorption indicators, ion exchange, nephlometry, precipitate filterability, chromatography, and decolorization (see also Vignette I.8)
Physical chemistry	Nucleation; superheating, supercooling, and supersaturation; and liquid crystals
Biochemistry and molecular biology	Electrophoresis; osmotic and Donnan equilibria and other membrane phenomena; viruses, nucleic acids, and proteins; and hematology (see also Vignettes I.2 and I.3)
Chemical manufacturing	Catalysis, soaps and detergents, paints, adhesives, and ink; paper and paper coating; pigments; thickening agents; and lubricants (see Vignettes I.9 and IX)
Environmental science	Aerosols, fog and smog, foams, water purification and sewage treatment; cloud seeding; and clean room technology (see also Vignettes I.1 and VIII)
Materials science	Powder metallurgy, alloys, ceramics, cement, fibers, and plastics of all sorts (see Vignettes I.5 and I.6 for some modern examples)
Petroleum science, geology, and soil science	Oil recovery, emulsification, soil porosity, flotation, and ore enrichment
Household and consumer products	Milk and dairy products, beer, waterproofing, cosmetics, and encapsulated products (see, for example, Vignette I.7)
Imaging technology	Photographic emulsions; xerography; printing inks; flat-panel displays (see, for example, Vignettes I.4 and XII)

1.2 THE IMPORTANCE OF THE SURFACE FOR SMALL PARTICLES

As we see in the rest of the book, many of the interesting properties of colloids are the result of their dimension, which lies between atomic dimensions and bulk dimensions. Two of the important consequences of the size range of colloids are (a) colloidal materials have enormous surface areas and "surface energies," and (b) the properties of colloidal "particles" are not always those of the corresponding bulk matter or those of the corresponding atoms or molecules. Let us use a simple exercise or a "thought experiment" to illustrate these points.

The contemporary science student is probably aware that the concept of the atom is traceable to early Greek philosophers, notably Democritus. More than likely, however, few have bothered to follow through the hypothetical subdivision process that led to the original concept of an atom. The time has come to remedy this situation since, as mentioned above, the colloidal size range lies between microscopic chunks of material and individual atoms.

1.2a Increase in Surface Area and Energy with Decrease in Size

Consider a spherical particle of some unspecified material in which the sphere has a convenient radius of, say, 1.0 cm. Let us "reapportion" this fixed quantity of material by subdividing it repeatedly into arrays of spheres, each with a radius half that of the original sphere. The

results of such an exercise are summarized in Table 1.2. These results are purely geometrical and therefore are (assumed to be) independent of any characteristic of the material. For example, it is implied in the calculations reported in Table 1.2 that the density of the material, whatever it may be, remains the same throughout the subdivision process. This is an assumption that breaks down when we reach sizes of the order of atomic dimensions; it is one of the points we consider below.

We can learn two lessons of importance in colloid and surface science from this simple exercise of subdividing a bulk particle. The first is that the properties of a material can be sensitive to the amount of sample under consideration. To see this point, assume that the material from which these spheres are made is water at a density of, say, 1.0 g cm^{-3}. Using this we can convert the particle volumes in Table 1.2 into the number of water molecules per sphere after each cut, as shown in Table 1.3 corresponding to the cuts used in Table 1.2. It is evident that by the time we reach spheres of radius 10^{-8} cm (i.e., typical atomic dimensions), we have begun chopping up the water molecules. As a matter of fact, calculating by this procedure we reach one water molecule per sphere after 25.62 halvings of all spheres, starting from one sphere of $R_s = 1.0$ cm. This corresponds to a radius of 0.193 nm for the water molecules if we use the formulas from Table 1.2 to evaluate the radius of the spheres. The radius of a water molecule estimated from the value of the van der Waals coefficient b (Atkins 1994) is about 0.145 nm, so a substantial discrepancy arises when a bulk property such as density is applied all the way down to molecular dimensions. It is not our purpose here to arrive at a precise estimate of molecular dimensions, but the above discrepancy does point out the fact that the characterization of a material may be sensitive to the size of the sample under consideration. A property such as density depends not only on the mass and volume of the molecules, but also on their packing in a bulk sample.

The second lesson we learn from this exercise is concerned with the extent of increase in surface area and the surface energy as we go to smaller and smaller particles. Let us first calculate the average number of water molecules that reside at the surface of the spheres in

TABLE 1.2 The Radius, Area, and Volume per Particle, Number of Particles, and Total Area for Any Array of Spheres After n "Cuts," Where a Cut is Defined to Be the Reapportionment of Materials into Particles With Radius that Is Half the Starting Value

Cut number	Radius (cm)	Number of spheres	Volume per sphere (cm^3)	Area per sphere (cm^2)	Total area (cm^2)
Original	1	1	4.19	1.26×10^{1}	1.26×10^{1}
Original, symbol	R_s	N_0	V_0	A_0	$A_{T,0}$
1	5×10^{-1}	8	5.24×10^{-1}	3.14	2.51×10^{1}
2	2.5×10^{-1}	6.4×10^{1}	6.55×10^{-2}	$7.86 \times {}^{-1}$	5.03×10^{1}
3	1.25×10^{-1}	5.12×10^{2}	8.18×10^{-3}	$1.96 \times {}^{-1}$	1.01×10^{2}
.	.	.	.	.	.
.	.	.	.	.	.
.	.	.	.	.	.
n	$\left(\frac{1}{2}\right)^n R_s$	$8^n N_0$	$\left(\frac{1}{8}\right)^n V_0$	$\left(\frac{1}{4}\right)^n A_0$	$2^n A_{T,0}$
.	.	.	.	.	.
.	.	.	.	.	.
.	.	.	.	.	.
19.93	10^{-6}	10^{18}	4.2×10^{-18}	1.26×10^{-11}	1.26×10^{7}
23.25	10^{-7}	10^{21}	4.2×10^{-21}	1.26×10^{-13}	1.26×10^{8}
26.58	10^{-8}	10^{24}	4.2×10^{-24}	1.26×10^{-15}	1.26×10^{9}

TABLE 1.3 Total Number of Water Molecules per Sphere and Number at Surface for Spheres of Water After n Cuts (Also Total Surface Energy of the Array of Spheres of Water)

Cut number	Radius (cm)	Number of water molecules per sphere	Number of water molecules at surface	Fraction of total water molecules at surface	Total surface energy (J)
0	1.0	1.38×10^{23}	1.26×10^{15}	9.13×10^{-8}	9.07×10^{-5}
1	5.0×10^{-1}	1.75×10^{22}	3.14×10^{15}	1.79×10^{-7}	1.81×10^{-4}
2	2.5×10^{-1}	2.18×10^{21}	5.03×10^{16}	3.64×10^{-7}	3.62×10^{-4}
3	1.25×10^{-1}	2.73×10^{20}	1.01×10^{17}	7.32×10^{-7}	7.27×10^{-4}
.	.	.	.	.	.
.	.	.	.	.	.
.	.	.	.	.	.
19.93	10^{-6}	1.40×10^{5}	1.26×10^{22}	9.13×10^{-2}	9.07×10^{1}
23.25	10^{-7}	1.40×10^{2}	1.26×10^{23}	9.13×10^{-1}	9.07×10^{2}
26.58	10^{-8}	1.40×10^{-1}	1.26×10^{24}	9.13	9.07×10^{3}

Tables 1.2 and 1.3 by assuming* that the area occupied by each molecule at the surface is about 0.10 nm^2. Using this estimate and the total area at various stages of subdivision from Table 1.2, we obtain an estimate of the number of molecules in the surface at each stage of the process (see Table 1.3). This quantity is also reported in the table as a fraction of the total number of molecules present. Note that this fraction approaches unity as molecular dimensions are approached. That is, as we decrease in size, a larger and larger number of atoms become surface atoms.

Now, let us consider the last column in Table 1.3. In this column the total surface energy of the array of spherical water droplets is reported at each stage of the process. We see in Chapter 6 that the surface tension of a substance is the energy required to make a unit area of new surface. For water, this quantity is about 72 mJ m^{-2} at room temperature. This value of the surface tension of water and the total areas from Table 1.2 can be used to obtain the values listed under "total surface energy" in Table 1.3. It should be recalled that a fixed amount of material (4.19 cm^3 water = 4.19 g water = 0.23 mole water) is involved throughout this entire process. It should also be noted that surface tension, like density, is a macroscopic property; its applicability is highly dubious for very small particles. Nevertheless, as the dimensions of the subdivided units decrease, the total energy associated with the formation of the surface takes on values comparable to other chemical energies. At $R_s = 10^{-4}$ cm, the surface energy is about 0.9 J/0.23 mole or about 4 J $mole^{-1}$. When we reach $R_s = 10^{-7}$ cm, this quantity equals 4 kJ $mole^{-1}$.

1.2b Specific Surface Area

The increasing importance of the surface area as the linear dimensions of particles decrease is stated concisely in a quantity known as the *specific surface area* A_{sp} of a substance. This quantity is determined as the ratio of the area divided by the mass of an array of particles. If the particles are uniform spheres, as we have assumed throughout this section, this ratio equals

$$A_{sp} = A_{tot}/m_{tot} = (n\,4\pi R_s^2)/[n(4/3)\pi R_s^3 \rho] \tag{1}$$

where n is the number of spheres having a radius R_s and made of a material of density ρ. Simplifying Equation (1)† leads to the result

*Our interest is in the order of magnitude of these quantities, so we need not worry about the cross-sectional shape or the surface packing efficiency of the water molecules in these calculations.

†This manner of referencing is used for equations occurring in the same chapter.

$$A_{sp} = 3/(\rho R_s) \tag{2}$$

This formula generalizes the conclusion reached in Tables 1.2 and 1.3. It shows clearly that for a fixed amount of material, the surface area is inversely proportional to the radius for uniform, spherical particles. At the same time, the formula reminds us that some lower limit for R_s must be imposed since the relationship is undefined for $R_s = 0$. If SI units were used consistently, A_{sp} would be expressed in $m^2\ kg^{-1}$; however, $m^2\ g^{-1}$ are the most commonly used units for this quantity. In the event of nonuniform or nonspherical particles, alternate expressions for Equation (2) have to be used. The following example considers the case of cylindrical particles.

* * *

EXAMPLE 1.1 *Variation of Specific Surface Area with Geometry.* A material of density ρ exists as uniform cylindrical particles of radius R_c and length L. Derive an expression for A_{sp} for this material and examine the limiting forms when either R_c or L is very small.

Solution: The area of each cylindrical particle equals the sum of the areas of both ends and the cylindrical surface: $A = 2(\pi R_c^2) + 2\pi R_c L$.

The volume of each cylindrical particle equals $\pi R_c^2 L$ and its mass is given by $\rho \pi R_c^2 L$.

For an array of n cylindrical particles, the total area per total mass equals A_{sp} and is given by

$$\begin{aligned} A_{sp} &= n(2\pi R_c^2 + 2\pi R_c L)/(n\rho\pi R_c^2 L) \\ &= [2(R_c^2 + R_c L)]/\rho R_c^2 L \\ &= (2/\rho)(1/R_c + 1/L) \end{aligned}$$

For a thin rod, $L \gg R_c$:

$$A_{sp} \approx \frac{2}{\rho}\frac{1}{R_c}$$

For a flat disk, $R_c \gg L$:

$$A_{sp} \approx \frac{2}{\rho}\frac{1}{L}$$

Note that in both of these limits, it is the *smaller* dimension that affects A_{sp}, with A_{sp} increasing as the smaller dimension decreases. ■

* * *

Both Example 1.1 and Equation (2) show that the surface plays an increasingly important role as the dimensions of the particles decrease.

The concept of specific area defined by Equation (1) is important because this is a quantity that can be measured experimentally for finely divided solids without any assumptions as to the shape or uniformity of the particles. We discuss the use of gas adsorption to measure A_{sp} in Chapter 9. If the particles are known to be uniform spheres, this measured quantity may be interpreted in terms of Equation (2) to yield a value of R_s. If the actual system consists of nonuniform spheres, an average value of the radius may be evaluated by Equation (2). Finally, even if the particles are nonspherical, a quantity known as the radius of an *equivalent sphere* may be extracted from experimental A_{sp} values. This often proves to be a valuable way of characterizing an array of irregularly shaped particles. We have a good deal more to say about average dimensions in this chapter and about equivalent spheres in Chapter 2.

In the discussion above we emphasized two-phase* colloidal systems, in which the concept of surface plays an important role. However, as we saw in Section 1.1, our definition of the colloidal range is based on the linear dimensions of particles, and there are numerous natural

*We occasionally use the term *two phase* in this chapter to refer to colloids when the colloidal particles and the fluid in which they are suspended are distinct (e.g., solid particles in a liquid or liquid droplets in another, immiscible liquid or in a gas).

and synthetic polymer molecules with dimensions that, considered individually, fall within this range, as we illustrate in the following section. It is clear when we deal with single molecules that the concept of surface is greatly different from when we consider a particle made of many molecules. If, in a given situation, we tend to concentrate on the surface characteristics of a material, then the focus is on *surface* science. If, however, we look at the subdivided sample as an array of particles, then the focus is colloid science. As far as we are concerned, these two fields differ primarily in point of view. Their mutual concern with finely subdivided material is the common denominator that connects the two disciplines.

1.3 CLASSIFICATION OF COLLOIDS BASED ON AFFINITY TO CARRIER FLUID

In the preceding section, we saw that either large molecules or finely subdivided bulk matter could be considered colloids inasmuch as both may consist of particles in the range 10^{-9} to 10^{-6} m in dimension. The difference between these two situations lies in the relationship that exists between the colloidal particle and the medium in which it is embedded. Macromolecular colloids are true *solutions* in the thermodynamic sense. Subdivided bulk matter, on the other hand, forms a two-phase (at least) system with the medium. We have already noted that the word *surface* connotes the existence of a phase boundary and therefore has a specific chemical meaning in the multiphase case; this is, of course, inapplicable to macromolecular colloids. The intent of this book is to discuss as wide a variety of colloidal phenomena as possible from a unified point of view and using a single set of terms. We use the words *continuous phase* and *dispersed phase* to refer to the medium and to the particles in the colloidal size range, respectively. This distinction is somewhat vague in the case of macromolecular solutions. (See Table 1.4 for a summary of the descriptive names for two-phase colloidal systems.)

Colloids are also often classified on the basis of the affinity of the *surfaces* of the particles to the continuous phase. This classification is also ambiguous in some respects, but deserves a brief mention.

The terms used to distinguish colloidal "particles" on the basis of their affinity to the fluid in which they are dispersed are *lyophilic* and *lyophobic*. These terms mean, literally, "solvent loving" and "solvent fearing," respectively. When water is the medium or solvent, the terms *hydrophilic* or *hydrophobic* are often used. This terminology is very useful when considering surface activity such as wettability of a surface; however, when used to classify colloids, the distinction is not always clear-cut. We consider these two types of colloids separately in the following subsections.

1.3a Lyophilic Colloids

The classical use of the term *lyophilic colloids* refers to soluble macromolecular materials in which the individual particles (macromolecules such as synthetic polymer chains or proteins)

TABLE 1.4 Summary of Some of the Descriptive Names Used to Designate Two-Phase Colloidal Systems

Continuous phase	Dispersed phase	Descriptive names
Gas	Liquid	Fog, mist, aerosol
Gas	Solid	Smoke, aerosol
Liquid	Gas	Foam
Liquid	Liquid	Emulsion
Liquid	Solid	Sol, colloidal solution, gel, suspension
Solid	Gas	Solid foam
Solid	Liquid	Gel, solid emulsion
Solid	Solid	Alloy

are of colloidal dimensions. However, there are macromolecules of colloidal dimensions containing both lyophobic and lyophilic components (e.g., proteins with hydrocarbon [hydrophobic] portions and hydrophilic [peptide and carboxyl groups] portions, as mentioned in Vignette I.2). Another such example is the case of micelles (Chapter 8), which are clusters of small molecules that form spontaneously in aqueous solutions (mostly) of certain compounds (therefore these are often called *association colloids*). The onset of micellization occurs at a well-defined concentration—known as the *critical micelle concentration*—which makes micelle formation very much like a phase separation. However, the individual small molecules retain their identity in the micelles, which are not covalent entities like polymers, so their surface is problematic. We see in Chapter 8 that both chemical equilibrium and phase equilibrium can be used to discuss micelle formation, so the classification of micelles remains fuzzy on that basis also. By the time we reach Chapter 8, however, we will be more experienced with colloids and less dependent on the "black and white" categories of this section.

Above we used the words *continuous phase* and *dispersed phase* to refer to the medium and to the particles, respectively, in the colloidal size range. It should be understood that these are solvent and solute in lyophilic systems. In micellar systems, the micelles are dispersed in an aqueous continuous phase. Furthermore, the system as a whole is generally called a *dispersion* when we wish to emphasize the colloidal nature of the dispersed particles. This terminology is by no means universal. Lyophilic dispersions are true solutions and may be called such, although this term ignores the colloidal size of the solute molecules.

VIGNETTE I.3 COLLOIDAL CARGO CARRIERS—FROM COSMETICS TO MEDICINE AND GENETIC ENGINEERING: Liposomes with a Molecular Cargo and a Mission

Wouldn't it be nice if we could package appropriate doses of medicine in physiologically friendly capsules that can deliver the medicine specifically to the organs that need it while at the same time keeping it away from areas that may find the medicine toxic? Or, perhaps, our interest is much more mundane and all we need—to satisfy our vanity—is a method to trap perfumes in our skins so that the fragrance lasts longer. These were hardly the questions that engaged the attention of Alec Bangham, a British scientist who discovered surfactant capsules known as *liposomes* in the early 1960s while studying the effect of lipid molecules on the clotting of blood.

Liposomes are colloidal-size containers made of lipid bilayers (Fig. I.3). A lipid molecule consists of a polar, hydrophilic head that is attached to (one or two) hydrophobic, hydrocarbon tails. At appropriate concentrations, the lipid molecules in water "self-assemble" to form bilayers since the hydrophobic tails like to avoid contact with the water. When such bilayers are broken up into small pieces, the fragments wrap themselves into closed structures known as liposomes and encapsulate some of the water inside. The potential applications of liposomes in cosmetics, pharmaceutical and medical technology, and genetic engineering (for studying basic properties of genes by isolating them inside a liposome or for developing schemes for gene- or protein-replacement therapies) are numerous.

Why should we be interested in using liposomes for drug delivery? In many cases, drugs administered in "free form" cause "side effects" because of their toxicity to areas of the body that are not affected by the disease or disorder. It appears possible to improve the effectiveness of drugs and minimize their toxicity by encapsulating the drugs in liposomes and delivering them efficiently and specifically to the affected organs. It is also possible to design liposomes that avoid detection* by "hunters" such as macrophages in the body so that

*These are known by their registered tradename, Stealth liposomes (see Lasic 1993, p. 284). As discussed by Lasic, the "stealth effect" is the result of a "steric" layer surrounding the liposome, and these liposomes are also known as sterically stabilized liposomes. We discuss steric stabilization in Chapter 13.

the drugs can be delivered to the cells that need treatment. Excellent descriptions of these and related issues in liposome technology are presented by Lasic in a lucidly written review in the *American Scientist* (Lasic 1992) and in a monograph (Lasic 1993).

The above discussion illustrates that liposome technology lies at the interface of colloid and surface chemistry, physics, biology, and medicine. There are a number of problems of interest to colloid scientists and engineers in this context, ranging from, for example, synthesis of new (short-chain as well as polymeric) surfactants for developing special membranes, artificial skins, pharmaceutical lotions, and (of course) drug delivery systems. Although many of these problems are outside the scope of our book, we consider some basic issues that are relevant to the above problems. For example, we discuss the formation of surfactant (e.g., lipid) layers in Chapter 7 and study some of their properties. Formation of surfactant micelles and the relation between the molecular architecture of the surfactants and the shapes of the self-assembled structures that result from the surfactants is covered in Chapter 8. In Chapter 8, we also touch on the special opportunities afforded by the microenvironments inside surfactant micelles and liposomes for studying catalysis and material synthesis.

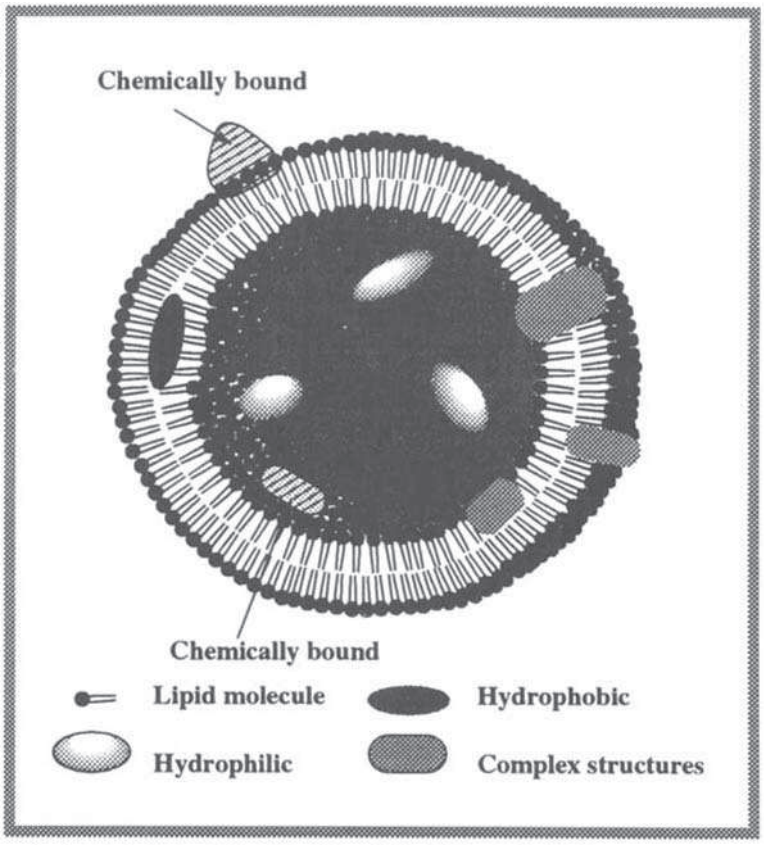


FIG. 1.3 Molecular cargo in a liposome. The cargo molecules are carried in different parts of the liposome depending on their chemical nature. Hydrophobic molecules are carried inside the hydrophobic part of the bilayer, whereas hydrophilic molecules reside in the interior. More complex molecules are wholly or partly embedded in the bilayer or chemically bound to the interior or exterior surface.

1.3b Lyophobic Colloids

Lyophobic colloids are known by a variety of terms, depending on the nature of the phases involved. Some of these are listed in Table 1.4. Some of the terms (e.g., aerosol, gel) are somewhat ambiguous, so the reader is warned to make certain that the system is fully understood, particularly when the original literature is consulted. Remember that a common feature of all systems we consider is that some characteristic linear dimension of the dispersed particles falls in the range defined in Section 1.1a. When we deal with two-phase colloids in this book, we are primarily concerned with systems in which the dispersed phase is solid and the continuous phase is liquid.

As in the case of lyophilic colloids, the use of the adjective *lyophobic* does *not* necessarily mean that the surfaces of the colloids are uniformly "liquid repelling." For example, ceramic sols such as silica and alumina powders in liquids do have surfaces with varying degrees of affinity to the liquid. Despite the ambiguities in the use of the terms *lyophobic* and *lyophilic*, such a classification is convenient.

Next, we consider another difference between lyophobic and lyophilic colloids in addition to the presence or absence of surfaces between the continuous and dispersed species. This difference deals with the "stability" of the dispersion, and we examine the meaning(s) of this term in more detail below.

VIGNETTE I.4 IMAGING SCIENCE AND TECHNOLOGY: Colloid-Based Electrophoretic Imaging Devices*

Colloids hold a considerable potential for applications that are unusual in the classical sense. Most of us are familiar with imaging devices such as the picture tube in a television set. These tubes are bulky and consume large amounts of electrical power. There is, therefore, a large incentive to develop compact imaging devices, known as *flat-panel devices*, that are easily portable and have lower power requirements. (Displays based on liquid crystal technology fall in this class.)

Another possible flat-panel device is one known as an *electrophoretic image display*, or EPID. EPIDs contain submicron-size particles of pigments dispersed in a liquid along with a dye that provides contrast. When an electrical potential is applied to the system, pigment particles are driven to the interface between the suspending liquid and a viewing plate, usually made of glass. There they can be seen under normal illumination. EPIDs have the potential of providing an image that has extremely high optical contrast under normal lighting, that is legible over a wide range of viewing angles, that is inherently retained on the display (as opposed to an image needing constant refreshing as on a TV picture tube), and that requires low voltage and power.

The EPID concept can be combined with other developments in imaging technology to produce optical devices such as light valves and x-ray imagers. An example of an electrophoretic x-ray imager is illustrated in Figure 1.4.

One of the most crucial aspects of this "chemical" display technology is the liquid dispersion of pigment particles and dye. The dispersion must be stable even when the particles are compressed (to concentrations as much as 10 times higher than in the bulk form) to form an image at the viewing plate. The pigment particles must be able to retain their charges after numerous switching operations. The fluid must allow a fast response time. Unwanted migration of particles when an electrical field is applied and other electrohydrodynamic effects must be controlled.

Many of the crucial problems for research and development in this area are the same as those encountered in other areas of colloid and surface science. There are questions that need to be addressed. How do particles interact when they are repeatedly and forcefully packed

*This vignette is a slightly revised version of the one prepared by one of the authors (RR) for a National Research Council report entitled, *Frontiers in Chemical Engineering: Research Needs and Opportunities*, published by the National Academy Press (Washington, DC, 1988).

and unpacked in electrical fields? How are particles, counterions, polymers, and surfactants transported across an EPID cell? How does fluid motion in an EPID device affect this overall transport? How fast do structures formed by charged particles dissipate in applied fields? These are questions central to many of the problems of interest in colloid and surface science.

1.4 CONCEPT OF STABILITY OF COLLOIDAL SYSTEMS

As mentioned in the last section, lyophilic colloids form true solutions, and true solutions are produced spontaneously when solute and solvent are brought together. In the absence of chemical changes or changes of temperature, a solution is stable indefinitely. Finely subdivided dispersions of two phases do not form spontaneously when the two phases are brought together. As a matter of fact, if such a dispersion is allowed to stand long enough, the reverse process would spontaneously occur. For example, oil and water can be vigorously mixed to form a nontransparent, heterogeneous mass; however, on standing, the mixture will separate into two clear, homogeneous layers.

We know from thermodynamics that spontaneous processes occur in the direction of decreasing Gibbs free energy. Therefore we may conclude that the separation of a two-phase dispersed system to form two distinct layers is a change in the direction of decreasing Gibbs

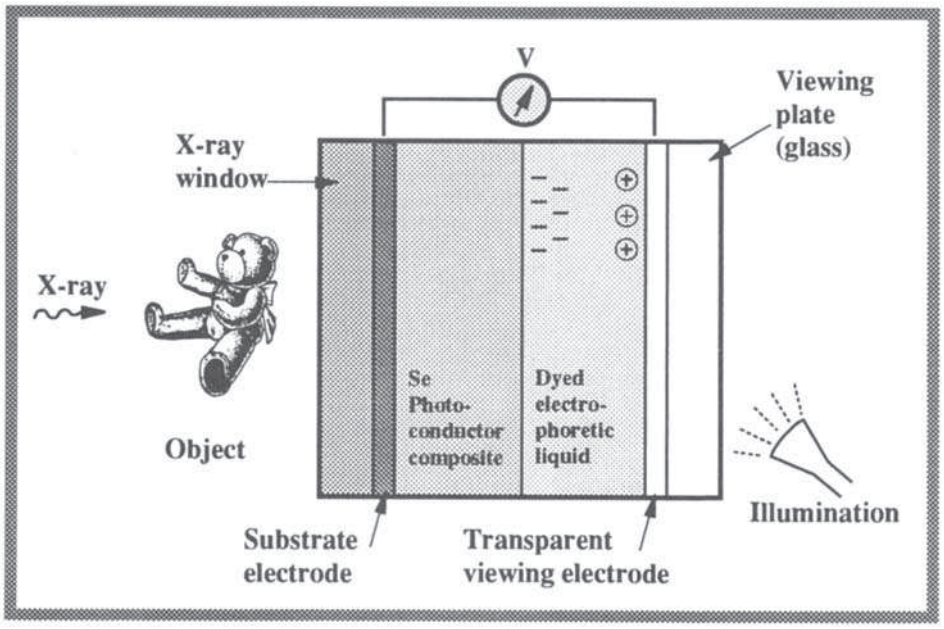


FIG. 1.4 A large-area, solid-state x-ray receptor with an electrophoretic image display. When a voltage is applied across the image cell, pigment particles and counterions in the liquid separate. Most of the voltage drop occurs across the Se layer. X-ray exposure under this condition leads to the creation of a charge-image at the photoconductor-composite/liquid interface due to the generation of x-ray-induced charges in the Se. After the x-ray exposure, the applied voltage is reduced to zero, and the pigment particles are driven to the viewing plate. The image becomes visible on illumination.

free energy. Using the logic employed in the discussion of Table 1.3, one can say that there is more surface energy in a two-phase system when the dispersed phase is in a highly subdivided state than when it is in a coarser state of subdivision. This suggests a correlation between the inherent instability of a highly dispersed lyophobic system and the thermodynamics of the surface. This is discussed in more detail in Chapter 6. For the present it is sufficient to note that lyophobic systems "dislike" their surroundings enough to want to separate out. Lyophilic systems, on the other hand, are perfectly "happy" in a solution. The two categories differ radically in their definitions of stability. We elaborate these individually in the following subsections.

1.4a Kinetic Stability

Remember that equilibrium thermodynamics has nothing to say about the rate at which processes occur. It is a fact that many two-phase dispersions appear unchanged over very long periods of time, but the situation is analogous to the thermodynamic instability of diamond with respect to graphite. The kinetics of the diamond-graphite reaction are slow enough that the thermodynamic instability is of very little practical consequence. Likewise, many colloidal dispersions have *kinetic* stability, even though they are unstable thermodynamically. This type of colloidal "solution" (Table 1.4) resembles a true solution. This is one way in which two-phase dispersions and solutions of macromolecules are very similar and explains in part how the two can be grouped together in our study of colloidal phenomena.

Above we noted that two-phase dispersions will always spontaneously change into a smaller number of large particles given sufficient time. However, many solutions of macromolecules do not undergo spontaneous separation into two phases. Common usage tempts us to describe the first as "unstable" and the second as "stable." Although these terms are used very frequently in colloid science, the reader should realize that *the words are meaningless unless the process to which they are applied has been clearly defined.* The situation is somewhat analogous to the insistence in thermochemistry that one always keep in mind the balanced chemical equation to which the thermodynamic quantities (ΔH, ΔC_p, etc.) apply. The coarsening described here is only one of a variety of possible processes that a dispersion might undergo.

The coarsening process of a thermodynamically unstable dispersion is called *coalescence* or *aggregation*. It is important that we differentiate between coalescence and aggregation. By *coalescence* we mean a process by which two (or more) small particles fuse together to form a single larger particle. The central feature of coalescence is the fact that total surface area is reduced. *Aggregation* is the process by which small particles clump together like a bunch of grapes (an aggregate), but do not fuse into a new particle. In aggregation there is no reduction of surface, although certain surface sites may be blocked at the points at which the smaller particles touch. The term *coagulation* is also used to describe the process of aggregation. A colloid that is stable against coalescence or aggregation is called *kinetically stable*. This is illustrated in Figure 1.5. The classical use of the term "colloid stability" thus represents kinetic stability. In this sense, the word *stability* describes the extent to which small particles remain uniformly distributed throughout a sample.

Incidentally, when small particles coalesce all evidence of the smaller particles is erased. Only the new, larger particle remains. With aggregation, however, the small particles retain their identity; only their kinetic independence is lost. The aggregate moves as a single unit. Likewise, the clusters that form as the products of the process may be called *aggregates*. The individual particles from which the aggregates are assembled are called *primary particles*. A system may be relatively stable in the kinetic sense with respect to one of these processes, say, coalescence, and be unstable with respect to the other, aggregation.

Finally, remember that whether a stable or unstable dispersion is needed depends on the application. A dispersion that is too fine to settle out may be a source of great frustration in some parts of a manufacturing process. On the other hand, a dispersion that settles too

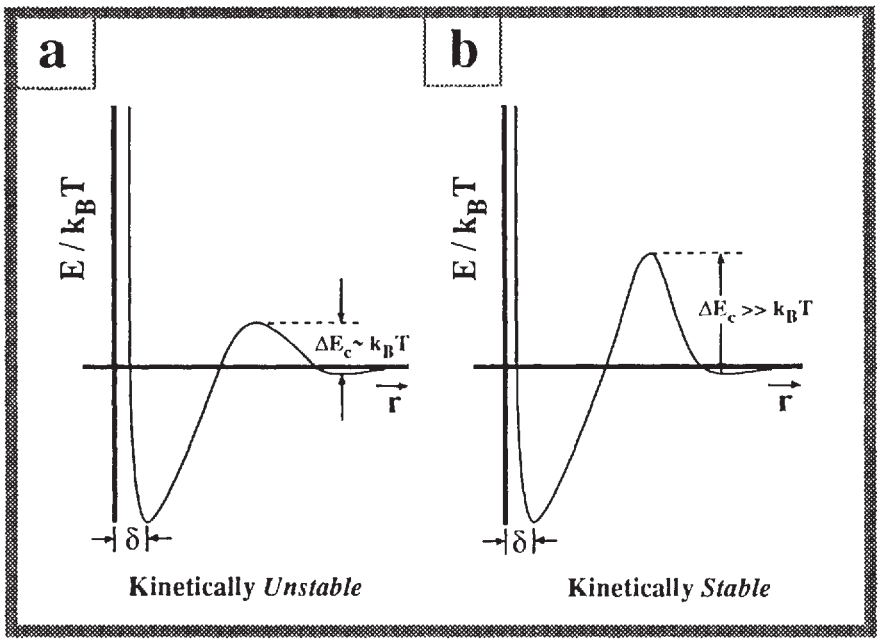


FIG. 1.5 Schematic illustration of kinetic stability of colloids. The figure shows the interaction energy (free energy) E as a function of the surface-to-surface separation r between two particles (k_B and T are the Boltzmann constant and the absolute temperature of the dispersion, respectively). (a) The free energy will reach the global minimum if the two particles can come close enough ($r = \delta$). However, the energy barrier against coagulation, ΔE_c, is of the order of the thermal energy k_BT, and therefore the dispersion is kinetically *unstable*. (b) The energy barrier $\Delta E_c >> k_BT$, and the dispersion is kinetically stable since the thermodynamically favored separation distance is not reachable. (See Chapters 11 and 13 for more details.)

rapidly may be equally troublesome to those who have to pump it around. The chemist doing gravimetric analyses wants coarse precipitates; one who uses adsorption indicators wants finely subdivided particles to form. We have seen an example in Vignette I.4 in which stability of the dispersion is essential. Vignette I.5 illustrates another example from colloidal processing of ceramics in which stability of the dispersion needs to be maintained even as we "consolidate" the dispersion to form dense structures.

VIGNETTE I.5 ADVANCED MATERIALS: Colloid Stability and Ceramic Processing

Considerable recent activity in the area of ceramic processing is aimed toward the formulation of materials with high strengths, comparable to the room temperature strength of metal alloys, at high temperatures (of the order of 2000 K). The impetus comes from the significant gains made in the last 20 years with materials formed from submicron powders of silicon nitride and silicon carbide and the promise of similar improvements in the near future.

The problems with the ceramics lie as much in reproducibility as in absolute strength. With brittle materials, microscopic flaws concentrate stress and cause failure at stresses, which vary according to the classical crack theory as $\delta^{-1/2}$, where δ is the characteristic dimension of the flaw. For materials formed from powders, the flaws represent structural

imperfections, introduced in the initial forming stage, that survive the subsequent processing steps. Hence variability in the initial powder or incomplete control of the fabrication steps lead to unreliable performance.

Of the many methods that are explored, one class of techniques relies on what is known as the *colloidal processing* route (Brinker and Scherer 1990). Here, one starts with a dispersion of precursor powders and densifies the dispersion to an appropriate consistency. The resulting slurry may be slip cast to produce a so-called green specimen in the required shape of the product. Slip casting employs a porous mold that imbibes the fluid but excludes the particles. The type of operation used may employ drain, pressure, or vacuum casting and is determined by the geometry desired and the time constraints. Regardless of the choice of casting method used, a pressure difference draws the fluid into the mold and leaves a dense solid casting adjacent to the wall and a fluid slip in the interior. The physical mechanism bears a strong resemblance to filtration and sedimentation. A broad outline of the above process is illustrated in Figure 1.6.

The structure of the green body depends on the processing technique and conditions as well as the underlying colloid science. For example, the dispersions are often stabilized using polymer additives (Chapter 13) so that the van der Waals attractive forces between the particles (Chapter 10) are minimized by polymer "brushes" adsorbed on the particle surfaces. This prevents "fractal" aggregates (Section 1.5b) that give rise to large voids and defects in the green body on consolidation.

The questions of interest to an engineer in this case are: How do the initial concentration, the particle size, and the nature of the interparticle potential affect the structure of the dispersion, the structure of the final specimen, and the processing time? How long does the process take? What kinds of chemical additives are suitable? The permeability and the capillary suction in the mold determine the rate of production of the specimens. How does one adjust the two to optimize production? These questions require a basic understanding of colloid and surface science and phenomena.

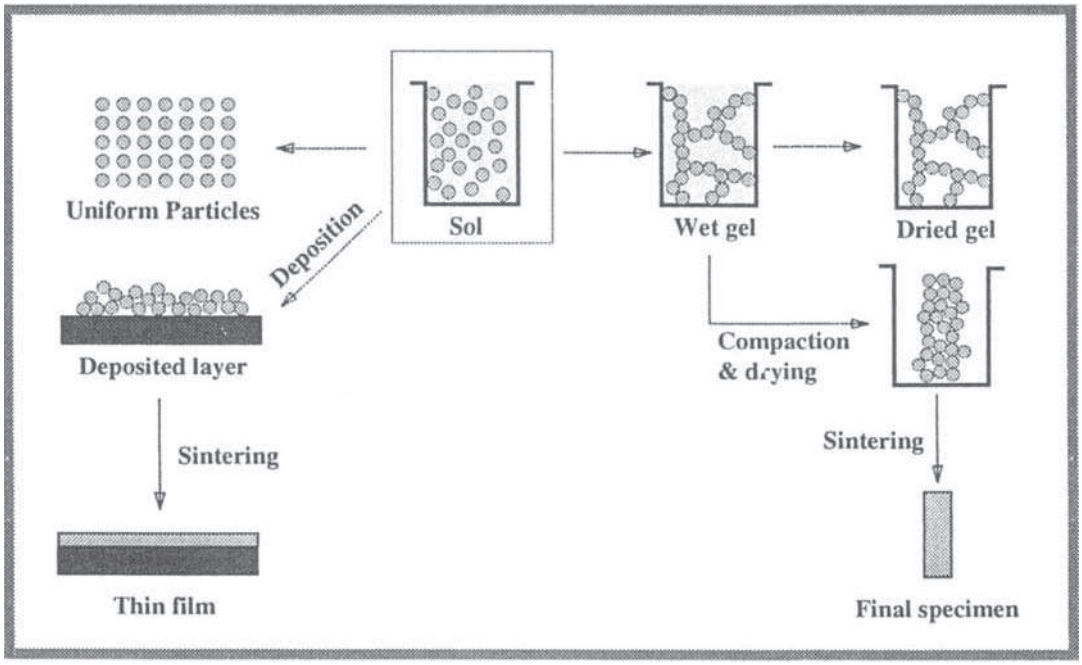


FIG. 1.6 A schematic diagram of colloidal processing of ceramic specimens. The figure illustrates some of the ways in which a dispersion is densified and transformed into porous or compact films or bulk objects. (Adapted and modified from Brinker and Scherer 1990.)

1.4b Thermodynamic Stability

In contrast to the above-described kinetic stability, colloids may also be *thermodynamically* stable. A stable macromolecular solution is an example we have already discussed. Formation of micelles beyond the critical micelle concentration is another example of the formation of a *thermodynamically stable* colloidal phase. However, when the concentration of the (say, initially spherical) micelles increases with addition of surfactants to the system, the spherical micelles may become thermodynamically *unstable* and may form other forms of (thermodynamically stable) surfactant assemblies of more complex shapes (such as cylindrical micelles, liquid-crystalline phases, bilayers, etc.).

Similarly, charged solid particles (such as latex spheres)—kinetically stable lyophobic colloids—may exist in colloidal crystalline phases (with body-centered or face-centered cubic structures) as a consequence of thermodynamically favored reduction in free energies (see Chapter 13). Even neutrally charged spherical particles ("hard spheres") undergo a phase transition from a "liquidlike" isotropic structure to face-centered cubic crystalline structures due to entropic reasons. In this sense, the stability or instability is of *thermodynamic* origin.

It is, therefore, important that we keep several things in mind when the word *stability* is used in colloid science. First, whenever we describe a two-phase dispersion in these terms, the words are being used relatively and often in a kinetic sense. Second, there is little unanimity among workers about the nomenclature of various processes. Finally, whether a stable or unstable system is desirable depends entirely on the context.

Other examples of thermodynamically stable colloidal structures are highlighted in Vignette I.6.

VIGNETTE I.6 POLYMER COMPOSITES: Polymer-Blend Composites as Thermodynamically Stable, Microstructured Materials

All it takes is a simple demonstration with a salad dressing to show that oil and vinegar do not mix. What one has in this case is an emulsion, discussed in Chapter 8. However, if we add a surfactant with molecules containing both oil-liking and vinegar-liking parts to the mixture, the surfactant molecules will preferentially position themselves at oil-vinegar interfaces, very much like the way a surfactant with hydrophilic and hydrophobic parts will behave in an oil-water mixture. One can then have a stable mixture consisting of very small oil droplets dispersed in vinegar or vice versa, depending on the proportion of oil to vinegar. Such *microemulsions*, also discussed in Chapter 8, can exist in very complicated and intricate structures of considerable significance in science and technology.

One can have the same type of situation in a blend of two mutually immiscible polymers (e.g., polymethylbutene [PMB], polyethylbutene [PEB]). When mixed, such homopolymers form coarse blends that are nonequilibrium structures (i.e., only kinetically stable, although the time scale for phase separation is extremely large). If we add the corresponding (PEB-PMB) *diblock copolymer* (i.e., a polymer that has a chain of PEB attached to a chain of PMB) to the mixture, we can produce a rich variety of microstructures of colloidal dimensions. Theoretical predictions show that cylindrical, lamellar, and bicontinuous microstructures can be achieved by manipulating the molecular architecture of block copolymer additives.

In fact, even in pure block copolymer (say, diblock copolymer) solutions the self-association behavior of blocks of each type leads to very useful microstructures (see Fig. 1.7), analogous to association colloids formed by short-chain surfactants. The optical, electrical, and mechanical properties of such composites can be significantly different from those of conventional polymer blends (usually simple spherical dispersions). Conventional blends are formed by quenching processes and result in *coarse* composites; in contrast, the above materials result from *equilibrium* structures and *reversible* phase transitions and therefore could lead to "smart materials" capable of responding to suitable external stimuli.

The colloidal structures described above are dictated by thermodynamics, and the resulting structures are thermodynamically stable. Similar thermodynamically stable structures can develop even in a copolymer *melt* (i.e., there is no other polymer or solvent). Such colloidal systems differ from kinetically stable lyophobic dispersions of the type discussed in Vignettes I.4 and I.5.

1.5 SOME PHYSICAL CHARACTERISTICS OF COLLOIDS

As has been emphasized in the sections above, one of the most important features of colloidal particles is their physical dimension, the defining characteristic of colloids. Many of the properties of colloids of scientific and industrial importance (e.g., specific surface area, viscosity, aggregation behavior, and microstructure) are strongly influenced by the dimension, as well as other characteristics such as surface charge and the chemical affinity of the particles to dissolved matter, among others. Clearly, uniform-size particles of spherical geometry are the easiest to deal with in this respect, but colloidal particles come in all sizes and shapes. Even in the case of spherical particles of uniform size, other physical and chemical properties (e.g., surface charge) may not be uniform due to surface heterogeneities at the molecular scale.

What these imply is that, at the minimum, we need to be aware of the complexities in representing quantitatively particle sizes, shapes, and the extent of their variations. It is also important to be aware of some of the standard experimental tools that are available for this purpose. In this section and the next, we consider some of these items. For convenience, we restrict ourselves primarily to some of the most common methods and techniques.

1.5a Particle Size and Shape

Figure 1.8 shows an electron micrograph of latex particles made from polystyrene cross-linked with divinylbenzene. Note that these latex particles are not the same as simple polystyrene molecules in a true solution. The particles shown in the figure display a remarkable degree of homogeneity with respect to particle size. Such a sample is said to be *monodisperse* (in *size*), in contrast to *polydisperse* systems, which contain a variety of particle sizes. We have a good

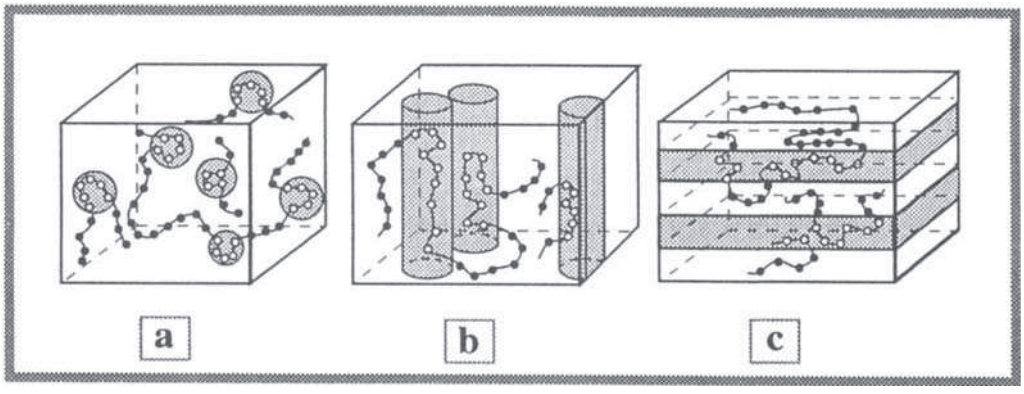


FIG. 1.7 Some of the microstructures produced by the self-association behavior of diblock copolymer solutions. The figure illustrates the (a) spherical, (b) cylindrical, and (c) lamellar structures (among others) that are possible in such solutions. Each diblock polymer chain consists of strings of white beads (representing one type of homopolymer) and strings of black beads (representing the second type of homopolymer). (Redrawn from A. Yu. Grosberg and A. Khokhlov, *Statistical Physics of Macromolecules*, AIP Press, New York, 1994.)

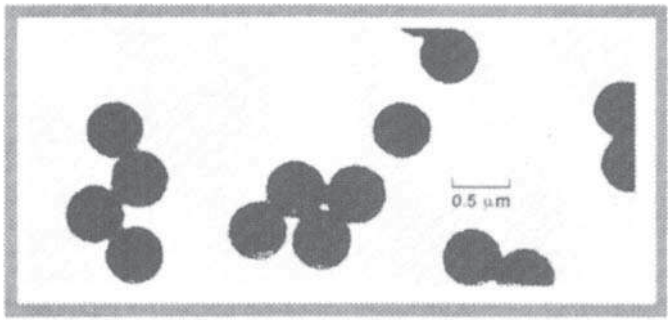


FIG. 1.8 Electron micrograph of cross-linked monodisperse polystyrene latex particles. The latex is a commercial product (d = 0.500 μm) sold as a calibration standard. (Photograph courtesy of R. S. Daniel and L. X. Oakford, California State Polytechnic University, Pomona, CA.)

deal more to say about polydisperse systems in other sections of this chapter. The particles of Figure 1.8 are also perfectly spherical. Except for the nature of the material involved, this figure could be a photograph of the hypothetical array of spheres discussed in Section 1.2.

Monodisperse spheres are not only uniquely easy to characterize, but also very rarely encountered. Polymerization under carefully controlled conditions allows the preparation of the polystyrene latex shown in Figure 1.8. Latexes of this sort are used as standards for the size calibration of optical and electron micrographs (also see Section 1.5a.3). However, in the majority of colloidal systems, the particles are neither spherical nor monodisperse, but it is often useful to define convenient "effective" linear dimensions that are "representative" of the sizes and shapes of the particles. There are many ways of doing this, and whether they are appropriate or not depends on the use of such dimensions in practice. There are excellent books devoted to this topic (see, for example, Allen 1990) and, therefore, we consider only a few examples here for the purpose of illustration.

1.5a.1 Characterizing Variations in Shape: Particles with a High Degree of Symmetry

Many solid particles are not actually spherical, but are characterized by a high degree of symmetry like a sphere and are often approximated as spheres. For example, a polyhedron approximates a sphere more and more closely as the number of its faces increases.

Figure 1.9 shows micrographs of carbon black particles. Broadly speaking, this material is soot, but a great deal of control over its properties may be accomplished by varying the conditions of its preparation. Figure 1.9a shows what is known as a thermal black in which both discrete and partially fused particles may be seen. Figure 1.9b shows the same carbon black preparation (not the same field of particles) after heat treatment at 2700°C in the absence of oxygen. The particles take on a distinctly polyhedral shape with this treatment, known as *graphitization*. The primary particles of the graphitized thermal black shown in Figure 1.9b are sufficiently symmetrical to be approximated as spheres. Likewise, many substances that display irregular but symmetrical particles are often described by a characteristic dimension called a *diameter*. This terminology does not necessarily mean that the particles are spherical.

The forms sketched in Figure 1.10a represent some irregularly shaped particles as they might be observed in a light or electron micrograph (see Section 1.6). The length of a line that bisects the projected area of a particle is a parameter known as *Martin diameter*. The direction along which the Martin diameter is measured is arbitrary, but it should be used consistently to avoid subjective bias. The lines sketched in the figure are intended to represent this quantity. The Martin diameter is most easily measured on micrographs, although movable crosshairs in the eyepiece of a microscope also permit such distances to be measured by direct observation.

Another method of characterizing irregular particles consists of reporting the diameter of a circle that projects the same cross section as the particle in question. This is done by

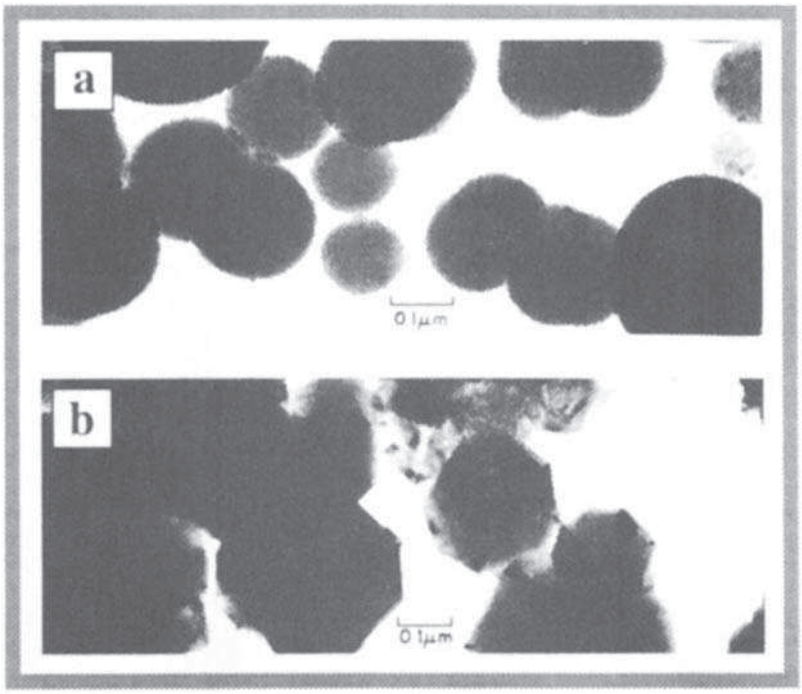


FIG. 1.9 Electron micrograph (150,000×) of carbon black particles: (a) before heat treatment; and (b) after heating to 2700°C in the absence of oxygen. (Adapted from F. A. Heckman, *Rubber Chem. Technol.* **37**, 1243 (1964.)

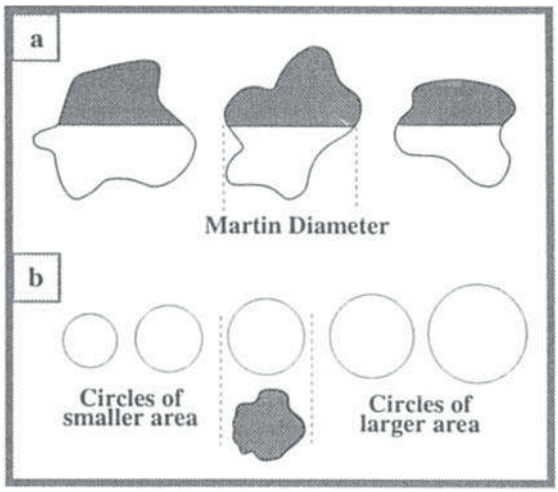


FIG. 1.10 Characterization of the "size" of irregular particles: (a) a schematic illustration of Martin diameters. (b) the use of a graticule to estimate the characteristic dimension.

inserting an object known as a *graticule* (an assortment of circles of different sizes etched on a transparent slide) into the eyepiece of a microscope and deciding which circle most closely approximates the projected area of the particle (see Fig. 1.10b). Both the graticule dimension and the Martin diameter are extremely tedious to evaluate since a large number of particles must be examined for the values to have any statistical significance, but instruments are available to classify images by size automatically. However, the fact that the labor can be done electronically does not decrease the importance of recognizing what is involved in particle sizing.

1.5a.2 Characterizing Variations in Shape: Particles with a Low Degree of Symmetry

The sphere is favored above all other shapes as a model for actual particles because it is characterized by a single parameter. Sometimes, however, the particles of a dispersion are so asymmetrical that no *single* parameter, however defined, can begin to describe the particle. In this case, the next best thing is to describe the particle as an ellipsoid of revolution. An ellipsoid of revolution is the three-dimensional body that results from the complete rotation of an ellipse around one of its axes. We define a as the radius of the ellipsoid measured along the axis of rotation and b as the radius measured in the equatorial plane. Obviously, if these two measurements of radius are equal for a particle, that particle is spherical. If $a > b$, the particle is called a *prolate ellipsoid*; if $a < b$, it is an *oblate ellipsoid*. These two geometries are illustrated in Figure 1.11.

The ratio (a/b), called the axial ratio of the ellipsoid, is frequently used as a measure of the deviation from sphericity of a particle. It plays an important role, for example, in our discussions of sedimentation and viscosity in Chapters 2 and 4, respectively. In the event that $a \gg b$, the prolate ellipsoid approximates a cylinder and, as such, is often used to describe rod-shaped particles such as the tobacco mosaic virus particles shown in Figure 1.12a. Likewise, if $a \ll b$, the oblate ellipsoid approaches the shape of a disk. Thus, even the irregular clay platelets of Figure 1.12b may be approximated as oblate ellipsoids.

One other particle geometry deserves mention. Suppose we were to take a length of string or other flexible material and allow it to tumble freely for a while in a large container. The string would certainly be expected to emerge from this treatment as a tangled jumble. Many long-chain molecules have sufficient flexibility to take on a random configuration like this under the influence of thermal jostling. This "random coil" is likely to be symmetrical rather than stretched out. We accordingly refer to the "radius" of such a coil. The random coil is discussed in detail in Chapter 2, Section 7.

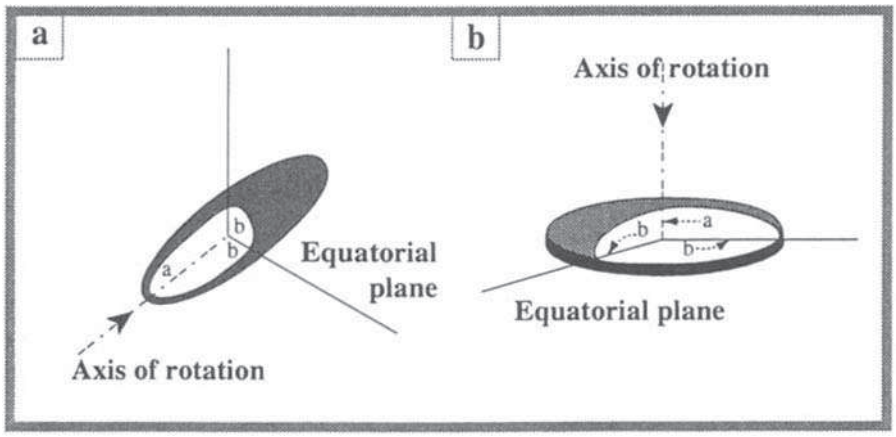


FIG. 1.11 Ellipsoids of revolution: (a) a prolate ($a > b$) ellipsoid; and (b) an oblate ($a < b$) ellipsoid. The figure shows the relationship between the semiaxes and the axis of revolution.

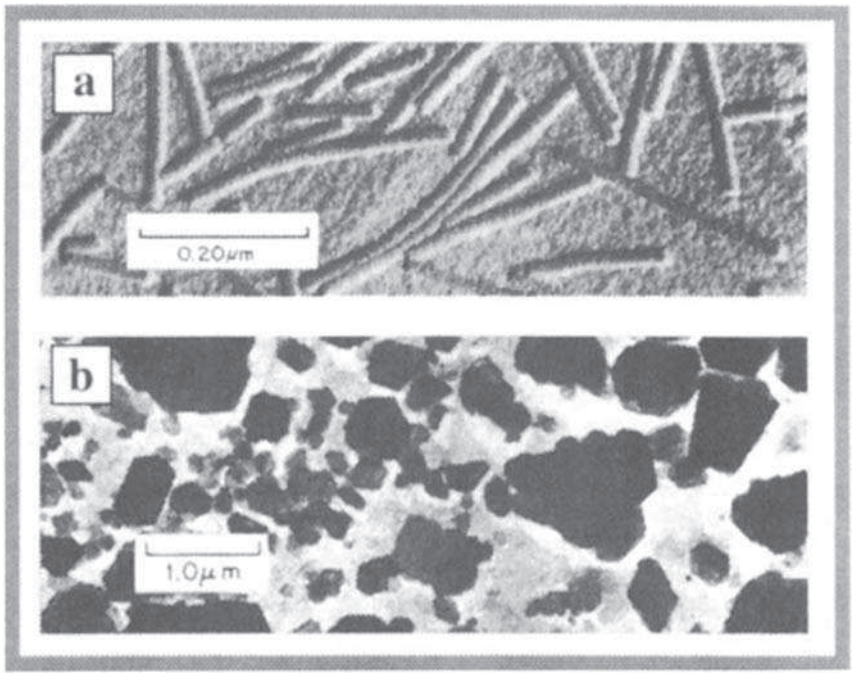


FIG. 1.12 Electron micrograph of two different types of particles that represent extreme variations from spherical particles: (a) tobacco mosaic virus particles (Photograph courtesy of Carl Zeiss, Inc., New York); and (b) clay particles (sodium kaolinite) of mean diameter 0.2 μm (by matching circular fields). In both (a) and (b), contrast has been enhanced by shadow casting (see Section 1.6a.2a and Figure 1.21). (Adapted from M. D. Luh and R. A. Bader, *J. Colloid Interface Sci.* **33**, 539 (1970).

1.5a.3 "Model Particles" of Various Shapes and Sizes

So far, we have focused on how to deal with the variations in the sizes and shapes of colloidal particles encountered in practice. However, in recent years another equally important and highly useful perspective in the use of colloids has emerged from viewing the variations in size and shape (and surface chemistry) as controllable parameters. The question is, Can we control the shapes and sizes of the particles to produce "model" particles that can be used to study properties of colloids and to develop new uses for colloids? The answer is "yes" (within reason).

We have already mentioned the use of the type of polystyrene latex particles shown in Figure 1.8 for the calibration of magnification in electron microscopy. Such latex particles are also used as "supports" in microbiological assay techniques. However, depending on the crystal (or amorphous) structure of the material used, we can also produce particles of many different shapes and sizes by controlling the nucleation and growth conditions. Two such examples are shown in Figures 1.13 and 1.14. More details on the preparation of such colloids are available in numerous review articles (see, e.g., Matijevic 1993 and references therein). Such colloids are often called *model colloids* since they are useful as model systems (i.e., because of their controlled size, shape, and surface chemistry) for studying a variety of phenomena both within and outside colloid science, as illustrated by the following partial list:

1. Spherical latex particles with a reasonably well-defined number of charges per particle can be synthesized and used to study the non-Newtonian behavior of charged dispersions and related electroviscous phenomena (described in Chapter 4). The surface

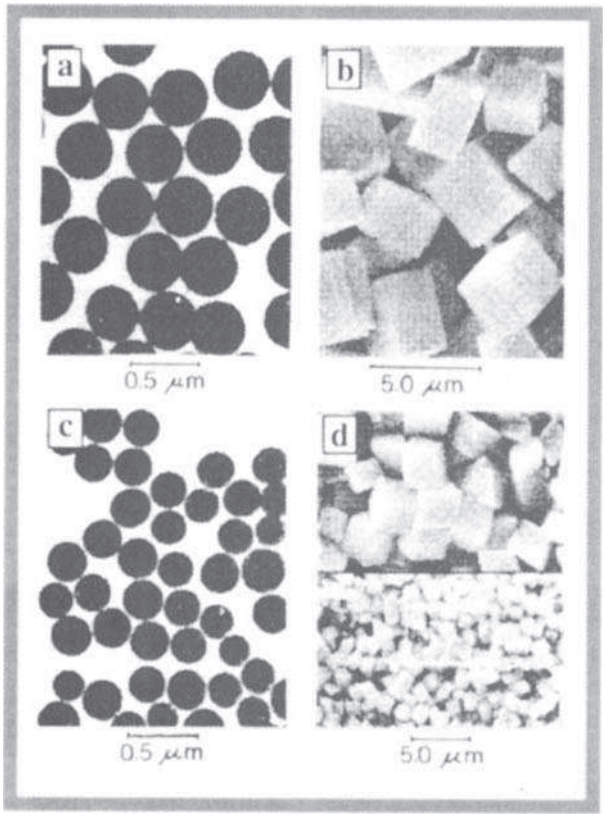


FIG. 1.13 Spherical and cubic "model" particles with crystalline or amorphous microstructure: (a) spherical zinc sulfide particles (transmission electron microscopy, TEM, see Section 1.6a.2a); x-ray diffraction studies show that the microstructure of these particles is crystalline; (b) "cubic" lead sulfide particles (scanning electron microscopy, SEM, see Section 1.6a.2a); (c) amorphous spherical particles of manganese (II) phosphate (TEM); and (d) crystalline cubic cadmium carbonate particles (SEM). (Reprinted with permission of Matijević 1993.)

chemistry of the particles and the chemical constituents in the solution can be controlled to produce neutral "hard-spherelike" particles or charged particles.

2. Such particles can also be used as models of atomic fluids or atomic solids of interest in physics, chemistry, and materials science (see, e.g., Section 13.2, Figs. 13.3 and 13.4; see also Murray and Grier 1995).
3. Since needlelike particles of controlled shape and size can be prepared (see Fig. 1.14, for example), the effects of asymmetry of the particles on flow and optical properties of dispersions can be studied in a systematic manner.
4. Model particles can also be used to develop processing methods for the fabrication of advanced "structural" ceramics and composites of high strength (see, e.g., Vignette 1.5).
5. The magnetic, optical, and electrical properties of materials often depend on the microstructural details and the morphology of materials. Even if the final state is not a colloid, many products pass through colloidal processing routes prior to the final stage. The availability of methods to produce model particles allows us to study and control the desired properties of the final product.

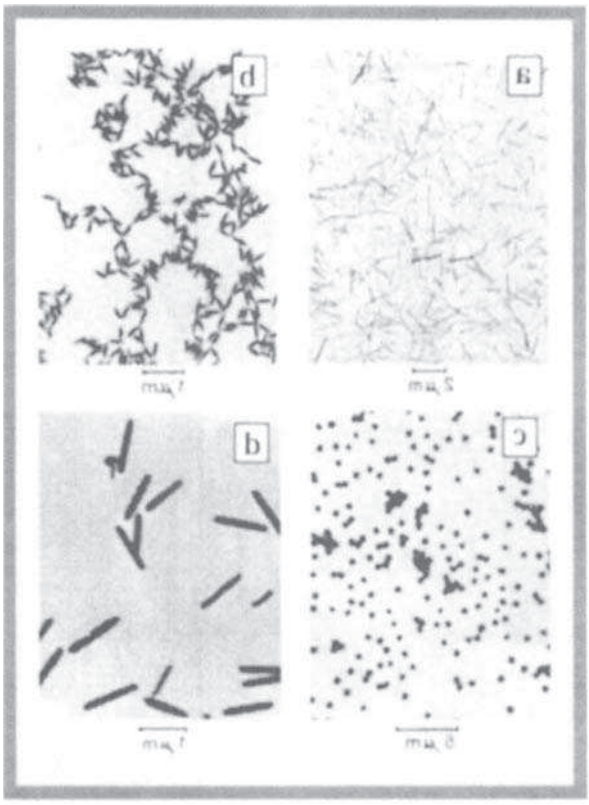


FIG. 1.14 "Model" particles of different shapes with the same or different chemical compositions: (a) rodlike particles of akageneite (β-FeOOH); (b) ellipsoidal particles of hematite (α-Fe_2O_3); (c) cubic particles of hematite; and (d) rodlike particles of mixed chemical composition (α-Fe_2O_3 and β-FeOOH). All are TEM pictures. (Reprinted with permission of Matijević 1993.)

Many additional examples of the above types and others are given in the references cited in Matijevic (1993).

1.5b Particle Aggregates

We have already introduced the idea that the primary particles of a dispersed system tend to associate into larger structures known as aggregates. The nature of the interparticle forces responsible for this aggregation is one of the most examined areas of colloid science. We defer our discussion of the aggregation (or coagulation) process until Chapter 13, but a few remarks about aggregates—the kinetic units that result from that process—and how their dimensions are represented quantitatively are in order at this time.

In many situations the dispersed phase is present as aggregates, not as primary particles. In such cases, it is the size, shape, and concentration of the aggregates that determine the properties of the dispersion itself. As a matter of fact, some substances—for example, those carbon blacks known as channel and furnace blacks—possess rigidly fused, aggregatelike structures as their primary particles.

Figure 1.15 shows some micrographs of aggregates of gold primary particles (with an average diameter of about 14 nm). The pictures are obtained using transmission electron microscopy (see Section 1.6a.2a), and the figure shows four aggregates at four different resolutions. Notice that these pictures show large amounts of open space within the aggregates

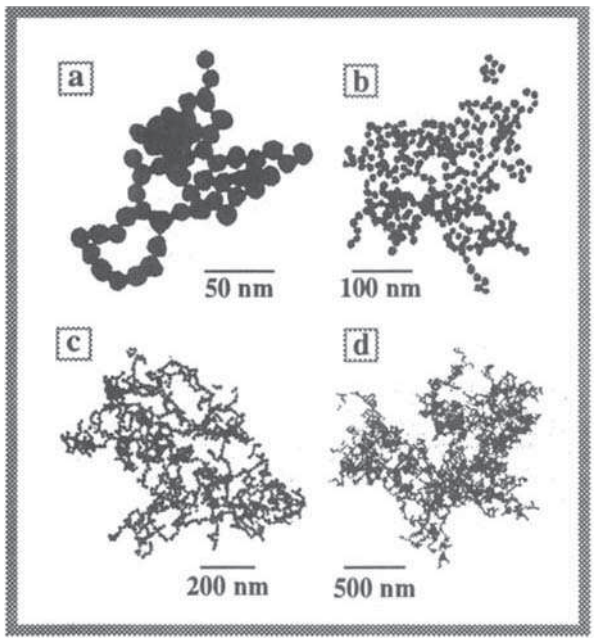


FIG. 1.15 Transmission electron micrographs of aggregates of gold particles. These aggregates were made from a gold colloid to study the relation between the kinetics of aggregation and the resulting structures of the aggregates (see also Section 1.5b.2). The a–d portions of the illustration show aggregates at various resolutions. (The pictures are not from the same aggregate.) (Adapted from D. A. Weitz and J. S. Huang, in *Kinetics of Aggregation and Gelation*, F. Family and D. P. Landau, Eds., Elsevier, Amsterdam, Netherlands, 1984.)

despite the fact that the pictures are two-dimensional projections of three-dimensional objects.* This open structure is related to what is known as the *fractal* nature of such aggregates. We consider this in the next subsection and discuss the relation between aggregation kinetics and the structure of the aggregates subsequently.

1.5b.1 Fractal Dimension of an Aggregate

Suppose we wish to measure the characteristic dimensions of aggregates or aggregatelike particles using micrographs. If the aggregation process results in irregular but, on the average, symmetrical particles, we might characterize the aggregates in terms of the dimension of an inscribing boundary, such as the Martin diameter or the diameter of an equivalent circle determined from a graticule. Then suppose we wish to evaluate the mass of the aggregate

*When an electron micrograph shows evidence of aggregation, we must remember that this may be an artifact arising from the preparation of the sample for microscopy. In other words, the amount of aggregation that a colloid displays in its dispersed state and the amount that appears in an electron micrograph made from the same preparation may be quite different. Optical microscopy is safer in this regard since the actual dispersion may be examined without first evaporating the continuous phase to dryness. Also, in both optical and transmission electron microscopy, it is the *projected* image of the particle that is observed, and microscopic observation alone is often inadequate to distinguish a particle such as that shown in Figure 1.15 from a true aggregate in which the structure is fairly readily disrupted. However, one can use other, "nondestructive" techniques to quantify the structure of aggregates; the use of light, neutron, and x-ray scattering is one such method and is illustrated in Section 5.6a.

enclosed within this boundary. To convert a linear dimension into a particle mass, the particle shape and density must be known.

We have already commented on the approximations involved in treating irregular particles as spheres. Here, however, we encounter an additional problem besides the geometrical approximation already discussed. The question is, What do we use for the density of an aggregate to convert the particle volume into its mass? If the aggregate dimensions have been measured, it is clearly the aggregate density that must be used. The latter is intermediate between the densities of the dispersed and continuous phases, the exact value depending on the structure of the aggregate. Of course, one might attempt to estimate the number of primary particles in an aggregate and then use their size and density as an alternate means of evaluating the mass of an aggregate. The main point of this, however, is the following. Whenever the dispersed particles are in an aggregated form, the properties of the dispersed units are intermediate between those of the two different phases involved. We encounter this difficulty again in Chapter 2, in which the density of the settling unit, whatever it may be, is involved in sedimentation.

Assume for the moment that the aggregate is a solid particle (i.e., without any empty interstitial space) with the same density ρ as that of the primary particles. Then, the mass $m(r)$ of the aggregate enclosed within a distance r from a suitably chosen center will be given by

$$m(r) = \rho(4/3)\pi r^3 \tag{3}$$

If we account for the fact that the region enclosed within the sphere of radius r is not completely filled by particles but contains empty spaces, then the actual mass will be less than that given by the above expression. In fact, it turns out that in many cases, one can write

$$m(r) \propto \rho(4/3)\ \pi\ r^{d_f} \tag{4}$$

where $d_f < 3$, the spatial dimension. The quantity d_f is known as the *fractal dimension* of the aggregate. In fact, it is known more precisely as the *mass fractal dimension* since it is the mass of the aggregate that we have used to specify the fractal dimension. This qualification is also meant to draw one's attention to the fact that other properties can be used to specify the fractal dimension (e.g., surface area), and one would generally expect the fractal dimension to depend on the property used to determine it. In Section 1.5c, we see that this situation frequently arises in characterizing particle distributions. Fractal objects are often called *self-similar* objects. The adjective self-similar draws attention to the fact that such objects look similar when viewed over a range of length scales (see Example 5.4 in Chapter 5). This can be seen from the pictures of aggregates presented in Figure 1.15, which show that the aggregates look similar for a range of resolutions. The fractal dimension, which is a measure of self-similarity, can be specified exactly for "mathematically" produced fractal objects, but for real objects such as the ones shown in Figure 1.15 one obtains the fractal dimension by fitting the mass (or an appropriate property) of the object with the size using an equation like Equation (4).

One way of measuring the fractal dimension of aggregates is discussed in Chapter 5 (See Section 5.6a and Example 5.4). In the example below, we illustrate the relation between the fractal structure of aggregates and the surface area of the aggregates.

* * *

EXAMPLE 1.2 *Surface Area of Fractal Aggregates.* An aerosol "reactor" is used to grow relatively large catalyst particles from monodispersed, spherical primary particles of diameter $d_p = 100$ nm. The specific gravity of the primary particles is 1.5. An examination of the aggregates using an optical microscope shows that the aggregates are essentially spherical. The aggregates produced by the reactor are sieved, and aggregates in five classes of diameters d_i (see Section 1.5c) are selected for measurement of surface areas. A cylindrical column of volume $V_{cyl} = 10\ \text{cm}^3$ is packed with fractionated aggregates having a narrow distribution with the average diameters shown below. Nitrogen adsorption experiments (see Chapter 9) are then conducted to measure the total area of the aggregates in each class.

$d_i \times 10^6$, m	10	50	100	500	1000
$A_{Total}(d_i)$, m^2	91	56	46	28	23

Assume that the actual volume of the aggregates in the nitrogen adsorption experiments is 60% of the total volume (with the rest accounting for pore space between aggregates in the packing). Calculate the fractal dimension of the aggregates.

An independent x-ray and light scattering analysis (see Section 5.6a and Example 5.4) of a dispersion of the aggregates suggests that the aggregates have a fractal structure with a fractal dimension of 2.65. Is this confirmed by your result?

Solution: Assume that the aggregates are fractals with the mass $m_a(d)$ of an aggregate of diameter d being given by

$$m_a(d) = Cd^{d_f}$$

where C is a constant and d is specified in meters. (Notice that the numerical value of the "constant" of proportionality [i.e., C] depends on the dimensions used for the diameter d.]

The number $N_a(d)$ of primary particles in an aggregate of diameter d is simply the mass of the aggregate divided by the mass of the primary particle:

$$N_a(d) = \frac{6C}{\pi d_p^3 \rho_p} d^{d_f}$$

where ρ_p is the density of the primary particle.

If we assume that the surface area πd_p^2 lost by tne contacts between the primary particles during the process of aggregation is negligible, the surface area of the aggregate $A_a(d)$ will simply be the number of primary particles in the aggregate times the surface area of the primary particle:

$$A_a(d) = \frac{6C}{d_p \rho_p} d^{d_f} = CA_{sp} d^{d_f}$$

where $A_{sp} = (6/d_p\rho_p)$ is the specific surface area of the primary particles, see Equation (2). The above equation shows that the surface area of a fractal aggregate is also a fractal of dimension d_f.

The total surface area of the aggregates in the volume V_{cyl} in the adsorption experiments is given by the number of aggregates in the cylinder times the area of each aggregate:

$$A_{Total}(d) = \frac{V_{cyl}(1 - \varepsilon)}{(\pi d_p^3/6)} A_a(d) = \frac{6V_{cyl}(1 - \varepsilon)}{\pi} CA_{sp} d^{d_f - 3}$$

For the given data, V_{cyl} is 10^{-5} m^3; ϵ, the porosity of the packing, is 0.4; and A_{sp} is $6/d_p\rho_p = 4 \cdot 10^4$ m^2/kg. The measured area can, therefore, be plotted against the diameter on a *log-log* graph to check if the aggregates are fractals and to determine the constant C. (Remember, the slope of a log-log plot gives the exponent.)

Figure 1.16 shows a plot of the given data. It is clear that ln A_{Total} is linear with respect to ln d, with a slope equal to -0.3, giving a d_f value of 2.7. Notice that if the "aggregates" had been smooth, solid spheres, the surface area of the aggregate would have been proportional to d^2, i.e., $d_f = 2$. The fact that $d_f > 2$ is an indication of the porous nature of the aggregates.

This result shows that the conclusion reached from scattering results is consistent with surface area measurements.

The intercept can be obtained by extrapolating the line to $d = 1$ m (i.e., at ln $d = 0$) and is 2.85 m$^{2.3}$. Therefore,

$$A_{Total}(d) = \frac{6V_{cyl}(1 - \varepsilon)}{\pi} CA_{sp} d^{d_f - 3} = 2.85\, d^{-0.3}, \qquad \text{with } d \text{ in meters.}$$

from which one gets C = 6.3 kg/m$^{2.7}$. ■

* * *

The fractal dimension 2.7 of the surface area of the aggregates in the above example is larger than 2, expected for a smooth, non-fractal surface. In contrast, the fractal dimension as defined in Equation (4) is less than the dimension of the space. Its value relative to the spatial dimension is a measure of the structure of the aggregate. For instance, the closer d_f is to 3 in Equation (4), the denser (i.e., the more tightly packed) is the aggregate. There are ways in which one can measure the fractal dimension experimentally (e.g., using light, x-ray, or neutron scattering techniques; see Chapter 5, Section 6), but we are not concerned with that in

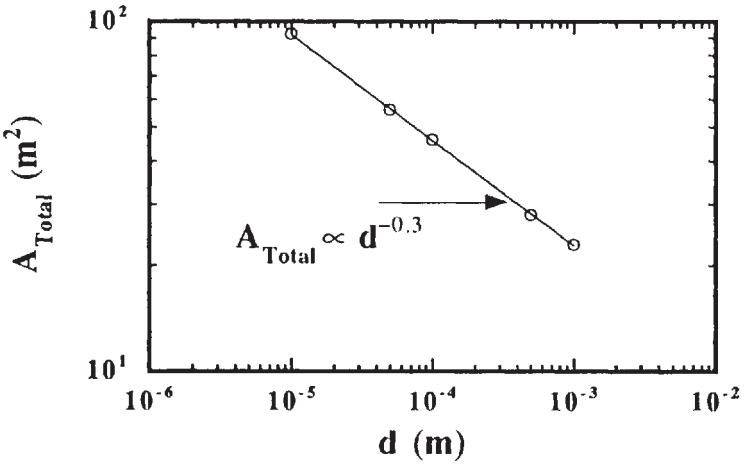


FIG. 1.16 The total area measured versus the diameter of the aggregate on a log-log scale for the data given in Example 1.2.

this book. It is, however, important to note that fairly successful attempts have been made to link the fractal dimension of aggregates to the processing techniques used in forming the aggregated particles (Brinker and Scherer 1990, Chapter 3).

Computer simulations combined with experiments have also shown that one can deduce from the fractal dimension the nature of nucleation and growth of particles and what chemical and physical mechanisms control the formation of particle aggregates. We consider this briefly before proceeding to other topics.

1.5b.2 Relating Aggregate Structure to Growth Mechanisms

A very interesting and fast-growing approach to research concerning aggregates is the technique of computer simulation. By this method, aggregates are "assembled" by a computer, which uses random numbers to determine the coordinates from which each primary particle approaches the growing aggregate. The model that has been most studied consists of spherical primary particles, although linear sets of spheres have been used to simulate asymmetrical primary particles.

The probability of adhesion on contact (say, p_a) can be made a variable quantity in such simulations. The probability of adhesion on contact can be thought of as the effectiveness of the "reaction" of forming an N-particle aggregate when a primary particle collides with an $(N - 1)$-particle aggregate.

One can also introduce in the simulation other physical mechanisms such as the rate of diffusion of the particles. If adhesion probability ("reaction") controls the rate of formation of the aggregate, the aggregation process is known as a *reaction-limited aggregation* (RLA); if diffusion controls the rate of aggregation (i.e., the rate of diffusional transport of a particle to another particle or an aggregate is slow relative to adhesion on contact), the process is known as *diffusion-limited aggregation* (DLA). In DLA, one assumes that $p_a = 1$ (the particles adhere on contact), whereas in RLA p_a is less than unity (i.e., the particles do not always stick to each other on contact). As might be expected intuitively, more open aggregates result when the probability of adhesion at initial contact is high. If, on the other hand, the added particle is permitted to roll along the surface of the growing aggregate before adhering, a more compact structure results.

Variations of the above concepts are also possible to simulate other situations. For example, one variation is to permit the joining together of small aggregates to form larger ones rather than restricting the addition to primary particles only. This leads to structures that are even more expanded than those resulting from the addition of primary particles alone (see Fig.

1.17). Such a procedure is known as *diffusion-limited cluster-cluster aggregation*, or DLCCA, if diffusion of the aggregates ("clusters") is the rate-limiting mechanism. (The above-described DLA is then more accurately described as *diffusion-limited monomer-cluster aggregation*, or DLMCA, since the aggregates in this case grow through contacts with single particles ["monomers"]). Figure 1.17 is a collection of examples of (three-dimensional) aggregates assembled in a number of different ways using computer simulations. The corresponding fractal dimensions are also shown in the figure.

The resemblance of the aggregates in Figure 1.17 to an actual aggregate depends on the growth mechanism. In fact, results of simulations such as the ones shown are often used to model or to interpret the growth mechanisms of real aggregates. For example, the analysis of the clusters shown in Figure 1.15 suggests that the gold aggregates shown result from diffusion-limited cluster-cluster aggregation (see Weitz et al. 1985).

These computer simulations permit the number density of primary particles within the aggregate to be evaluated, important information for relating the properties of the aggregate to its composition. As might be expected, however, it is difficult to know a priori what model to use for a particular system. However, this technique does allow some interesting a posteriori interpretations of known structures to be made. Another closely related problem that has been

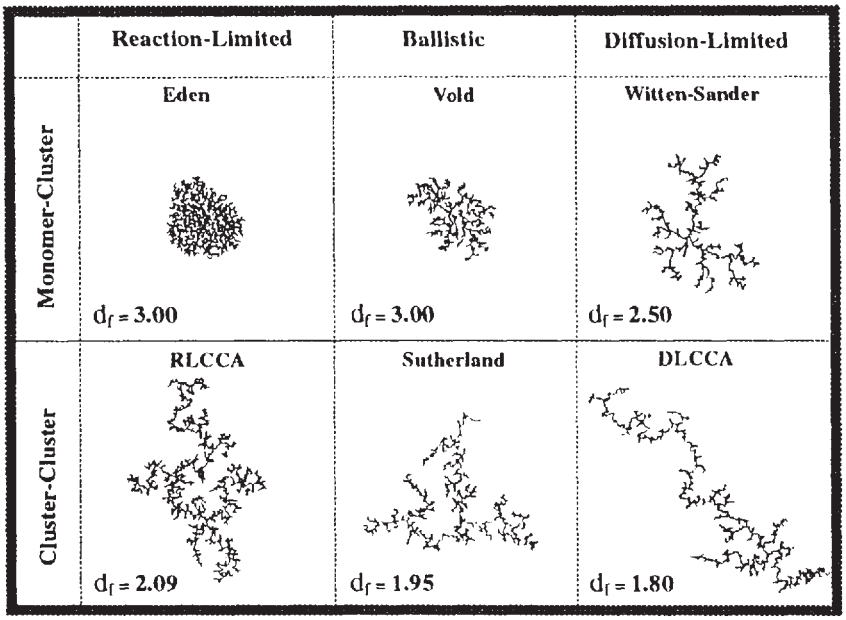


FIG. 1.17 Aggregates obtained through computer simulations using various growth models. The figure shows typical aggregates produced in the simulations under a number of conditions. The results show two-dimensional renditions of three-dimensional simulations. The column headings identify the "controlling step" in the aggregation process (i.e., the type of particle motion and probability of adhesion p_a). (*Reaction limited* implies $p_a < 1$; *ballistic* implies that the "particle" motion is rectilinear [with the added assumption that $p_a = 1$]; *diffusion limited* implies that the "particle" motion is a random walk [with the assumption that $p_a = 1$].) The row labels specify which type of collision is considered (i.e., monomer-cluster or cluster-cluster). The names associated with the models are also shown. For example, diffusion-limited monomer-cluster aggregation (DLMCA) is known as the Witten-Sander model. (RLCCA signifies reaction-limited cluster-cluster aggregation.) (Redrawn from D. W. Schaefer, *MRS Bulletin* **8**, 22 (1988). Simulations are from P. Meakin, in *On Growth and Form*, H. E. Stanley and N. Ostrowsky, Eds., Martinus-Nijhoff, Boston, 1986.)

studied by computer simulation is the volume occupied by a sediment. As with aggregates, it is found that sediments become more voluminous as the probability of adhesion on contact increases.

This discussion of aggregates leads us to another important characteristic of dispersions we have not yet considered in sufficient detail: *polydispersity*. Monodisperse systems are the exception rather than the rule. Even in those rare cases in which a monodisperse system exists, any aggregation that occurs will result in a distribution of particle sizes because of the random nature of the aggregation process.

1.5c Polydispersity

It is difficult to imagine an array of particles that would be easier to describe than the latex particles of Figure 1.8. A single parameter such as the radius of the spheres is sufficient to characterize the dispersed phase in terms of a linear dimension. However, the only realistic attitude to take toward the dispersed systems we are interested in is to assume that they are polydisperse (see, e.g., Vignette I.7, which presents an extreme case). Even the particles in Figure 1.8, which appear remarkably uniform, have a narrow distribution of particle sizes. There are rare cases in which the distribution of dimensions is of negligible width, but, generally speaking, a statistical approach is required to describe a colloidal dispersion. A brief introduction to statistical analysis of polydispersed colloids, definitions of moments, and theoretical distribution functions is given in Appendix C. Here, we are mainly concerned with illustrating some of the essential ideas.

VIGNETTE I.7 FOOD SCIENCES: Structure and Processing of Food Products

Most food products and food preparations are colloids. They are typically multicomponent and multiphase systems consisting of colloidal species of different kinds, shapes, and sizes and different phases. Ice cream, for example, is a combination of emulsions, foams, particles, and gels since it consists of a frozen aqueous phase containing fat droplets, ice crystals, and very small air pockets (microvoids). Salad dressing, special sauce, and the like are complicated emulsions and may contain small surfactant clusters known as micelles (Chapter 8). The dimensions of the "particles" in these entities usually cover a rather broad spectrum, ranging from nanometers (typical micellar units) to micrometers (emulsion droplets) or millimeters (foams). Food products may also contain macromolecules (such as proteins) and gels formed from other food particles aggregated by adsorbed protein molecules. The texture (how a food feels to touch or in the mouth) depends on the structure of the food.

One of the major concerns of a food chemist or technologist is the stability of food products with time (Dickinson 1992). Requirements change with the type of food and may depend on nutritional requirements, governmental regulations, or economy of production. For example, products such as margarine, cream, and spreads require specific textural and flow characteristics. Beers may require additives in order to produce froth on dispensing. Sauces and gravies need to maintain their structure at fairly high temperatures. In practical terms, these imply that one is concerned with the microstructure of the constituents; how it varies with chemical additives, temperature, etc.; and how to produce products with appropriate "shelf life" and structural integrity.

In general terms these problems are not unlike those encountered in the manufacture of other colloidal products such as paints, face creams, and printing inks and toners, for example. The difference, of course, is that the inks and toners are hardly appetizing and one seldom cares about how they taste! (Of course, physical stability alone does not determine the taste or the worthiness of a food product. Chemical reactions play a role in taste and in determining the structure, and other considerations such as nutritional value and esthetic appeal may also apply.)

TABLE 1.5 A Hypothetical Distribution of 400 Spherical Particles[a]

Class boundaries $<d<$ (μm)	Class mark d_i (μm)	Number of particles n_i	Fraction of total number in class $f_{n,i}$	Total number with $d < d_i$, $n_{T,i}$
0-0.1	0.05	7	0.018	7
0.1-0.2	0.15	15	0.038	22
0.2-0.3	0.25	18	0.045	40
0.3-0.4	0.35	28	0.070	68
0.4-0.5	0.45	32	0.080	100
0.5-0.6	0.55	70	0.175	170
0.6-0.7	0.65	65	0.163	235
0.7-0.8	0.75	59	0.148	294
0.8-0.9	0.85	45	0.113	339
0.9-1.0	0.95	38	0.095	377
1.0-1.1	1.05	19	0.048	396
1.1-1.2	1.15	4	0.010	400

[a]The data are classified into twelve classes: the class marks, numbers, and fractions of particles per class and the total number of particles up to and including each class are listed. See text for clarification.

As described in Appendix C, if we have a discrete set of measurements of a quantity such as the diameters of particles, one can represent the results in terms of a few useful quantities like the average diameter and standard deviation. Table 1.5 represents the frequency distribution for a hypothetical array of spheres; all the numerical examples of this section are based on this sample of 400 particles. The particles have been sorted into categories called *classes* with a narrower range of dimensions. Each class is represented by the midpoint of the interval, a quantity called the class mark, symbolized by d_i for class i. Similarly, we define the number of particles in each class as n_i.

A common graphical representation of a frequency distribution is the *histogram*, a bar graph in which the class marks are plotted as the abscissa and the height of the bar is proportional to the number of particles in the class. Sometimes the ordinate is defined as the fraction of particles in the class $f_{n,i}$, with the first subscript n representing the fact that the fraction is based on the number of particles. Figure 1.18a is a plot of the histogram of the data in Table 1.5. Obviously, as the number of classes approaches infinity, the width of each interval approaches zero and the histogram approaches a smooth curve. Analytical distribution functions give the equation for such smooth curves. However, in practice, a bar graph is a convenient approximation to the smooth function.

Another way in which these kinds of data are sometimes represented is as a *cumulative* curve in which the total number (or fraction) of particles $n_{T,i}$ having diameters less (sometimes more) than and including a particular d_i are plotted versus d_i. Figure 1.18b shows the cumulative plot for the same data shown in Figure 1.18a as a histogram. The cumulative curve is equivalent to the integral of the frequency distribution up to the specified class mark. Cumulative distribution curves are used in Chapter 2 in connection with sedimentation.

1.5c.1 Average Diameters and Standard Deviation

Although the histogram is a convenient pictorial way to present data, a more concise representation is often required. Most students will quickly identify the *average* as such a representation, and we calculate this quantity here for the data in Figure 1.18. First, however, let us write a very general definition of "average" since this word will take on a considerably wider meaning than usual before we are finished with it. For classified data of the type presented in Table 1.5, average may be defined as follows:

$$\text{“average”} = \sum_i (\text{weighting factor})_i\,(\text{quantity being averaged})_i \tag{5}$$

This formulation may look somewhat unfamiliar, but we shall see that both the weighting factor and the quantity being averaged can mean many different things in colloid science. Appendix C presents a more detailed presentation of basic statistical concepts for both discrete and continuous distributions for those who desire more information on these topics.

For the data in Table 1.5, the first kind of average we examine is what is known as the *number average* $\bar{d}_n$ because the weighting factors $f_{n,i}$ are the *number fraction* of particles in each class:

$$f_{n,i} = n_i / \sum_i n_i \tag{6}$$

In evaluating the number average diameter $\bar{d}_n$, the d_i values are the quantities being averaged. With these substitutions, Equation (5) becomes

$$\bar{d}_n = \sum_i \frac{n_i}{\sum_i n_i} d_i \tag{7}$$

This quantity is also known as the *mean* of the distribution. For the data in the table, $\bar{d}_n$ = 0.64 μm; this value has been marked in Figure 1.18 (and is denoted by $\bar{d}$).

In experiments, often a single property is measured for a large number of particles—such as the surface area for a fine powder—rather than for individual particles as in microscopy, and an effective size may be “backed out” from the quantity measured. This is the situation, for example, when Equation (2) is used to evaluate the radius of an equivalent sphere from experimental A_{sp} values.

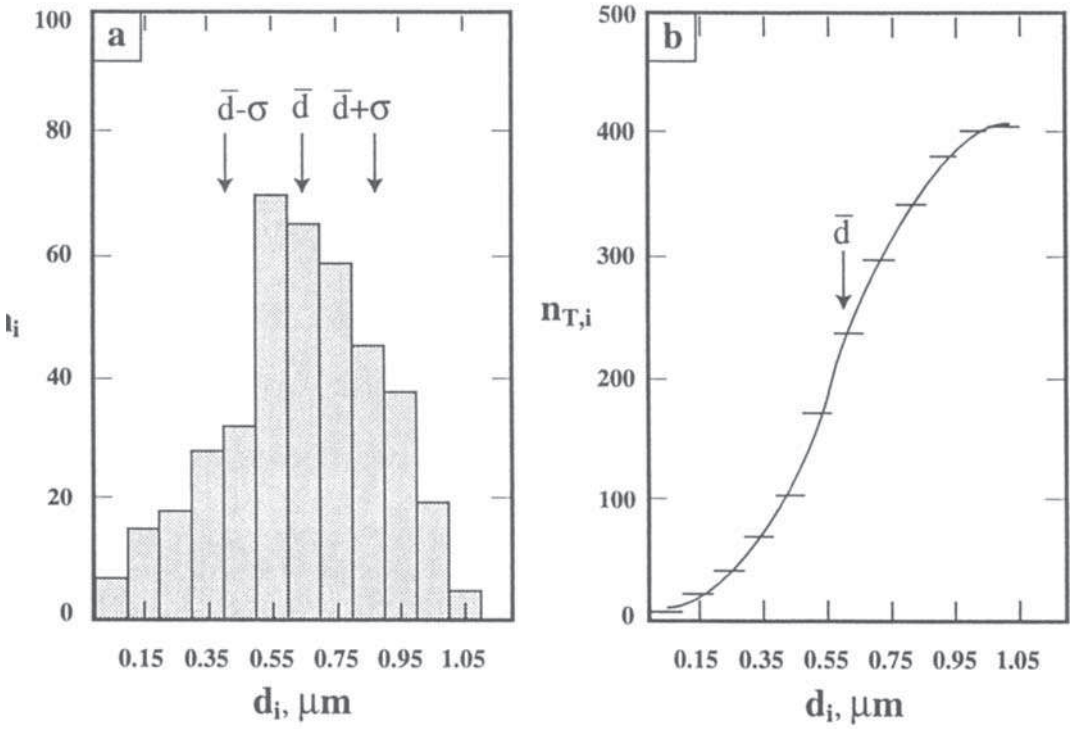


FIG. 1.18 Graphical representation of the data in Table 1.5. Data are presented as (a) a histogram and (b) a cumulative distribution curve.

When polydispersity is present, we have no problem labeling the calculated equivalent radius an "average". We have used quotation marks around this label to alert us to the possibility that this average may be something other than the mean.

To explore this further, we present some additional data about the 400 spheres in Table 1.5, namely, that the sample possesses a total surface area of $5.85 \cdot 10^2\ \mu m^2$ or $5.85 \cdot 10^2/400 = 1.46\ \mu m^2$ per particle. Likewise, the total volume of the 400 spheres is $76\ \mu m^3$ or $0.19\ \mu m^3$ per particle. We see in Chapter 9 how the surface areas of actual powdered samples are measured and the volume is readily available from mass when the density of the bulk material is known. Now let us calculate the "average" diameter of an equivalent sphere from these data.

1. Since $A = \pi d^2$, the "average" diameter based on average area of the particle $= (A/\pi)^{1/2} = 0.68\ \mu m$.
2. Since $V = (\pi/6)d^3$, the "average" diameter based on average volume of the particle $= (6V/\pi)^{1/3} = 0.72\ \mu m$.
3. For this same population of spheres, the "average" diameter based on the number fraction $f_{n,i}$ of the particles in each class $i = \bar{d}_n = 0.64\ \mu m$, as we have already seen.

These calculations illustrate a very general feature of polydisperse systems: Different experimental approaches (e.g., measuring d's versus measuring areas) give different averages for what nominally (hence the quotation marks) might be called the same quantity. This divergence between "average" values determined by different methods can be a cause of consternation for the uninitiated, who might expect corroboration of a previous result determined by a different method. In fact, the only time these different procedures would ever result in the same value would be in the event that all spheres in the sample had the same diameter.

Herein lies the value of these different "averages": the divergence between the "averages" calculated by different methods offers a clue as to the breadth of the distribution of particle sizes. Remember, the average, however evaluated, is only one measure of the distribution of sizes. A fuller description requires some measure of the width of the distribution as well. For classified data, the *standard deviation* (see Appendix C) is routinely used for this purpose. For characterizations based on macroscopic experiments such as we have been discussing, it is quantities such as $\bar{d}_s/\bar{d}_n$ or $\bar{d}_v/\bar{d}_s$ that quantify this spread. (The "averages" $\bar{d}_s$ and $\bar{d}_v$ are defined below and are also discussed in Appendix C.)

The disparity among the values calculated above reveals the inadequacy of simply calling them "average" (and justifies the quotation marks). What is needed now is a clearer understanding of why these differences arise . . . and better terminology. Using the area as an example, we realize that the total area of the spheres can be represented by $\Sigma_i n_i (\pi d_i^2)$ and the "per particle" quantity $\bar{A}$ is just the total area divided by $\Sigma_i n_i$ or

$$\bar{A} = \frac{\sum_i n_i(\pi d_i^2)}{\sum_i n_i} = \pi \sum_i \frac{n_i}{\sum_i n_i} d_i^2 \tag{8}$$

Comparing this result with Equation (5) shows that $\bar{A}/\pi$ is simply the average value of d^2; it is the square root of this quantity that was evaluated above. This kind of average is called the "surface average diameter," $\bar{d}_s$, to distinguish it from the "number average" diameter $\bar{d}_n$. (This terminology leaves a little bit to be desired since both $\bar{d}_s$ and $\bar{d}_n$ use number fractions as weighting factors.) An identical argument would lead to the result for volume "per particle":

$$\bar{V} = \frac{1}{6} \frac{\sum_i n_i(\pi d_i^3)}{\sum_i n_i} = \frac{\pi}{6} \sum_i \frac{n_i}{\sum_i n_i} d_i^3 \tag{9}$$

so that $6\bar{V}/\pi$ is recognized as the number average value of d^3. The cube root of this quantity, evaluated above, is called the "volume average diameter," $\bar{d}_v$. These parameters are summa-

TABLE 1.6 Some of the More Widely Encountered Size "Averages" in Surface and Colloid Science

Name	Symbol	Definition	Quantity averaged	Weighting factor
Number average, or mean	$\bar{d}_n$ or $\bar{d}$	$\frac{\Sigma_i n_i d_i}{\Sigma_i n_i}$	Diameter	Number in class
Second moment about origin	$\overline{d^2}$	$\frac{\Sigma_i n_i d_i^2}{\Sigma_i n_i}$	Square of diameter	Number in class
Surface average	$\bar{d}_s$	$(\overline{d^2})^{1/2}$	Square of diameter	Number in class
Third moment about origin	$\overline{d^3}$	$\frac{\Sigma_i n_i d_i^3}{\Sigma_i n_i}$	Cube of diameter	Number in class
Volume average	$\bar{d}_v$	$(\overline{d^3})^{1/3}$	Cube of diameter	Number of class
Radius of gyration	$(\overline{R_g^2})^{1/2}$	$\left(\frac{\Sigma_i m_i r_i^2}{\Sigma_i m_i}\right)^{1/2}$	Square of radius	Mass in class

rized in Table 1.6. Note that the relative magnitudes of the number, surface, and volume averages are given by the sequence

$$\bar{d}_n < \bar{d}_s < \bar{d}_v \tag{10}$$

for a polydisperse system. Only for a monodisperse system would all three parameters have identical values. Therefore, the divergence from unity of the ratio of any two of these, measured independently, is often taken as an indication of the polydispersity of the dispersion. We see in the next section that this comparison of averages evaluated by different techniques finds particular application in the characterization of molecular weight distributions.

1.5c.2 Radius of Gyration

Table 1.6 also lists the radius of gyration. This is an average dimension often used in colloid science to characterize the spatial extension of a particle. We shall see that this quantity can be measured for polydisperse systems by viscosity (Chapter 4) and light scattering (Chapter 5). It is therefore an experimental quantity that quantifies the dimensions of a disperse system and deserves to be included in Table 1.6. Since the typical student of chemistry has probably not heard much about the radius of gyration since general physics, a short review seems in order.

We assume that the particle with the radius of gyration under discussion may be subdivided into a number of volume elements of mass m_i. Then, the moment of inertia I about the axis of rotation of the body is given by

$$I = \sum_i m_i r_i^2 \tag{11}$$

where r_i is the distance of the ith volume element from the axis of rotation.

Regardless of the shape of the particle, there is a radial distance at which the entire mass of the particle could be located such that the moment of inertia would be the same as that of the actual distribution of the mass. This distance is the *radius of gyration* R_g. According to this definition, it is clear that

$$R_g^2 \sum_i m_i = I = \sum_i m_i r_i^2 \tag{12}$$

Therefore, we see that the value of R_g for an array of volume elements is the average value of r^2, with mass fraction (rather than number fraction) as the weighting factor:

$$R_g^2 = \frac{\Sigma_i m_i r_i^2}{\Sigma_i m_i} \tag{13}$$

We see in the following subsection that this type of weighting factor gives one of the common molecular weight averages for polydisperse systems.

The radius of gyration is a parameter that characterizes particle size without the need to specify particle shape. However, the relationship between R_g and the actual dimensions of a particle depends on the shape of the particle. Such relationships are derived in most elementary physics texts for rigid bodies of various geometries. To translate a radius of gyration into an actual geometrical dimension, some shape must be assumed. For example, for a sphere of radius R_s, $R_g = (3/5)\, R_s^2$. It should be emphasized, however, that the radius of gyration in itself is a perfectly legitimate way of describing the dimensions of a particle. The specification of particle geometry is really optional.

1.5c.3 Molecular Weight Averages

The averages discussed above can also be extended to molecular weights. In Section 1.6b.2, we will discuss a technique called *size-exclusion chromatography* (SEC) that can be used to obtain number fractions and weight fractions of particles in various molecular weight classes. From this sort of information, the two most common molecular weight averages, number-average molecular weight and weight-average molecular weight, can be readily determined.

1.5c.3a Number-Average Molecular Weight. Suppose a dispersion is classified into a set of categories in which there are n_i particles of molecular weight M_i in the ith class. Then the number-average molecular weight equals

$$\bar{M}_n = \frac{\Sigma_i n_i M_i}{\Sigma_i n_i} = \sum_i f_{n,i} M_i \tag{14}$$

where, as before, $f_{n,i}$ is the number fraction of particles in class i.

1.5c.3b Weight-Average Molecular Weight. Alternatively, we could define the average in such a way that the weight of particles in each class w_i rather than their number is used as the weighting factor. This results in an average known as the *weight-average molecular weight* $\bar{M}_w$

$$\bar{M}_w = \frac{\Sigma_i w_i M_i}{\Sigma_i w_i} = \sum_i f_{w,i} M_i \tag{15}$$

Note that $f_{w,i} = w_i/\Sigma_i w_i$ is the *weight fraction* of particles in class i. Since the weight of material in a particular size class is given by the product of the number of particles in the class and their molecular weight, Equation (15) may be written as

$$\bar{M}_w = \frac{\Sigma_i (n_i M_i) M_i}{\Sigma_i n_i M_i} = \frac{\Sigma_i n_i M_i^2}{\Sigma_i n_i M_i} = \frac{\Sigma_i f_{n,i} M_i^2}{\Sigma_i f_{n,i} M_i} \tag{16}$$

We shall see that measurement of the osmotic pressure of a polydisperse system permits the experimental evaluation of $\bar{M}_n$ (Chapter 3), and light scattering experiments enable us to measure $\bar{M}_w$ (Chapter 5). It follows from the definition of these various averages that

$$\frac{\bar{M}_w}{\bar{M}_n} \geq 1 \tag{17}$$

by analogy with the inequalities in Equation (10). The ratio given in Equation (17) is known as the *polydispersity index* of the material. Only when the system is monodisperse does the equality apply in Equation (17).

Here, too, the deviation of this ratio from unity may be taken as a measure of polydispersity. The relationship between the ratio $\overline{M}_w/\overline{M}_n$ and the standard deviation of the molecular weight distribution is easily seen as follows. From Equation (16), it is clear that

$$\sum_i f_{n,i} M_i^2 = \overline{M}_w \sum_i f_{n,i} M_i = \overline{M}_w \overline{M}_n \tag{18}$$

From the general procedure for defining the mean, the left-hand side of Equation (18) may also be written as $\overline{M^2}$. Substituting this result into Equation (C.5) of Appendix C (with M in place of ϕ), we can write the standard deviation σ of the molecular weight distribution as

$$\sigma = (\overline{M}_n \overline{M}_w - \overline{M}_n^2)^{1/2} = \overline{M}_n \left(\frac{\overline{M}_w}{\overline{M}_n} - 1 \right)^{1/2} \tag{19}$$

Therefore the square root of the amount by which the molecular weight ratio exceeds unity measures the standard deviation of the distribution relative to the number average molecular weight.

Example 1.3 illustrates these relationships for a hypothetical polymer.

* * *

EXAMPLE 1.3 *Polydispersity of a Synthetic Polymer.* Columns (1) and (2) of Table 1.7 list the number of moles and the molecular weight, respectively, for eight fractions of a synthetic polymer. Calculate $\overline{M}_n$ and $\overline{M}_w$ from these data and evaluate σ using

$$\sigma = \left(\frac{\Sigma_i n_i (M_i - \overline{M}_n)^2}{\Sigma_i n_i} \right)^{1/2} \tag{A}$$

(see Appendix C) and Equation (19).

Solution. For each class of molecules calculate the quantities listed in columns (3)–(6) in Table 1.7.

The sum of the values in column (1) equals $\Sigma_i n_i = 0.136$ mole.

The product of columns (1) and (2) is w_i and values for this quantity are listed in column (3). The sum of the entries in column (3) equals $n_i M_i = 6700$ g.

Dividing 6700 g by 0.136 mole gives $\overline{M}_n = 49{,}300$ g mole^{-1}.

The product of columns (2) and (3) equals $w_i M_i$, and values for this quantity are listed in column (4); $\Sigma_i w_i M_i = 337.8 \cdot 10^6$ g^2 mole^{-1}.

Dividing $3.378 \cdot 10^8$ by 6700 gives $\overline{M}_w = 50{,}400$ g mole^{-1}.

The square of the difference between M_i *and* $\overline{M}_n$ is given in column (5), and column (6) lists $n_i(M_i - \overline{M}_n)^2$;

$$\Sigma_i n_i (M_i - \overline{M}_n)^2 = 7.68 \cdot 10^6$$

TABLE 1.7 Number of Moles and Molecular Weights for Eight Classes of a Hypothetical Fractionated Polymer (Remaining Quantities Calculated in Example 1.3)

(1) n_i (mole)	(2) M_i (g mole^{-1})	(3) w_i (g)	(4) $w_i M_i \times 10^{-6}$	(5) $(M_i - \overline{M})^2 \times 10^{-8}$	(6) $n_i(M_i - \overline{M})^2 \times 10^{-6}$
0.003	30,000	90	2.70	3.72	1.12
0.007	35,000	245	8.58	2.04	1.43
0.015	40,000	600	24.00	0.86	1.29
0.024	45,000	1080	48.60	0.18	0.43
0.040	50,000	2000	100.00	0.00	0.00
0.032	55,000	1760	96.80	0.32	1.02
0.010	60,000	600	36.00	1.14	1.14
0.005	65,000	325	21.10	2.50	1.25
$\Sigma = 0.136$		$\Sigma = 6700$	$\Sigma = 337.78$		$\Sigma = 7.68$

TABLE 1.8 The Most Common Molecular Weight Averages, Their Definitions, and Their Methods of Determination

Average	Definition	Methods
$\overline{M}_n$	$\frac{\Sigma_i n_i M_i}{\Sigma_i n_i}$	Osmotic pressure; other colligative properties
$\overline{M}_w$	$\frac{\Sigma_i f_{n,i} M_i^2}{\Sigma_i f_{n,i} M_i}$	Light scattering; sedimentation velocity
$\overline{M}_{vis}$	$\left(\frac{\Sigma_i n_i M_i^{1+a}}{\Sigma_i n_i M_i}\right)^{1/a}$	Intrinsic viscosity

Dividing $7.68 \cdot 10^6$ by 0.136 gives σ^2 by Equation (A); therefore $\sigma^2 = 5.65 \cdot 10^7$ and $\sigma = 7500$.

The ratio $\overline{M}_w/\overline{M}_n = 1.022$ and, by Equation (19), $\sigma/\overline{M}_n = (0.022)^{1/2}$, or $\sigma = 0.148(49{,}300) = 7300$.

The discrepancy between the two values of σ is a matter of significant figures and is not meaningful. ■

* * *

As in the case of particle sizes, averages of molecular weights can also be defined in terms of properties of the dispersions. Here we define one such average based on viscosities of the dispersions.

1.5c.3c Viscosity-Average Molecular Weight. The *viscosity average molecular weight* $\overline{M}_{vis}$ is written as

$$\overline{M}_{vis} = \left(\frac{\Sigma_i n_i M_i^{1+a}}{\Sigma_i n_i M_i}\right)^{1/a} \tag{20}$$

The logic of this definition follows the logic used in defining average diameters based on surface area or volume discussed above and, as its name implies, $\overline{M}_{vis}$ is a molecular weight average determined from viscosity measurements on polydisperse polymer samples. The exponent a in this definition is called the *Mark-Houwink coefficient*, generally $0.5 < a < 1.0$. The exponent is characteristic of the polymer-solvent-temperature conditions of the experiment. We see how these experiments are conducted and the significance of the Mark-Houwink coefficient in Chapter 4. For now, we may take Equation (20) as merely defining another kind of molecular weight average. Note that $\overline{M}_{vis} = \overline{M}_w$ when $a = 1$.

The various molecular weight averages we have discussed, their definitions, and the experimental methods that measure them are listed in Table 1.8. All these different "averages" are admittedly confusing. Without this information, however, it would be far more confusing to try to rationalize the discrepancy between two different molecular weight determinations on the same sample by methods that yield different averages. The divergence between such values is a direct consequence of polydispersity in the sample. As we have seen, it is not only unavoidable but also informative as to the extent of polydispersity.

1.6 SOME CLASSICAL AND EMERGING EXPERIMENTAL TOOLS

In this section, we present a few examples of instruments available for visual observation and imaging of colloids and surfaces, for measurement of "sizes" and for surface force measurements. Such a presentation can hardly be comprehensive; in fact, that is not our purpose here. Throughout the book, we discuss numerous other techniques such as osmotic pressure measurements, light and other radiation scattering techniques, surface tension measurements,

and electrokinetic measurements, among others. There are many more tools for which there is no space here. Some of the techniques described in other chapters can provide the type of information described in this section or serve as adjuncts to the techniques described here.

The main objective of the present section, however, is to begin with a very standard technique such as optical microscopy and to use it to illustrate why colloids are difficult to see and what modern developments have emerged in recent years to allow us to "see" and do things that were considered impossible until a decade ago. We also use this opportunity to review briefly some new techniques that are currently available to measure interaction forces between particles directly. We appeal to some of these techniques in other chapters when we discuss colloidal forces.

1.6a "Visual" Observation and Imaging

1.6a.1 Optical Microscopy

Because of the particle sizes involved, classically the optical microscope has been the instrument of choice especially for lyophobic colloids. Excellent books and manuals are available (Bradbury 1991; Cherry 1991; Schaeffer 1953) on the numerous variations of optical microscopy, and we do not go into all the details. Our purpose here is merely to point out some very elementary principles that make this method ideally suited for direct examination of colloids. We also use this introduction as a first step in pointing out modern techniques that fall under the class of "microscopy" but use principles (e.g., electron tunneling; see Vignette I.8) and radiation (e.g., electron or x-ray) other than those used in optical microscopy.

1.6a.1a Contrast. For a particle to be visible optically, there must be an acceptable difference between its refractive index and that of its surroundings; this is known as *contrast*. This requirement has nothing to do with particle size: A glass rod can be made to "disappear" by immersing it in a liquid of matching refractive index. The fact that a technique is not applicable to all possible systems does not invalidate it; this only means that one must have access to more than one technique in order to deal with a variety of problems. Contrast is a generic term, and how the contrast between a particle and its surrounding arises depends on

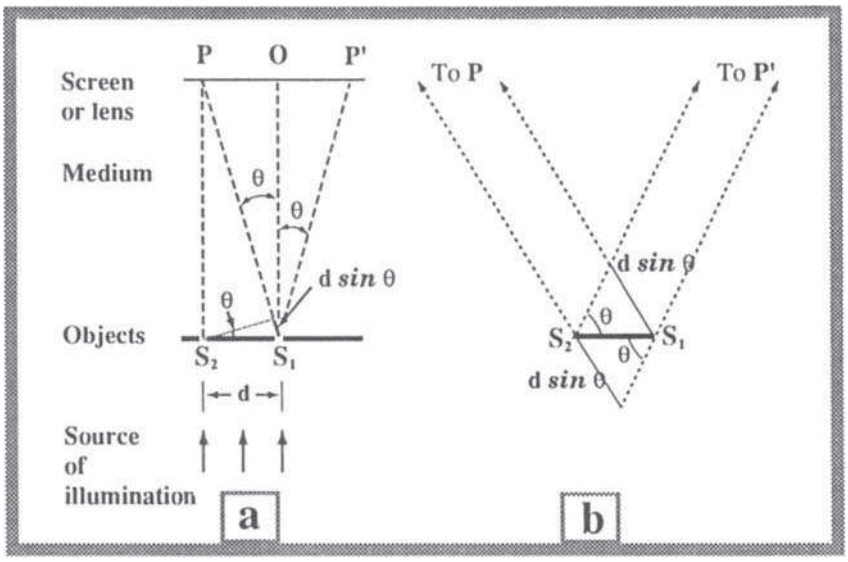


FIG. 1.19 Basic optical principle governing the operation of an optical microscope: (a) the geometry on which the resolving power d of a microscope is based; (b) detail showing how light from both sources must be intercepted by the lens to become part of the image.

the radiation and probing technique used. For example, in the case of scattering techniques, discussed in Chapter 5, contrast arises from the refractive index difference between the particle and surroundings when light scattering is used, whereas in neutron scattering it is the difference in the interaction between the neutrons and the atomic nuclei that give rise to the contrast.

1.6a.1b Resolution and Magnification. The first thing we tend to think of in connection with microscopes is the magnification they achieve. More important, however, is a quantity known as the *resolving power*, or *limit of resolution*, of the microscope. Magnification determines the size of an image, but the resolving power determines the amount of distinguishable detail. Enlargement without detail is of little value. For example, a row of small spherical particles will appear simply as a line if the row is enlarged with an instrument of poor resolving power. Further magnification would increase the thickness of the line, but would not reveal its particulate nature. If the resolving power is increased, however, the individual spheres would be discernible. One may then choose a magnification that is convenient. Both the depth and area of the in-focus field decrease as the magnification is increased, so one pays a price for enlargement even though the amount of perceptible detail is not affected much by the magnification.

A beam of light is always diffracted at the edges of an object to produce a set of images of the edge known as a diffraction pattern. The diffracted light is what our eye receives from an object and from which our image of the object is constructed. The diffracted light contains sufficient information to assemble an image; any light that is not incorporated into the image will result in a loss of detail in the image. To estimate the efficiency with which an image reproduces the object, let us consider the situation in which all first-order diffracted light is intercepted by the lens, but light from all higher orders of diffraction is assumed to be lost.

Suppose we have a set of pinholes in an opaque shield and that this shield is illuminated by a source far enough away that the incident light may be regarded as a set of parallel rays, as sketched in Figure 1.19a. To an observer on the opposite side of the shield, each pinhole will function as a light source from which a hemispherical wave front seems to emerge. If adjoining pinholes are separated by a distance d, the wave from one source, say S_1, must travel a longer distance than light from S_2 to reach a distant screen at point P. If the distance between the shield and the screen is large compared to the wavelength, the extra distance may be equated with $d \sin \theta$. Positive reinforcement through diffraction occurs whenever such an extra distance equals some integral number of wavelengths of light. For first-order diffraction, then, we require

$$d \sin \theta = \lambda' \tag{21}$$

where λ' is the wavelength in the medium between the screen and the shield. If the medium has a refractive index n, then $\lambda' = \lambda/n$ where λ is the wavelength under vacuum. Therefore, Equation (21) becomes

$$d \sin \theta = \lambda/n \tag{22}$$

Equation (21) is called the *Bragg equation* after the father-and-son team of W. H. and W. L. Bragg (Nobel Prize, 1915); it is the underlying relationship for all diffraction phenomena. We encounter the Bragg equation again in Chapter 9 when we discuss the diffraction of low-energy electrons by surface atoms (see Section 9.8b).

Recall that our objective is to consider an image constructed from only first-order diffracted light. To do this, we identify the perforated shield S_1S_2 as the object, the screen as the objective lens of the microscope, and the distance PP' as the diameter of the objective. It is clear from Figure 1.19b that light originating at S_1 must travel a distance $d \sin \theta$ longer than the light from S_2 to form an image at P; likewise, light originating at S_2 must travel a distance $d \sin \theta$ longer than light from S_1 to form an image at P'. Thus, to be intercepted by the lens and thereby become part of the image, rays from different parts of the source must travel paths that differ by $2d \sin \theta$. In order for this to happen and still consist of only first-order diffracted light, the difference in path lengths must equal λ'. The resolving power is defined to be the magnitude of the separation between objects that is required to produce discernibly

different images when the angle subtended by the microscope is 2θ. Therefore, the resolving power is identical to d in Figure 1.19 and is estimated to be

$$d = \lambda/(2n \sin \theta) \tag{23}$$

which is sometimes written

$$d = \lambda/2(NA) \tag{24}$$

where NA (i.e., $n \sin \theta$) is called the *numerical aperture* of the lens.

The significance of Equation (23) is that any points closer together than the distance d will produce a first-order diffraction image with light having a wavelength equal to λ/n at some angle greater than θ. This light would not be intercepted by the lens PP', so a significant amount of detail about distances of this magnitude is lost. It is the wave nature of light that imposes a limit to the amount of detail an image may possess. Equation (23) shows that the resolving power is decreased by increasing θ or by decreasing λ/n. The subtended angle 2θ is increased by increasing the diameter of the lens and by decreasing the distance between the object and the lens; the design of the lens limits the range of these parameters. The ratio λ/n may be decreased by decreasing λ or by increasing n. Although shorter wavelengths improve resolving power, visible light is almost always used in microscopy, primarily because of the absorption of shorter wavelengths by glass. Since the refractive index of some oils is 50% higher than that of air, a significant improvement in the resolving power is achieved by filling the gap between object and lens with so-called immersion oil.

We may use Equation (23) to estimate the particle size that will be clearly resolved microscopically. As a numerical example, we may take λ to be 500 nm, $n = 1.5$, and $2\theta = 140°$, all of which are attainable and are optimal values for these quantities. This leads to a resolving power of 142 nm, which represents the lower limit of resolved particle size under completely ideal circumstances. Under closer to average circumstances, a figure twice this value may be more typical. These figures reveal that direct microscopic observation is feasible only for particles at the upper end of the colloidal size range and then only if the particle contrasts sufficiently in refractive index with its surroundings. This is the reason for exploring the use of other radiations such as x-ray and neutron and other techniques to "visualize" submicroscopic particles and features of interest in colloid science. We briefly consider a rather straightforward extension of optical microscopy, i.e., electron microscopy, in Section 1.6a.2.

1.6a.1c Dark-Field Microscopy. There are many variations of optical microscopic techniques in use today. One variation of direct microscopic examination extends the range of microscopy considerably but at the expense of much detail. By this technique, known as *dark-field microscopy*, particles as small as 5 nm may be detected under optimum conditions, with about 20 nm as the lower limit under average conditions. In dark-field microscopy, the sample is illuminated from the side rather than from below as in an ordinary microscope. If no particles were present, no light would be deviated from the horizontal into the microscope, and the field would appear totally dark. The presence of small dispersed particles, however, leads to the scattering of some light from the horizontal into the microscope objective, a phenomenon sometimes called the *Tyndall effect*. The presence of colloidal particles is indicated by minute specks of light in an otherwise dark field.

In dark-field microscopy, the particles are only a blur; no details are distinguishable at all. Some rough indication of the symmetry of the particles is afforded by the twinkling that accompanies the rotation of asymmetrical particles, but this is a highly subjective observation. However, the technique does permit the rate of particle diffusion to be observed. We see in Chapter 2 how to relate this information to particle size and shape. The number of particles per unit volume may also be determined by direct count once the area and depth of the illuminated field have been calibrated. This is an important technique for the study of coagulation kinetics, a topic we discuss in Chapter 13.

In dark-field, as in direct observation microscopy, the refractive index difference between

the particles and the medium is one of the crucial factors that influences the feasibility of the method. Very good resolving power is obtained with metallic particles, which offer maximum contrast in refractive index. The technique of dark-field microscopy played an important historic role in colloid science, particularly in the study of metallic colloids. R. Zsigmondy (Nobel Prize, 1925), for example, made extensive use of this technique in his study of colloidal gold. Note that it is the deviation of a beam of light or scattered light that is utilized in dark-field microscopy. We have a good deal more to say about light scattering in Chapter 5.

1.6a.2 Electron Microscopy and Scanning Probe Microscopies

There are numerous modern developments that have made atomic-scale resolution possible in recent years. In fact, some of these developments in instruments can also be used to measure forces between particles and surfaces. These developments for force measurements are discussed briefly in Section 1.6c and in Vignette I.8. In this section, we review electron and scanning probe microscopies (SPMs), which allow atomic-scale visualization of surfaces and particles.

1.6a.2a Transmission and Scanning Electron Microscopy. It is clear from Equation (23) that the best prospect for extending the range of microscopy lies in the extension by orders of magnitude of the wavelength used to produce the image. The wave-particle duality principle of modern physics shows us how to make this extension. According to the *de Broglie equation* (see Atkins 1994), the wavelength of a particle is inversely proportional to its momentum:

$$\lambda = h/mv \tag{25}$$

where h is Planck's constant and m and v are the particle mass and velocity, respectively. In an electron microscope, a beam of electrons replaces light in producing an image. An electron beam is produced by a hot filament, accelerated by an electron gun, and focused by electric or magnetic fields that function as lenses. In this section, we consider primarily a technique based on the transmitted electron beam. Accordingly, the method is called *transmission electron microscopy* (TEM). Some brief remarks are made near the end of this section about a technique known as *scanning electron microscopy*.

A schematic comparison of light and electron microscopes is shown in Figure 1.20. Although wavelengths on the order of 10^{-12} m (1 picometer or 1 pm) are easily achieved in electron microscopes, the numerical apertures of such microscopes are low. Accordingly, the resolving power of a conventional electron microscope is generally about 1 nm. However, because of this remarkable resolution, small features may be enormously magnified without loss of detail.

The intensity of the electron beam that is transmitted through the specimen under observation in an electron microscope depends on the thickness of the sample and the concentration of atoms in the sample. Thus, unless there are large differences in these characteristics between the particles of the sample and the support on which they rest, a very low contrast image will be produced. A poor-quality image results from these low-contrast situations; the image is analogous to a landscape viewed from an airplane with the sun directly overhead. The most common way of overcoming this difficulty is by means of a technique known as *shadow casting*. The sample to be examined in the electron microscope is placed in a chamber, which is then evacuated. Next, some gold or other metal is vaporized in the same chamber. The vapor condenses on all cooler surfaces, including the surface of the sample. If the vapor source is positioned to the side of the sample, the condensed vapor will deposit unevenly, effectively casting a shadow over the sample. The situation is shown schematically in Figure 1.21; Figure 1.12 shows actual photomicrographs enhanced by this technique. With shadow casting, the field shows much more detail, just as the view from an airplane does when the sun is lower in the sky.

When the angular position of the vapor source relative to the sample is known, the thickness of the particle may be evaluated from the length of its shadow by a simple trigono-

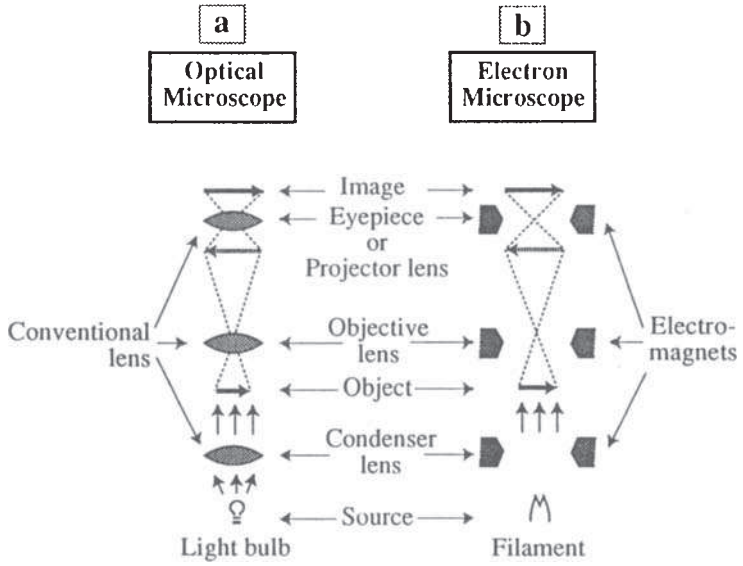


FIG. 1.20 Schematic comparison of (a) light and (b) electron microscopes showing components that perform parallel functions in each.

metric calculation. Alternatively, the length of the shadow cast by a spherical reference particle, introduced for calibration, may be used as a standard to calculate the thickness of particles.

In contrast to the TEM, the scanning electron microscope (SEM) works somewhat like a TV cathode-ray tube. An electron gun is used to scan a surface with the help of deflection coils and focusing "lenses" in a raster pattern, and the "reflected" electron signal is used synchronously to generate an image of very high magnification on a screen. The SEM has been in

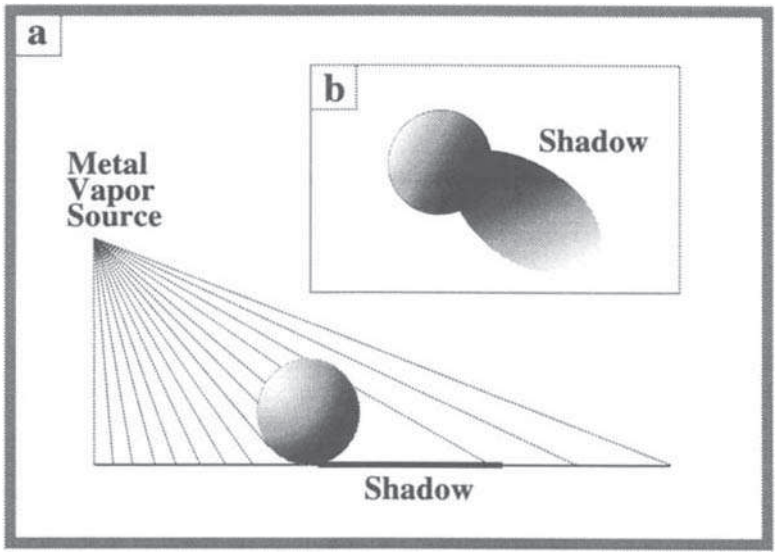


FIG. 1.21 Shadowing of a spherical particle using metal vapor: (a) side view; and (b) top view.

routine use since the 1960s and can be combined with spectroscopic methods to provide a powerful array of techniques for surface analysis (Adamson 1990; Hubbard 1995).

1.6a.2b Scanning Probe Microscopy. At the beginning of this section, the microscope was proposed as an ideal tool for the study of colloidal particles. The light microscope is limited in range, but the electron microscope clearly has access to the entire colloidal size range. However, a very real disadvantage still persists even with the conventional electron microscope. The transmission and scanning electron microscopes require that the sample be placed in an evacuated enclosure for examination by the electron beam. Therefore, a dispersion must be evaporated to dryness before examination. The result, although still informative, bears about as much resemblance to the original dispersion as a pressed flower bears to a blossom on a living plant. These limitations have served in recent years as incentives for the development of newer techniques such as environmental scanning electron microscopy (E-SEM) and scanning probe microscopy (SPM), some of which are highlighted in Vignette I.8 (see also Hubbard 1995). Each of these has its own advantages as well as disadvantages, but they all have increased our ability to "see" details that were not possible to see as recently as a decade ago.

For example, the E-SEM, introduced around 1988, is an outcome of a serendipitous discovery of a new method of signal detection and combines the high resolution and the depth of field of the SEM with the flexibility and the ease of use of the optical microscope (Baumgarten 1989). It allows the use of high pressures (over 20 torr) and the examination of specimens in their natural state so that biological specimens and dynamic processes such as wetting and drying can be studied in real time. Some of the SPMs, on the other hand, require very low pressure environments, but others allow natural environments (see Vignette I.8 and Section 1.6c).

In subsequent chapters, we discuss some *in situ* techniques for the characterization of colloidal particles, especially with respect to particle size, structure, and molecular weight.

VIGNETTE I.8 NEW EXPERIMENTAL TOOLS: Scanning Probe Microscopes for Surface Analysis at the Atomic Scale

Speculations about the existence of atoms are often traced to ancient times, but a fierce (and sometimes vitriolic) debate was being waged in scientific circles about the reality of atoms even as recently as the dawn of the 20th century. In fact, Ludwig Boltzmann's depression that drove him to suicide in 1906 is partly attributed to the criticism he faced from his detractors for his staunch advocacy of the atomic theory of matter. Who, then, would have thought that before the end of the century we would have the capability to not only "see" atoms, but to manipulate and move them one by one on a surface? Yet, that is what has been made possible by the ingenuity of Gerd Binnig and Heinrich Rohrer, two physicists at the IBM Zürich Research Laboratory, who invented what is known as the *scanning tunneling microscope* (STM) in 1982—an invention for which they shared the 1986 Nobel Prize in Physics with Ernst Ruska of Germany, the inventor of the electron microscope.

When a probe, usually a metal tip down to the size of an atom, is brought to within a nanometer from the surface of a conductor or a semiconductor (see Fig. 1.22), electrons can be made to jump from the surface to the tip by applying what is known as a *bias potential*; this is called the *tunneling effect*, and the resulting current is known as the *tunneling current*. By keeping the tunneling current constant as one moves the probe along a surface, one can follow the surface contour with atomic-scale precision! In the constant current mode of operation, for example, the vertical displacement of the probe can be mapped with angstrom-level precision using the deflections of the piezoelectric ceramic piece on which the metal tip is mounted. The data can be converted to images to allow one to see the surface (see Fig. 1.22, inset).

In fact, what is mapped is not the topography of the surface but contours of constant electron densities; that is, the results are sensitive to the type of atom seen on the surface by

the probe. This implies that the STM can serve as a powerful and highly specific and sensitive local probe to study a surface (and materials adsorbed on a surface) with high precision.

Vacuum tunneling of electrons is not the only option, however. The invention of STM has triggered the development of a whole family of techniques known collectively as *scanning probe microscopy* (SPM) that are not restricted to tunneling as the mechanism or to only conducting or semiconducting surfaces. One type of SPM, known as the *atomic force microscopy* (AFM; see Section 1.6c.2), can be used to measure forces such as van der Waals forces, ion-ion interactions, and hydration forces on polymer-coated surfaces, biologically relevant bilayers, and so on. In the force-measuring mode, an SPM is often referred to as a surface *force* microscope (SFM), but the label SPM signifies the more general capability of these devices. For example, atoms on surfaces can be manipulated and rearranged individually using STM and SPM, and conformation of polymers and biological macromolecules on surfaces can be probed.

In addition, the STM can be combined with suitable spectroscopic techniques to design powerful local probes of a surface to a precision not possible through other techniques; these are known as *scanning probe spectroscopy*. Because of its precision, the STM/SPM family can provide details that are often missed by other techniques. For example, we discuss a diffraction technique known as *low-energy electron diffraction* (LEED) in Chapter 9 (Section 9.8) for mapping the arrangement of atoms on a surface. Sharp diffraction spots obtained through LEED are taken to be indicative of ordered arrangements of the atoms on a surface. However, we know now from STM studies that even a surface showing clear and sharp diffraction spots in LEED experiments can in fact be highly defective. (This illustrates why more than one experimental technique [especially independent techniques] is often necessary in research.)

The development of SPMs has significantly changed the "landscape" of the instrumentation available to colloid and surface scientists, and the possibilities are endless!

1.6b Size and Molecular Weight Measurements

1.6b.1 Particle Size Measurement

Particle size measurement is one of the essential requirements in almost all uses of colloids. However, our discussion in Section 1.5 makes it clear that this is no easy task, especially since even the definition of *particle size* is difficult in many cases. A number of techniques have been developed for measuring particle size and are well documented in specialized monographs (e.g., Allen 1990). Optical and electron microscopy described in the previous section can be used when *ex situ* measurements are possible or can be acceptable, but we also touch on a few nonintrusive methods such as static and dynamic light scattering (Chapter 5) and field-flow fractionation (see Vignette II; Chapter 2) in other chapters.

In this chapter, we restrict our discussion to a chromatographic technique normally used for molecular weight measurements. The chromatographic concept can also be used for direct size (instead of molecular weights) measurement in the case of rigid particles, as we illustrate in our description of field-flow fractionation methods in Chapter 2.

1.6b.2 Molecular Weights

Lyophilic colloids, especially those of synthetic origin, also possess a distribution of particle sizes, but these are generally characterized by molecular weight rather than by linear dimension (except when R_g is measured for such systems). We discuss the experimental methods used for molecular weight determination of particles in the colloidal size range in Chapters 2–5. These methods yield the kinds of molecular weight averages discussed in Section 1.5c.3. Here, we consider a colloidal solution, say polystyrene in toluene, and discuss a chromatographic method for its fractionation.

1.6b.2a Chromatographic Techniques. To begin, a polymer like polystyrene is formed by the addition reaction in which styrene monomers $CH_2{=}CHC_6H_5$ combine into long chains that may be thousands of repeat units long. This reaction is one of random addition, so not all

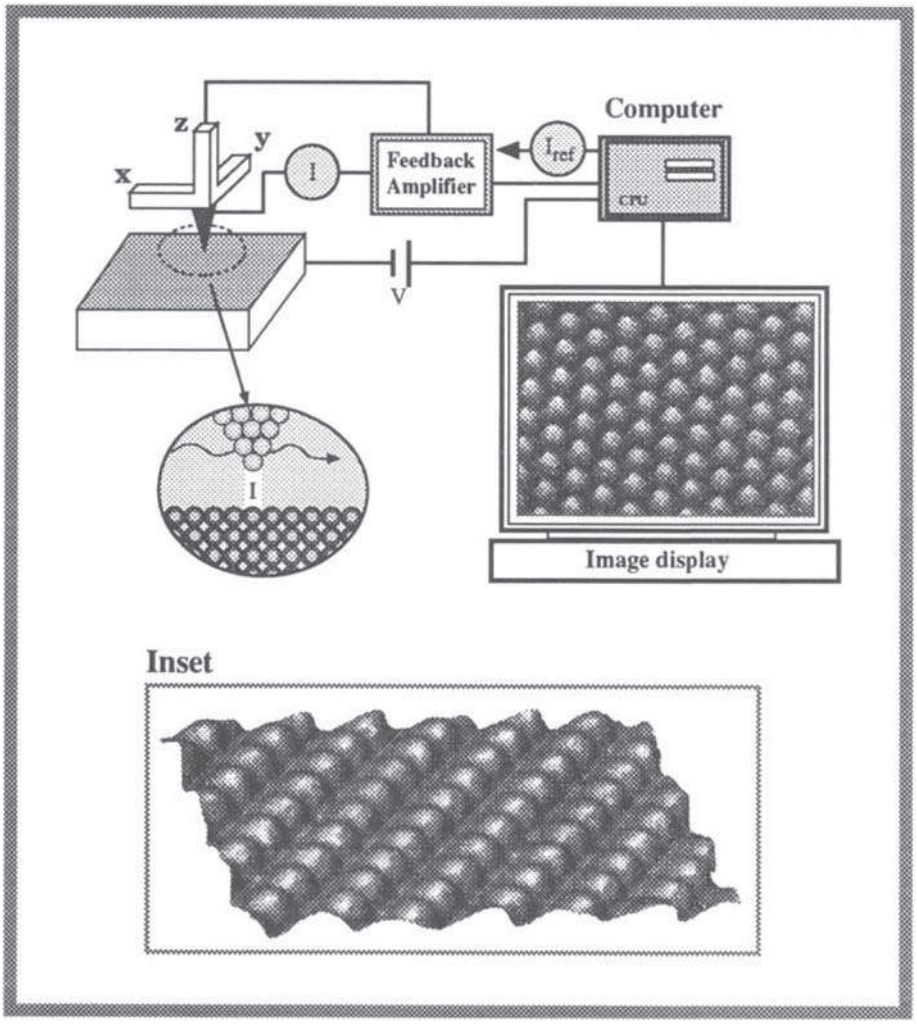


FIG. 1.22 One type of operation of a scanning tunneling microscope (STM). A tunneling current I flows between the sharp tip of the probe and the surface when a bias voltage V is applied to the sample. A computer monitors the tunneling current and adjusts the distance between the probe and the surface such that a constant tunneling current I_{ref} is maintained. The resulting changes in the position of the tip are then recorded and converted to an image such as the one shown on the monitor or the one shown in the inset. The image shown in the inset is that of an atomically smooth nickel surface. The periodic arrangement of the atoms on the surface can be seen clearly in the STM image. (Adapted with permission of C. M. Lieber, *Chem. & Eng. News* **72**(16), 28 (1994). The inset is adapted from T. A. Hoppenheimer, in *A Positron Named Priscilla: Scientific Discovery at the Frontier*, M. Bartusiak, B. Burke, A. Chaikin, A. Greenwood, T. A. Heppenheimer, M. Hoffman, D. Holzaman, E. J. Maggio, and A. S. Moffat, Eds., National Acad. Press, Washington, DC, 1994.)

chains are identical in length. Furthermore, chains containing slightly different numbers of repeat units will not differ enough in their properties to allow complete separation. Nevertheless, procedures exist in which an initially polydisperse system is divided into fractions of narrower molecular weight distribution. One such method is called *size-exclusion chromatography* (SEC) because it segregates molecules in the distribution on the basis of their spatial extensions (see Section 1.6b.2b).

In chromatography, a cylindrical column is packed with porous particles and then filled with solvent. Next, a portion of the colloidal solution is layered on the solvent, and the liquid is allowed to pass through the column. The eluted liquid is monitored for the colloid by an appropriate detection method; spectrophotometry and refractive index are probably the most widely used methods of detection. What results is a *chromatogram* showing the detector output as a function of the volume of liquid eluted through the column. For a synthetic colloid like polystyrene, a broad peak results. Samples of biological origin often contain several components of quite different molecular weight. Under optimum conditions, these may emerge as distinct peaks in the chromatogram.

The volume at which a particular colloidal fraction emerges from the column is called the *retention volume* V_R of that fraction. Normally, a particular experimental system is calibrated with a colloidal solute of known molecular weight M. A plot of log M *versus* V_R is generally linear over several orders of magnitude in molecular weight. Not only the solute but also the solvent, the column packing, and the operating conditions affect the calibration. Therefore, the calibration is set up on a system-by-system basis using molecular weights that have been determined by one of the absolute methods we discuss elsewhere in this book. Thus, the calibration process can be time consuming, but, once set up, chromatography provides a rapid method for fractionating polydisperse systems.

Figure 1.23 shows how log M and the detector output vary with V_R. By reading from the chromatogram to the calibration curve, the molecular weight of the fraction emerging at a

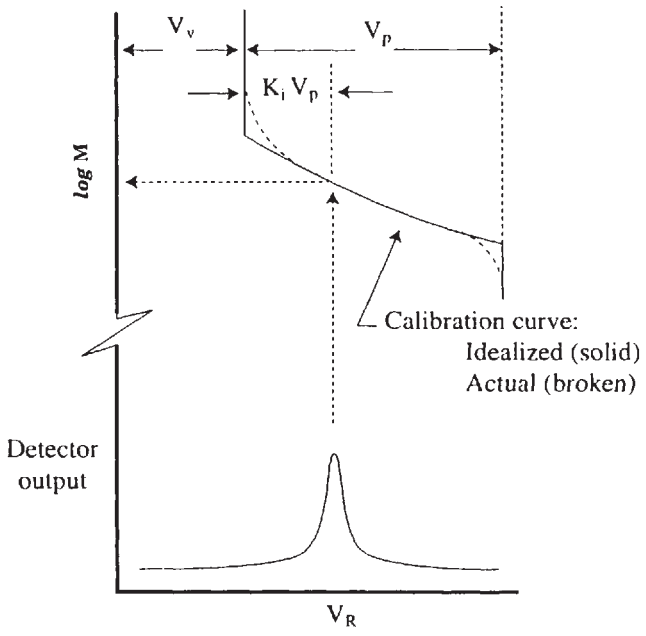


FIG. 1.23 Plot of log M and detector output versus retention volume for size-exclusion chromatography. Also shown is the relation among V_R, V_V, V_P, and $K_i V_p$ as discussed in the text. (Redrawn with permission of P. C. Hiemenz, *Polymer Chemistry: The Basic Concepts*, Marcel Dekker, New York, NY, 1984.)

particular value of V_R can be determined. Furthermore, for a molecular weight fraction M_i, the height h of the detector output signal is directly proportional to the mass m_i of the solute eluted in that particular fraction. Since $m_i \propto h_i$, $\Sigma_i h_i \propto \Sigma_i m_i$ and $m_i/\Sigma_i m_i = h_i/\Sigma_i h_i$. Thus, the weight fractions of different molecular weight classes in the distribution are readily obtained by this method. Likewise, if n_i is the number of moles of component i, $h_i/M_i \propto n_i$ and $(h_i/M_i)/\Sigma_i h_i/M_i = n_i/\Sigma_i n_i$. If we assume that suitable calibration exists, the chromatogram produced by SEC can be interpreted in terms of either number fractions or weight fractions of the various molecular weight classes. This kind of information can be used to calculate molecular weight averages (see Section 1.5c.3).

Incidentally, this method is known by various names among workers in different fields. In biochemistry, the technique is called *gel filtration chromatography* (GFC), and in polymer chemistry, it is *gel permeation chromatography* (GPC).

1.6b.2b Size-Exclusion Chromatography. Now, however, let us briefly consider the mechanism by which molecules are fractionated by SEC. The basic idea is very simple. Large molecules cannot penetrate into the pores of the packing medium (the gel) and hence are eluted first. The smaller particles are distributed in the pores as well as in the voids between the gel particles and emerge later.

The total volume of solvent in the column can be divided into two categories: V_v is the volume of solvent in the voids between the gel particles, and V_p is the volume of solvent in the pores. Not all of the pore volume is accessible to particles in a given molecular weight class. We define $K_i V_p$ as the pore volume into which molecules from the ith class can permeate. Thus, K_i describes what fraction of a pore is accessible to molecules in class i. The fraction K_i is zero for very large particles and unity for very small particles and varies over this range for particles of intermediate size. The retention volume for molecules in class i is given by

$$V_R = V_v + K_i V_p \tag{26}$$

Figure 1.23 shows the relationship among V_R, V_v, V_p, and K_i.

It is fairly easy to arrive at a theoretical expression for K_i based on a simple model of the permeation process. We picture the colloidal particle as a sphere of radius R_s and the pore as a cylinder of radius R_c and length ℓ, as shown in Figure 1.24a (Giddings 1991). The center of the solute particle cannot approach any closer than a distance R_s from the wall of the pore. Therefore, the radius of the pore that is accessible to the colloidal particle is $(R_c - R_s)$. The layer of solution adjacent to the walls of the pore is "off limits" to the solute, so the concentration of the colloid in the pore is only a fraction of its value in the bulk solution. This fraction is K_i for the sphere of radius R_s and equals the ratio of the accessible volume to the total volume of the pore:

$$K_i = \frac{\pi(R_c - R_s)^2\,(\ell - R_s)}{\pi R_c^2 \ell} \approx \left(\frac{R_c - R_s}{R_c}\right)^2 = \left(1 - \frac{R_s}{R_c}\right)^2; \ell >> R_s \tag{27}$$

Equation (27) is an approximation valid when the pore is long enough so that $(\ell - R_s) \approx \ell$, so that the fraction on the right-hand side of the equation ranges between zero and unity as required for R_s equal to R_c and 0, respectively, and thus gives the parameter K_i as a function of particle size.

The next stage of developing a theory would be to relate the radius of the particle to the molecular weight. Thus expanded, Equations (26) and (27) would provide a relationship between M and V_R that—if the theory were successful—would eliminate the need for empirical calibration of SEC. For some systems, such as rigid globular proteins, establishing the connection between R_s and M is not difficult. For synthetic polymers, however, the flexible chains exist as random coils with an extension that depends on the interaction between polymer and solvent. In this case, relating R_s to M is a complicated matter, so we are content with empirical calibration to interpret V_R in terms of M. This, of course, is not the only modification that needs to be examined in a fully developed theory for K_i. Models based on more complicated geometries for both the colloidal particles and the pores have been considered, but we shall

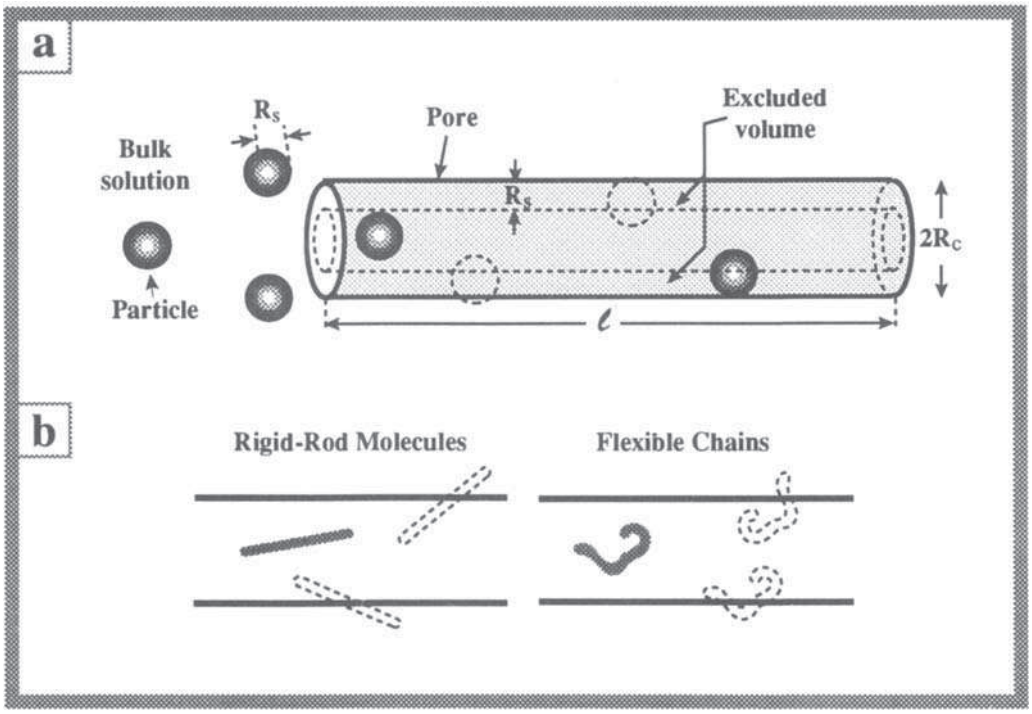


FIG. 1.24 Size exclusion of a particle in a pore: (a) exclusion of a spherical particle in a cylindrical pore; and (b) exclusion of macromolecules. Particles shown in dashed lines indicate positions or orientations that are excluded. (Redrawn with modifications from Giddings 1991.)

not pursue these. Yau et al. (1979) discuss additional theoretical models and experimental procedures in detail for those desiring more information on this topic.

It is important to note that the implication of "size exclusion" is more general than its use in SEC might imply. The exclusion of particles based on size and shape (as illustrated for a simple case in Fig. 1.24b) finds applications in other important contexts. For instance, specially designed zeolite catalysts are used in industry for accomplishing highly selective reactions, and the accessibility of the reactants to the sites on the surfaces of such catalysts is controlled using size-exclusion (and shape-exclusion) principles. This is illustrated in Vignette I.9. The size-exclusion and shape-exclusion properties of the zeolites of the type described in Vignette I.9 are also taken advantage of in industrial separation processes (the zeolites are often called *molecular sieves* for this reason).

VIGNETTE I.9 SIZE EXCLUSION IN HETEROGENEOUS CATALYSIS AND SEPARATION PROCESSES: Zeolites—Molecular Cages for Selectivity in Separation and Catalysis

The adjective *heterogeneous* in heterogeneous catalysis draws attention to the fact that an adsorbent in the form of one phase (e.g., solid) assists in reactions between molecules in another phase (e.g., gas). When a molecule adsorbs on a surface, whether by chemisorption or physisorption (see Chapter 9), one or more of its atoms form strong (chemisorption) or

relatively weak (physisorption) bonds with the surface. This leads to breaking or weakening of other bonds within the adsorbed molecules, thus promoting reactions with other species present in the system. The adsorbent in this instance contributes to the disassembly of the adsorbate and becomes a catalyst (see Vignette IX in Chapter 9).

In addition to a large catalytically active surface and good capacity for adsorption, an efficient catalyst requires *selectivity*, that is, preferential affinity for the appropriate reactants. Nonselectivity is a source of significant problems in catalysis, particularly in the petrochemical industry, which has to deal with hydrocarbons of various types and isomers in a single stream. This situation has provided a strong incentive for the development of artificial (man-made) catalysts that offer the type of selectivity unthinkable on metal surfaces discussed in the last section of Chapter 9 (see, for example, Ball 1994) and illustrates another example of molecular design (or "molecular engineering") of advanced materials for use in science and industry.

Zeolites, aluminosilicates originally found as minerals in nature, can now be synthesized in various forms and have intricate cagelike labyrinths (see Fig. 1.25) with shape- and size-selective properties. The channels in these materials can be enlarged, with remarkable precision, by substituting phosphorus for silicon. Precise selectivity is achieved by a number of mechanisms, the simplest being the result of the size of the channels, as illustrated in Figure 1.25. Molecular shape of the reactants and products also plays a major role. These mechanisms influence the selectivity of absorption (notice the spelling; *absorption* refers to the amount of gas molecules taken in by a material and not necessarily to the ability of a surface to adsorb the gas). But *adsorptive* selectivity is imparted through the chemical elements used in the material, as different chemical constituents have different "talents" (Ball 1994) for adsorption and for promoting reactions.

The zeolites are also known as *molecular sieves* because of their capacity to discriminate between molecules; they find numerous uses in separation and catalytic processes. Although they appear to be "solid" particles to the naked eye, they are highly porous, with a typical specific surface area of about 1000 m^2/g. Catalysis is discussed in Chapter 9, but the scope of that chapter does not permit detailed discussions of the various types of catalysts and the role of physisorption and chemisorption in catalysis; this vignette provides a glimpse of the rationale used in the molecular design of new materials of interest in surface chemistry and how the concepts introduced in Chapter 1 and Chapter 9 fit into the larger scheme.

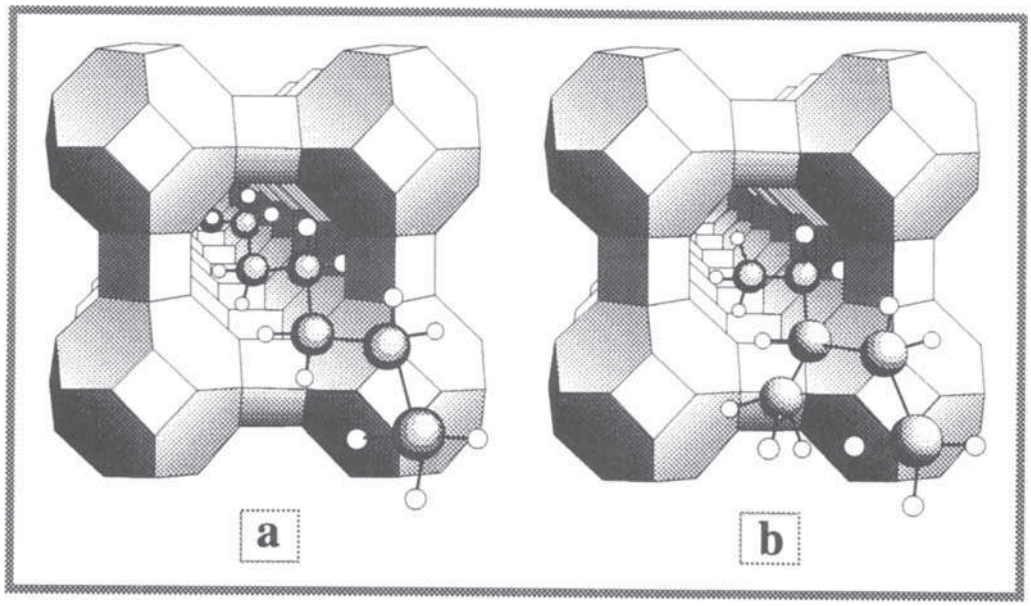


FIG. 1.25 Shape selectivity of a zeolite cage. The cage allows (a) straight-chain hydrocarbons to snake their way into the pores while (b) preventing branched-chain hydrocarbons from entry. (Adapted from Ball 1994.)

1.6c Force Measurements

We have already seen in the sections above and in the vignettes that interaction forces between surfaces and among particles are important in colloidal systems. Stability of colloids, adhesion and wetting phenomena, xerography and printing, and the like depend on the types of forces that exist between surfaces and how they vary with physical and chemical conditions. Such forces may be approached or understood on at least two different levels. The first is a coarser level, e.g., the magnitude and strength of the surface forces as they manifest themselves in macroscopic phenomena. Examples of these include the magnitude of the repulsive barrier needed to stabilize a dispersion from coagulation and the strength of adhesive force needed for an adhesive film to meet the performance requirement. The second is on a finer scale, as in obtaining a detailed "force law" for interaction between two surfaces as a function of distance of separation. We discuss examples of these in Chapters 10 and 11.

What is needed or sufficient depends on the application. For engineering applications, the first level might often be sufficient, whereas for developing new materials or understanding chemical and biological phenomena involving surfaces (see the vignette in Chapter 12), a more detailed understanding is necessary.

A number of techniques have been developed over the years to determine colloidal and surface forces as well as interatomic and intermolecular forces of interest in colloid and surface chemistry. These methods can be divided into two groups: (a) indirect methods and (b) direct methods. We discuss examples of both in other chapters. It is therefore useful to consider these methods here briefly.

1.6c.1 Indirect Methods

Indirect methods are those in which we use some macroscopic property or phenomenon to deduce information about (colloidal or intermolecular) interaction forces. These methods are known as *indirect* since additional assumptions or approximations about the relation between the forces and the property or phenomenon measured are needed. Some examples of these methods are illustrated in Figure 1.26 and are described below (Israelachvili 1991).

1. Figures 1.26a and 126b illustrate particle detachment and peeling experiments for estimating information on attractive short-range forces that are important in particle adhesion and adhesion energies between solid surfaces in contact. Such experiments provide information of importance in powder technology, xerography, ceramic processing, and the making of adhesive films.
2. As we see in Chapter 6, surface tension and contact angle measurements provide information on liquid-liquid and solid-liquid adhesion energies (Fig. 1.26c). Contact angles measured under different atmospheric environments or as a function of time provide valuable insights into the states of surfaces and adsorbed films and of molecular reorientation times at interfaces.
3. The thicknesses of free soap films and liquid films adsorbed on surfaces (Figs. 1.26d and 1.26e), which can be measured using optical techniques such as reflected intensity, total internal reflection spectroscopy, or ellipsometry as functions of salt concentration or vapor pressure, can provide information on the long-range repulsive forces stabilizing thick wetting films. We see an example of this in Chapter 11.
4. More sophisticated methods such as nuclear magnetic resonance (NMR), light scattering, x-ray scattering, and neutron scattering can be used to measure dynamic interparticle separations and motions in liquids (Figs. 1.26f and 1.26g). In such experiments, the "particles" can be globular or spherical (e.g., micelles, vesicles, colloidal particles, latex particles, viruses), sheetlike (e.g., clays, lipid bilayers), or rodlike (e.g., DNA). The interparticle forces can be studied as functions of the solution conditions and the mean separation between particles by changing the hydrostatic or osmotic pressure via a semipermeable membrane. Notable among these techniques is the compression cell or osmotic pressure technique from which the interaction force between particles can be obtained from the deviations from ideality in the PVT data (see Chapter 10

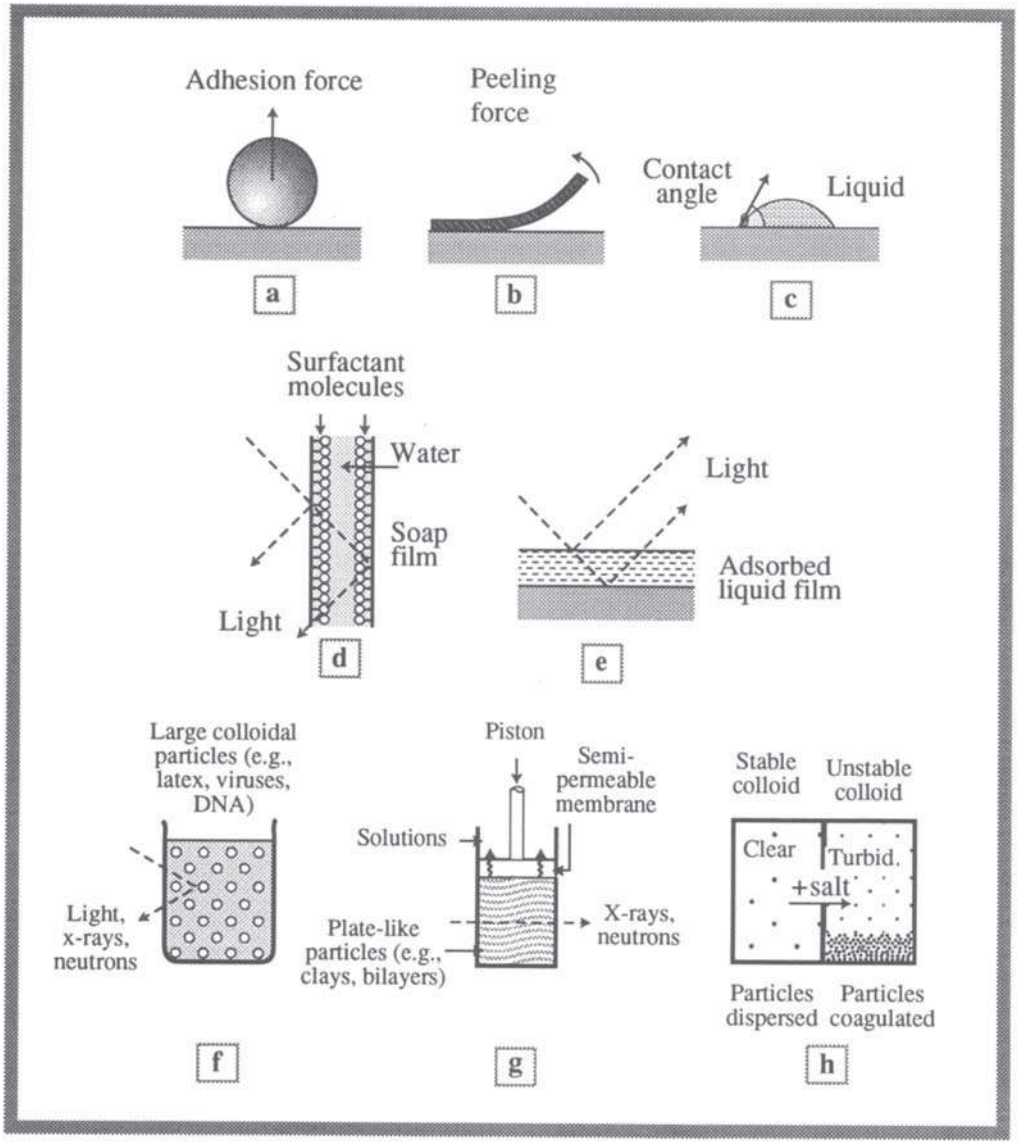


FIG. 1.26 Examples of various types of measurements that provide information on the forces between particles and surfaces: (a) adhesion measurements; (b) peeling measurements; (c) contact angle measurements (see Chapter 6); (d) equilibrium thickness of thin free films; (e) equilibrium thickness of thin adsorbed films (examples of practical applications include wetting of hydrophilic surfaces by water, adsorption of molecules from vapor, protective surface coatings and lubricant layers, photographic films; see Chapters 6 and 9); (f) interparticle spacing in liquids (examples of applications include colloidal suspensions, paints, pharmaceutical dispersions; see Chapter 13); (g) sheetlike particle spacings in liquids (examples of practical applications include clay and soil-swelling behavior, microstructure of soaps and biological membranes); (h) coagulation studies (e.g., basic experimental technique for testing the stability of colloidal preparations; see Chapter 13). (Redrawn with permission of Israelachvili 1991.)

for an example using van der Waals equation of state). These techniques are usually limited to measuring only the repulsive parts of a force law.

5. Coagulation of colloidal dispersions (Fig. 1.26h) as a function of salt concentration, pH, or temperature of the suspending liquid medium can also be used to obtain information on the interplay of repulsive and attractive forces between particles in pure liquids as well as in surfactant and polymer solutions.

The above methods are not usually sufficient when detailed information on the strength of interaction forces as a function of separation distance between surfaces is needed. Moreover, methods such as osmotic pressure measurements involve collective interactions of many molecules or particles so that the data obtained are of a thermodynamic nature. Translating this information to interaction force as a function of distance is not straightforward and is often ambiguous. Direct force measurements have been made possible in recent years by developments such as the scanning force microscopy (SFM) mentioned in Vignette I.8 and discussed further in Section 1.6c.2b and have contributed to a significant change in the field of colloid and surface science. We consider two important examples of direct force measurement techniques below.

1.6c.2 Direct Methods

As pointed out by Israelachvili (1991), the principle of direct force measurements is usually very straightforward, but the challenge is in measuring very weak forces at very small intermolecular or surface separations that must be controlled and measured to within 0.1 nm. Following Israelachvili (1991), we divide our description into two parts, namely, surface force measurements and interatomic force measurements.

1.6c.2a Surface Forces: Surface Force Apparatus. The surface force apparatus (SFA) was originally developed in the late 1960s by Tabor, Winterton, and Israelachvili (see Israelachvili 1991) for measuring the van der Waals forces between molecularly smooth mica surfaces in air or vacuum. The apparatus, which has since been modified for making measurements in liquids, consists of two curved surfaces of mica, the interaction forces between which are measured using highly sensitive force-measuring springs (see Fig. 1.27a). The two surfaces are in a crossed cylinder configuration and geometrically correspond locally to a sphere near a flat surface or to two spheres close together. The separation between the two surfaces can be measured by use of an optical technique (which employs multiple beam interference fringes), from microns down to molecular contact, usually to better than 0.1 nm. In addition, the exact shapes of the two surfaces and the refractive index of the liquid (or material) between them can also be measured. The ability to measure refractive index allows one to determine the quantity of material (e.g., lipid or polymer) deposited or adsorbed on the surfaces.

The force is measured through the expansion or contraction of the piezoelectric crystal by known amounts and by determining optically how much the two surfaces have actually moved. The difference in the two values is then multiplied by the stiffness of the force-measuring spring to obtain the force difference between the initial and final positions. Thus, both repulsive and attractive forces can be measured as functions of the distance between interacting surfaces, with a sensitivity of the order of 10^{-8} N. Figure 1.27b illustrates an example of force measurement for the case of attraction.

In the past, mica has been the material of choice for the interacting surfaces because of the ease of handling and since molecularly smooth surfaces can be fabricated; mica surface coated with a thin film of other materials (e.g., lipid monolayers or bilayers, metal films, polymer films, or other macromolecules such as proteins) can also be used. The use of alternative materials such as molecularly smooth sapphire and silica sheets and carbon and metal oxide surfaces is also being explored.

The invention and refinement of the SFA have been among the most significant advances in experimental colloid science and have allowed researchers to identify and quantify most of the fundamental interactions occurring between surfaces in aqueous solutions as well as nonaqueous liquids. Attractive van der Waals and repulsive electrostatic double-layer forces, oscillatory (solvation or structural) forces, repulsive hydration forces, attractive hydrophobic

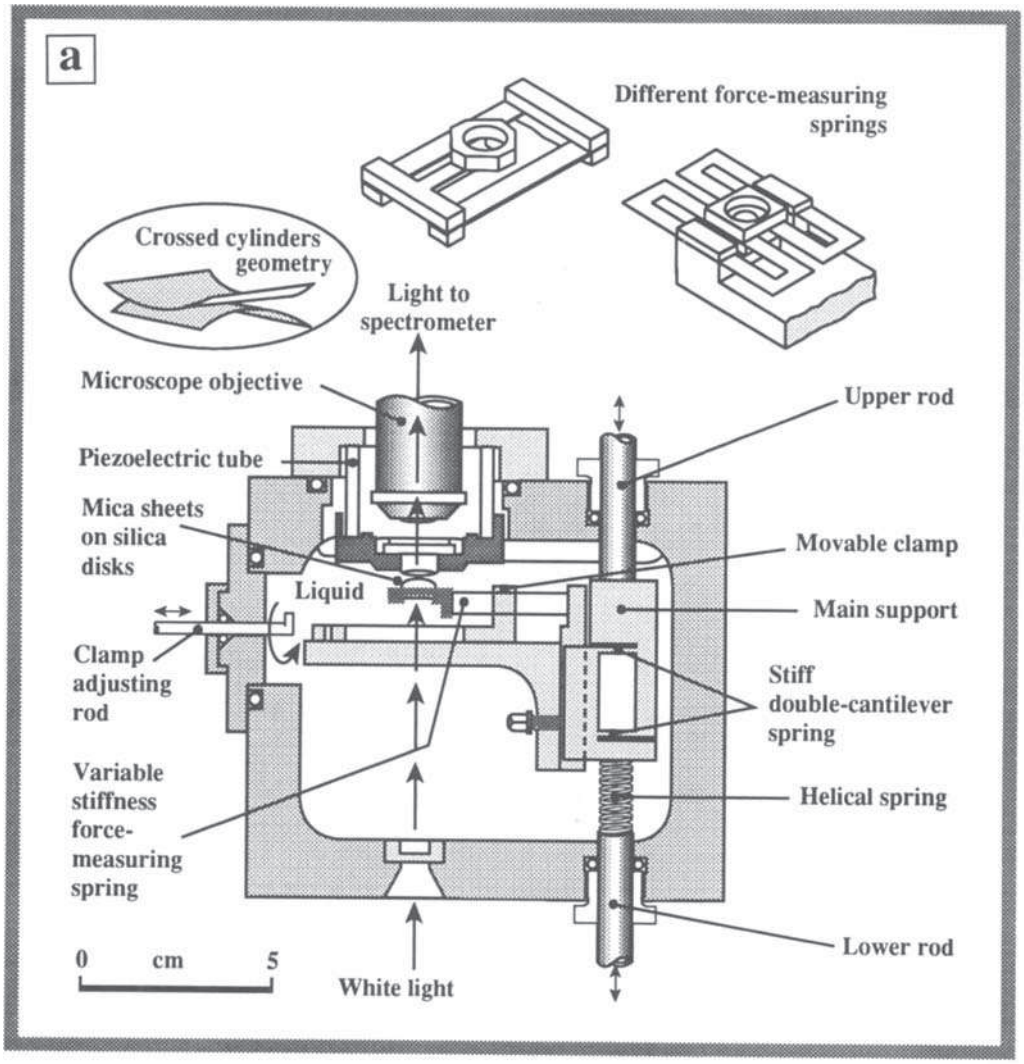


FIG. 1.27 The surface force apparatus (SFA) and an example of direct force measurement. (a) The surface force apparatus for direct measurement of the forces between surfaces in liquids or vapors at the angstrom resolution level. With the SFA, two atomically smooth surfaces immersed in a liquid are brought toward each other, with the surface separation being controlled to 0.1 nm, and the forces between the two surfaces can be measured. Moreover, the surfaces can be moved laterally past each other and the shear forces also measured during sliding. (Redrawn with permission of Israelachvili 1991.) (b) A schematic illustration of how the force between two surfaces is measured. As the two surfaces (inset) are brought together from separations x larger than x_2, the spring deflects toward contact to reach mechanical equilibrium (i.e., the spring force matches the attractive force between the surfaces). However, beyond the point x_2 at which the gradient of the interaction force equals the spring constant K, mechanical equilibrium is no longer possible, and the separation distance jumps to x_4. When the movement is reversed, the same thing happens again and x jumps from x_3 to x_1. By using springs of different spring constant K, one can also measure the force between the surfaces for $x_2 > x > x_3$. (Adapted from W. B. Russel, *The Dynamics of Colloidal Systems*, University of Wisconsin Press, Madison, WI, 1987.)

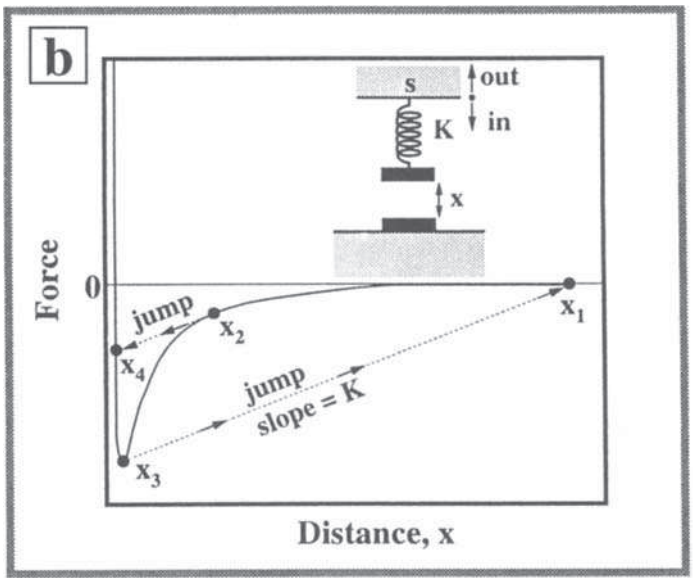


FIG. 1.27 Continued

forces, steric interactions involving polymeric systems, and capillary and adhesion forces have all been measured using this apparatus or variations of it (we see an example in Chapter 10). Moreover, the scope of phenomena that can be studied using the SFA has also been extended to measurements of dynamic interactions and time-dependent effects such as the viscosity of liquids in very thin films, shear and frictional forces, and the fusion of lipid bilayers.

The impact of the SFA and similar direct measurement forces goes beyond simple testing of theories of intermolecular forces. For instance, such measurements are also useful for explaining more complex phenomena such as the unexpected stability of certain colloidal dispersions in high salt, the crucial role of hydration and ion correlation forces in clay swelling and ceramic processing, and the deformed shapes of adhering particles and vesicles (see Israelachvili 1991). Further, both static (i.e., equilibrium) and dynamic (e.g., viscous) forces can now be studied with remarkable precision and accuracy for obtaining information on the structure of liquids adjacent to surfaces and related interfacial phenomena. Such studies demonstrate that properties of ultrathin films are profoundly different from those of bulk liquids.

1.6c.2b Atomic Forces: Scanning Force Microscope. As we have already seen in Vignette 1.8, the SFM is an offshoot of the scanning tunneling microscope (STM) and is a hybrid between a surface profilometer and the SFA (Wiesendanger 1994). The difference between an SFM and the SFA is that forces at the atomic level can be measured using the former with the help of an atomically sharp tip that scans the surface (for this reason, the SFM with atomic resolution is often called the *atomic force microscope* [AFM]). The tip radii can be as small as one atom or larger than 1 mm, but the smaller tips are more popular since they allow one to measure directly the force between an individual atom and a surface or even between two individual atoms.

In SFM, the probe tip is mounted on a highly sensitive, cantilever-type spring. The force of interaction between the sample and the tip can be calculated from the spring constant and the measured deflection of the cantilever. The deflection is sensed using the STM principle (Vignette 1.8) or capacitance or optical methods. The SFM can be operated in the "contact regime" or like the SFA. In the latter mode, one can measure van der Waals forces (see Chapter 10), ion-ion repulsion forces (see Chapter 11), and capillary forces and frictional forces, among others. In contrast to STM, the SFM can be used for both conductors and

insulators. The fact that nonconducting samples can be studied is very significant to colloid science since biological and polymeric surfaces can be probed using SFM.

Further, a whole new technology has arisen devoted to fabricating highly sensitive, micron-size, force-sensing devices for SFM work. For example, force measurements between a flat surface and a sphere in aqueous salt solutions out to surface separations of 60 nm have been made recently by attaching a micron-size quartz sphere to the end of an SFM tip (Israelachvili 1991).

Numerous research publications and monographs are available on the various versions of SFM (and STM) and on the interpretation of the data (Bonnell 1993; Wiesendanger 1994), but for our purpose here the above description is sufficient to illustrate the possibilities in force measurements at the atomic level and the opportunities they open up in more advanced studies of the materials discussed in subsequent chapters of this book.

1.7 AN OVERVIEW OF THE BOOK

The rest of the book can be divided into roughly three groups of chapters. A brief overview of these is given below.

1.7a Basic Phenomena and Techniques

Chapters 2–6 and Chapter 9 focus on many of the basic phenomena and techniques relevant to colloid and surface chemistry. These provide additional theoretical concepts and experimental tools routinely employed in the area. These include

1. The competition between diffusion and sedimentation and how this is used to measure particle size, solvation of particles, etc. (Chapter 2)
2. Osmometry and Donnan equilibrium (Chapter 3)
3. Flow and rheology of dispersions and viscosity of polymer solutions and how viscosity is used to characterize dispersions (Chapter 4)
4. Turbidity measurements, static and dynamic light scattering, and a brief introduction to x-ray and neutron scattering (Chapter 5)
5. Surface tension and contact angle, wetting phenomena, effects of the curvature of the surface on capillarity and phase equilibria, and porosimetry (Chapter 6)
6. Adsorption at solid-gas surfaces, derivation of adsorption isotherms, surface area measurement, and structural analysis of surfaces and adsorbed layers using low-energy electron diffraction (Chapter 9)

1.7b Association Colloids

Chapters 7 and 8 discuss surfactant solutions:

1. In the Chapter 7, formation of monolayers in air-liquid interfaces and the resulting film pressure and phase transitions are discussed. This chapter also includes a brief discussion of adsorption on solid surfaces from solutions.
2. Chapter 8 discusses self-assembly of surfactants to form micelles, models of micellization, use of micelles in catalysis and solubilization, and oil-in-water and water-in-oil microemulsions.

1.7c Colloidal Forces, Electrokinetic Phenomena, and Colloidal Stability

The final group deals with colloidal forces and their applications to colloid stability and deposition phenomena. These include

1. The van der Waals forces at the atomic level as well as those between macroscopic bodies (Chapter 10)
2. Theories of electrical double layers and forces due to double-layer overlap (Chapter 11)

3. Electrokinetic phenomena such as electroosmosis, streaming potential, and viscoelectric effects (Chapter 12)
4. The Derjaguin-Landau-Verwey-Overbeek (DLVO) theory of coagulation, polymer-induced forces, and steric stabilization (Chapter 13)

As mentioned above, each chapter also includes a vignette that provides a glimpse of advanced concepts or techniques that are beyond the scope of this book.

REVIEW QUESTIONS

1. What is the definition of a *colloid*?
2. Name at least five industrial products in colloidal form and discuss what colloidal properties are essential for each of them.
3. Name at least five examples in which a large surface area is desirable either from the processing standpoint or from the functionality of the end product.
4. What are *lyophilic* colloids? What are *lyophobic* colloids? Give some examples.
5. What are the difficulties with the characterization of colloids as *lyophilic* or *lyophobic*?
6. Why is a *foam* considered a colloid?
7. What is the meaning of *stability* as used in the context of colloids? What is the difference between *kinetic stability* and *thermodynamic stability* of a colloid?
8. Give at least five examples of colloids that are *thermodynamically* stable.
9. Define *coalescence* and *aggregation*. Give examples of processes or products in which aggregation (i.e., coagulation) is desirable. Give some examples for which coagulation is not desirable.
10. What is meant by the *fractal dimension* of an aggregate? What is the upper bound of the fractal dimension? Name at least one use of the fractal dimension of an aggregate.
11. What are some of the measures of *polydispersity*?
12. Why is the average diameter of a collection of polydisperse spheres different from the average diameter determined from the average of a property such as area or volume?
13. What is meant by the term *polydispersity index* in the case of molecular weights of polymers?
14. What is the *limit of resolution* of an optical microscope? Why?
15. What is *dark-field microscopy*? How does it compare in terms of resolution and contrast with conventional optical microscopy?
16. What is the limit of resolution of an *electron microscope*? Why?
17. What is the difference between *atomic* forces and *surface* forces? Can you deduce the details of atomic forces from surface force measurements? Why or why not?
18. Describe some of the recent developments for imaging surfaces and particles? What principles do they use? Name some advantages and limitations of each of the techniques.
19. What is meant by *scanning probe microscopy*?
20. Why is the measurement of atomic forces and surface forces important? Name some techniques for measuring atomic and surface forces.
21. Explain how scanning probe techniques can be modified to provide spectroscopic information.

REFERENCES

General References (with Annotations)

Atkins, P., *Physical Chemistry*, 5th ed., W. H. Freeman, New York, 1994. (Undergraduate level. This is an excellent, introductory-level book on physical chemistry and should be valuable for reviewing basic concepts of physical chemistry not described in detail in this chapter and elsewhere in this book.)

Ball, P., *Designing the Molecular World: Chemistry at the Frontier*, Princeton University Press, Princeton, NJ, 1994. (Undergraduate level. A highly readable, popular introduction in the *Scientific American* style to the many facets of molecular design in chemistry. Chapter 2 discusses chemical reactions, heterogeneous catalysis, and separation. Chapter 7 is an excellent introduction to self-organization of surfactant molecules into micelles, vesicles, Langmuir layers, and biological cells. This chapter also discusses the formation of polymer gels.)

Dickinson, E., *An Introduction to Food Colloids*, Oxford University Press, Oxford, England, 1992. (Undergraduate level. A readable introduction to the application of colloids in food science. Many of the topics should be accessible to undergraduate students.)

Giddings, J. C., *Unified Separation Science*, Wiley, New York, 1991. (Undergraduate level. A textbook on chromatographic and other analytical and preparative separation techniques. Some of the techniques discussed are relevant to the concepts we introduce in Chapters 2 and 3 of this book.)

Israelachvili, J. N., *Intermolecular and Surface Forces*, 2d ed., Academic Press, New York, 1991. (Undergraduate and graduate levels. The best reference available currently on the topic. Many examples of the application of surface forces in biological systems. The links between molecular forces and surface forces and the relation between molecular forces and bulk properties of materials are discussed in a manner accessible to advanced undergraduate students.)

Kruyt, H. R. (Ed.), *Colloid Science*, Vols. 1 and 2, Elsevier, Amsterdam, Netherlands, 1949, 1952. (Undergraduate and graduate levels. One of the classic references on colloids.)

Murray, C. A., and Grier, D. G., Colloidal Crystals. *American Scientist*, **83**, 238 (1995). (Graduate-level concepts, but some of the general ideas presented are accessible to undergraduates. This article is a moderately technical, but qualitative, overview of the use of charged latex particles for studying phase transitions in atomic solids and alloys. The color illustrations presented provide a glimpse of the numerous new avenues made possible by "model colloids" for studying structure of materials.)

Myers, D., *Surfaces, Interfaces, and Colloids: Principles and Applications*, VCH Publ., New York, 1991. (Undergraduate level. A qualitative overview of topics in colloid science. Numerous applications are discussed.)

Shaw, D. J., *Introduction to Colloid and Surface Chemistry*, 4th ed., Butterworth-Heinemann, Oxford, England, 1992. (Undergraduate level. A somewhat condensed treatment of many of the topics discussed in this book.)

Tanford, C., *Ben Franklin Stilled the Waves*, Duke University Press, Durham, NC, 1989. (A nontechnical, popular introduction to surfactant films and their applications in biology by a pioneer in the so-called hydrophobic effect.)

Vold, R. D., and Vold, M. J., *Colloid and Interface Chemistry*, Addison-Wesley, Reading, MA, 1983. (Graduate and undergraduate levels. A textbook at a more advanced level than this book.)

Other References

Adamson, A. W., *Physical Chemistry of Surfaces*, 5th ed., Wiley, New York, 1990.

Alberts, B., Bray, D., Lewis, J., Raff, M., Roberts, K., and Watson, J. D., *Molecular Biology of the Cell*, 2d ed., Garland Publishing, New York, 1989.

Allen, T., *Particle Size Measurement*, 4th ed., Chapman and Hall, New York, 1990.

Baumgarten, N., *Nature*, **341**, 81 (1989).

Bergethon, P. R., and Simons, E. R., *Biophysical Chemistry: Molecules to Membranes*, Springer-Verlag, New York, 1990.

Bonnell, D. A., *Scanning Tunneling Microscopy and Spectroscopy: Theory, Techniques, and Applications*, VCH Publ., New York, 1993.

Bradbury, S., *Basic Measurement Techniques for Light Microscopy*, Oxford University Press, Oxford, England, 1991.

Briggs, A., *Acoustic Microscopy*, Oxford University Press, Oxford, England, 1992.

Brinker, C. J., and Scherer, G. W., *Sol-Gel Science: The Physics and Chemistry of Sol-Gel Processing*, Wiley, New York, 1990.

Cherry, R., *New Techniques of Optical Microscopy and Microspectroscopy*, CRC Press, Boca Raton, FL, 1991.

Goodsell, S., *The Machinery of Life*, Springer-Verlag, New York, 1993.

Hubbard, A. T. (Ed.), *The Handbook of Surface Imaging and Visualization*, CRC Press, Boca Raton, FL, 1995.

Lasic, D., *American Scientist*, **80**, 20 (1992).

Lasic, D., *Liposomes: From Physics to Applications*, Elsevier Scientific, Amsterdam, Netherlands, 1993.

Matijevic, E., *Chem. Mater.*, **5**, 412 (1993).

Morowitz, H. J., *Beginnings of Cellular Life: Metabolism Recapitulates Biogenesis*, Yale University Press, New Haven, CT, 1992.

Schaeffer, H. F., *Microscopy for Chemists*, Dover, New York, 1953.

Weitz, D. A., Huang, J. S., Lin, M. Y., and Sung, J., *Phys. Rev. Lett.*, **54**, 1416 (1985).

Wiesendanger, R., *Scanning Probe Microscopy and Spectroscopy: Methods and Applications*, Cambridge University Press, Cambridge, England, 1994.
Yau, W. W., Kirkland, J. J., and Bly, D. D., *Modern Size Exclusion Liquid Chromatography*, Wiley, New York, 1979.

PROBLEMS*

1. The specific area of dust particles from the air over Pittsburgh, Pennsylvania, has been determined by gas adsorption†

Treatment	$A_{sp}(m^2g^{-1})$
4 h under vacuum at 200°C	5.61
8 h under vacuum at 25°C	2.81

   a. Calulate the radius of these particles if they are assumed to be uniform spheres of density 2.2 g cm^{-3}.
   b. Propose an explanation for the effect of degassing on particle size.
   c. What kind of average for the radius is obtained by this procedure?

2. Colloidal palladium particles in an alumina matrix catalize the hydrogenation of ethene to ethane. The following data describe various catalyst preparations:‡

Diameter of Pd particles (Å)	55	75	75	115	145
Parts per million Pd in catalyst	170	250	200	250	250
Percentage conversion per 25 mg catalyst	50	45	40.5	38.5	29

   a. Calculate the activity of these catalysts on the basis of the weight of Pd and the area of Pd in the preparations.
   b. Does the catalytic role of Pd seem to be a bulk or surface phenomenon?

3. The accompanying table shows how the trace metal content of coal-ash aerosols depends on particle size:¶

Range of diameters (μm)	Trace element (μg/g ash)					
	Pb	Tl	Sb	Cd	Se	As
30–40	300	5	9	< 10	< 15	160
5–10	820	20	25	< 10	< 50	800
1.1–2.1	1600	76	53	35	59	1700

   Discuss the implications of these results on human health in view of the following considerations:
   a. The intrinsic toxicity of these trace elements
   b. small particles travel to the lung, larger particles are stopped in the nose, pharynx, and so on
   c. Trace elements are absorbed from the alveoli 7–10 times more efficiently than from the upper respiratory spaces
   d. A possible mechanism for the effect of particle size on trace element content

4. Select a field containing about 10 particles from Figure 1.12b and measure the Martin diameters of the population parallel to the bottom of the micrograph (a photocopy can be made and

*The data for many of the problems in this book are taken from graphs appearing in the original literature. As a result, the values given do not necessarily reflect the accuracy of the original experiments. Likewise, the number of significant figures cited may not be justified in terms of the approximations involved in graph reading.

†Corn, M., Montgomery, T. L., and Reitz, R. J. *Science* **159**, 1350 (1967).

‡Turelevich, J., and Kim, G. *Science*, **169**, 873 (1970).

¶Natusch, D.F.S., Wallace, J. R., and Evans, C. A., Jr. *Science* **183**, 202 (1974).

the particles checked off to avoid duplication and omissions). Classify the data and calculate the mean and standard deviation. Repeat, measuring the Martin diameter parallel to the side of the micrograph. Discuss the agreement or discrepancy between the two means in terms of:
a. Bias in the choice of the field
b. Systematic orientation effects
c. The size of the population

5. Suppose that the particles in Figure 1.8 were actually oblate ellipsoids (all in their preferred orientation) rather than spheres. Would their volume be over- or underestimated if the particles were assumed to be spheres? In terms of their axial ratio, calculate the factor by which the mass is under- or overestimated when the particles are assumed to be spheres. (Consult a handbook for the volume of an ellipsoid.)

6. Mixtures of 50 g of ZnO-TiO_2 are each shaken with 250 ml of water and allowed to settle. After 14 days of equilibration, it is found that the sediment volume is 1.65 times larger when the weight ratio of ZnO to TiO_2 is 1.0 than when the ratio is 100.* Some particle characteristics are the following:

	Diameter (μm)	Density (g ml^{-1})	Charge in water
ZnO	1.0	5.6	Positive
TiO_2	2.2	4.2	Negative

a. Assuming the particles are uniform spheres, calculate the ratio of the number of particles of ZnO to TiO_2 for each of the weight ratios given.
b. Propose an explanation for the more voluminous sediment that results when the weight ratio is 1.0 than when it is 100.

7. The following data describe the particle size distribution in a dispersion of copper hydrous oxide particles in water:†

d_i (μm)	0.426	0.401	0.376	0.351	0.326	0.301	0.276	0.251	0.226	0.201
n_i	1	0	6	6	17	14	11	12	6	6

Calculate $\bar{d}_n$, σ, $\bar{d}_s$, and $\bar{d}_v$ for this dispersion.

8. A graticule was used to size sand particles and glass spheres.‡ The percentage by weight of particles less than the stated size was found to be as follows:

d (μm)	0.4	0.8	1.6	2.4	3.0	4.0	8.0	12.0
Sand	0.01	0.07	0.23	0.56	1.23	2.35	11.77	18.06
Glass	0.01	0.11	0.26	0.43	0.72	1.43	17.84	28.07
d (μm)	16.0	20.0	30.0	40.0	60.0	75.0	90.0	120.0
Sand	24.62	32.11	52.33	64.14	83.60	–	98.7	100.0
Glass	36.89	47.68	59.47	61.78	91.49	100.0	–	–

Plot the results on probability and log-probability coordinates. Use the best representation to evaluate (the appropriate) average and standard deviation for the samples (see Appendix C).

9. Particle size distributions were measured on aerosols collected in a New York highway tunnel, and the following results were obtained:¶

Cumulative % mass $< d_i$ (μm)	30	40	50	60	70	80
d_i weekend	0.5	1.0	2.50	5.0	–	–
d_i, weekday	–	–	0.07	0.2	0.9	4.0

*Princen, L. H., and DeVena, M. J. *J. Am. Oil Chem Soc.* **39**, 269 (1962).
†McFadgen, P., and Matijević, E. *J. Colloid Interface Sci.* **44**, 97 (1973).
‡Fairs, G. L. *Chem. Ind.* **62**, 374 (1943).
¶Lee, R. E., Jr., *Science* **178**, 567 (1972).

Plot these results on log-probability coordinates and estimate the mean and standard deviation for each distribution (see Appendix C). What kind of "averages" are these quantities? How do the weekend and weekday particle size distributions compare with respect to the location and width of the maximum? The weekend results are attributed to automobile exhaust, whereas the weekday results are assumed to be "diluted" by aerosols from outside the tunnel.

10. The mass of bull sperm heads in a sample was determined by interference microscopy, and the following results were obtained:*

n_i	4	2	27	37	32	26	20	8	3	3	1
$w_i \times 10^{12}$ (g)	5	6	7	8	9	10	11	12	13	14	15

Calculate the number average and weight average weights of these particles.

11. Yau et al.† used polystyrene samples of known molecular weight in tetrahydrofuran to calibrate the retention volume of a size-exclusion chromatograph. They obtained the following results:

$M \times 10^{-3}$ (g mole^{-1})	4.0	50.0	110	179	390	1800
V_R (ml)	16.3	14.5	13.9	13.6	13.1	12.1

Do these results display the expected correlation between M and V_R? The flow rate in the instrument used in this research was 1.0 ml min^{-1}; how long would it take for samples of molecular weight 10^4, 10^5, and 10^6 to emerge from the column? Is this order of elution times qualitatively consistent with Equations (26) and (27)?

---

*Bahr, G. F., and Zeitler, E. *J. Cell Biol.* **21**, 175 (1964).

†Yau, W. W., Jones, M. E., Ginnard, C. R., and Bly, D. D. In *Size Exclusion Chromatography* (T. Provder, Ed.), American Chemical Society, Washington, DC, 1980.

2

Sedimentation and Diffusion and Their Equilibrium

Even if you had completed your third year . . . in the University, and were perfect in the theory of the subject, you would still find that there was need of many years of experience, before you could move in a fashionable crowd without jostling against your betters.

From Abbott's *Flatland*

2.1 INTRODUCTION

2.1a Sedimentation and Diffusion: Why Are They Important?

Sedimentation and diffusion affect many colloidal phenomena. Sedimentation occurs under the influence of gravity and, considerably faster, in a centrifuge. Many analytical and preparative techniques in colloid science (particularly in biophysics and biochemistry) take advantage of centrifugal force for physical characterization or purification of materials in the colloidal size range. In fact, because of the small sizes of the particles or macromolecules involved, the use of centrifugation is indispensable. In view of these, basic principles of (gravitational and centrifugal) sedimentation are an essential part of colloid science.

Diffusion, of molecular species as well as colloidal particles, plays perhaps a more dominant role in many topics of interest to us. For example, without diffusion of ions we will not have the diffuse electrical double layers next to charged surfaces (discussed in Chapter 11). At the colloidal level, diffusion plays a central role in the transport and collision of particles in colloidal stability (discussed in Chapter 13). There are many more such examples.

The equilibrium between gravitational or centrifugal force and diffusion is routinely taken advantage of in colloid science, as illustrated in Vignette II. Our objective in this chapter is to examine the effects of sedimentation and diffusion, first taken separately and then combined, on particles in the colloidal size range.

VIGNETTE II ANALYTICAL SEPARATIONS AND CHROMATOGRAPHY: Sedimentation Field Flow Fractionation

Were it not for the never-ending, gentle tussle between gravity and diffusion, our planet would not have an atmosphere, nor would we be here to reflect on it! The *barometric equation*, which describes this "balance of power" between the above two well-known phenomena, is derived in most introductory physical chemistry books and is mentioned in the closing paragraph of this chapter. There are many more life-sustaining processes that are

affected by or rely on sedimentation and diffusion, but frequently it is the more mundane "practical" consequences of these phenomena that attract our attention.

One such consequence is their use in the physical characterization of colloidal dispersions and macromolecular solutions. Let us highlight one such application through one element of a class of analytical separation techniques known as *field flow fractionation* (FFF).

The name field flow fractionation stands for a family of techniques, invented in the 1960s, that take advantage of the response of colloids and macromolecules to electrical, thermal, flow, or centrifugal fields to produce a chromatography-type separation of the particles (Giddings 1966, 1991, 1993). In a typical set up, a suitable force field is applied in a direction normal to the axis of a thin chamber that contains the dispersion. The field forces the particles against one of the walls of the chamber, and, at steady state, a concentration profile is set up in the direction of the applied field as a consequence of the differences in the responses of the various species in the dispersion to the applied field (see Fig. 2.1a). The particles are then eluted by flowing an elutant fluid through the chamber. The fluid velocity decreases progressively from the axis toward the accumulation wall because of friction at the wall (see Chapter 4, Section 4a). As a consequence, the particles are carried along the axis at different velocities depending on their distance from the accumulation wall. For example, the component closest to the accumulation wall lags behind the one near the center of the chamber (see Fig. 2.1b). Samples can now be eluted through a detector or collection device. The detection is usually based on changes in standard properties such as refractive index or light absorption.

One of the more advanced of the FFF techniques is sedimentation FFF (SdFFF), in which the applied field is a centrifugal force (see Fig. 2.1b). A typical separation achieved through SdFFF is also illustrated in Figure 2.1b. The SdFFF is suitable for species with molecular weights larger than about 10^6 and has proved useful for a large number of biocolloids (e.g., subcellular particles), polymers, emulsions, and natural and industrial colloids (Giddings 1991).

It is our objective in this chapter to outline the basic concepts that are behind sedimentation and diffusion. As we see in this chapter, gravitational and centrifugal sedimentation are frequently used for particle-size analysis as well as for obtaining measures of solvation and shapes of particles. Diffusion plays a much more prevalent role in numerous aspects of colloid science and is also used in particle-size analysis, as we see in Chapter 5 when we discuss dynamic light scattering. The equilibrium between centrifugation and diffusion is particularly important in analytical and preparative ultracentrifuges.

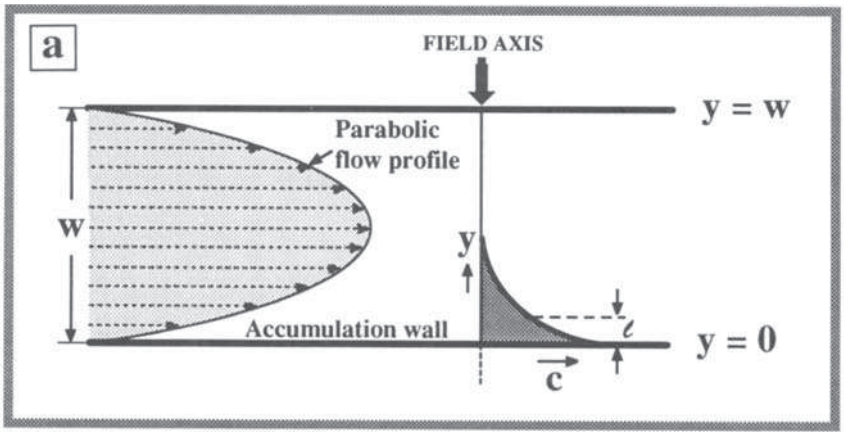


FIG. 2.1 Sedimentation field flow fractionation (SdFFF): (a) an illustration of the concentration profile and elutant velocity profile in an FFF chamber; and (b) a schematic representation of an SdFFF apparatus and of the separation of particles in the flow channel. A typical fractionation obtained through SdFFF using a polydispersed suspension of polystyrene latex spheres is also shown. (Adapted from Giddings 1991.)

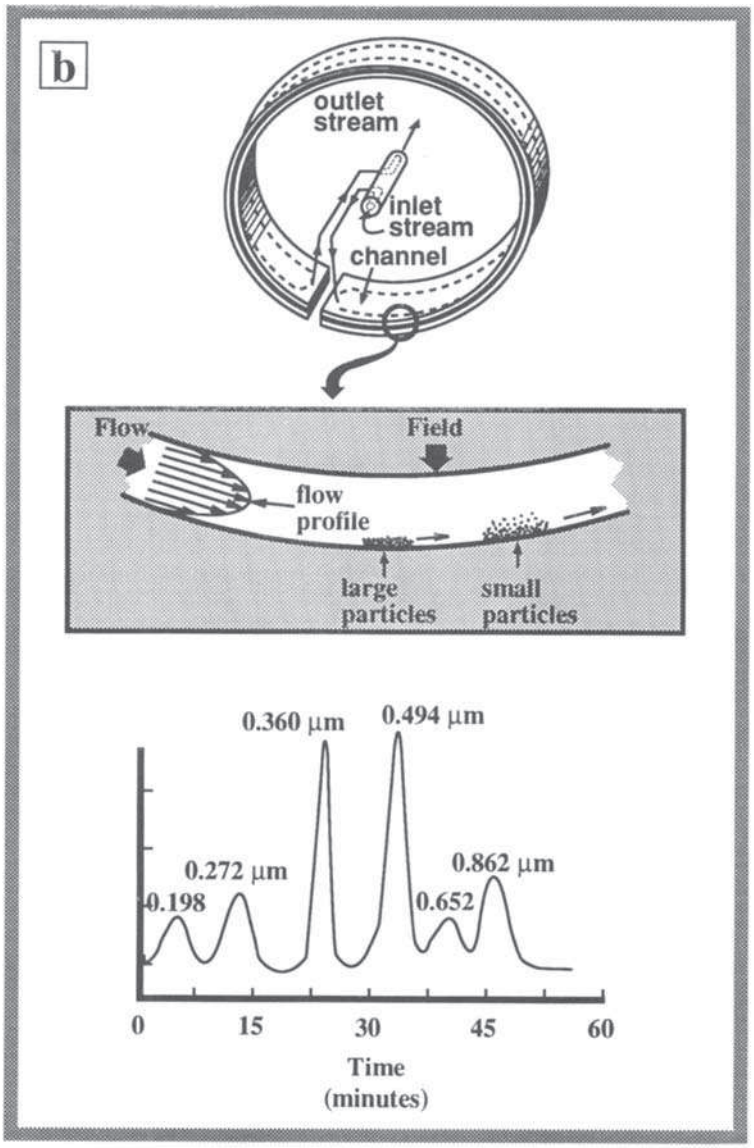


FIG. 2.1 Continued

2.1b Purpose and Content of This Chapter

The purpose of this chapter is to introduce the basic concepts that govern sedimentation and diffusion of colloidal particles. The material presented is organized into roughly four major groups.

1. The first group, consisting of Sections 2.2–2.4, covers sedimentation. After some preliminaries, we discuss Stokes's law, a hydrodynamic equation that will appear again when we discuss electrokinetic phenomena in Chapter 12 and the kinetics of coagulation in Chapter 13. Stokes's law is a key relationship in understanding the rate of sedimentation and is used in the derivation of the sedimentation equation for spherical particles. Following this, the equation for the sedimentation coefficient, a

standard measure of the steady-state velocity of a particle in a centrifugal field, is given and its use is illustrated through an example.

2. The second major topic, diffusion, is covered in Section 2.5 from a classical perspective, through Fick's laws of diffusion and thermodynamic arguments. In discussing diffusion, we focus our attention on the diffusion coefficient, which is defined experimentally by Fick's laws and theoretically by two equations derived by Einstein.
3. The statistical basis of diffusion requires arguments that may be familiar from kinetic molecular theory. Elementary concepts from the theory of random walks and its relation to diffusion form the third topic, which is covered in Section 2.6. As is well known, the random walk statistics can also be used for describing configurational statistics of macromolecules under some simplifying assumptions; this is outlined in Section 2.7.
4. Finally, we turn our attention to equilibrium between sedimentation and diffusion (Section 2.8). As our educated intuition might lead us to expect, larger particles sediment more rapidly and diffuse more slowly than smaller particles. The effects of sedimentation and diffusion, therefore, are not of comparable magnitude for all sizes of particles. There is a range of particle sizes, however, for which the two are comparable and equilibrium between them is established. This equilibrium between sedimentation and diffusion has been studied extensively by means of the ultracentrifuge. Since many of the particles thus investigated are of biological importance, we frequently use protein molecules as our examples in this chapter.

Much of our discussion of sedimentation alone and of sedimentation combined with diffusion focuses on the determination of the mass or molecular weight of the dispersed particles.

2.2 SEDIMENTATION: SOME BASIC CONSIDERATIONS

To see how sedimentation occurs, consider the gravitational forces that operate on a particle of volume V and density ρ_2 that is submerged in a fluid of density ρ_1. The situation is shown in Figure 2.2a for a spherical particle, but the discussion is independent of the actual shape of

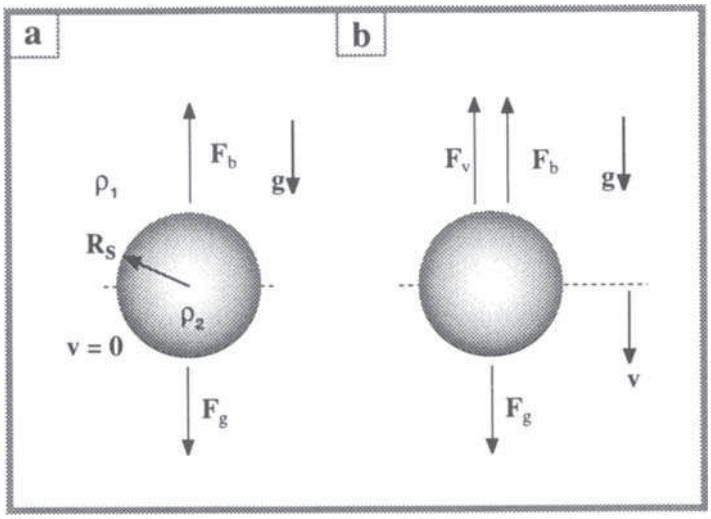


FIG. 2.2 The forces acting on a spherical particle due to gravity and viscosity: (a) due to gravity alone; and (b) due to gravity and the viscosity of the medium ($\rho_2 > \rho_1$).

the particle. The particle experiences a force $\mathbf{F}_g$ due to gravity, taken to be positive in the downward direction. At the same time a buoyant force $\mathbf{F}_b$ acts in the opposite direction. A net force equal to the difference between these forces results in the acceleration of the particle:

$$\mathbf{F}_{net} = \mathbf{F}_g - \mathbf{F}_b = V(\rho_2 - \rho_1)\mathbf{g} \tag{1}$$

This force will pull the particle downward—that is, $\mathbf{F}_{net}$ will have the same sign as $\mathbf{g}$—if $\rho_2 > \rho_1$, and the particle is said to sediment. If, on the other hand, $\rho_1 > \rho_2$, then the particles will move upward, which is called *creaming*.

2.2a Friction Factor and Stationary Settling Velocity

As the net velocity of the particle is increased, the viscous force $\mathbf{F}_v$ opposing its motion also increases. Soon this force, shown in Figure 2.2b, equals the net driving force responsible for the motion. Once the forces acting on the particle balance, the particle experiences no further acceleration and a stationary state velocity is reached. It may be shown that, under stationary state conditions and for small velocities, the force of resistance is proportional to the stationary state velocity $\mathbf{v}$:

$$\mathbf{F}_v = f\mathbf{v} \tag{2}$$

where the proportionality constant f is called the *friction factor*. The friction factor has the dimensions mass time^{-1}, kg s^{-1} in SI units. We consider some aspects of the proof of Equation (2) for spherical particles in the following section. However, Equation (2) is independent of any particular geometry. The stationary state velocity is positive for sedimentation and negative for creaming.

Since the net force of gravity and the viscous force are equal under stationary state conditions, Equations (1) and (2) may be equated to give

$$V(\rho_2 - \rho_1)\mathbf{g} = f\mathbf{v} \tag{3}$$

The stationary state is quite rapidly achieved, so Equation (3) describes the velocity of a settling particle during much of its fall. This velocity is known as the *stationary settling velocity*.

2.2b Mass-to-Friction-Factor Ratio

Equation (3) may also be written

$$m\left(1 - \frac{\rho_1}{\rho_2}\right)\mathbf{g} = f\mathbf{v} \tag{4}$$

where m is the mass of the particle ($\rho_2 V$). This equation has the following features:

1. It is independent of particle shape.
2. It assumes that the bulk density of the pure components applies to the settling units (i.e., no solvation).
3. It permits the evaluation of $\mathbf{v}$ for a situation in which the mass-to-friction-factor ratio (m/f) is known.
4. It permits the evaluation of (m/f) in a situation for which v is known.

The stationary sedimentation velocity of a particle is an experimentally accessible quantity for some systems, so item 4 summarizes much of our interest in sedimentation. Unfortunately, it is the ratio (m/f) rather than m alone that is obtained from sedimentation velocity in the general case of particles of unspecified geometry. The situation is comparable to the result of the classical experiment of J. J. Thomson in which the charge-to-mass ratio of the electron was determined.

What is needed is a method for arriving at one of the quantities in the ratio independently. This information, in addition to the ratio, permits the evaluation of both the mass and the

friction factor of the particle. In general, it takes two experiments to evaluate numerically these two parameters that characterize the dispersed particles. There are two ways of proceeding to overcome this impasse:

1. The particle may be assumed to be a sphere, in which case its friction factor may be calculated theoretically and thus eliminated from Equation (4). We discuss the friction factor of spherical particles in Section 2.3b.
2. An experiment may be conducted that permits the evaluation of f from measured quantities. Diffusion studies are ideally suited for this purpose, as we see further in this chapter.

2.2c Sensitivity of Sedimentation to Density Differences

Before turning to the two procedures above to eliminate f from the sedimentation equation, one other consideration inherent in the use of Equation (4) should be discussed. As noted above, Equation (4) uses the bulk density of the pure components. If the continuous and bulk phases are totally noninteracting, this may be justified. However, for aggregates or solvated lyophilic particles, the density of the settling unit is intermediate between the densities of the two pure components. In these cases, choosing an appropriate density for the settling particle can be a real problem.

Equation (3) shows that the sedimentation velocity increases with the density difference between the particle and the medium. Any situation that brings the density of the settling unit closer to that of the solvent will decrease the sedimentation velocity. To an observer who is unaware of its derivation, however, the smaller velocity would be interpreted by Equation (4) as indicating a smaller value of (m/f). Since the actual mass of colloidal material is unaffected by the solvation, it is more correct to attribute the reduced sedimentation velocity to an increase in the value of the friction factor.

In the next section, we see how to deal quantitatively with solvation. Until now, the friction factor has been merely a proportionality factor of rather ill-defined origin. We shall not undertake a derivation of Equation (2) in any general sense. In the next section, however, we outline the derivation of an important result due to G. G. Stokes—the friction factor for an unsolvated sphere.

2.3 GRAVITATIONAL SEDIMENTATION

2.3a Stokes's Law for Spheres

We begin our discussion of the Stokes equation by considering a single spherical particle of radius R, in a state of relative motion with respect to the surrounding fluid. For the purposes of this derivation, it does not matter whether the fluid is stationary and the particle is moving through it or whether the particle is stationary with the fluid moving around it. Although the former describes sedimentation, it is somewhat more convenient to discuss the phenomenon in terms of the fluid flowing in the $+z$ direction with a velocity v_z; this is the same as the particle moving in the $-z$ direction with a velocity of the same magnitude. This situation is represented schematically in Figure 2.3.

If the spherical particle were not present in Figure 2.3, the volume elements of the flowing fluid would move upward in straight lines. In the presence of the particle, however, the flow profile is distorted around the sphere in the manner suggested by Figure 2.3. It is apparent that the velocity of any volume element passing the sphere is a function of both time and location and must be described as such in any quantitative treatment. The trajectory of such a volume element is called the flow *streamline* function. For spherical particles, this was analyzed by G. G. Stokes in 1850.

The streamline function is the solution of a differential equation, the details of which we will not pursue. We will have more to say about flow streamlines in our discussion of viscosity in Chapter 4. For the present, however, it is sufficient to note that an important part of

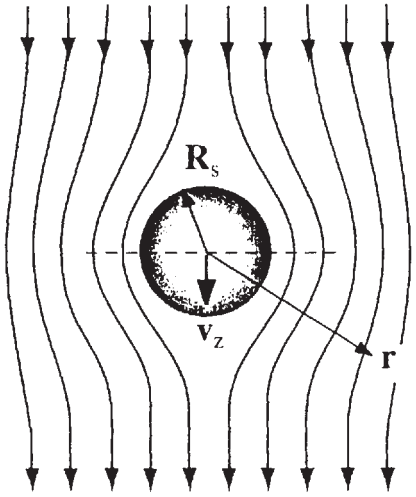


FIG. 2.3 Distortion of flow streamlines around a spherical particle.

solving any differential equation is the adequate incorporation of the boundary conditions that describe the problem. Since the disturbance of the flow streamlines in Figure 2.3 is centered at the particle, it is convenient to discuss the boundary conditions in terms of the spatial variable r originating at the center of the sphere. Several aspects of the problem can be described in terms of this variable.

1. The thickness of each successive layer in the fluid is the infinitesimal mathematical increment dr. Since the streamlines roughly follow the outline of the particle and are thin compared to the radius of the sphere R_s, we can think of each flow layer as moving tangentially to the surface of the sphere.
2. The velocity varies from layer to layer in the fluid near the surface of the sphere; this variation can be described in terms of a velocity gradient dv/dr. Any integration over this gradient must be able to assign two different values of velocity to two different r values.
3. When $r = \infty$, the disturbing influence of the particle has been damped away and the streamlines behave as if the particle was not present.
4. When $r = R_s$, the velocity is zero. This is described as a *nonslip* (or *no-slip*) condition between the stationary surface of the particle and the layer of fluid adjacent to it. There is nothing particularly self-evident about the nonslip condition between the solid and the fluid, but it is an experimental fact. Some commonplace evidence that suggests this is the layer of dust that accumulates on the blades of a fan. However stiff the breeze may be some distance in front of the blades, the air is still and travels with the surface of the blades.

The fact that the velocity of a fluid changes from layer to layer is evidence of a kind of friction between these layers. The layers are mathematical constructs, but the velocity gradient is real and a characteristic of the fluid. The property of a fluid that describes the internal friction or resistance to flow is the viscosity of the material. Chapter 4 is devoted to a discussion of the measurement and interpretation of viscosity. For now, it is enough for us to recall that this property is quantified by the coefficient of viscosity η of a material. The coefficient of viscosity has dimensions of mass length^{-1} time^{-1}, kg m^{-1}s^{-1} in SI units. In actual practice, the cgs unit of viscosity, the poise (P), is widely used. Note that pure water at 20°C has a viscosity of about 0.01 P $= 10^{-3}$ kg m^{-1} s^{-1}

Because of internal friction, some of the translational kinetic energy of flow dissipates over a period of time. For the physical situation pictured in Figure 2.3, Stokes was able to show that the rate of energy dissipation dE/dt is given by

$$dE/dt = 6\pi\eta v_z^2 R_s \tag{5}$$

The product of the viscous force F_v exerted by the particle on the fluid and the velocity v_z of the fluid also equals the rate at which work is done on the fluid by the particle. Remember, a force times a distance equals energy, so a force times a velocity equals the rate of energy dissipation. Thus, we have two different ways of expressing the rate of energy dissipation associated with the presence of a spherical particle in the flowing liquid. Equating these two expressions gives

$$F_v v_z = 6\pi\eta v_z^2 R_s \tag{6}$$

It is clear that the force acting on the fluid acts in opposition to the flow. Therefore, if the fluid is flowing in the positive direction, the sphere resists the flow by a force in the negative direction. Conversely, if the particle is moving in one direction, the fluid resists the movement by a force in the opposite direction. This force is given by *Stokes's law*,

$$\mathbf{F} = 6\pi\eta R_s \mathbf{v} \tag{7}$$

where the subscripts for **F** and **v** are no longer needed, provided we remember that **F** and **v** are in opposing directions. (Note that **F** and **v** are vectors and, therefore, a direction as well as a magnitude must be specified for each.)*

2.3b Sedimentation Equation

Equation (7) expresses, for a spherical particle, the general result presented as Equation (2). Two aspects of the result are important:

1. For small, stationary-state velocities, the viscous force on a particle is directly proportional to the velocity.
2. For a spherical particle, the friction factor is given by

$$f = 6\pi\eta R_s \tag{8}$$

It was just noted that the case of spherical particles is one in which the friction factor can be eliminated from Equation (4), yielding a result that permits the mass of a spherical particle to be evaluated from sedimentation data alone. To see how this works, we return to Equation (3), using Equation (8) to evaluate f and $(4/3)\pi R_s^3$ as a substitution for V. Thus, for a spherical particle, Equation (3) becomes

$$\frac{4}{3}\pi R_s^3 (\rho_2 - \rho_1)\,\mathbf{g} = 6\pi\eta R_s \mathbf{v} \tag{9}$$

Equation (9) may be solved for the stationary-state sedimentation velocity

$$\mathbf{v} = \frac{2}{9}\frac{R_s^2(\rho_2 - \rho_1)\mathbf{g}}{\eta} \tag{10}$$

and for the radius of the spherical particle

$$R_s = \left(\frac{9\eta v}{2(\rho_2 - \rho_1)g}\right)^{1/2} \tag{11}$$

Since the particle to which it applies is a sphere, Equation (11) may also be used to calculate the mass and friction factor of the particle:

*Vectors are denoted by bold letters and their magnitudes by corresponding roman letters. Such a differentiation is not used when only the magnitudes are important.

$$m = \rho_2 \frac{4}{3} \pi R_s^3 = \frac{4}{3} \pi \rho_2 \left(\frac{9\eta v}{2(\rho_2 - \rho_1)g} \right)^{3/2} \tag{12}$$

and

$$f = 6\pi\eta \left(\frac{9\eta v}{2(\rho_2 - \rho_1)g} \right)^{1/2} \tag{13}$$

Equation (9) is an important result since it describes the relationship among R_s, v, η, and $\Delta\rho$, the density difference. Any one of these quantities may be evaluated by Equation (9) when the other three are known. Thus, Equation (9) can be used to determine the density difference between two phases or to determine the viscosity of a liquid. In this chapter, however, our interest is in the characterization of colloidal particles by means of observations of their sedimentation behavior. Therefore, we are primarily concerned with Equations (11) and (12), which are specifically directed toward this objective.

2.3c Effects of Nonsphericity and Solvation

Stokes's law and the equations developed from it apply to spherical particles only, but the dispersed units in systems of actual interest often fail to meet this shape requirement. Equation (12) is sometimes used in these cases anyway. The lack of compliance of the system to the model is acknowledged by labeling the mass, calculated by Equation (12), as the mass of an "equivalent sphere." As the name implies, this is a fictitious particle with the same density as the unsolvated particle that settles with the same velocity as the experimental system. If the actual settling particle is an unsolvated polyhedron, the equivalent sphere may be a fairly good model for it, and the mass of the equivalent sphere may be a reasonable approximation to the actual mass of the particle. The approximation clearly becomes poorer if the particle is asymmetrical, solvated, or both. Characterization of dispersed particles by their mass as equivalent spheres at least has the advantage of requiring only one experimental observation, the sedimentation rate, of the system. We see in sections below that the equivalent sphere calculations still play a useful role, even in systems for which supplementary diffusion studies have also been conducted.

Actual particles can deviate from the Stokes model by being either solvated, asymmetrical, or both, in which case f increases so that a particle of mass m will display a smaller sedimentation velocity than it would if it were an unsolvated sphere. The justification for this statement is based on Figure 2.4, which shows an unsolvated sphere and represents its radius as R_0. If this particle were solvated, it would swell to a larger radius R_s associated with a larger friction factor. Small asymmetrical particles rotate through every possible orientation during settling because of Brownian motion. Figure 2.4 shows that an encompassing sphere drawn around the tumbling particle will also have a radius larger than R_0.

An increase in f is associated with this situation as well. To deal with these ideas quantitatively, it is convenient to consider the following ratios:

$$f/f_0 = (f/f^*)(f^*/f_0) \tag{14}$$

1. In Equation (14), f is the friction factor of the actual particle.
2. The f_0 is the friction factor for an unsolvated sphere given by Stokes's law to equal $6\pi\eta R_0$ in the present notation. This is the lowest friction factor possible for a particle with the required mass.
3. The ratio f/f_0 measures the amount by which the actual friction factor exceeds the minimum value.
4. The f^* is the friction factor for a spherical particle having the same volume as the *solvated* particle of mass m.
5. The ratio f^*/f_0 measures the increase in f due to solvation.
6. The ratio f/f^* measures the increase in f due to asymmetry.

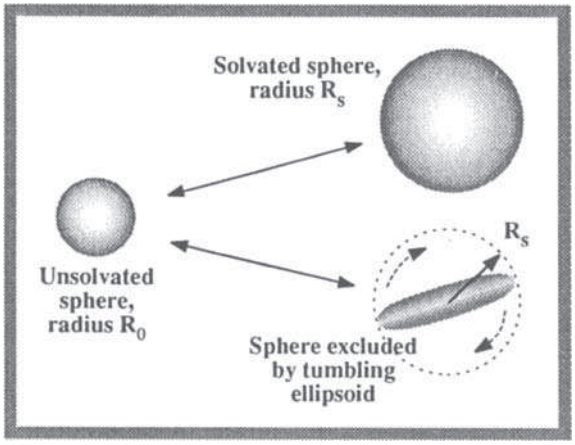


FIG. 2.4 The effects of solvation and asymmetry on the effective radius of a particle. The figure illustrates how solvation and asymmetry have qualitatively equivalent effects in increasing the effective radius of an unsolvated sphere of radius R_0. (Redrawn with permission of P. C. Hiemenz *Polymer Chemistry: The Basic Concepts*, Marcel Dekker, New York, 1984.)

We see presently that both of the ratios on the right-hand side of Equation (14) can be expanded in terms of appropriate models, so the experimental f value—compared to f_0 for a particle of the same mass—can be analyzed in terms of solvation and asymmetry.

Prior to this, however, let us consider how the mass of an equivalent sphere can be extracted from experimental data since this quantity is an important concept in its own right.

2.3d Sedimentation Measurements

We begin our discussion of the experimental aspects of sedimentation by considering settling caused by gravity; in Section 2.4, we examine centrifugation. Substitution of numerical values into Equation (10) reveals that the range of particle sizes that can be studied by sedimentation under gravity is quite limited. For example, taking $\Delta\rho/\eta = 100$ g cm^{-3} P^{-1} yields values of v between $2.2 \cdot 10^{-12}$ and $2.2 \cdot 10^{-6}$ m s^{-1} for spherical particles of R_s between 1.0 nm and 1.0 μm. This value of $\Delta\rho/\eta$ corresponds physically to sulfur particles in water at 20°C, in which $\Delta\rho$ and v are both positive, and to gas bubbles in water, in which $\Delta\rho$ and v are negative. Although sedimentation under gravity is feasible only at the upper size limit of the usual colloidal range, there are still a number of important applications of the technique. We consider only two of the procedures that have been developed for such studies.

If a particle is large enough to be visible to the unaided eye or in a traveling microscope, its velocity can be measured directly; however, smaller particles present more of a problem. If a dispersion consists of particles of uniform size, all will settle at the same rate, so their positions relative to each other do not change until they reach the bottom of the container. This means that a sharp boundary will exist between the domain occupied by the settling particles and that part of the system already swept free of particles. Even though the individual particles may not be visible, this boundary may be visible if the particles differ sufficiently in color or refractive index from the continuous phase. The velocity of the boundary and the velocity of the particles are the same. If the particles differ in size, no such boundary will develop. Each size fraction in a polydisperse system will settle at its own characteristic velocity, so that at any given time one end of the column may be free of some size fractions but not of others. The "boundary zone" between the domain that clearly contains settling particles and the domain that is totally free of particles will appear to be a diffuse region across which the concentration of the disperse phase gradually diminishes. Since no sharp boundary develops, some other method must be devised to follow sedimentation velocity. Clearly, we would

like as much information as possible about the particle size distribution since the system is polydisperse. The following procedure shows how this information may be obtained.

Suppose the pan of a balance is positioned at some convenient location below the surface of a dispersion. As sedimentation occurs, the settled material collects on the balance pan. The total weight W of the material on the pan is measured at various times t, either by adding counterweights to a second pan or by noting the displacement of a calibrated fiber that supports the pan. The following example suggests how this kind of data is analyzed.

* * *

EXAMPLE 2.1 *Analyzing Cumulative Sedimentation Data for Most Probable Settling Velocity.* The following data show—as a function of time—the weight (as percentage of total) of suspended clay particles W, which has accumulated on a plate submerged 20 cm beneath the surface in a sedimentation experiment (Oden 1915).

Time (s)	34	57	115	175	254	294	337	360	406
W (%)	7.42	11.13	16.7	21.44	28.7	33.04	38.85	41.75	47.55
Time (s)	458	503	558	648	705	773	862	982	1112
W (%)	53.35	57.7	62.05	67.85	70.75	73.65	76.55	79.45	82.35
Time (s)	1245	1318							
W(%)	83.8	84.52							

Criticize or defend the following proposition: The data give the time required for particles to fall 20 cm, making it easy to convert time to sedimentation velocity for each point. Equation (11) may then be used to convert the velocity into the radius of an equivalent sphere. The resulting graph of W versus radius is a cumulative distribution function similar to that shown in Figure 1.18b.

Solution: The proposition is on the right track, but needs refinement to be quantitatively correct. Remember, at any given time the weight of material that has accumulated equals the weight of all particles large enough to have settled onto the pan in the time of the experiment. Such particles come from two categories; they include all those particles large enough to have fallen through the full length of the container and a fraction of smaller particles that have only fallen from a fraction of the column but still land on the pan. Both of these contribute to the total weight observed at any time.

The two contributions can be calculated as follows: The quantity dW/dt at any instant t is the slope of the cumulative curve at a particular point. It represents the rate of deposition of particles in a size range that has not already settled out completely in the bottom during the time t. Since this size range has been settling at the same rate dW/dt from the beginning of the experiment (i.e., $t = 0$), the weight of the particles in this size range that has collected on the pan until time t is given by $t(dW/dt)$. Therefore, the remainder of the particles on the pan, that is, $W - [t(dW/dt)]$, represents the weight w corresponding to larger particles that have already settled out by time t. That is, the total weight W at any time t can be written as

$$W = w + t\frac{dW}{dt}$$

The weight w represents the *cutoff size* corresponding to time t, and dW/dt represents the rate of settling of particles smaller than the above cutoff size.

Figure 2.5 shows this graphically; note that the intercepts of the tangent lines, drawn at different times, give w, the weight of particles in the size class associated with time under consideration. It is a plot of this last quantity against time that gives a representation similar to Figure 1.18b.

In this example, about one-third of the solids are coarse enough to have settled out in 5 minutes. The remaining particles have velocities of $6.7 \cdot 10^{-4}$ m s^{-1} ($= 20 \cdot 10^{-2}$ m/300 s) or less, corresponding to equivalent spheres with radii of 12 μm or less. Incidentally, the equivalent sphere is a poor model for these clay particles, as can be seen from the electron micrographs in Figure 1.12b. ■

* * *

A slight variation of this method for analyzing particle size distributions is particularly convenient if the system contains particles of several discretely different particle sizes rather than a

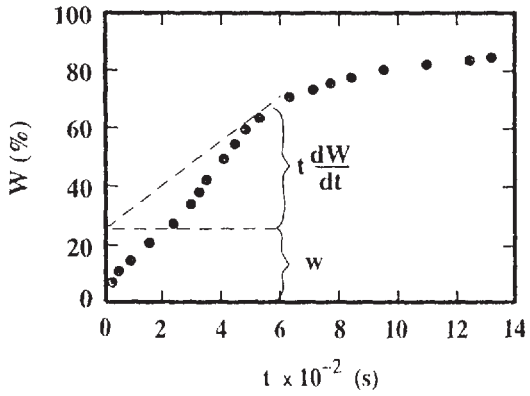


FIG. 2.5 Sedimentation of clay particles. Cumulative weight (as percentage) versus time for the data given in Example 2.1. The figure also shows the graphical construction explained in the example for obtaining the cutoff weight w associated with any time t. (Data from Oden, 1915.)

continuous distribution of sizes. A thin band of such a dispersion is layered on top of a column of pure solvent. Particles in the different size categories will settle through the solvent at different rates so that, if the column is long enough, the first fraction collected at the bottom of the column will contain only the largest particles, with smaller sizes showing up in progressively later fractions. The clear solvent is literally the column within which the various particle sizes are resolved. The "resolving power" of such an arrangement increases with the length of the column, at least up to some optimum length. It might be noted that there is a certain formal similarity between this situation and column chromatography.

Provided the column is long enough to separate the different sizes of particles, this method gives w versus time directly. Any additional interpretations of the results are made in a manner entirely analogous to the one just described. Since only a narrow band of the dispersion is used in this method, the weight of the dispersed phase in each fraction will be relatively low. Fairly sensitive analytical techniques are required for the fractions collected.

Sedimentation runs should be conducted at a constant temperature, not only so that $\Delta\rho$ and η are known, but also to minimize disturbances due to convection. Any sort of disturbance will obviously disrupt the segregation of the particles by size that has occurred as a result of sedimentation. An intrinsic difficulty with the balance method lies in the fact that the liquid below the balance pan is less dense than the liquid with dispersed particles above the pan. Thus, there is a tendency for a counterflow of pure solvent to arise, which would introduce an error in the particle size analysis.

The methods just discussed are only two of a wide variety of techniques that provide essentially the same kinds of information. In general, any measurement that gives (a) the amount of suspended material a fixed distance below the surface at various times or (b) the amount of material at various depths at any one time can be interpreted in terms of particle size distribution. Pressure, density, and absorbance are additional measurements that have been analyzed this way.

Instruments are commercially available that automatically carry out the analysis presented above. Although much of the tedium of the experiment is thereby relieved, automation does not alter the basic assumptions and/or approximations of the method. In one commercial instrument, a collimated beam of light or low-energy x-rays scans the settling compartment, measuring the absorbance of the suspended particles at various depths. Because the detector does not have to wait for settling particles to arrive at a fixed position, this method is more rapid than direct observation of sedimentation. A built-in computer converts the absorbance at a particular distance below the top of the sample to concentration after some time has elapsed from the start of settling. The distance-time combination is reduced to a velocity,

which is converted into the radius of an equivalent sphere by Equation (11). All of this is calculated automatically, and a recorder plots the cumulative weight percent of settling material versus the equivalent radius. The sample compartment is thermostated, and suspending media with a range of known densities and viscosities are used; the characteristics of the media are part of the input data of the operation. By utilizing the full range of operating parameters, this instrument allows equivalent diameters from 0.1 to 100 μm to be analyzed. Example 2.2 considers the application of Stokes's law to the output of such a particle analyzer.

* * *

EXAMPLE 2.2 *Use of the Sedimentation Equation for Particle Size Determination.* A titanium dioxide pigment of density 4.12 g cm $^{-3}$ is suspended in water at 33°C. At this temperature, the density and viscosity of water are 0.9947 g cm $^{-3}$ and 7.523 · 10 $^{-3}$ P, respectively. A particle size analyzer (SediGraph, Micromeritics Instruments Corp., Norcross, GA 30093) plots the following data for cumulative weight percent versus equivalent spherical radius:

Cumulative wt %										
95	90	80	70	60	50	40	30	20	10	5
Radius of an equivalent sphere (μm)										
0.60	0.45	0.33	0.37	0.28	0.27	0.24	0.22	0.21	0.18	0.15

To what sedimentation velocity does the most abundant component in this distribution correspond?

Solution: These cumulative percentages are of the same form as in Figure 2.5; therefore, the particle size distribution peaks where the cumulative curve increases most steeply. This occurs at about 0.29 μm for this distribution. Equation (10) permits the sedimentation velocity for particles of this size to be calculated:

$$v = \frac{2}{9}\frac{R_s^2(\rho_2 - \rho_1)g}{\eta} = \frac{2(2.9\cdot 10^{-5})^2(4.12 - 0.99)\,980}{9(7.52\cdot 10^{-3})}$$
$$= 7.62\cdot 10^{-5}\,cm\,s^{-1} = 7.62\cdot 10^{-7}\,m\,s^{-1}$$

The distance-time combination corresponding to this weight percent is determined by the particle size analyzer and converted to R_s. ■

* * *

A basic limitation of all these methods is the narrow range of particle sizes that can be investigated by sedimentation under gravity. Therefore, we turn next to a consideration of centrifugation, particularly the ultracentrifuge, as a means of extending the applicability of sedimentation measurements.

2.4 CENTRIFUGAL SEDIMENTATION

A well-known fact from elementary physics is that a particle traveling in a circular path of radius r at an angular velocity ω (in radians per second) is subject to an acceleration in the radial direction equal to $\omega^2 r$. Since there are 2π radians per revolution, ω equals the number of revolutions per second (rps) times 2π, or the number of revolutions per minute (rpm) times the ratio $2\pi/60$. It is not particularly difficult to produce accelerations by centrifugation that are many times larger than the acceleration due to gravity. It is conventional, in fact, to describe the radial acceleration achieved by a centrifuge as some multiple of the standard gravitational acceleration or as being so many g's, where g is about 9.8 m s $^{-2}$.

2.4a Ultracentrifuge

The ultracentrifuge is an instrument in which a cell is rotated at very high speeds in a horizontal position. As we see below, the gravitational acceleration is easily increased by a factor of 10^5 in such an apparatus. Accordingly, the particle size that may be studied by sedimentation

is decreased by the same factor. Two types of ultracentrifuges, analytical and preparative, are common, particularly in biological analyses. The analytical ultracentrifuges require very small amounts of samples (about 0.1 to 1 ml) and use optical systems designed to measure concentration or concentration gradients in the samples directly. Preparative ultracentrifuges typically use larger volumes of samples (5 to 100 ml) and are run in "batch mode," that is, the contents are removed from the centrifuge after a fixed amount of time, the purpose being fractionation for preparation or purification of samples. The ultracentrifuge has been used extensively for the characterization of colloidal materials, particularly those of biological origin, such as proteins, nucleic acids, and viruses (Cantor and Schimmel 1980).

Because of the extreme importance of the ultracentrifuge, it seems appropriate to describe it in some detail. Although the particulars differ from instrument to instrument, the essential features are present in all ultracentrifuges. The actual sedimentation takes place in a cell mounted within an aluminum or titanium rotor. The cell is sector shaped; its side walls converge toward the center along radial lines. Since the radial acceleration is proportional to the distance from the axis of rotation, we see that this quantity varies from top to bottom in the cell. Although this variation is considered explicitly in a section below, it is sufficient for the present to consider the average acceleration at the midpoint of the cell, which is typically located $6.5 \cdot 10^{-2}$ m from the center of the rotor. For speeds of 10,000, 20,000, and 40,000 rpm, accelerations of

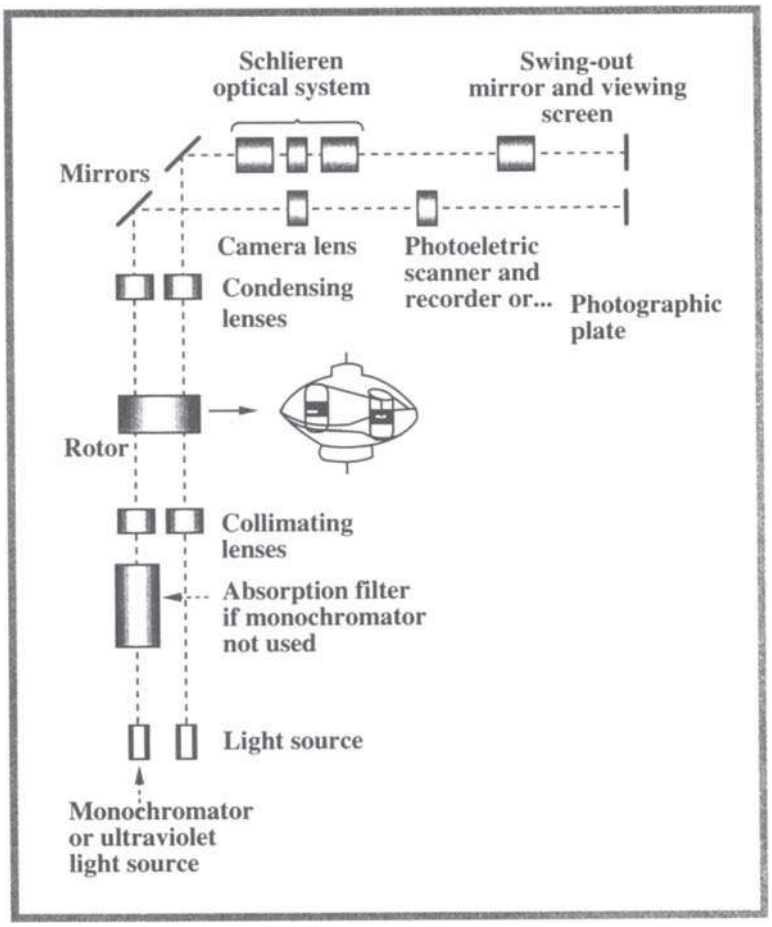


FIG. 2.6 Schematic of optical systems in the Spinco Model E Ultracentrifuge (Beckman Instruments, Inc., Spinco Division, Palo Alto, CA).

$7.13 \cdot 10^4$, $2.85 \cdot 10^5$, and $1.14 \cdot 10^6$ m s^{-2} are produced at this radius. These correspond to $7.27 \cdot 10^3$, $2.91 \cdot 10^4$, and $1.16 \cdot 10^5$ times the acceleration of gravity, respectively.

An important part of the ultracentrifuge is the optical system that makes observation during operation a possibility; it is shown schematically in Figure 2.6. The sample compartment fits into a hole in the rotor and is positioned to intersect the light path of the optics. Those faces of the cell perpendicular to the light path are transparent to visible and ultraviolet light so that various optical methods of chemical analysis may be employed to measure the distribution of material along the sedimentation path. Schlieren refractometry, interferometry, and spectrophotometry are all employed for this purpose, the last being particularly useful when very low concentrations are involved. The location of the settling particles may be followed by one of these methods, and the results recorded either photographically or on a chart recorder. Figure 2.7a shows how the concentration profile varies with radial location in the ultracentrifuge as the particles settle. Although the boundary between the solution and the solvent that has been swept free of solute seems clear in this representation, the use of schlieren refractometry makes the sedimentation even more easy to visualize.

The schlieren system of optics is an analytical method that is particularly well suited to following the location of a chemical boundary with time. It is routinely employed in ultracentrifuges and also in electrophoresis experiments, as we see in Chapter 12. Schlieren optics produces an effect that depends on the way the refractive index varies with position, that is, the refractive index *gradient* rather than on the refractive index itself. Therefore, the schlieren effect is the same at all locations along the axis of sedimentation, except at any place where the refractive index is changing. In such a region, it will produce an optical effect that is proportional to the refractive index gradient. The boundary between two layers is thus per-

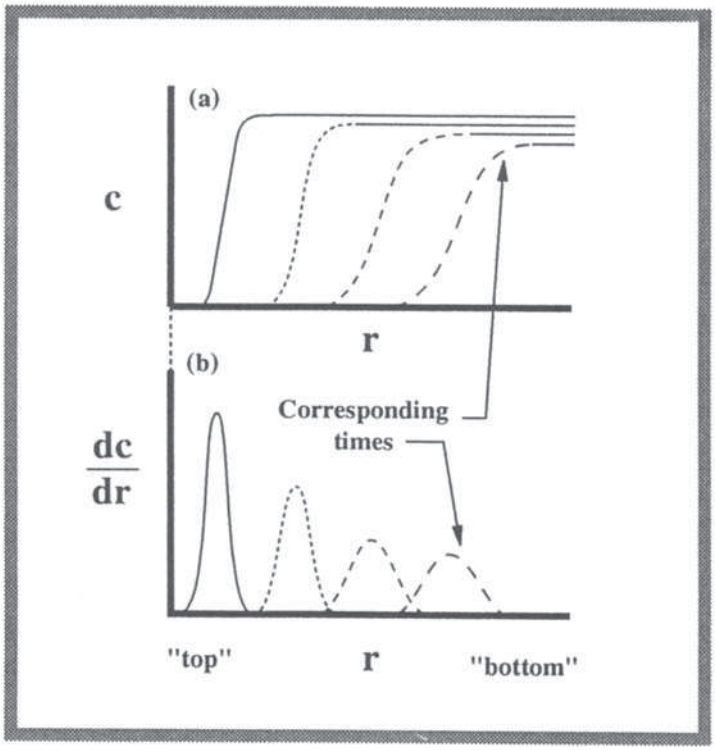


FIG. 2.7 Location of the boundary for particles settling in an ultracentrifuge: (a) concentration profile; and (b) derivative profile as revealed by schlieren optics. (Redrawn with permission of P. C. Hiemenz, *Polymer Chemistry: The Basic Concepts*, Marcel Dekker, New York, 1984.)

ceived by schlieren optics as a sharp peak on a flat baseline. The displacement of such a peak with time measures the velocity of the boundary. Likewise, a band of settling particles is seen as a broad schlieren peak with a width and velocity that measure the width and velocity of the band. The schlieren optical system works by using a diaphragm to cut off from a photographic plate the light that is deviated from the optical path by the refractive index gradient. The same effect may be produced on a ground-glass screen for instantaneous viewing by an ingenious system that combines a diagonal slit and a cylindrical lens to produce an image of the refractive index gradient. Figure 2.7b shows how the same data presented in Figure 2.7a would appear with schlieren optics. It is thus an easy matter to monitor the peak location as the sedimentation proceeds.

The usual precautions regarding temperature and vibration control also apply to the ultracentrifuge. To overcome air resistance and the attendant frictional heating, the compartment in which the rotor spins is evacuated and may be thermostated over a wide range of temperatures. The rotor is mounted on a flexible drive shaft that minimizes the need for precise balancing as a requirement for vibration-free operation. Finally, the rotor assembly is enclosed in an armored steel chamber for safety. At these speeds, a runaway rotor is deadly!

2.4b Sedimentation Coefficient

The results of a sedimentation experiment in a centrifugal field are conventionally reported in terms of what is known as a *sedimentation coefficient*. This quantity is defined as

$$s = \frac{dr/dt}{\omega^2 r} \tag{15}$$

that is, it equals the sedimentation velocity per unit of centrifugal acceleration. In the SI system, this ratio has the units m s^{-1}/m s^{-2} or seconds. In practice, the quantity 10^{-13} s is defined to be 1.0 *svedberg* (1.0 S) after T. Svedberg, the originator of the ultracentrifuge and pioneer in its use (Nobel Prize, 1926). Sedimentation coefficients are usually reported in this unit. If the location of a particle along its settling path is measured as a function of time, the sedimentation coefficient is readily evaluated by integrating Equation (15) using the following limits of integration. The component is at radial position r_1 at time t_1 and at r_2 at t_2. Therefore,

$$\omega^2 s \int dt = \int \frac{dr}{r} \tag{16}$$

or

$$s = \frac{\ln (r_2/r_1)}{\omega^2 (t_2 - t_1)} \tag{17}$$

The sedimentation coefficient depends on concentration; consequently, it is usually measured at several different concentrations, and the results are extrapolated to zero concentration. It is customary to designate this limiting value by a superscript zero. Experimental values are also generally labeled with respect to temperature, so a value listed as s^0_{20} corresponds to a sedimentation coefficient measured at (or corrected to) 20°C and extrapolated to zero concentration. Under stationary state conditions, the force due to the centrifugal field and the viscous force of resistance will be equal. Therefore, $\omega^2 r$ replaces g in Equation (4) to give

$$m\left(1 - \frac{\rho_1}{\rho_2}\right)\omega^2 r = f\frac{dr}{dt} \tag{18}$$

where f is the friction factor for the settling particle. Equations (18) and (15) may be combined to yield

$$\frac{m}{f}\left(1 - \frac{\rho_1}{\rho_2}\right) = s \tag{19}$$

which shows that the sedimentation coefficient is directly proportional to the ratio of the mass to the friction factor. As with sedimentation under the acceleration of gravity, any further interpretation of m/f depends either on independent determination of the friction factor from diffusion or on the assumption of spherical particles, with Equation (8) used to evaluate f. Thus, experimental sedimentation coefficients may be analyzed to yield the mass, radius, and friction factor of an equivalent sphere. In the absence of supplementary data, this is as far as sedimentation alone can be interpreted. Example 2.3 illustrates the use of these relationships.

* * *

EXAMPLE 2.3 *Sedimentation in an Ultracentrifuge.* What should be the speed of an ultracentrifuge so that the boundary associated with the sedimentation of a particle of molecular weight 60,000 g mole^{-1} moves from $r_1 = 6.314$ cm to $r_2 = 6.367$ cm in 10 min? The densities of the particle and the medium are 0.728 and 0.998 g cm^{-3}, respectively, and the friction factor of the molecule is $5.3 \cdot 10^{-11}$ kg s^{-1}.

Solution: First we calculate the particle mass by dividing the molecular weight by Avogadro's number:

$$m = 60{,}000/6.02 \cdot 10^{23} = 9.97 \cdot 10^{-20} \text{ g molecule}^{-1}$$
$$= 9.97 \cdot 10^{-23} \text{kg molecule}^{-1}$$

Equation (19) allows us to calculate the sedimentation coefficient:

$$s = (m/f)[1 - (\rho_1/\rho_2)] = (9.97 \cdot 10^{-23}/5.3 \cdot 10^{-11})[1 - (0.998/0.728)]$$
$$= 6.98 \cdot 10^{-13} s = 6.98 \text{ S}$$

Equation (17) can now be solved for ω^2

$$\omega^2 = \frac{\ln(r_2/r_1)}{s(t_2 - t_1)} = 1.997 \cdot 10^7 s^{-2}$$

or $\omega = 4.47 \cdot 10^3$ rad s^{-1}. Dividing by 2π converts this to revolutions per second: $\omega = 711$ revolutions per second, or 42,700 rpm. ■

* * *

In the preceding sections of this chapter, we have considered sedimentation as if it were the only process that influenced the spatial distribution of particles. If this were the case, all systems of dispersed particles, even gases, would eventually settle out. In practice, convection currents arising from temperature differences keep many systems well stirred. Even in carefully thermostated laboratory samples, however, there is another factor operating that prevents the complete sedimentation of small particles, namely, *diffusion*.

Diffusion and sedimentation are opposing processes inasmuch as the former tends to keep things dispersed, whereas the latter tends to collect them in one place. Diffusion is more important for smaller particles: remember that sedimentation is negligible for gases. For larger particles, diffusion is negligible. Of course, there is a range of particle sizes for which both effects are comparable; we examine the combined effects of sedimentation and diffusion in Section 2.8. Even for very small particles, the balance between diffusion and sedimentation becomes important if "sedimentation" is induced through centrifugal forces, as in the case of ultracentrifuges or separation techniques such as the sedimentation field flow fractionation described in Vignette II. For the present, however, let us examine the phenomenon of diffusion.

2.5 DIFFUSION

If external forces such as gravity can be neglected, the composition of a single equilibrium phase will be macroscopically uniform throughout. This means that the concentration or density is constant throughout the phase. It should be noted that we are talking about the macroscopic description of the phase. At the molecular level, there will be local fluctuations from the mean value, a fact important in, for example, light scattering, as we see in Chapter 5 when we discuss Rayleigh scattering from solutions (Section 5.3b). However, for the present,

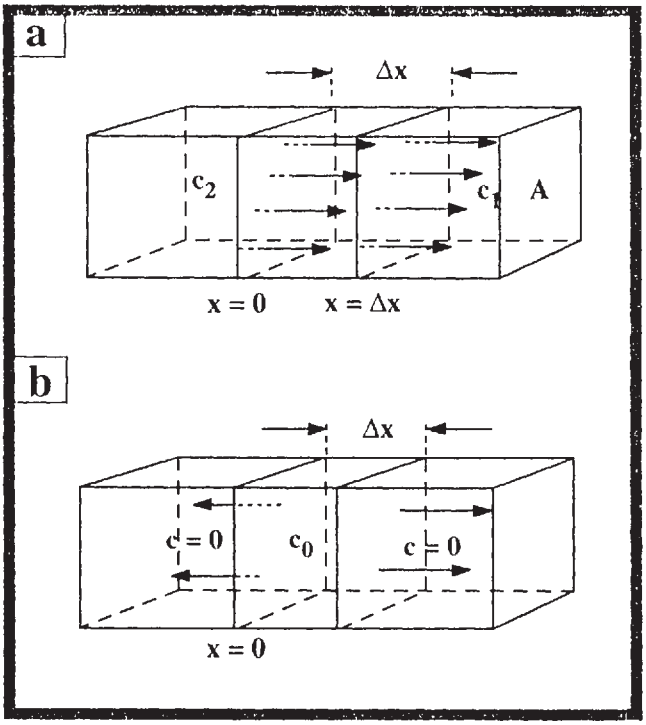


FIG. 2.8 Schematic of a diffusion process through a volume element of thickness Δx: (a) the volume element separates two solutions of concentrations c_1 and c_2; and (b) the element contains a solution at concentration c_0 and separates two regions of pure solvents.

we restrict ourselves to the macroscopic description. Fundamentally, it is the second law of thermodynamics that is responsible for the uniform distribution of matter at equilibrium since entropy is maximum when the molecules are distributed randomly throughout the space available to them. If for some reason a nonuniform distribution of matter should exist, the particles of the system will experience a force that tends to distribute them uniformly. Consider, for example, the situations shown in Figure 2.8a, in which two solutions of different concentrations are separated by a hypothetical porous plug of thickness dx. We can set $dx = 0$; in this case, there will be a migration of solute from the high-concentration side to the low-concentration side until a uniform distribution is obtained, that is, equilibrium is reached. The resulting transport of matter is called *diffusion*. In the following sections, we set up the equations that describe the diffusional flux and the consequent readjustment of concentration and will develop an equation that relates the diffusion coefficient to the friction coefficient.

2.5a Fick's Laws of Diffusion

Now let us define the following quantities. Suppose Q represents the amount of material that flows through a cross section of area A in Figure 2.8. The rate of change in the quantity Q/A is called the flux of solute across the boundary; that is, the flux J is defined as

$$J = \frac{d(Q/A)}{dt} \tag{20}$$

Therefore, the amount of material that crosses A in a time interval Δt is

$$\Delta Q = AJ\Delta t \tag{21}$$

The phenomenological equation that relates the flux of material across the boundary to the concentration gradient at that location is given by *Fick's first law*:

$$J = -D\frac{\partial c}{\partial x} \tag{22}$$

where D is defined as the *diffusion coefficient* of the solute. With the amount of material measured in the same units in both J and c, Equations (20) and (22) show that the SI units of D are square meters per second.

Now, instead of a boundary of zero thickness, let us consider the concentration changes that occur within a zone of cross section A, which has a thickness Δx as shown in Figure 2.8a. Any change in the amount of material in this zone will equal the difference between the amount of material that enters the zone and the amount of material that leaves it:

$$\Delta Q = Q_{in} - Q_{out} = (J_{in} - J_{out})A\Delta t \tag{23}$$

The quantity ΔQ also equals the product of the volume of the zone and the concentration change that occurs in it:

$$\Delta Q = A\Delta x\Delta c \tag{24}$$

Equating Equations (23) and (24) and substituting Fick's first law from Equation (22), we obtain

$$-D\frac{[(\partial c/\partial x)_{x=0} - (\partial c/\partial x)_{x=\Delta x}]}{\Delta x} = D\frac{\Delta(\partial c/\partial x)}{\Delta x} = \frac{\Delta c}{\Delta t} \tag{25}$$

assuming that the cross section is uniform throughout the compartment and that D is independent of small changes of concentration. Finally, if we take the limit of Equation (25) as Δx and Δt approach zero, we obtain

$$\frac{\partial c}{\partial t} = D\frac{\partial^2 c}{\partial x^2} \tag{26}$$

a result known as *Fick's second law*.

Equation (26) is a differential equation with a solution that describes the concentration of a system as a function of time and position. The solution depends on the boundary conditions of the problem as well as on the parameter D. This is the basis for the experimental determination of the diffusion coefficient. Equation (26) is solved for the boundary conditions that apply to a particular experimental arrangement. Then, the concentration of the diffusing substance is measured as a function of time and location in the apparatus. Fitting the experimental data to the theoretical concentration function permits the evaluation of the diffusion coefficient for the system under consideration.

Rather than getting deeply involved in the mathematics of differential equations, we use a statistical model to find a solution to Equation (26) for a system with simple boundary conditions. This will be sufficient to illustrate the experimental technique by which diffusion coefficients are determined and will also lead to a better understanding of the random processes underlying diffusion. This statistical discussion and the experimental procedure it suggests are addressed in Sections 2.6c and 2.6d.

Next, it will be helpful to anticipate a description of experimental procedures and consider the magnitude of measured diffusion coefficients. The self-diffusion coefficients for ordinary liquids with small molecules are of the order of magnitude 10^{-9} m^2 s^{-1}; for colloidal substances, they are typically of the order 10^{-11} m^2 s^{-1}. In the next section, we see that for spherical particles the diffusion coefficient is inversely proportional to the radius of the sphere. Therefore, every increase by a factor of 10 in size decreases the diffusion coefficient by the same factor. Qualitatively, this same inverse relationship applies to nonspherical particles as well. Once again, we see that diffusion decreases in importance with increasing particle size, precisely those conditions for which sedimentation increases in importance. For larger particles, for which D is very small, the diffusion coefficient also becomes harder to measure. For

spherical particles, the time required for the particle to diffuse a unit distance is directly proportional to its radius. For small molecules, this time is experimentally accessible; it becomes experimentally inconvenient for particles at the upper end of the colloidal range.

2.5b Diffusion Coefficient and Friction Factor: Thermodynamic Description of Diffusion

As already noted, the driving force underlying diffusion can be thought of as thermodynamic in origin. A very general way to describe a force is to write it as the negative gradient of a potential. In the context of diffusion, the potential to be used is the chemical potential μ_i, the partial molal Gibbs free energy of the component of interest. Thus, the magnitude of the driving force (per particle) diffusion may be written as

$$F_{diff} = -\frac{1}{N_A}\frac{\partial \mu_i}{\partial x} \tag{27}$$

This is often called the thermodynamic force for diffusion. It is necessary to divide by Avogadro's number N_A since μ_i is a molar quantity. Thermodynamics show that

$$\mu_i = \mu_i^0 + RT \ln a_i = \mu_i^0 + RT \ln (\gamma_i c_i) \tag{28}$$

where a_i, c_i, and γ_i are the activity, concentration, and activity coefficient, respectively, of the ith component. Since we are interested in infinitely dilute systems, the activity coefficient may be assumed to equal unity. Substituting Equation (28) into Equation (27) gives

$$F_{diff} = k_B T \frac{\partial \ln c_i}{\partial x} = -\frac{k_B T}{c_i}\frac{\partial c_i}{\partial x} \tag{29}$$

where k_B *is Boltzmann's constant* R/N_A. Under stationary state conditions, this force will be equal to the force of viscous resistance, given by fv according to Equation (2). Therefore, the magnitude of the velocity of diffusion equals

$$v = -\frac{k_B T}{fc}\frac{\partial c}{\partial x} \tag{30}$$

where the subscript has been omitted from the concentration of the solute c because this is now the only quantity involved in the relationship. Finally, we make the following observation. The flux of material through a cross section equals the product of its concentration and its diffusion velocity:

$$J = cv_{diff} \tag{31}$$

Combining Equations (30) and (31) and comparing with Equation (22) leads to the important result

$$D = \frac{k_B T}{f} \tag{32}$$

It should be noted that this derivation contains no assumptions about the shape of the particles. However, when the particles are assumed to be spherical, we can substitute Equation (8) for f, and the resulting equation for the diffusion coefficient is the well-known *Stokes-Einstein relation*.

Many of the relationships of this chapter have involved the friction factor f, which, until now, has been an unknown quantity except for spherical particles. Equation (32) breaks this impasse and points out the complementarity between sedimentation and diffusion measurements. For example, substitution of Equation (32) into (4) gives

$$m = \frac{k_B T v}{D[1 - (\rho_1/\rho_2)]g} \tag{33}$$

and substituting Equation (32) into Equation (19) yields

$$m = \frac{k_B T s}{\mathcal{D}[1 - (\rho_1/\rho_2)]} \tag{34}$$

Equations (33) and (34) show that diffusion studies combined with sedimentation studies, either under the force of gravity or in a centrifuge, yield information about particle masses with no assumptions about the shape of the particle.

Human hemoglobin, for example, has a sedimentation coefficient of 4.48 S and a diffusion coefficient of $6.9 \cdot 10^{-11}$ m^2 s^{-1} in aqueous solution at 20°C. The density of this material is 1.34 g cm^{-3}. Substituting these values into Equation (34) shows the particle mass to be

$$\begin{aligned} m &= (1.38 \cdot 10^{-23})(293)(4.48 \cdot 10^{-13})/(6.9 \cdot 10^{-11})(1 - 1.0/1.34) \\ &= 1.03 \cdot 10^{-22} \text{ kg particle}^{-1} \end{aligned} \tag{35}$$

or, in terms of molecular weight,

$$M = (1.03 \cdot 10^{-22})(6.02 \cdot 10^{23}) = 62{,}300 \text{ g mol}^{-1} \tag{36}$$

Since the friction factor was measured experimentally, this value is correct regardless of the state of solvation or ellipticity of the hemoglobin molecules in solution. We see below how the combination of these two experimental approaches may be interpreted further to yield some information about solvation and ellipticity.

2.5c Effect of Solvation

We have already seen that the ratio f/f_0 describes the effect on the friction factor of either solvation, ellipticity, or both. This ratio equals unity for an unhydrated sphere and increases with both the amount of bound solvent and the axial ratio of the particles. We are now in a position to see how this ratio may be evaluated experimentally. The steps of the procedure are the following:

1. The diffusion coefficient and sedimentation velocity (or sedimentation coefficient) are used to evaluate m by Equation (33) or (34).
2. The friction factor is evaluated from the diffusion coefficient by Equation (32).
3. If we assume the particle to be unsolvated, we can determine its volume by dividing its mass by the density of the dry material.
4. The radius of the particle is calculated from its volume if we assume the particle to be a sphere.
5. The friction factor f_0 of the equivalent unsolvated sphere is evaluated by Equation (8).
6. The ratio of the experimental friction factor to f_0 is determined.

As noted, the larger this ratio is than unity, the more the particle deviates from the unsolvated spherical shape assumed to calculate f_0. Although this statement is qualitatively accurate, we realize that the f/f_0 value reflects some actual quantitative condition of the particles, and we search for additional ways to interpret this ratio.

The factoring of f/f_0 into two contributions in Equation (14) was introduced with this interpretation in mind. It turns out that the effect of solvation is easily handled by a physically reasonable model through f^*/f_0. Dealing with particle asymmetry through f^*/f is more complicated, but this has also been analyzed according to a model that is plausible for many situations. Let us now consider each of these developments in turn.

When particles are solvated, a certain volume of the solvent must be counted as part of the dispersed phase rather than the continuous phase. In dilute solutions, the effect of this reclassification of some solvent is negligible for the remaining solvent (component 1), but the effect on the solute (component 2) may be considerable. The effect of the attached solvent on the volume of the solute particles may be calculated if some model is assumed for the mode of attachment. To assume the solvation occurs uniformly throughout the particle is a plausible

model for the solvation of lyophilic colloids. For example, protein molecules have ionic and polar substituent groups distributed at random along the polymer chain. Such groups are expected to be extensively hydrated in an aqueous solution.

If we assume that solvation occurs at numerous positions along the molecule, the volume of the solvated particle may be written as the sum of the volume of the unsolvated particle and the volume of the bound (subscript b) solvent:

$$\text{volume of solvated particle} = V_2 + V_{1,b} \tag{37}$$

This can be rearranged as

$$V_2 + V_{1,b} = V_2\left(1 + \frac{V_{1,b}}{V_2}\right) = V_2\left[1 + \frac{m_{1,b}}{m_2}\left(\frac{\rho_2}{\rho_1}\right)\right] \tag{38}$$

where $m_{1,b}$ is the mass of bound solvent and the last equality requires the density of the bound solvent to be identical to that of the free solvent. In the event that the solvated specie differs in density from its bulk counterpart, the ratio ρ_2/ρ_1 may be replaced by $(\bar{v}_1/\bar{v}_2)$, where the $\bar{v}_i$'s are the partial specific volumes of the components, and measure the volume per unit mass of the indicated components in the solution at the indicated concentration and hence reflect any volume changes that accompany the interaction. We continue to discuss solvation in terms of Equation (38), using the more familiar densities.

We are explicitly looking for the effect of solvation on the friction factor of a sphere; particle asymmetry is handled by the second ratio f/f^*. By Stokes's law, the ratio of the friction factors of solvated to unsolvated spheres f^*/f_0 is equal to the ratio of their radii a_{solv}/a_0. This ratio equals the cube root of the volume ratio of the solvated to unsolvated spheres, or

$$\frac{f^*}{f_0} = \left[1 + \frac{m_{1,b}}{m_2}\left(\frac{\rho_2}{\rho_1}\right)\right]^{1/3} \tag{39}$$

The ratio f^*/f_0 equals unity when $m_{1,b} = 0$ and increases with increasing solvation, as required. We examine an application of this relationship below.

Asymmetry as well as solvation can cause a friction factor to have a value other than f_0. Next let us consider the ratio f/f^*, which, according to Equation (14), accounts for the effect of particle asymmetry on the friction factor. We saw in Section 1.5 that ellipsoids of revolution are reasonable models for many asymmetric particles.

Jean Perrin derived expressions for the ratio f/f^* for ellipsoids of revolution in terms of the ratio of the equatorial semiaxis to the semiaxis of revolution b/a. The following expressions were obtained:

First, for prolate ellipsoids ($b/a < 1$),

$$\frac{f}{f^*} = \frac{\left[1 - \left(\frac{b}{a}\right)^2\right]^{1/2}}{\left(\frac{b}{a}\right)^{2/3} \ln\left(\frac{1 + [1 - (b/a)^2]^{1/2}}{b/a}\right)} \tag{40}$$

Second, for oblate ellipsoids ($b/a > 1$),

$$\frac{f}{f^*} = \frac{\left[\left(\frac{b}{a}\right)^2 - 1\right]^{1/2}}{\left(\frac{b}{a}\right)^{2/3} \tan^{-1}\left[\left(\frac{b}{a}\right)^2 - 1\right]^{1/2}} \tag{41}$$

Equations (40) and (41) allow the ratio f/f^* to be evaluated for any axial ratio. The reverse problem, finding an axial ratio that is consistent with an experimental f/f^* ratio, is best accomplished by interpolating from plots of Equations (40) and (41). These two equations, along with Equations (14) and (39), allow the states of solvation and ellipticity compatible

with an experimental friction factor to be established. Example 2.4 considers this for the human hemoglobin molecule.

* * *

EXAMPLE 2.4 *Solvation and Ellipticity of Human Hemoglobin from Sedimentation Data.* The diffusion coefficient of the human hemoglobin molecule at 20°C is $6.9 \cdot 10^{-11}$ m^2 s^{-1}. Use this value to determine f for this molecule. Evaluate f_0 for hemoglobin using the particle mass calculated in Equation (35). Indicate the possible states of solvation and ellipticity that are compatible with the experimental f/f_0 ratio.

Solution: According to Equation (32), $f = k_B T/D = (1.38 \cdot 10^{-23})(293)/6.9 \cdot 10^{-11} = 5.86 \cdot 10^{-11}$ kg s^{-1}.

Equation (35) shows that the mass of one of these molecules is equal to $1.03 \cdot 10^{-22}$ kg.

Dividing this by the density (1.34 g cm^{-3}) gives $7.69 \cdot 10^{-26}$ m^3 as the particle volume. The radius of an unsolvated sphere of this volume is given by $(3V/4\pi)^{1/3} = 2.64 \cdot 10^{-9}$ m.

Using the viscosity of water, 0.01 P, the friction factor f_0 is calculated by Equation (8):

$$f_0 = (6\pi)(10^{-3})(2.64 \cdot 10^{-9}) = 4.98 \cdot 10^{-11} \text{ kg s}^{-1}$$

Since these calculations involve only empirical data, we can say the experimental value of $f/f_0 = 5.86 \cdot 10^{-11}/4.98 \cdot 10^{-11} = 1.18$.

Next we recall Equation (14), which interprets f/f_0 as the product

$$(f/f^*)(f^*/f_0)$$

If $f/f^* = 1.00$, the particle is a sphere and $f^*/f_0 = 1.18$. Using the bulk density of water and the protein and Equation (39), we can show that $m_{1,b}/m_2 = 0.48$ g water per g hemoglobin.

If $f^*/f_0 = 1.00$, the particle is unsolvated and $f/f^* = 1.18$. This corresponds to an axial ratio $b/a = 0.24$, according to Equation (40), if the particle is a prolate ellipsoid, or to $b/a = 4.0$, according to Equation (41), if the particle is oblate. These values are summarized in Table 2.1. ■

* * *

It is apparent that there are many combinations of f/f^* and f^*/f_0 that form the product 1.18, which we attempted to explain in Example 2.4. Setting either one of these contributions equal to unity merely establishes an upper limit for the other. Intermediate cases between solvated spheres and unsolvated ellipsoids can also be calculated that are consistent with a given axial ratio.

TABLE 2.1 A Summary of the Characterization of Human Hemoglobin at 20°C Based on Sedimentation and Diffusion Measurements

Quantity	Value	Determination
Sedimentation coefficient, s	4.48×10^{-13} s	Experimental
Diffusion coefficient, D	6.9×10^{-11} m^{2}s^{-1}	Experimental
Density, ρ	1.34 g cm^{-3}	Experimental
Mass of particle, m	1.03×10^{-22} kg molecule^{-1}	Eq. (35)
Molecular weight	6.23×10^4 g mol^{-1}	Eq. (36)
f	5.86×10^{-11} kg s^{-1}	Eq. (32)
Volume of particle	7.69×10^{-26} m^3	$V_{unsolv} = m/\rho$
Radius of equivalent sphere	2.64×10^{-9} m	$R_{sph} = (3V/4\pi)^{1/3}$
f_0	4.98×10^{-11} kg s^{-1}	Eq. (8)
f/f_0	1.18	—
$m_{1,b}/m_2$, if spherical	0.48 g H_2O (g protein)$^{-1}$	Eq. (39)
a/b if unsolvated and prolate	4.0	Eq. (40)
a/b if unsolvated and oblate	0.24	Eq. (41)

Source: Data from E. J. Cohn, and J. T. Edsall, *Proteins, Amino Acids and Peptides*, American Chemical Society Monograph, reprinted by Hafner, New York, 1965.

Figure 2.9 is a plot of possible combinations of hydration and asymmetry for protein particles in water. Similar curves could be drawn for other materials as well. For the human hemoglobin molecule discussed in Table 2.1, the combination of sedimentation and diffusion measurements gives an f/f_0 value that lies within the domain defined by the 1.15 and 1.20 contours of Figure 2.9. The current picture of the structure of human hemoglobin, deduced from x-ray diffraction studies, suggests that the molecule may be regarded as an ellipsoid with height, width, and depth equal to 6.4, 5.5, and 5.0 nm, respectively. Applying these dimensions to the dispersed unit leads us to describe the particle as being hydrated to the extent of about 0.4–0.5 g water (g protein)$^{-1}$.

Sedimentation and diffusion data allow for the unambiguous determination of particle mass and also allow the suspended particles to be placed on a contour in a plot such as that of Figure 2.9. This is as far as these experiments can take us toward the characterization of the particles. Of course, additional data from other sources, such as the x-ray diffraction results just cited, may lead to still further specification of the system. One such source of information is intrinsic viscosity data for the same dispersion. In Chapter 4 we discuss the complementarity between viscosity data and sedimentation-diffusion results (see Section 4.7b).

2.6 BROWNIAN MOTION AND DIFFUSION

A liquid that is totally homogeneous on a macroscopic scale undergoes continuous fluctuations at the molecular level. As a result of these fluctuations, the density of molecules at any location in the liquid varies with time and at any time varies with location in such a way that the mean density of the sample as a whole has its bulk value. This pattern of "flickering" molecular densities will produce continually varying pressures on the surface of any particle

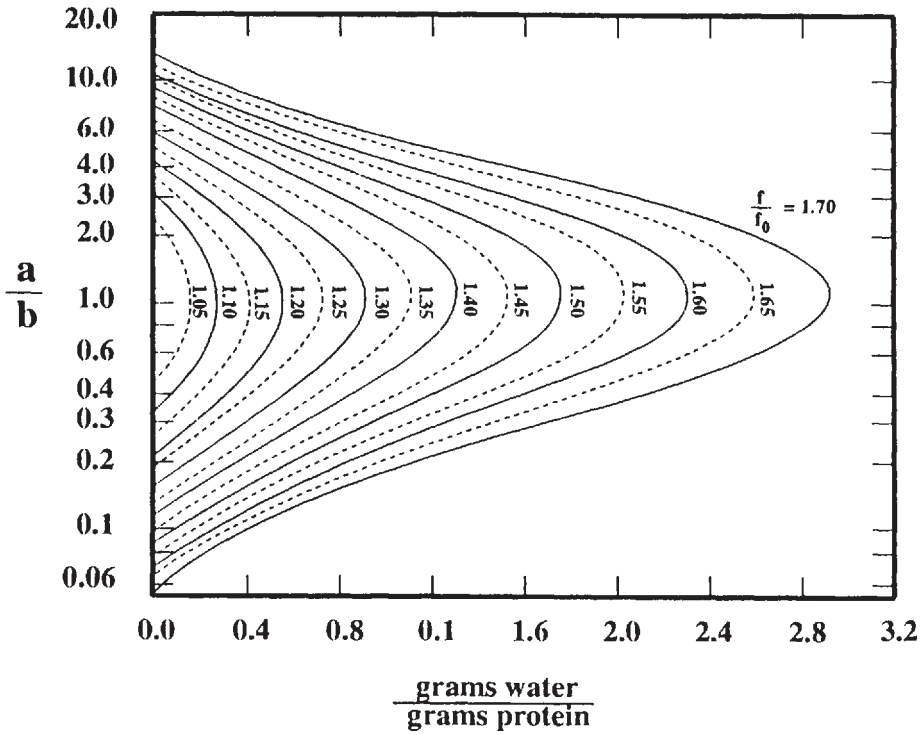


FIG. 2.9 Variations of the ratio f/f_0 with asymmetry and hydration for aqueous protein dispersions. (Redrawn from L. Oncley, *Ann. NY Acad. Sci.* **41**, 121 (1941).

submerged in the liquid. Since the fluctuations are confined to domains of the order of molecular dimensions, this randomly variable pressure is quite small. A small particle will be displaced, however, by the resulting force imbalance at its surface. The pattern of its displacements will also be a totally random thing, a reflection of the fluctuations that cause the motion. As the size of the submerged particles increases, the effect of fluctuations on them decreases. In this section we consider the trajectory of a particle in the colloidal size range that is engaged in this pattern of motion. Such movements have been studied microscopically and are called *Brownian motion* after Robert Brown, a British biologist who described them in 1828.

In this section we review random walk statistics and their relation to diffusion. More elaborate discussions of these and related topics are available in Berg (1993); the collection of the original papers of Einstein (reprinted in Einstein 1956) is another excellent source of this material.

2.6a Random Walk and Random Walk Statistics

In general, a dispersed particle is free to move in all three dimensions. For the present, however, we restrict our consideration to the motion of a particle undergoing random displacements in one dimension only. The model used to describe this motion is called a one-dimensional *random walk*. Its generalization to three dimensions is straightforward.

To begin, the statistical nature of this phenomenon should be apparent. We might watch the pattern of displacements of thousands of otherwise identical particles and find no uniformity in the zigzag steps they follow. Only statistical quantities such as the "average" displacement after a certain number of steps or after a certain elapsed time make any sense. Let us consider how to calculate such a quantity.

Suppose we consider a game in which a marker is moved back and forth along a line, say, the x axis, in a direction determined by the toss of a coin. The rules of the game provide that we move the marker a distance ℓ in the plus direction every time "heads" is tossed, and a distance ℓ in the negative direction every time "tails" is tossed. If n_+ and n_- represent the number of heads and tails, respectively, in a game consisting of a total of n tosses, then we may write

$$n_+ + n_- = n \tag{42}$$

We may also write

$$x = (n_+ - n_-)\ell \tag{43}$$

where x is the net displacement of the marker after n tosses. These two equations may be solved simultaneously for n_+ and n_- to give

$$n_+ = \frac{1}{2}\left(n + \frac{x}{\ell}\right) \tag{44}$$

and

$$n_- = \frac{1}{2}\left(n - \frac{x}{\ell}\right) \tag{45}$$

The problem of interest may now be expressed as follows. What is the probability $P(n,x)$ that the marker will be at position x after n moves? The answer to this problem is supplied by the well-known binomial distribution formula

$$P(n,x) = \frac{n!}{n_+!\, n_-!}\, p_+^{n_+} p_-^{n_-} \tag{46}$$

in which p_+ and p_- represent the probability of throwing a head or a tail, respectively, in a single toss. Although this formula is often used, its validity is not always fully appreciated. Accordingly, let us briefly examine the origin of Equation (46).

Each time we toss an unbiased coin, the outcome is independent of all other tosses. The

probability of tossing n_+ heads is therefore $p_+^{n_+}$ since the probabilities compound by multiplication when we require a specified set of outcomes. Therefore n_- events, each of which occurs with a probability p_+, will occur with a probability $p_+^{n_+}$. If we specify the outcome further by requiring n_- tails in addition to the heads already specified, then the probability is given by $p_+^{n_+} \cdot p_-^{n_-}$. What we have calculated is the probability of a particular, fully specified sequence of outcomes such as the following for $n = 6$: HHHTTT. However, the same net number of heads and tails could come about in other ways. For example, HTTTHH, HHTHTT, and HTHTHT are also consistent with $n = 6$ and $n_+ = n_- = 3$, along with the distribution previously given and other possibilities. Therefore we must multiply the probability of one specified sequence by the number of other sequences that have the same net composition of heads and tails.

At first glance, this factor may appear to be $n!$, the number of different ways the n items can be rearranged, giving a probability of $n!\, p_+^{n_+} p_-^{n_-}$. This quantity overcounts, however, since not all rearrangements are recognizably different. For example, we can interchange the first two members of the sequence TTHHHT to produce the sequence TTHHHT; the interchange obviously leaves the sequence unaltered. Therefore we must divide $n!\, p_+^{n_+} p_-^{n_-}$ by the number of ways that identical members of the series can be interchanged without altering the sequence. Any of the n_+ heads (or the n_- tails) may be interchanged among themselves with this result. There are $n_+!$ permutations of the heads and $n_-!$ permutations of the tails. Therefore these are the factors by which the previous result overcounted. Dividing by these factors leaves us with the binomial equation. The student should be so familiar with this argument as to be able to justify each factor in Equation (46).

Since the probability of tossing either a head or a tail is equal to 1/2 for a fair coin, Equation (46) may be rewritten

$$P(n,x) = \frac{n!}{\left|\frac{1}{2}\left(n + x/\ell\right)\right|! \left|\frac{1}{2}\left(n - x/\ell\right)\right|!} \left(\frac{1}{2}\right)^n \tag{47}$$

by incorporating Equations (44) and (45). The length ℓ by which the marker is moved in our hypothetical game is equivalent to the displacement of a particle in a single fluctuation. Because of the nature of the fluctuations underlying the whole process, these individual steps are very small. Observable diffusion is always the result of a very large number of such steps. For the case in which n is a large number, the factorials of Equation (47) may be expanded according to what is known as *Sterling's approximation*:

$$\ln y! = y \ln y - y \tag{48}$$

which is valid for large values of y. Taking the logarithm of Equation (47) and applying Approximation (48) leads to the result

$$-\ln P = \frac{n\ell + x}{2\ell} \ln\left(1 + \frac{x}{n\ell}\right) + \frac{n\ell - x}{2\ell} \ln\left(1 - \frac{x}{n\ell}\right) \tag{49}$$

The net displacement x is always small compared to the total distance traveled $n\ell$ since a good deal of the back-and-forth motion cancels out. Therefore the logarithmic terms in Equation (49) may be expanded as a power series (see Appendix A) in $(x/n\ell)$, with all terms higher than second order in $x/n\ell$ regarded as negligible. This leads to the result

$$\ln P \approx -\frac{x^2}{2\, n\, \ell^2} + \ldots \tag{50}$$

or

$$P(n, x) = C \exp\left(-\frac{x^2}{2\, n\, \ell^2}\right) \tag{51}$$

where the factor C represents a normalization constant.

A well-behaved probability function adds up to unity when the probabilities for all possi-

TABLE 2.2 Integrals Related to Gamma Functions Encountered in This Chapter

$$\int_0^\infty x^n e^{-ax^2} dx =$$

$\frac{1}{2}\sqrt{\pi/a}$ if $n = 0$	$\frac{1}{2a}$ if $n = 1$
$\frac{1}{4a}\sqrt{\pi/a}$ if $n = 2$	$\frac{1}{2a^2}$ if $n = 3$
$\frac{3}{8a^2}\sqrt{\pi/a}$ if $n = 4$	$\frac{1}{a^3}$ if $n = 5$

ble outcomes are totaled. Therefore, in order to evaluate the constant C in Equation (51), we should integrate Equation (51) over all possible values of x—which is equivalent to summing the probabilities—and set the result equal to unity. This leads to the expression

$$C = \left| \int_{-\infty}^{\infty} \exp\left(-\frac{x^2}{2\,n\,\ell^2}\right) dx \right|^{-1} \tag{52}$$

The integral is a gamma function, the value of which can be determined from the list of related integrals in Table 2.2. Evaluating Equation (52) gives

$$C = (2\pi n\ell^2)^{-1/2} \tag{53}$$

Substitution of this result into Equation (51) gives a continuous analytical expression for $P(n,x)$, which equals the binomial result for large values of n:

$$P(n, x) = (2\,\pi\,n\,\ell^2)^{-1/2} \exp\left(-\frac{x^2}{2\,n\,\ell^2}\right) \tag{54}$$

Strictly speaking, we should multiply both sides of Equation (54) by dx. The equation now expresses the probability of a displacement between x and $(x + dx)$ after n random steps of length ℓ.

2.6b Diffusion Coefficient from Random Walk Statistics

The application of Equation (54) to at least one set of boundary conditions for a diffusion problem is easy. We let a modification of Figure 2.8a describe the system. This time, instead of having a solution on one side of the barrier and pure solvent on the other side, suppose we imagine that both sides of the barrier are filled with solvent. Furthermore, suppose that the solute under investigation is introduced into the system in the pores of the plug, assuming the latter to be infinitesimally thin. Thus at the beginning of the experiment all the material is present at $x = 0$, in the notation of the derivation above, at a concentration c_0. With the passage of time, the material will gradually diffuse outward; the number of diffusion steps taken will be directly proportional to the elapsed time:

$$n = Kt \tag{55}$$

Equation (54) may be transformed into an expression that gives the probability as a function of x and t by substituting Equation (55) into Equation (54):

$$P(x, t)\,dx = (2\,\pi\,K\,t\,\ell^2)^{-1/2} \exp\left(-\frac{x^2}{2\,K\,t\,\ell^2}\right) dx \tag{56}$$

This suggests that the concentration as a function of x and t in the diffusion cell just described is given by multiplying c_0 by $P(x,t)$:

$$c(x, t)\, dx = c_0\, (2\, \pi\, K\, t\, \ell^2)^{-1/2} \exp\left(-\frac{x^2}{2\, K\, t\, \ell^2}\right) dx \tag{57}$$

Since $P(x,t)$ is normalized, its integral over all values of x equals unity. Likewise, integrating Equation (57) over all values of x gives c_0: The same quantity of solute is present at all times whether it is concentrated at the origin or is spread out from $-x$ to $+x$. The function given by Equation (57) is a solution to Fick's second law, shown in Equation (26), for a one-dimensional problem in which all the material is present initially at $x = 0$ and at a concentration c_0.

The only ambiguity remaining at this point is the value of K in Equations (55)–(57). This constant is easily evaluated as follows. If Equation (57) is indeed a solution to Equation (26), then the right- and left-hand sides of Equation (26) must be equal when the indicated differentiations are carried out. Differentiating Equation (57), substituting into Equation (26), and simplifying lead to the result

$$K = \frac{2D}{\ell^2} \tag{58}$$

Substituting this expression into Equation (57) yields

$$\frac{c(x, t)}{c_0}\, dx = (4\, \pi\, D\, t)^{-1/2} \exp\left(-\frac{x^2}{4\, D\, t}\right) dx \tag{59}$$

This relationship describes the diffusion of a solute in the x direction when it is concentrated initially in an infinitesimally thin layer at the origin.

Equation (59) may be developed somewhat further as follows. We define a parameter z to be

$$z = \frac{x}{(2\, D\, t)^{1/2}} \tag{60}$$

and then rewrite Equation (59) in terms of this quantity:

$$\frac{c(x, t)}{c_0}\, dx = (2\, \pi)^{-1/2} \exp\left(-\frac{z^2}{2}\right) dz = P(z)\, dz \tag{61}$$

Thus the concentration ratio c/c_0 is seen to be described at all times as a function of the single parameter z. The function $P(z)$ defined by Equation (61) is the normal or Gaussian distribution function, Equation (C.10).* Example 2.5 considers how the concentration profile of the diffusing species changes with time according to the normal distribution function.

* * *

EXAMPLE 2.5 *Unsteady State Variation of Concentration Profiles Due to Diffusion: Gaussian Distribution.* By consulting tables of the normal distribution function, draw curves that show the broadening of a band of material with time if the substance is initially at concentration c_0 and in a plug of infinitesimal thickness at $x = 0$. Assume that the diffusion coefficient has the value $5 \cdot 10^{-11}$ m^2 s^{-1} for this material. Use $t = 10^6$ and $t = 3 \cdot 10^6$ s to see how the concentration profile changes with time.

Solution: Tables of the normal distribution function list values of $P(z)$ for different values of z. As examples, we cite the following entries from the tables (Beyer 1987):

z	0.3	0.7	1.0	1.3	1.7	2.0	2.3	2.7
$P(z)$	0.3814	0.3123	0.2420	0.1714	0.0941	0.0540	0.0283	0.0104

*This manner of referencing is used to describe equations occurring in other chapters or in the appendices. The number (or the letter) to the left of the point is the chapter number (or the appendix).

Actual displacements x are obtained by multiplying z by $(2Dt)^{1/2}$; this means that the ordinate must be divided by this factor so that the area under the curve remains normalized:

For $t = 10^6$ s, $(2Dt)^{1/2} = [2(5 \cdot 10^{-11})10^6]^{1/2} = 10^{-2}$ m

Therefore $x = 10^{-2} z$.

$x(m)$	0.003	0.007	0.010	0.013	0.017	0.020	0.023	0.027
$P(x) \cdot 10^{-2}$	0.381	0.312	0.242	0.171	0.094	0.054	0.028	0.010

For $t = 3 \cdot 10^6$ s, $(2Dt)^{1/2} = 1.73 \cdot 10^{-2}$ m. Again, z is multiplied by this quantity to give x, and $P(z)$ is divided by the same factor to give the relative value of the ordinate, $P(x)$:

$x(m)$	0.0052	0.0121	0.0173	0.0225	0.0294	0.0346	. . .
$P(x) \cdot 10^{-2}$	0.221	0.181	0.140	0.099	0.054	0.031	

These values are plotted in Figure 2.10, in which the factor 10^{-2} for $P(x)$ has been used as a scale factor for the ordinate to show the relative shapes of the curves. ■

* * *

Figure 2.10 indicates a gradual approach toward equilibrium by the fact that the concentration tends to be more nearly uniform as time increases. In Appendix C the function z is defined to be δ/σ, where δ is the deviation of a particular value from the mean of a distribution and σ is the standard deviation of the distribution. Since the net displacement of a diffusing particle is analogous to δ, we may infer that the quantity $\sqrt{2Dt}$ is also analogous to the standard deviation. The point of this is the following.

It is well known that the width of the normal error curve at its inflection point (where the

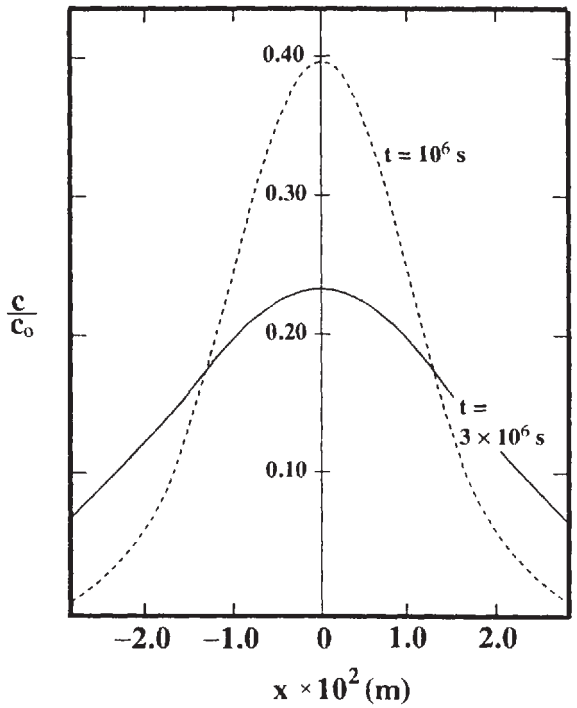


FIG. 2.10 Variation of concentration with time. The figure shows the variations of c/c_0 with distance from the origin at $t = 10^6$ and $t = 3 \cdot 10^6$ s if $D = 5 \times 10^{-11}$ m^2 s^{-1} for the data developed in Example 2.5.

second derivative changes sign or the curve changes from concave to convex) is equal to σ. Therefore we can conclude that the width of the curves shown in Figure 2.10, measured at their inflection points, increases in proportion to $\sqrt{t}$. This is one method whereby D could be measured in an experiment that corresponds to the boundary conditions specified above.

2.6c Random Walk Statistics and Experimental Measurement of the Diffusion Coefficient

With the mathematics of the one-dimensional random walk as background, we may visualize the following experimental arrangement by which D could be measured. Suppose that a narrow band of dispersion is layered between two portions of solvent in a long tube. We shall not worry about the practical difficulties in doing this since, in practice, other initial conditions that are easily obtained are actually used. The narrow band we have pictured, however, approximates the initial state of the system described by Equation (59). We might further imagine observing this band by means of a schlieren optical system (Section 2.4a). The two edges of the band, where the refractive index gradient is large, would define the positively and negatively sloped branches of the schlieren peak. With the passage of time, the material in the band diffuses outward; the schlieren peak would also broaden. In other words, the schlieren pattern observed at successive times would generate a family of curves—for example, Figure 2.7b—that very much resemble the theoretical curves of Figure 2.10.

The curves in Figure 2.10 are drawn for an arbitrary value of the diffusion coefficient. The experimental profiles produced by the schlieren optics are characterized by the diffusion coefficient of the experimental system. The remaining question is how to extract the appropriate D value from the experimental observations.

Remember that the normal distribution function has an inflection point (where the second derivative changes sign) at $z = 1.0$. Therefore the x value at which the inflection point occurs at any time equals $\sqrt{2Dt}$ according to Equation (60). By locating the inflection point at different times during a diffusion experiment, the appropriate D value may be evaluated for the diffusing species. For example, on the $t = 10^6$ s curve in Figure 2.10, the inflection point lies at $x = 0.010$ m. Substituting $x = 0.010$ m and $t = 10^6$ s when $z = 1$ into Equation (60) enables us to calculate the value of the diffusion coefficient that was used ($D = 5 \cdot 10^{-11}$ m^2 s^{-1}) to draw the curves in Figure 2.10. A similar analysis may be conducted on experimental curves obtained by the schlieren method.

As an experimental procedure, this method is less precise than others developed for the evaluation of D. It does point out, however, the intimate connection between diffusion and the random events at the molecular level that cause it.

A more practical experimental method for the determination of D is based on Figure 2.8. The theoretical arrangement represented by Figure 2.8 is implemented experimentally in an apparatus like that sketched in Figure 2.11. One side of the sintered glass barrier contains solution; the other side contains pure solvent. The entire apparatus is thermostated and both compartments are magnetically stirred. Samples are withdrawn at various times, and the quantity of material that has diffused into the solvent compartment is measured. The primary difficulty with this procedure is the tendency of the pores to plug owing to the entrapment of air bubbles, clogging by solid particles, or adsorption of the diffusing molecules themselves. Nevertheless, what is obtained experimentally is a record of the approach toward a uniform distribution of material on both sides of the barrier.

The situation represented by the apparatus in Figure 2.11 has also been analyzed theoretically. The function $c(x,t)$, which satisfies Fick's second law when there is a solution of concentration c_0 on one side of a boundary at $t = 0$ and pure solvent on the other side is given by

$$c = \frac{c_0}{2}\left[1 - \int_0^z P(z)\, dz\right] = \frac{c_0}{2}\left[1 - 2\, Erf\,(z)\right] \tag{62}$$

We can verify the plausibility of this expression as follows. Recall that z and t are inversely related and that the integral gives the area under one-half of the error curve between its

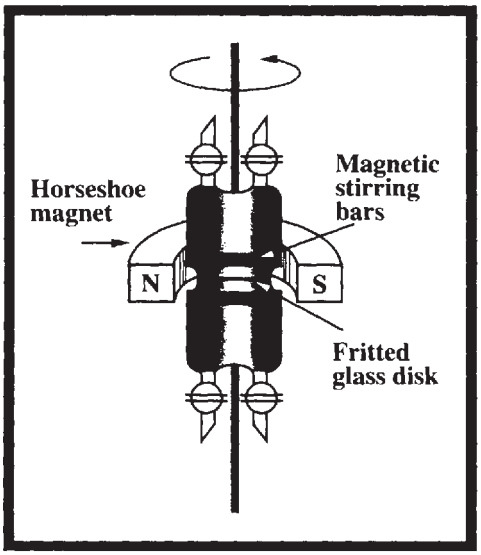


FIG. 2.11 Laboratory apparatus equivalent to that in Figure 2.8. The entire apparatus is rotated between the poles of a magnet for stirring.

midpoint ($z = 0$) and some specified value of z. Therefore, at $t = 0$, z is infinite, the integral equals 1/2, and $c = 0$. At $t = \infty$, z is zero, the integral equals zero, and $c = 1/2\ c_0$. Thus Equation (62) makes sense at either extreme. The detailed profile of c/c_0 versus z is obtained by reading values of the integral in Equation (62) from tables that give the area under the normal error curve. The results of this procedure are plotted in Figure 2.12, in which c/c_0 is shown versus x at several different times. The approach toward a uniform distribution of material is evident.

As in the case of diffusion from an initially thin layer, experimental concentration data and theoretical concentration profiles may be compared. From this comparison, the value of

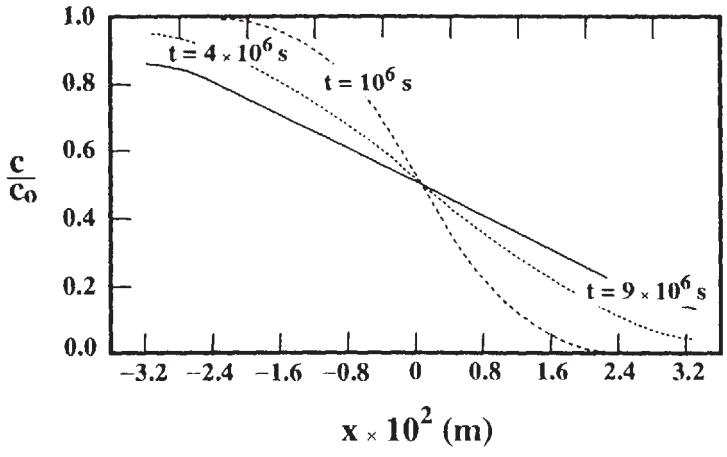


FIG. 2.12 Diffusion of solute from solution into pure solvent. Here c/c_0 is plotted versus distance at $t = 10^6$, $4 \cdot 10^6$, and $9 \cdot 10^6$ s for $D = 5 \cdot 10^{-11}$ m^2 s^{-1}. The plot has been drawn according to Equation (62).

D that is consistent with the observed diffusion behavior may be evaluated. The diffusion coefficient is a function of concentration; therefore, it is measured at a series of different concentrations and extrapolated to zero concentration; the limiting value is given the symbol D^0. Diffusion coefficients are temperature dependent, and it is sometimes necessary to adjust a value measured at one temperature to some other temperature. Example 2.6 considers how this is done.

* * *

EXAMPLE 2.6 *Temperature Dependence of Diffusion Coefficients.* Suppose the diffusion coefficient of a material is measured in an experiment (subscript *ex*) at some temperature T_{ex} at which the viscosity of the solvent is η_{ex}. Show how to correct the value of D to some standard (subscript *s*) conditions at which the viscosity is η_s. Take 20°C as the standard condition and evaluate D^0_{20} for a solute that displays a D^0 value of $4.76 \cdot 10^{-11}$ m^2 s^{-1} in water at 40°C. The viscosity of water at 20 and 40°C is $1.0050 \cdot 10^{-2}$ and $0.6560 \cdot 10^{-2}$ P, respectively.

Solution : By Equation (32), D is proportional to T/f and, by Equation (8), f is proportional to η. Therefore $D \propto T/\eta$, or

$$\frac{D^0_{20}}{D^0_{40}} = \frac{T_s/\eta_s}{T_{ex}/\eta_{ex}} = \frac{T_s}{T_{ex}} \frac{\eta_{ex}}{\eta_s}$$

Applying this relationship to the data provided gives

$$D^0_{20} = (293/313)\,(0.6560/1.0050)\,(4.76 \cdot 10^{-11}) = 2.91 \cdot 10^{-11}\ \text{m}^2\,\text{s}^{-1}.$$

The temperature dependence of the viscosity is seen to have the largest effect on the temperature dependence of D. ■

* * *

2.6d "Average" Displacements from Random Walk Statistics

Having examined the connection between the phenomenological equations of diffusion and the statistics of the random walk, let us now return to the random walk as a model for *Brownian motion*. The problem we wish to consider is the "average" displacement of a marker after an n-step, one-dimensional random walk. The foregoing discussion already supplies the answer to this problem. We have seen that the probability function for the one-dimensional walk is symmetrical about the origin, implying a mean displacement of zero. This simply reflects the fact that on the average the number of heads and tails will be equal. While we accept this result, we feel that there is something unsatisfactory about it. The average displacement is zero because positive and negative displacements are equally probable and effectively cancel one another. It is not because the marker scarcely moves from its initial location, a conclusion suggested by the value of the mean displacement.

This shows that the mean displacement is simply not a useful parameter to characterize the trajectory of the particle and suggests we should seek an alternate quantity. Instead of averaging the displacements directly, suppose we first square them to eliminate the differences in sign, then average them and take the square root. This quantity, called the *root mean square* (rms) *displacement*, will give a better measure of the meanderings of the marker since the sign differences have been eliminated.

The calculation of the rms displacement is quite simple. Any average is calculated by multiplying the quantity to be averaged by an appropriate probability function and then integrating the result over all possible values of the variable. This is a direct extension of Equation (C.3) using a continuous function for the weighting factor. Applying this procedure to the problem at hand gives the result

$$\overline{x^2} = \int_{-\infty}^{\infty} x^2\, P(n, x)\, dx \tag{63}$$

in which $P(n,x)$ is given by Equation (54). Making this substitution gives

$$\overline{x^2} = (4\pi Dt)^{-1/2} \int_{-\infty}^{\infty} x^2 \exp\left(-\frac{x^2}{4Dt}\right) dx \tag{64}$$

the value of which is found in Table 2.2. Evaluating the integral leads to the conclusion that

$$\overline{x^2} = 2Dt \tag{65}$$

Note that we have taken the initial position to be $x = 0$. Therefore, the left-hand side of Equation (65) is the average of the square of the *displacement* from the initial position. This important equation, also derived by Einstein, provides us with a means for measuring the diffusion coefficient for particles that are visible in a microscope. This is a particle size range for which measurement of D by following concentration changes with time is very difficult. Instead, the actual displacement of a particle in time t is measured microscopically. The rms displacement, evaluated from a statistically meaningful number of observations, permits D to be calculated. Jean Perrin (Nobel Prize, 1926) interpreted observations of a Brownian motion in terms of Equations (8), (32), and (65) as a means of determining the first precise value of Avogadro's number.*

Equation (65) also permits us to assign a physical interpretation to the diffusion coefficient in addition to the macroscopic meaning it has from Fick's laws. Rearranging and factoring in a way that admittedly ignores the averaging procedure, we write Equation (65) as

$$D = \frac{1}{2}\left(\frac{x}{t}\right) x \tag{66}$$

Since D is a constant for a particular substance, the ratio x/t must vary inversely with x. The distance traveled by a diffusing particle divided by the time required for the displacement x/t gives the apparent diffusion velocity of the particle. Equation (66) shows that the diffusion velocity is inversely proportional to the length of the path over which it is measured. This apparently paradoxical conclusion becomes less mysterious when we concentrate on the distinction between the diffusion velocity and the actual velocity of the particle as it travels along its zigzag path. The actual velocity reflects the average translational kinetic energy of the particle; its average value depends on the absolute temperature but is independent of the distance traveled. The diffusion velocity, on the other hand, decreases as the distance traveled increases. This simply means that large displacements are so much less probable than small displacements that they require disproportionately longer times to occur. Note that if $x = 1$ m, D has the significance of being equal to half the diffusion velocity measured over that distance. This is a result that was anticipated at the end of Section 2.5a.

2.7 THE RANDOM COIL AND RANDOM WALK STATISTICS

The dimensions of a randomly coiled polymer molecule are a topic that appears to bear no relationship to diffusion; however, both the coil dimensions and diffusion can be analyzed in terms of random walk statistics. Therefore we may take advantage of the statistical argument we have developed to consider this problem.

2.7a End-to-End Distance in a Polymer Coil

Suppose we visualize a polymer molecule as consisting of n segments of length ℓ connected by a completely flexible linkage at each joint. Imagine, furthermore, that the placement of each successive segment is determined by some sort of purely random criterion. We could anchor

*Jean Perrin's 1913 monograph, *Les Atomes*, on Brownian motion and some of the related topics on the molecular nature of matter has been reprinted recently (Perrin 1990) and is an interesting source of information on the evolution of ideas on diffusion and determination of Avogadro's number. This monograph also contains some of Perrin's sketches of random walks executed by colloidal spheres in his experiments.

one end of the chain to the origin of a hypothetical coordinate system, for example, and position successive segments as follows. Suppose we toss a single die and agree to orient the next segment in the $+x$ direction if the die shows a 1 and in the $-x$ direction if a 6 shows. Likewise, 2, 5, 3, and 4 correspond to $+y$, $-y$, $+z$, and $-z$, respectively. Our problem is this: On the *average*, what will be the end-to-end distance L for a chain of n segments? To calculate this, we must assume that n is very large, and we must ignore the sites excluded by previously positioned segments.

We recognize first that the end-to-end distance can be resolved into x, y, and z components that obey the relationship

$$L^2 = x^2 + y^2 + z^2 \tag{67}$$

The probability that the loose end of the chain is in a volume element located at x, y, and z is given by

$$P(x,y,z)\,dx\,dy\,dz = P(x)P(y)P(z)\,dx\,dy\,dz \tag{68}$$

where $P(x)$ is the probability that the x coordinate has a value between x and $x + dx$, with similar definitions for $P(y)$ and $P(z)$. Equation (54) gives the expression for $P(x)$, except we must remember that now only $n/3$ segments will be aligned with the x axis since each of the directions is independent and equally probable. Since the x, y, and z directions are equivalent, the same expression holds for $P(y)$ and $P(z)$ with the appropriate change of variables. Therefore, on incorporating Equation (54), Equation (68) becomes

$$P(x,\,y,\,z)\,dx\,dy\,dz = \left(2\pi\,\frac{n}{3}\ell^2\right)^{3/2} \exp\left(-\frac{3\,L^2}{2\,n\,\ell^2}\right) dx\,dy\,dz \tag{69}$$

This expression gives the probability that the loose end of the molecule is in a volume element located at some particular values of x, y, and z, as shown in Figure 2.13a. Our interest is not in any specific x, y, z coordinates, but in all combinations of x, y, z coordinates that result in the end of the chain being a distance L from the origin. This can be evaluated by changing the volume element in Equation (69) to spherical coordinates and then integrating over all angles.

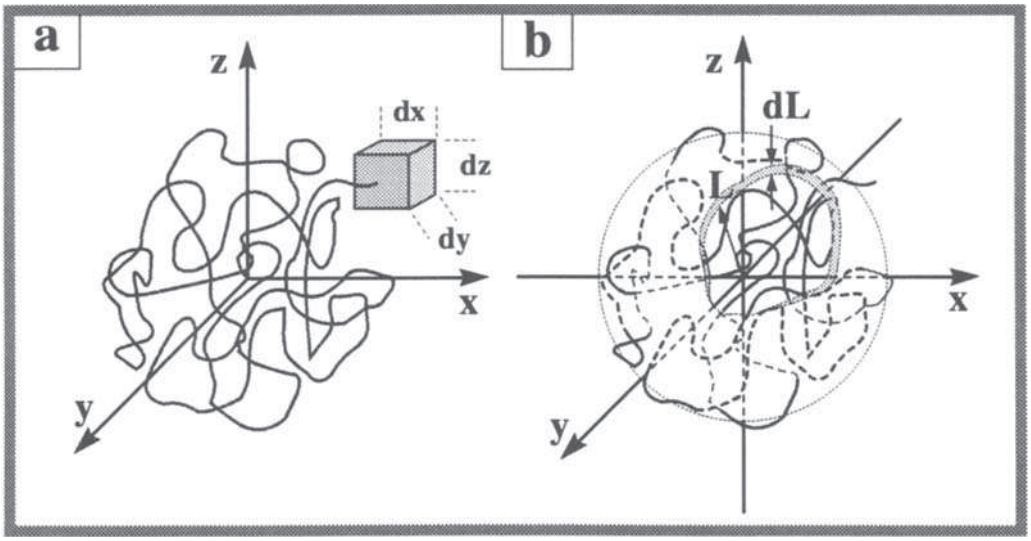


FIG. 2.13 Coordinate systems for a flexible random coil. One end of a flexible random coil lies at the origin and the other end is (a) in a volume element $dx\,dy\,dz$ and (b) in a spherical volume element $4\pi L^2 dL$.

This amounts to replacing the volume element in Equation (69) with the volume of a spherical shell of radius L and thickness dL:

$$dx\,dy\,dz \rightarrow 4\pi L^2 dL \tag{70}$$

A geometrical representation of this situation is shown in Figure 2.13b. Incorporating this expression into Equation (69) gives the probability that the loose end of the molecule lies between L and $(L + dL)$, irrespective of the direction:

$$P(L)\,dL = 4\pi\left(2\pi\frac{n}{3}\ell^2\right)^{-3/2} L^2 \exp\left(-\frac{3L^2}{2n\ell^2}\right) dL \tag{71}$$

With this distribution function, it is an easy matter to calculate average quantities. Again, since positive and negative displacements are equally probable, we evaluate the average of L^2. Following the same procedure as used in Equation (63), we write

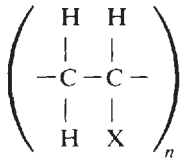
$$\overline{L^2} = \left(\frac{2\pi n\ell^2}{3}\right)^{-3/2} 4\pi \int_0^\infty L^4 \exp\left(-\frac{3L^2}{2n\ell^2}\right) dL \tag{72}$$

Evaluation of this integral, using Table 2.2, leads to the result

$$\overline{L^2} = n\ell^2 \tag{73}$$

which is the desired quantity.

Equation (73) shows that the rms end-to-end distance in a polymer coil equals the square root of the number of steps times the length of each step. We might ask, therefore, what is the physical significance of n and ℓ for a polymer molecule? A general formula for a typical vinyl polymer is



where n is called the degree of polymerization. Since this quantity measures the number of repeated segments in the chain, it seems reasonable to equate this with the number of steps in the three-dimensional random walk. If M represents the molecular weight of the polymer and M_0 is the molecular weight of the vinyl monomer, then M/M_o equals n. Since M_o and ℓ are constants, we see that the radius of the coil is predicted to be proportional to $M^{1/2}$.

The physical significance of ℓ in Equation (73) is somewhat harder to define. At first glance it appears to be the length of the repeating unit, about 0.25 nm for a vinyl polymer. We must remember, however, that the derivation of Equation (73) assumed that the coil was connected by completely flexible joints. Molecular segments are attached at definite bond angles, however, so an actual molecule has less flexibility than the model assumes. Any restriction on the flexibility of a joint will lead to an increase in the dimensions of the coil. The effect of fixed bond angles on the dimensions of the chain may be incorporated into the model as follows.

A 360° rotation around any carbon-carbon bond in a vinyl chain will cause the next bond in the chain to trace a cone with one carbon atom at the apex and the other carbon atom along the rim of the cone. Ignoring hindered rotation for the moment, we see that each position on the rim of such a cone is an equally probable site for the apex of the cone generated by the next bond in the chain. This situation is illustrated schematically in Figure 2.14. This effect has been shown to increase the actual length of the repeating unit ℓ_0 by the factor

$$\ell^2 = \ell_0^2\left(\frac{1 - \cos\theta}{1 + \cos\theta}\right) \tag{74}$$

θ is the bond angle. For a tetrahedral bond angle, $\cos\theta = -(1/3)$, so the additional factor in Equation (74) equals 2.

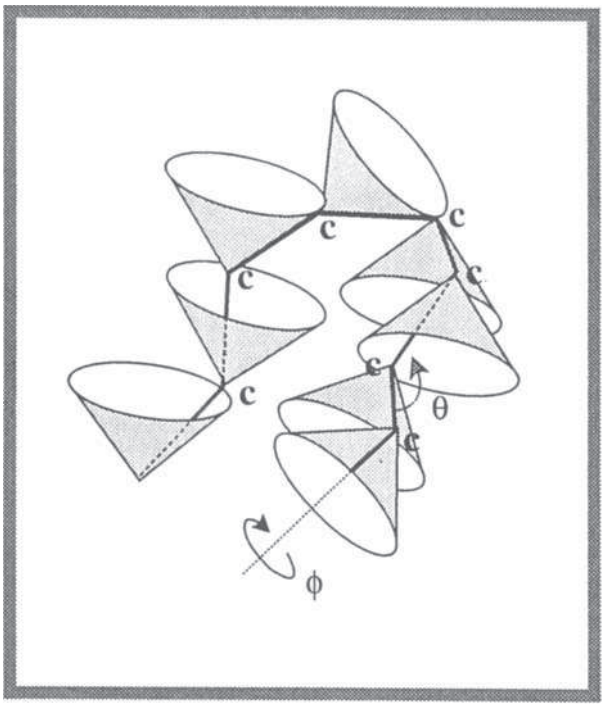


FIG. 2.14 The effect of fixed bond angle in restricting the flexibility of a polymer coil. For tetrahedral bonds, $\theta = 109°$.

In Figure 2.14, it is implied that the terminal carbon atom can occupy any position on the rim of the cone; that is, there is assumed to be perfectly free rotation around the penultimate carbon-carbon bond. This is equivalent to saying that all values of ϕ, the angle that describes the rotation (see Fig. 2.14), are equally probable. Any hindrance to free rotation will block certain configurations, expanding the coil dimensions still further.

Finally, interactions between the polymer segments and the solvent molecules may introduce a bias that tends to position segments as close or as far as possible from other segments, depending on the nature of the interaction. As a matter of fact, those properties of polymer solutions that are sensitive to the dimensions of the molecules, such as viscosity, vary widely with solvent "goodness." A good solvent may be defined as one in which polymer-solvent contacts are favored; a poor solvent is one in which polymer-polymer contacts are favored. Therefore the value of ℓ in Equation (74) will be increased in a good solvent and decreased in a poor solvent. In discussing size-exclusion chromatography in Section 1.6b, we anticipated this conclusion by noting that polymer-solvent interactions complicate the relationship between M and L for a random coil.

2.7b Radius of Gyration of a Polymer Coil

The statistical relationships of this section can also be applied to Equation (1.13), which defines the *radius of gyration*. If the masses in that equation are identified with the mass m_0 of the repeat units of the polymer—of which there are n in the molecule—then Equation (1.13) can be written as

$$\overline{R_g^2} = \frac{m_0 \Sigma_i r_i^2}{n\, m_0} = \frac{1}{n} \Sigma_i r_i^2 \tag{75}$$

Next, the summation over discrete values of r_i^2 is replaced by the integral of $r^2P(r)dr$. This sounds exactly like the procedure that results in Equation (72), but there are some differences. To evaluate the radius of gyration, we consider the masses relative to the center of mass of the swarm. We proceed in two steps. First, an arbitrary segment, say, segment j, is considered to be the center of mass, and r in Equation (75) is measured outward in both directions from j. Using k to index the segments outward from j (on either side) one obtains the following expression for the radius of gyration relative to segment j:

$$(R_g^2)_j = \frac{1}{n}\left(\sum_{k=1}^{j} \int P(r_k)\, r_k^2 dr_k + \sum_{k=1}^{n-j} \int P(r_k) r_k^2 dr_k\right) \tag{76}$$

where r_k denotes the distance of the kth segment (mass point) from segment j and $P(r_k)dr_k$ is the probability that r_k lies in the interval $[r_k, r_k + dr_k]$. The integral, in essence, accounts for the various possible configurations of the chain. Finally, this quantity must be totaled for all values of j between 1 and n since any one of the segments might lie at the center of mass of the coil. This summation must also be divided by n since we are effectively determining the average value for j. The result must also be divided by 2 since, otherwise, we will be counting each distance twice. (For example, if the focus is on two segments p and q, the distance between p and q is counted when we take q to be the center and later when p is taken as the center.) Mathematically, the above procedure is fairly complicated, but when all of these steps are carried out—using Equation (69) as the expression for $P(r)$—the radius of gyration for the random coil may be shown to equal

$$\overline{R_g^2} = \frac{1}{6} n \ell^2 \tag{77}$$

which is simply one-sixth the value of the mean of L^2 as given by Equation (73). Since the radius of gyration can be measured by viscosity and light scattering for random coils, it is apparent that Equation (77) may be used to measure ℓ. In this way, the effects of hindered rotation around bonds and of imbibed solvent may be evaluated quantitatively.

2.7c Relation to the Kinetic Theory of Gases

Finally, we note that all the statistical equations of this chapter could have been borrowed directly from the kinetic theory of gases by simply changing the variables. We illustrate this now by going in the opposite direction. For example, if we replace the quantity $3/n\ell^2$ by m/k_BT and replace L by v in Equation (69), we obtain the *Boltzmann distribution* of molecular velocities in three dimensions. If we make the same substitutions in Equation (73), we obtain an important result from kinetic molecular theory:

$$\overline{v^2} = \frac{3\, k_B T}{m} \tag{78}$$

This shows that Equation (73) occupies a position for the random coil that is analogous to the position of average kinetic energy in the kinetic molecular theory. This is not just a fortuitous similarity, but a reflection of the statistical basis of both. As a little self-test: Did you recognize the similarity between the formalisms of the last few sections and kinetic theory as we went along?

2.8 EQUILIBRIUM BETWEEN SEDIMENTATION AND DIFFUSION

We have already noted that sedimentation and diffusion are opposing processes, the first tending to collect and the second to scatter. Let us now consider the circumstances under which these two tendencies equal each other. Once this condition is reached, of course, there will be no further macroscopic changes; the system is at equilibrium. In order to formulate this problem, consider the unit cross section shown in Figure 2.15, in which the x direction is in the direction of either a gravitational or a centrifugal field. Suppose this field tends to pull

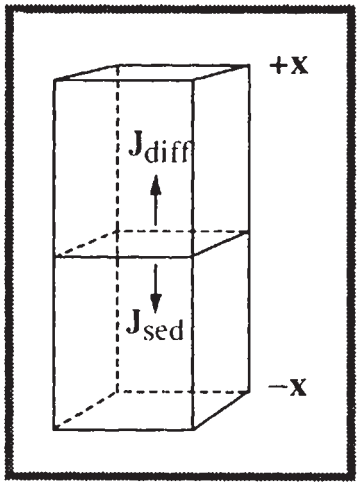


FIG. 2.15 The relationship between the flux due to sedimentation and that due to diffusion. At equilibrium, the two are equal.

the particles in the $-x$ direction. Gradually the concentration of the particles will increase in the region below the cross section of interest.

Back-diffusion occurs at a rate that increases with the buildup of a concentration gradient. When equilibrium is finally reached, we may write

$$J_{sed} = -J_{diff} \tag{79}$$

where J_{sed} is the flux across the area due to sedimentation and J_{diff} is the flux due to diffusion. The J_{diff} is given by Equation (22), and J_{sed} by

$$J_{sed} = vc \tag{80}$$

in which v is the sedimentation velocity and c is the concentration at the plane. Substituting Equations (22) and (80) into Equation (79) gives

$$v\,c = D\frac{dc}{dx} \tag{81}$$

If we substitute the value for the rate of sedimentation under gravity, Equation (4), we obtain

$$\frac{m}{f}\left(1 - \frac{\rho_1}{\rho_2}\right)g\,c = D\frac{dc}{dx} \tag{82}$$

If the sedimentation occurs in a centrifugal field, on the other hand, g must be replaced by $\omega^2 x$ in Equation (4):

$$\frac{m}{f}\left(1 - \frac{\rho_1}{\rho_2}\right)\omega^2 x\,c = D\frac{dc}{dx} \tag{83}$$

Equations (82) and (83) are easily integrated to produce expressions that give c as a function of x at equilibrium. Defining c_1 and c_2 to be the equilibrium concentrations at x_1 and x_2, respectively, and then integrating, we obtain for Equation (82)

$$\frac{m}{fD}\left(1 - \frac{\rho_1}{\rho_2}\right)g\,(x_2 - x_1) = \ln\left(\frac{c_2}{c_1}\right) \tag{84}$$

and for Equation (83)

$$\frac{m}{2fD}\left(1 - \frac{\rho_1}{\rho_2}\right)\omega^2 (x_2^2 - x_1^2) = \ln\left(\frac{c_2}{c_1}\right) \tag{85}$$

Finally, we recall Equation (32), which permits us to substitute k_BT for Df, giving

$$\ln\left(\frac{c_2}{c_1}\right) = \frac{m}{k_BT}\left(1 - \frac{\rho_1}{\rho_2}\right) g\ (x_2 - x_1) \tag{86}$$

for sedimentation equilibrium under gravity and

$$\ln\left(\frac{c_2}{c_1}\right) = \frac{m}{2\ k_BT}\left(1 - \frac{\rho_1}{\rho_2}\right)\omega^2 (x_2^2 - x_1^2) \tag{87}$$

for sedimentation equilibrium in a centrifuge. Note that sedimentation equilibrium studies permit the evaluation of particle mass with no assumptions about particle shape. More elaborate mathematical formulations of sedimentation/diffusion equilibria may be found in advanced textbooks (see, for example, Probstein 1994).

It should now be apparent why the ultracentrifuge is such an important tool in molecular biology. Although the method is in no way restricted to particles of biological significance, these particles are of a size and density that are ideally suited to the ultracentrifuge. An ultracentrifuge permits the evaluation of particle mass through equilibrium studies—see Equation (87)—and the evaluation of the ratio m/f through studies of the rate of sedimentation—see Equation (18). Combining these data permits the separate evaluation of f. From the mass and density of the material, the volume and radius of an equivalent sphere, and hence f_0, can be calculated. Then, Figure 2.9 may be consulted to determine particle characterizations that are consistent with the ratio f/f_0. Although both utilize the same instrument, sedimentation rate and sedimentation equilibrium are two different experiments that complement one another very nicely.

It might be noted that sedimentation equilibrium is approached very slowly; however, techniques that permit equilibrium conditions to be estimated from preequilibrium measurements have been developed by W. J. Archibald. Equations (86) and (87) predict a linear semilogarithmic plot of c versus x or x^2 for gravitational and centrifugal studies, respectively. The slope of such a plot is proportional to the mass of the particles involved. Remember that monodispersity was assumed in the derivation of these equations. If this condition is not met for an experimental system, the plot just described will not be linear. If each particle size present is at equilibrium, however, each component will follow the equations and the experimental plot will be the summation of several straight lines. Under certain conditions these may be resolved to give information about the polydispersity of the system. In any event, nonlinearity implies polydispersity once true equilibrium is reached.

We conclude this chapter with a final observation about Equation (86). If the particles in question are gas molecules instead of suspended particles, then the concentration ratio equals the ratio of pressures measured at two locations, and the particle mass requires no correction for buoyancy. Under these conditions, Equation (86) becomes

$$\ln\left(\frac{p_2}{p_1}\right) = \frac{m\ g\ (x_2 - x_1)}{k_BT} = \frac{M\ g\ (x_2 - x_1)}{RT} \tag{88}$$

This familiar equation gives the variation of barometric pressure with elevation and is known as the *barometric equation*. Once again we are reminded of the connection between the material of this chapter and kinetic molecular theory.

REVIEW QUESTIONS

1. List a few examples from your experience of phenomena in which sedimentation and diffusion play major roles, individually or together.
2. Describe the reasons for hydrodynamic friction. What is the physical significance of the friction factor? What are its units?
3. What is *Stokes's law*? Under what conditions is it valid?
4. Describe what types of information can be obtained from the sedimentation equation? What are the restrictions on the applicability of the equation derived in the text?
5. How is the expression for the sedimentation coefficient derived? What are the units of the sedimentation coefficient?
6. State *Fick's laws of diffusion*. Describe at least two different techniques to use Fick's laws to measure the diffusion coefficient.
7. Sketch qualitatively the variation of concentration profiles with time that you would expect from the solution of Fick's second law. Assume suitable boundary and initial conditions.
8. How would you combine sedimentation and diffusion measurements to determine the mass of a particle without any information on the shape of the particle.
9. Describe how the random walk statistics are used to relate the random walk to the diffusion coefficient.
10. What assumptions are made in the development of the random walk statistics discussed in the text, and what do they imply for the correspondence between random walk and diffusion?
11. What is the difference between the mean of the displacement and rms displacement? Can they be equal under any circumstance?
12. What is the relation between the diffusion coefficient and the rms displacement?
13. Describe an experiment to determine Avogadro's number from the average root mean square displacement of a particle due to random walk.
14. What is the *Stokes-Einstein equation*? Suggest at least two uses for that equation.
15. Why can one use random walk statistics to derive expressions for the end-to-end distance of a polymer chain? Under what conditions can this be done?
16. What is the difference between the radius of gyration of a polymer chain and the end-to-end distance?
17. What is the physical significance of the segment length used in the derivation of polymer chain statistics?
18. Describe the relation between sedimentation/diffusion equilibrium and the barometric equation.

REFERENCES

General References (with Annotations)

Berg, H. C., *Random Walks in Biology*, expanded ed. Princeton University Press, Princeton, NJ, 1993. (Undergraduate level. A lucid introduction to random walk statistics. Discusses translational and rotational diffusion, self-propelled motion, random walk, and sedimentation. Relevant to biological species of colloidal dimensions, but no background in biology is needed.)

Bird, R. B., Stewart, W. E., and Lightfoot, E. N., *Transport Phenomena*, Wiley, New York, 1960. (Undergraduate and graduate levels. A classic textbook on the formulation and solutions of equations of transport phenomena in engineering. Contains both undergraduate-level and graduate-level topics and problems.)

Einstein, A. *Investigations in the Theory of the Brownian Movement.*, Dover, New York, 1956. (Undergraduate level [now!]. A collection of the original papers of Einstein on Brownian movement. Accessible to undergraduates. A window to an era in which atomic theory of matter was still under hot debate.)

Giddings, J. C. *Unified Separation Science*, Wiley, New York, 1991. (Undergraduate level. A good introduction to many examples of separation processes in which diffusion plays an important role.)

Perrin, J. *Atoms*, Ox Bow Press, Woodbridge, CT, 1990. (Undergraduate level today; a research monograph when it originally appeared! The translation of the 1913 French original, *Les Atomes*, by the French physicist recognized by the 1926 Nobel Prize in physics for experimental

observation of Brownian motion. The first five chapters deal with atomic theory, Brownian motion, fluctuations, and experimental observation of Brownian motion. Another window to the past for those interested in history of science.)

Probstein, R. F. *Physicochemical Hydrodynamics: An Introduction*, 2d ed., Wiley-Interscience, New York, 1994. (Primarily graduate level. Despite the confusing title, this is a textbook on topics in colloidal phenomena.)

Other References

Beyer, W. H. (Ed.). *CRC Handbook of Mathematical Sciences*, 6th ed., CRC Press, Boca Raton, FL, 1987.

Cantor, C. R., and Schimmel, P. R. *Biophysical Chemistry. Part II: Techniques for the Study of Biological Structure and Function*, W. H. Freeman, San Francisco, CA, 1980.

Giddings, J. C. *Separation Science*, **1**, 123 (1966).

Giddings, J. C. *Science*, **260**, 1456 (1993).

Oden, S. *Proc. R. Soc. Edinburgh*, **36**, 219 (1915).

Svedberg, T., and Pederson, K. O. *The Ultracentrifuge*, Oxford University Press, Oxford, England, 1940.

PROBLEMS

1. The following results describe the rate of accumulation of rutile (TiO_2; $\rho = 4.2$ g cm^{-3}) particles on a submerged balance plan.*

Time t (min)	W (g)	Time t (min)	W (g)
0.0	0.000	35	0.310
1.5	0.045	60	0.403
2.0	0.060	84	0.480
2.5	0.075	127	0.610
3.0	0.090	159	0.702
4.0	0.110	226	0.880
5.5	0.125	274	0.998
8.0	0.160	327	1.110
14.0	0.210	384	1.240
24.0	0.260	420	1.320

   (a) Plot W versus t and use this to evaluate w as a function of time. Prepare a plot of w versus t.
   (b) At what time (t_{max}) does the greatest increment in w appear to occur?
   (c) Calculate the radius of an equivalent sphere corresponding to t_{max} ($\rho_{soln} = 0.997$ g cm^{-3}, $\eta = 0.00894$ P, $h = 12$ cm).

2. A preparation of reduced and carboxymethylated Mouse-Elberfeld virus protein particles in water reached sedimentation equilibrium at 25°C after 40 hr at 12,590 rpm. The following data show the recorder displacement (proportional to concentration) versus r for this protein:†

c (arbitrary units)	2.29	2.51	2.79	3.09	3.51	3.89	4.47	5.01	5.89	6.61	7.41	8.51
r (cm)	6.55	6.58	6.60	6.65	6.67	6.69	6.71	6.74	6.76	6.79	6.81	6.84

*Sullivan, W. F., and Jacobson, A. E. *Symposium on Particle Size Measurements*, ASTM Publication No. 234, 1959.

†Rueckert, R. R. *Virology* **26**, 345 (1965).

‡Vinograd, J., Bruner, R., Kent, R., and Weigle, J. *Proc. Nat. Acad. Sci., USA* **49**, 902 (1963).

(a) Use these results to evaluate the mass of the particles present ($\rho_{protein}$ = 1.370 g cm^{-3}) and estimate R_0 and f_0 for the molecules. Does the sample appear to be monodisperse?
(b) The sedimentation coefficient is known to be 2.7 S for this preparation. Evaluate f and f/f_0.
(c) What can be said about the possible axial ratio-hydration combinations of this protein in terms of Figure 2.9?

3. Southern bean mosaic virus (SBMV) particles are centrifuged at 12,590 rpm and the absorbance at 260 nm is measured along the settling direction as a function of time. The center of the absorption band varies with distance from the center of the rotor as follows:‡

t (min)	R (cm)
16	6.22
32	6.32
48	6.42
64	6.52
80	6.62
96	6.72
112	6.82
128	6.92
144	7.02

4. Phosphatidylcholine micelles are spherical particles having a molecular weight of 97,000 g $mole^{-1}$. Assuming that the density of the dry lipid (ρ = 1.018 g cm^{-3}) applies to the micelles, calculate the radius R_s and the diffusion coefficient D for these particles in water at 20°C. The experimental value of the diffusion coefficient is 6.547 × 10^{-7} cm^2 s^{-1} under these conditions.[#] Evaluate f/f^* and estimate the hydration of the lipid.

5. Calculate the diameter of a spherical particle (ρ = 4 g cm^{-3}) for which the rms displacement due to diffusion at 25°C is 1% the distance of sedimentation in a 24-hr period through a medium for which ρ = 1 g cm^{-3} and η = 9 × 10^{-3} P. For what diameter is the diffusion distance 10% of the settling distance?

6. The diffusion of alkyl ammonium ions into clay pellets has been studied by bringing the pellet into superficial contact with an isotopically labeled salt and then, after a suitable time, using a microtome to slice the pellet. The radioactivity is then measured in successive thin slices of the pellet. Assuming that Equation (62) describes the diffusion process, estimate how long it would take for 1% of the initial activity of each of the ions to appear in the 15th slice inward from the exposed surface of the dry clay if each slice is 40-μm thick. The diffusion coefficients for the methyl and trimethyl ammonium cations under these conditions are 7.03 × 10^{-12} and 2.65 × 10^{-11} cm^2 s^{-1}, respectively.†

7. Suppose two reservoirs 4-cm apart are cut into an agar gel in a petri dish. Solutions of $Pb(NO_3)_2$ and Na_2CrO_4 are introduced simultaneously into the two reservoirs. At what distance into the gel does $PbCrO_4$ precipitate if the diffusion coefficients of Pb^{2+} and CrO_4^{2-} in agar are 0.657 × 10^{-5} and 0.752 × 10^{-5} cm^2 s^{-1}, respectively?‡ Where would Prussian blue precipitate if the reservoirs contained Fe^{3+} (D = 0.434 × 10^{-5} cm^2 s^{-1}) and $Fe(CN)_6^{4-}$ (D = 0.557 × 10^{-5} cm^2 s^{-1})?

[#]Hasser, H. In *Water, a Comprehensive Treatise*, Vol. 4 (F. Franks, Ed.), Plenum, New York, NY, 1975.

†Gast, R. G., and Morfland, M. M. *J. Colloid Interface Sci.* **37**, 80 (1971).

‡Lee, R. E., and Meeks, F. R. *J. Colloid Interface Sci.* **35**, 584 (1971).

8. Use a compass to inscribe a circle around the aggregate shown on the cover of the book. Such a "graticule" might be used to define an average diameter for the aggregate. Compare the radius of the inscribed circle to that of the primary particles in the aggregate. How does the ratio of these radii compare with the value that would be obtained for the case $n = 76$ if the aggregate were built up according to a random walk model? Compare the features of the two models with the objective of accounting for the relative magnitude of the aggregate radius relative to the primary particle radius in the two cases.

9. The molecular weights and sedimentation coefficients of human plasminogen and plasmin ($\rho = 1.40$ g cm^{-3}) are as follows:*

	Plasminogen	Plasmin
M (g mole^{-1})	81,000	75,400
s at 20°C (S)	4.2	3.9

   (a) Calculate the diffusion coefficient for each.
   (b) Prepare a graph such as that of Figure 2.10 showing quantitatively how an initially thin band of these proteins widens with time. Show at least three different times.

10. Colloids (casein micelles) of two different particle sizes are isolated from skim milk by centrifugation under different conditions. The sedimentation and diffusion coefficients of the two preparations are as follows:†

Preparation conditions	$D \times 10^8$ (cm^2 s^{-1})	$s \times 10^{13}$ (s)
5 min at 5,000 rpm	0.97	2200
40 min at 20,000 rpm	2.82	800

    Calculate the mass per particle and the gram molecular weight of the two micelle fractions, assuming $\rho = 1.43$ g cm^{-3} for the dispersed phase.

11. The following data give the number of gold particles (as $\log_{10}$) versus depth beneath the surface for an aqueous dispersion allowed to reach sedimentation equilibrium under the influence of gravity:‡

Depth (mm)	4.44	5.06	5.67	6.30	6.90	7.53	8.15	8.65
log n	10.36	10.51	10.63	10.75	10.89	11.05	11.22	11.39

    Calculate the radius of the gold particles ($\rho_{Au} = 19.3$ g cm^{-3}), treating the dispersed units as equivalent spheres.

12. At equilibrium at 20°C the concentration of tobacco mosaic virus (TMV) shows a linear semilogarithmic graph when plotted against the square of the distance to the axis of rotation in a centrifuge. Evaluate the molecular weight of the TMV particles if $\ln c = -1.7$ at $r^2 = 44.0$ cm^2 and $\ln c = -2.8$ at $r^2 = 41.2$ cm^2 when the dispersion is centrifuged at 6.185 rps. The density of the TMV is 1.36 g cm^{-3}.§

13. Verify that the expansion (see Appendix A) of Equation (49) leads to Equation (50). Retain no terms higher than the second order in expansions. Verify that the integration (see Table 2.2) of Equations (64) and (72) leads to Equations (65) and (73), respectively.

---

*Barlow, G. H., Summaria, L., and Robbins, K. C. *J. Biol. Chem.* **244**, 1138 (1969).
†Morr, C. V., Lin, S. H. C., Dewan, R. K., and Bloomfield, V. A. *J. Dairy Sci.* **56**, 415 (1973).
‡McDowell, C. M., and Usher, F. L. *Proc. R. Soc. London* **138A**, 133 (1932).
§Weber, F. N., Jr., Elton, R. M., Kim, H. G., Rose, R. D., Steere, R. L., and Kupke, D. W. *Science* **140**, 1090 (1963).

3

Solution Thermodynamics

Osmotic and Donnan Equilibria

All faults or defects, . . . Pantocyclus attributed to some deviation from perfect Regularity in the bodily figure, caused perhaps . . . by some collision in a crowd; by neglect to take exercise, or by taking too much of it; or even by a sudden change of temperature.

From Abbott's *Flatland*

3.1 INTRODUCTION

3.1a What Are Osmotic Pressure and Donnan Equilibria?

Osmotic pressure is a property of solutions of organic or inorganic solutes and appears in a number of contexts in the study of colloidal systems. As we see below, it is usually defined using a real or thought experiment involving a solution separated from the pure solvent by a membrane permeable only to the solvent. In such a setup, the pure solvent would flow through the membrane in an attempt to equalize the concentration difference. The pressure necessary to stop this flow is known as the *osmotic pressure* (from the Greek word *osmos*, meaning "to push or thrust"). In thermodynamic terms, the concentration difference created by the membrane represents a difference in the chemical potentials in the two compartments separated by the membrane, and the osmotic pressure counteracts the difference in the chemical potential.

It is important to note that the concept of osmotic pressure is more general than suggested by the above experiment. In particular, one does not have to invoke the presence of a membrane (or even a concentration difference) to define osmotic pressure. The osmotic pressure, being a property of a solution, always exists and serves to counteract the tendency of the chemical potentials to equalize. It is not important how the differences in the chemical potential come about. The differences may arise due to other factors such as an electric field or gravity. For example, we see in Chapter 11 (Section 11.7a) how osmotic pressure plays a major role in giving rise to repulsion between electrical double layers; here, the variation of the concentration in the electrical double layers arises from the electrostatic interaction between a charged surface and the ions in the solution. In Chapter 13 (Section 13.6b.3), we provide another example of the role of differences in osmotic pressures of a polymer solution in giving rise to an effective attractive force between colloidal particles suspended in the solution.

A related phenomenon occurs when the membrane in the above-mentioned experiment is permeable to the solvent and small ions but not to a macroion such as a polyelectrolyte or charged colloidal particles that may be present in a solution. The polyelectrolyte, prevented from moving to the other side, perturbs the concentration distributions of the small ions and gives rise to an ionic equilibrium (with attendant potential differences) that is different from what we would expect in the absence of the polyelectrolyte. The resulting equilibrium is known as the *Donnan equilibrium* (or, the *Gibbs-Donnan equilibrium*) and plays an important role in

a number of problems of interest in colloid and surface chemistry, electrochemistry, and biology, as illustrated through an example in Vignette III.

3.1b Importance of Osmotic Pressure and Donnan Equilibrium

The osmotic pressure can be measured accurately for colloidal solutes, and one molecular parameter of interest that is readily determined by osmometry is the number average molecular weight of the solute. Molecular weights determined by osmometry are absolute values; no calibration with known standards or any assumed theoretical models is required. Even the assumption of solution ideality is not a problem, since results are extrapolated to zero solute concentration before calculations are made.

Equilibrium thermodynamics of solutions of colloidal solutes determines the quantitative details of osmotic and Donnan equilibria and occupies our concern in this chapter. In particular, our discussion here centers on the osmotic pressure.

VIGNETTE III BIOPHYSICS AND PHYSIOLOGY: Donnan Equilibrium and the "Resting States" of Nerve Cells

A major function of biological cell membranes is to regulate the transport of ionic and nonionic nutrients and reject metabolic wastes while maintaining the intracellular ionic atmosphere necessary for life-sustaining functions. The transfer of materials often has to be accomplished *against* concentration gradients. Understanding how the cells regulate such transport is essential in health care.

A cell is enclosed by a lipid bilayer known as the plasma membrane. In Vignette I.2 in Chapter 1 we discussed an example of a membrane, a complex structure with a mosaic of embedded or adsorbed moieties such as proteins. It is these membranes that protect the intracellular contents from the exterior environment of the cells and regulate the transport of materials into and out of the cells. They can also act as signal transducers and control the electrical excitation in the nervous system by altering the (membrane) permeability to particular ions in response to stimuli. Such electrical activities can propagate over long distances and represent one of the most spectacular of the membrane functions.

For our purpose here, we are interested in the following question: How do the membranes maintain the ionic environments in the cell over long periods of time while simultaneously allowing transmembrane passage of the needed ions? The mechanism that makes this possible is the so-called Donnan equilibrium; we discuss the details of this equilibrium in this chapter, but a qualitative picture of the Donnan effect can be obtained with the help of a simplified model of a cell shown in Figure 3.1 (See, also, Section 3.5a.).

Figure 3.1a shows a membrane that is permeable to water and K^+ and Cl^- ions but impermeable to colloidal electrolytes (polyelectrolytes such as charged proteins). Let α denote the interior of the cell and β the extracellular region. In the absence of the polyelectrolyte, water, K^+ and Cl^- partition themselves into the two sides such that the chemical potentials of each species are the same inside as well as outside, as thermodynamics would demand. Moreover, the requirement of electroneutrality in both α and β demands that the concentrations of each species K^+ and Cl^- be the same on either side of the partition.

Now, what can we expect if polyelectrolyte ions P^{z-}, each with z negative charges, are present in α? Since the electroneutrality in α is disturbed by the presence of the polyelectrolyte, some of the Cl^- ions will be pushed out to the extracellular region; similarly, there will be an accompanying increase in the concentration of the K^+ ions inside the cell. Eventually the system will reach a state illustrated in Figure 3.1b, with a higher concentration (hence, higher chemical potential) of Cl^- outside and a higher concentration of K^+ inside. This effect is the *Donnan effect* and is behind the maintenance of what is known as the "resting state" of the cell, that is, the polyelectrolyte maintains the concentrations in its environment against the chemical potential gradients. This is one of the reasons why cell and molecular biology books begin with a description of the Donnan equilibrium and the attendant osmotic pressure effects in their discussions of cellular functions.

Elementary and advanced treatments of such cellular functions are available in specialized monographs and textbooks (Bergethon and Simons 1990; Levitan and Kaczmarek 1991; Nossal and Lecar 1991). One of our objectives in this chapter is to develop the concepts necessary for understanding the Donnan equilibrium and osmotic pressure effects. We define osmotic pressures of charged and uncharged solutes, develop the classical and statistical thermodynamic principles needed to quantify them, discuss some quantitative details of the Donnan equilibrium, and outline some applications.

3.1c Focus of This Chapter

As mentioned above, the primary focus of this chapter is on osmotic pressure and its basis in solution thermodynamics. We consider both classical and statistical thermodynamic interpretations of osmotic pressure. The next three sections are devoted to this. The last two sections describe osmotic effects in charged systems and a few applications of osmotic phenomena.

1. The thermodynamic preliminaries and concepts needed for defining osmotic pressure are discussed in Sections 3.2a–c. The nonideality of colloidal solutions can be appreciable since the solvent and solute particles are so different in size. Classical thermo-

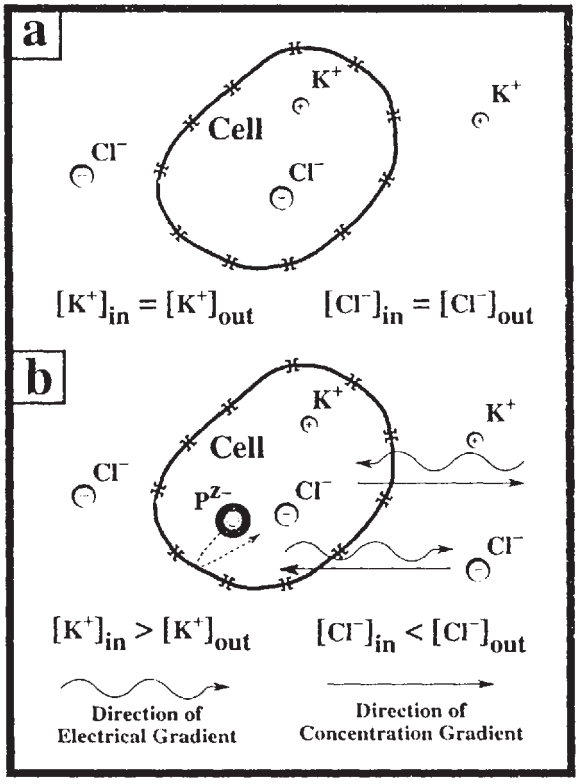


FIG. 3.1 Donnan equilibrium and regulation of ionic concentrations resulting from the presence of a semipermeable membrane: (a) conditions in the absence of polyelectrolyte molecules; and (b) ionic concentrations and electrical gradients in the presence of the polyelectrolyte. The membrane is permeable to water and K^+ and Cl^-, but not to the polyelectrolyte P^{z-}.

dynamics allows this nonideality to be quantified in terms of the so-called virial coefficients, as outlined in Section 3.2d.

2. Some of the experimental details of osmometry and the use of osmometry for determining molecular weights and second virial coefficients are discussed in Section 3.3.
3. If we turn from phenomenological thermodynamics to statistical thermodynamics, then we can interpret the second virial coefficient in terms of molecular parameters via a model. We pursue this approach for two different models, namely, the excluded-volume model for solute molecules with rigid structures and the Flory-Huggins model for polymer chains, in Section 3.4.
4. We conclude the chapter with a discussion of the Donnan equilibrium and the thermodynamic behavior of charged colloids, particularly with respect to osmotic pressure and molecular weight determination (Section 3.5), and some applications of osmotic phenomena (Section 3.6).

Although this is the first place in this book that we have devoted any attention to charged particles, it is not the last. Chapters 11–13, in particular, devote a good deal of attention to such systems. A part of Chapter 4 is devoted to the effects of charges on particles on viscous behavior of dispersions.

3.2 OSMOTIC PRESSURE: THERMODYNAMIC FOUNDATIONS

3.2a Gibbs Free Energy and Chemical Potential

An extremely useful quantity in the thermodynamic treatment of multicomponent phase equilibria is the chemical potential. The chemical potential for component i, μ_i, is the partial molal Gibbs free energy with respect to component i at constant pressure and temperature:

$$\mu_i = \left(\frac{\partial G}{\partial n_i}\right)_{p,T,n_{j\neq i}} \tag{1}$$

Although the notation of this mathematical definition of chemical potential is somewhat cumbersome, the physical significance is fairly clear. The chemical potential is the coefficient that describes the way the Gibbs free energy of a system changes per mole of component i if the temperature, pressure, and number of moles of all components other than the ith are held constant. Although it is expressed on a molar basis, it is important to note that μ_i is a differential quantity; that is, it represents the local slope of the line that shows the variation of G with n_i. The line itself arises from slicing across the complex surface that describes G at the specified values of p, T, and n_j. For a pure substance, μ_i is identical to the Gibbs free energy per mole of that substance. For the present we regard the pure substance as the standard state for μ_i and represent it by the symbol μ_i^o.

The great utility of the chemical potential in phase equilibrium problems arises in the following way. In open systems in which the number of moles of any component may increase or decrease, any change in the Gibbs free energy of the system as a whole may be expressed as the sum of the following contributions:

$$dG = \left(\frac{\partial G}{\partial T}\right)_{p,n} dT + \left(\frac{\partial G}{\partial p}\right)_{T,n} dp + \sum_i \left(\frac{\partial G}{\partial n_1}\right)_{p,T,n_j} dn_i \tag{2}$$

This equation may be written

$$dG = -S\,dT + V\,dp + \sum_i \mu_i dn_i \tag{3}$$

by substituting into Equation (2) the definition of chemical potential and the familiar relationships

$$\left(\frac{\partial G}{\partial T}\right)_{p,n} = -S \tag{4}$$

and

$$\left(\frac{\partial G}{\partial p}\right)_{T,n} = V \tag{5}$$

For any equilibrium that occurs at constant temperature and pressure, the first two terms from the right-hand side of Equation (3) equal zero, reducing the equation to the form

$$dG = \sum_i \mu_i dn_i \tag{6}$$

If the total system consists of several different phases, which we designate by Greek subscripts $\alpha, \beta, \ldots$, then we may also write

$$dG = dG_\alpha + dG_\beta + \ldots \tag{7}$$

Finally, the equilibrium condition requires that

$$dG = 0 \tag{8}$$

Therefore, the equilibrium between two phases means that

$$dG = 0 = dG_\alpha + dG_\beta = \sum_i \mu_{i,\alpha} dn_{i,\alpha} + \sum_i \mu_{i,\beta} dn_{i,\beta} \tag{9}$$

That is, for each of the components, the following holds:

$$\mu_{i,\alpha} dn_{i,\alpha} + \mu_{i,\beta} dn_{i,\beta} = 0 \tag{10}$$

If the entire system consists of only two phases, the hypothesis here, the conservation of matter requires that any substance lost from one phase must appear in the other:

$$dn_{i,\alpha} = -dn_{i,\beta} \tag{11}$$

Combining these last two results leads to the conclusion that the condition for phase equilibrium at constant temperature and pressure is

$$\mu_{i,\alpha} = \mu_{i,\beta} \tag{12}$$

It is important to realize that the chemical potential of each component must be the same in all equilibrium phases, although the value for this quantity will, in general, be different for each component.

3.2b Thermodynamic Activities of Components

Perhaps the most basic equation involving the chemical potential is the one that relates this quantity to the activity of component i in the solution, a_i:

$$\mu_i = \mu_i^0 + RT \ln a_i \tag{13}$$

For the present a_i is expressed in mole fraction units. We see, therefore, that μ_i approaches μ_i^0 as a_i approaches unity. Furthermore, since μ_i is the partial molal Gibbs free energy, Equation (5) also applies to μ_i provided we replace V with $\overline{V}_i$, the partial molal volume of component i:

$$\left(\frac{\partial \mu_i}{\partial p}\right)_{T,n} = \overline{V}_i \tag{14}$$

Suppose we apply these relationships to the equilibrium of a liquid mixture and its vapor. At equilibrium, μ_i must have the same value for each component in both the liquid and vapor phases. Therefore

$$\mu_{i,L} = \mu_{i,V} \tag{15}$$

or

$$\left(\frac{\partial \mu_{i,l}}{\partial p}\right)_{T,n} = \left(\frac{\partial \mu_{i,v}}{\partial p}\right)_{T,n} = \bar{V}_i = \frac{RT}{p_i} \tag{16}$$

where the last relationship assumes the vapor to behave ideally. If Equation (13) is used to evaluate the left-hand side of this equation, we obtain

$$RT\,\partial \ln a_i = RT\frac{\partial p_i}{p_i} \tag{17}$$

Recalling that $p_i = p_i^0$, the *normal vapor pressure* of the pure liquid, when $a_i = 1$, we may integrate Equation (17) to give

$$a_i = \frac{p_i}{p_i^0} \tag{18}$$

This relationship constitutes the basic definition of the activity. If the solution behaves ideally, $a_i = x_i$ and Equation (18) define *Raoult's law*. Those four solution properties that we know as the *colligative properties* are all based on Equation (12); in each, solvent in solution is in equilibrium with pure solvent in another phase and has the same chemical potential in both phases. This can be solvent vapor in equilibrium with solvent in solution (as in vapor pressure lowering and boiling point elevation) or solvent in solution in equilibrium with pure, solid solvent (as in freezing point depression). Equation (12) also applies to osmotic equilibrium as shown in Figure 3.2.

3.2c Osmotic Equilibrium: Relating Chemical Potentials and Compositions to Osmotic Pressure

Figure 3.2 shows schematically two liquid phases—one solution, the other pure solvent—separated by a partition known as a *semipermeable* membrane. In many ways this membrane is the central feature of osmometry; we have more to say about it, but for now it is sufficient to define a semipermeable membrane as one that is permeable to the solvent and impermeable

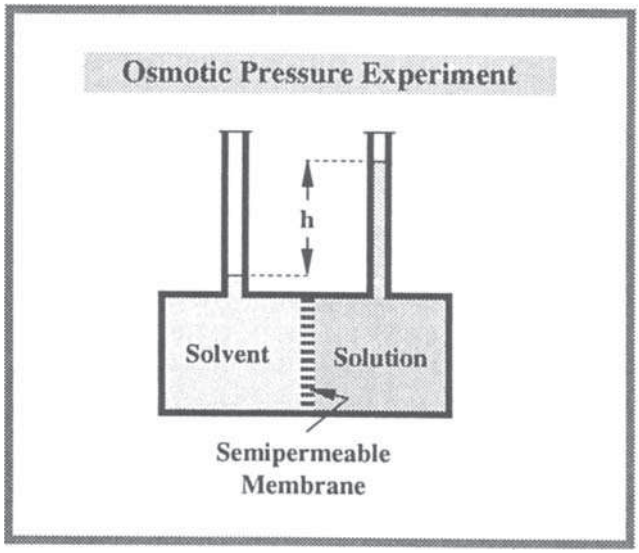


FIG. 3.2 Schematic representation of an osmotic pressure experiment.

to the solute. Thus the semipermeable membrane in Figure 3.2 prevents the two liquids from mixing and at the same time allows both sides of the membrane to come to equilibrium.

In order for solvent and solution to be in equilibrium in an apparatus such as that shown in Figure 3.2, the solution side must be at a higher pressure than the solvent side. This excess pressure is what is known as the *osmotic pressure* of the solution. If no external pressure difference is imposed, solvent will diffuse across the membrane until an equilibrium hydrostatic pressure head has developed on the solution side. In practice, to prevent too much dilution of the solution as a result of the solvent flow into it, the column in which the pressure head develops is generally of a very narrow diameter. We return to the details of osmotic pressure experiments in the next section. First, however, the theoretical connection between this pressure and the concentration of the solution must be established.

Since the two sides of the membrane are in true isothermal equilibrium in an osmotic pressure experiment, the chemical potential of the solvent must be the same on both sides of the membrane. On the side containing pure solvent, μ_i equals μ_i^0. On the solution side of the membrane, the chemical potential of the solvent must also equal the same value according to the equilibrium criterion of Equation (12).

Equation (13) reminds us that the chemical potential has its greatest value, μ_i^0, for a pure substance. Any value of a_i less than unity will cause μ_i to be altered from μ_i^0 by an amount RT ln a_i, which will be negative for $a_i < 1$. Second, any pressure on a liquid that exceeds p_i^0 increases μ_i above μ_i^0. This is seen from the combination of Equations (13) and (18). Thus consideration of the chemical potential of the solvent makes it clear how osmotic equilibrium comes about. The presence of a solute lowers the chemical potential of the solvent. This is offset by a positive pressure on the solution, the osmotic pressure π, so that the net chemical potential on the solution side of the membrane equals that of the pure solvent on the other side of the membrane. This is summarized by the expression

$$\mu_1^0 = \mu_1^0 + RT \ln a_i + \int_{p_1^0}^{p_1^0 + \pi} \overline{V}_1 dp \tag{19}$$

In which subscript 1 indicates the solvent. We use the subscripts 1 and 2 to indicate solvent and solute, respectively, throughout this chapter. In order to relate π to the concentration of the solution, then, we must find a way to integrate Equation (19). The easiest way of doing this is to assume that $\overline{V}_1$ is constant. This approximation is justified because the solution is a condensed phase and shows negligible compressibility. Making this assumption and integrating Equation (19) gives

$$\ln a_i = -\frac{\pi \overline{V}_1}{RT} \tag{20}$$

Combining Equations (13) and (20) permits us to express the chemical potential in terms of osmotic pressure instead of activity:

$$\mu_1 = \mu_1^0 - \pi \overline{V}_1 \tag{21}$$

This relationship will be useful in Chapter 5.

Equation (20) provides the relationship we have sought between osmotic pressure and concentration. If the solution is ideal, we may replace activity by mole fraction. Then Equation (20) becomes

$$\ln x_1 = -\frac{\pi \overline{V}_1}{RT} = \ln (1 - x_2) \approx -x_2 - \frac{x_2^2}{2} - \cdots \tag{22}$$

where the approximation arises from the expansion of the logarithm (see Appendix A). Therefore for ideal, two-component solutions, we write

$$x_2 = \frac{\pi \overline{V}_1}{RT} \tag{23}$$

Since real solutions tend toward ideality as the solute concentration decreases,

$$x_2 = \frac{n_2}{n_1 + n_2} \approx \frac{n_2}{n_1} \tag{24}$$

in which the n's are the number of moles of the indicated component. The approximate form of Equation (24) applies in the case of dilute solutions for which $n_2 \ll n_1$. Introducing the dilute solution approximation of Equation (24) into Equation (23) yields

$$n_2 = \frac{n_1 \pi \bar{V}_1}{RT} = \frac{\pi V_1}{RT} \tag{25}$$

where V_1 is simply the volume of the solvent in the solution, $n_1 \bar{V}_1$. This relationship, known as the *van't Hoff equation* (after J. H. van't Hoff, recipient of the first Nobel Prize in chemistry, 1901), is analogous in form to the ideal gas law and, like the ideal gas law, is a limiting law that applies perfectly only in the limit of $n_2 \to 0$ or $n_2/V_1 \to 0$. Equation (25) shows that osmotic pressure experiments provide a means of measuring the number of solute particles in a solution. If the weight of solute in the solution is also known, this information may be used to evaluate molecular weights. We consider the application of osmometry to problems of molecular weight determination in Example 3.1. For now it is sufficient to note that even high molecular weight solutes—for which the number of molecules per weight of sample is orders of magnitude less than for low molecular weight compounds—produce appreciable osmotic pressures. For example, if a 1% aqueous solution is assumed to be ideal at 25°C, then Equation (25) shows that solutes of molecular weight 10^3, 10^4, 10^5, and 10^6 have osmotic pressures of 2530, 253, 25.3, and 2.53 mm of solution, respectively. Thus even very high molecular weight solutes generate pressures that result in easily measured liquid columns. We see below how to get around the assumption of ideality in Equation (25) so this result can be applied with confidence to real solutions.

3.2d Nonideality: Virial Expansion for the Osmotic Pressure

After reaching Equation (20) we abandoned a completely general discussion of osmotic pressure in favor of the simpler assumption of ideality. The ideal result, Equation (25), applies to real solutions in the limit of infinite dilution. The objective of this section is to examine the extension of Equation (20) to nonideal solutions or, more practically, to solutions with concentrations that are greater than infinitely dilute.

Since both the osmotic pressure of a solution and the pressure-volume-temperature behavior of a gas are described by the same formal relationship of Equation (25), it seems plausible to approach nonideal solutions along the same lines that are used in dealing with nonideal gases. The behavior of real gases may be written as a power series in one of the following forms for n moles of gas:

$$\frac{pV}{nRT} = 1 + B\rho + C\rho^2 + \ldots \tag{26}$$

or

$$\frac{pV}{nRT} = 1 + B\left(\frac{n}{V}\right) + C\left(\frac{n}{V}\right)^2 + \ldots \tag{27}$$

Equations of this type are known as virial equations, and the constants they contain are called the virial coefficients. It is the second virial coefficient B that describes the earliest deviations from ideality. It should be noted that B would have different but related values in Equations (26) and (27), even though the same symbol is used in both cases. One must be especially attentive to the form of the equation involved, particularly with respect to units, when using literature values of quantities such as B. The virial coefficients are temperature dependent and vary from gas to gas. Clearly, Equations (26) and (27) reduce to the ideal gas law as $p \to 0$ or as $n/V \to 0$. Finally, it might be recalled that the second virial coefficient in Equation (27) is related to the van der Waals constants a and b as follows:

$$B = b - \frac{a}{RT} \tag{28}$$

This last relationship points out that for gases the second virial coefficient arises both from the finite size and from the interactions of the molecules of the gas since these are the origins, respectively, of b and a. We have more to say about this in Chapter 10 (see, for example, Sections 10.4 and 10.5).

As we extend these ideas to nonideal solutions, a similar set of statements will apply to the resulting power series:

1. The second virial coefficient is our primary concern since we focus attention on the first deviations from ideality.
2. The value of B will depend in part on the units chosen for concentration, as well as on the temperature and the nature of the system.
3. The virial equation for osmotic pressure must reduce to the van't Hoff equation in the limit of infinite dilution.
4. The second virial coefficient may be expected to reflect both the finite size and the interactions of the molecules.

Each of these points is taken up in the following discussion.

The easiest way to extend these considerations to the osmotic pressure of nonideal solutions is to return to Equation (22), which relates π to a power series in mole fraction. This equation applies to ideal solutions, however, since ideality is assumed in replacing activity by mole fraction in the first place. To retain the form and yet extend its applicability to nonideal solutions, we formally include in each of the concentration terms a correction factor defined to permit the series to be applied to nonideal solutions as well:

$$\frac{\pi \bar{V}_1}{RT} = A' x_2 + \frac{1}{2} B' x_2^2 + \cdots \tag{29}$$

The coefficients A', B', . . . must all equal unity in ideal solutions in order to recover Equation (22). Since the van't Hoff equation is a limiting law, the coefficient A' must equal unity in all solutions. Therefore Equation (29) becomes

$$\frac{\pi V_1}{RT} = x_2 + \frac{1}{2} B' x_2^2 + \cdots \tag{30}$$

Next, let us consider the transformation of mole fraction concentration units into other units more appropriate for use with solutes with molecular weights that might be unknown. In working with unknowns, mass volume^{-1} units offer the greatest flexibility since only a balance and a volumetric flask are needed to quantitatively characterize a solution in these units. We represent this unit of concentration by c in this chapter. For the reasons presented here, this same concentration unit is used in viscosity and light scattering work with solutes of unknown molecular weight. If V is the volume of a solution of components 1 and 2, it may be written as $n_1\bar{V}_1 + n_2\bar{V}_2$, where the $\bar{V}_i$'s are partial molar volumes. Remember, the partial molar volumes give the volume occupied by a mole of the indicated component in a mixture; their precise value depends on the concentration of the solution. Therefore, if m_2 is the mass of solute in the solution,

$$c = \frac{m_2}{n_1\bar{V}_1 + n_2\bar{V}_2} \approx \frac{m_2}{n_1\bar{V}_1} \tag{31}$$

where the approximation applies to dilute solutions in which $n_2 \ll n_1$. Since m_2 is equal to the product of n_2 and M_2,

$$c = \frac{n_2 M_2}{n_1\bar{V}_1} \approx x_2 \frac{M_2}{\bar{V}_1} \tag{32}$$

in which the dilute solution form of Equation (24) has been used to introduce x_2. Substituting this result into Equation (30) gives

$$\frac{\pi \bar{V}_1}{RT} = \frac{\bar{V}_1}{M_2} c + \frac{1}{2} B' \left(\frac{\bar{V}_1}{M_2}\right)^2 c^2 + \cdots \tag{33}$$

Ignoring higher order terms, we can rearrange Equation (33) to yield

$$\frac{\pi}{RTc} = \frac{1}{M_2} + \frac{1}{2} \frac{B' \bar{V}_1}{M_2^2} c = \frac{1}{M_2} + Bc \tag{34}$$

The quantity π/c is called the *reduced osmotic pressure*, and $(\pi/c)_0$ the limiting reduced osmotic pressure. Osmotic pressure results are most commonly encountered as plots of reduced osmotic pressure—with or without the RT—versus c. This form suggests that a plot of π/RTc versus c should be a straight line, the intercept and slope of which have the following significance:

$$\text{intercept} = \left(\frac{\pi}{RTc}\right)_0 = \frac{1}{M_2} \tag{35}$$

$$\text{slope} = B = \frac{1}{2} \frac{B' \bar{V}_1}{M_2^2} \tag{36}$$

Note that the value of the intercept, the value of π/RTc at infinite dilution, obeys the van't Hoff equation, Equation (25). At infinite dilution even nonideal solutions reduce to this limit. The value of the slope is called the second virial coefficient by analogy with Equation (27). Note that the second virial coefficient is the composite of two factors, B' and $(1/2)\ \bar{V}_1/M_2^2$. The factor B' describes the first deviation from ideality in a solution; it equals unity in an ideal solution. The second cluster of constants in B arises from the conversion of practical concentration units to mole fractions. Although it is the nonideality correction in which we are primarily interested, we discuss it in terms of B rather than B' since the former is the quantity that is measured directly. We return to an interpretation of the second virial coefficient in Section 3.4.

3.3 OSMOMETRY: SOME APPLICATIONS

3.3a Experimental Considerations

To carry out an osmotic pressure experiment, we need to prepare a solution, find a suitable semipermeable membrane, achieve isothermal equilibrium, and measure the equilibrium pressure. Aside from noting that the pressures produced by colloidal solutes are large enough to be measurable, we have not yet considered any of the experimental aspects of osmometry. This is our present task.

First, a suitable solvent and membrane must be found. The solvent must dissolve enough solute to produce an adequate pressure. The results of measurements made at relatively high concentrations may be extrapolated, as discussed in Section 3.2d and illustrated below, to zero concentration, so we need not worry about the effects of nonideality. As low a solvent viscosity as possible is desirable to minimize the time required for equilibration.

It is important that the materials be free of contaminants in osmotic pressure experiments. Suppose, for example, that the solvent contains a small amount of impurity that, like the colloidal solute, is retained by the membrane. Then, as far as the osmotic pressure is concerned, that impurity will contribute to the osmotic pressure in the same way that the colloid does. Since the osmotic pressure responds to the number of solute particles present, a low molecular weight impurity in extremely small amounts may contain as many or more particles as a dilute solution of a colloidal solute of very high molecular weight. Quite large errors in molecular weight may arise in this way. The confusion may be compounded if the same system is investigated using a different membrane material. It is conceivable that another membrane

would be permeable to the impurity. This would result in a different apparent molecular weight for the colloid. We have approached the issue of impurities from the viewpoint of the solvent. Actually, the colloid is more likely to be the source of the impurity since these materials are often difficult to purify. We discuss some additional aspects of this problem in the sections on average molecular weight (Section 3.3c), charged colloids (Section 3.5b), and dialysis (Section 3.6a).

The membrane is the source of most of the difficulties in osmometry. There is no general way to select a membrane material that will be permeable to one component and impermeable to another for any conceivable combination of chemicals. Very high molecular weight components are generally more easily retained, however, so we have a slight advantage in this regard. The membrane must be sufficiently thin to permit equilibration at a reasonable rate. At the same time, it must be strong enough to withstand the considerable pressure differences that may exist across it. This problem may generally be overcome, at least in part, by suitable mechanical support of the membrane. A more serious problem is the preparation of thin membranes that are free from minute imperfections that would constitute a "leak" between the two compartments. Such a leak would totally invalidate the experimental results. One peculiarity of membranes is their tendency to display what is known as an asymmetry pressure; that is, an equilibrium pressure difference may exist across a membrane even when there is solvent on both sides. This must be measured and subtracted as a "blank correction."

Equilibrium osmometry is a thermodynamic phenomenon. As such, it makes no difference what mechanism the membrane uses to retain the solute, nor do we learn anything about the mechanism from equilibrium studies. It is easy to visualize that some solutes, especially those in the colloidal size range, are retained by a sieve effect; that is, the molecules are simply too large to pass through the pores in the membrane material. Another possibility is a mechanism by which the membrane displays selective solubility. This means that the membrane dissolves the solvent but not the solute. In this way the solvent can pass through the membrane, while the solute is retained. An analogous mechanism for charged particles may arise by the membrane repelling (and thus retaining) particles of one particular charge.

A great many different materials have been used in osmotic pressure experiments. Various forms of cellophane and animal membranes are probably the most common membrane materials. Various other polymers, including polyvinyl alcohol, polyurethane, and polytrifluorochloroethylene, have also been used along with such inorganic substances as $CuFe(CN)_6$ precipitated in a porous support.

Once a suitable membrane and solvent are selected, an experimental arrangement must be devised that measures the equilibrium pressure under isothermal conditions. Many variations in apparatus design have been studied. Two particularly instructive pieces of apparatus are shown in Figure 3.3. The assembly shown in Figure 3.3a consists of an inner solution compartment with a relatively large opening at the membrane end and a capillary at the small end. The entire solution chamber is then immersed in a tube containing the solvent. Once assembled, the entire apparatus is placed in a constant temperature bath for equilibration.

Another osmometer design is shown assembled in Figure 3.3b and in detail in Figure 3.3c. The membrane is pressed between two grooved faces that contain solvent on one side and solution on the other. The grooves are attached to filling tubes and to capillaries in which the pressure head develops. This apparatus permits a large contact area between liquids with a minimum volume of liquids involved.

Several times in this discussion we have noted the importance of experimental conditions that permit as rapid an equilibration as possible. The implication of these remarks is that osmotic equilibrium is reached slowly. In some cases as much as one week may be required for equilibrium to be achieved. To shorten this time, procedures based on measuring the rate of approach to equilibrium have been developed. The osmometer of Figure 3.3b is especially suited for this procedure.

In successive runs, the capillary on the solution side of the membrane is filled with solution to some initial setting that will be above or below the equilibrium location of the meniscus. At various times after the initial settings, the height of the liquid column is mea-

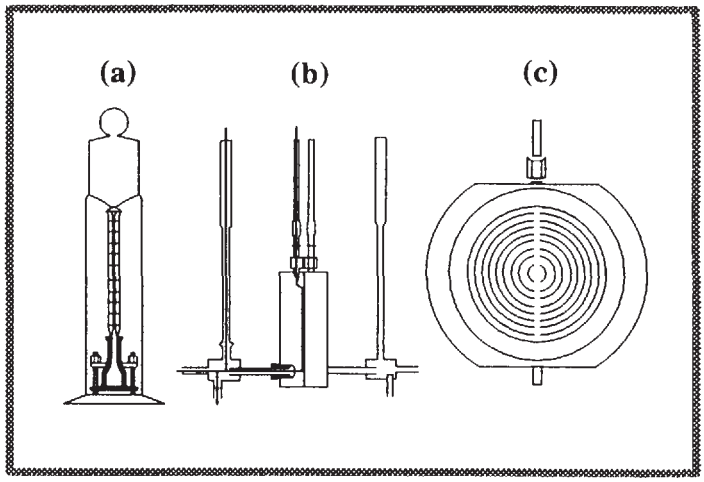


FIG. 3.3 Two types of osmometers: (a) the solution compartment is submerged in solvent. (Adapted from D. M. French and R. H. Ewert, *Anal. Chem.*, **19**, 165 (1947)); and (b) the solution and solvent occupy grooves on opposite faces of the central unit, as shown in detail in (c) (Adapted from R. M. Fuoss and D. J. Mead, *J. Phys. Chem.*, **47**, 59 (1943)).

sured. It is found that the ascending and descending branches of the curve converge to the same point—a value that equals the equilibrium osmotic pressure. This is shown in Figure 3.4. The rate at which equilibrium is achieved decreases as equilibrium is approached. Therefore it is desirable to bracket the true value as narrowly as possible to take full advantage of this approach-to-equilibrium extrapolation procedure. With some judicious planning, the time for an osmotic experiment may be shortened considerably by this dynamic method.

Once equilibrium has been reached, the height difference between the two liquid surfaces is all that remains to be measured. The primary factor to note here is that capillaries are used to minimize the dilution effects. This means that corrections for capillary rise must be taken into account unless the apparatus allows the difference between two carefully matched capillaries to be measured. We discuss capillary rise in Chapter 6, Sections 6.2 and 6.4. Finally, there is an extremely important practical reason, in addition to the theoretical requirement of isothermal conditions, for good thermostating in osmometry experiments. The apparatus consists of a large liquid volume attached to a capillary and therefore has the characteristics of a liquid thermometer: The location of the meniscus is quite sensitive to temperature fluctuations.

3.3b Measurement of Molecular Weight Using Osmometry

We now illustrate the use of the equations developed in Section 3.2 for interpreting osmotic pressure data. Figure 3.5 shows examples of two plots of π/RTc versus concentration. In Figure 3.5a the data all describe different molecular weight fractions of the same solute, cellulose acetate, in acetone solutions. Since the lines in this plot all have essentially the same slope, B must be the same for each.

Figure 3.5b shows data for a sample of nitrocellulose in three different solvents. All show the same intercept corresponding to a single molecular weight as required. Note, however, that the slopes are different, including even a negative slope, indicating that wide variations in B are possible. We examine the factors that determine B in Section 3.4.

Example 3.1 considers the molecular weights of the polymers in Figure 3.5 in terms of Equation (35).

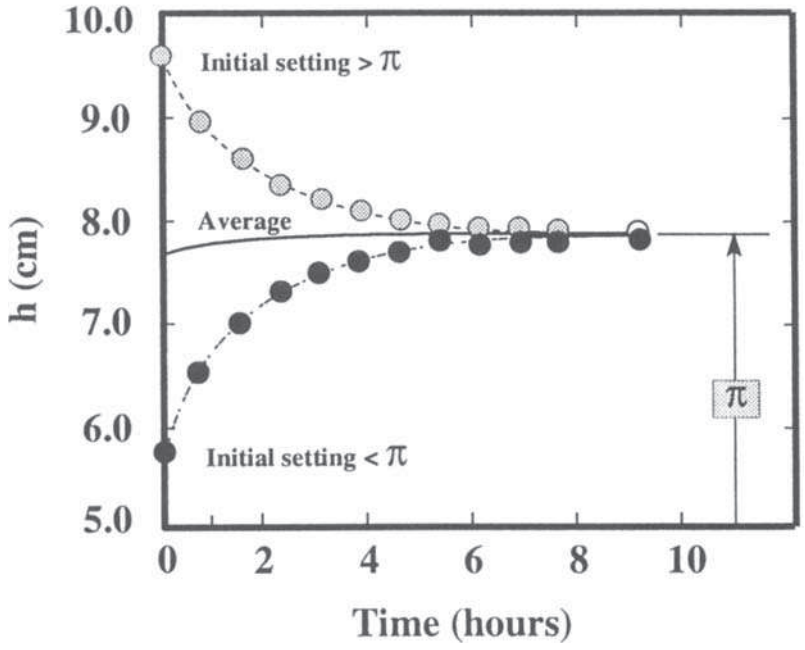


FIG. 3.4 Data showing the approach to osmotic equilibrium from initial settings above and below the equilibrium column height. (Adapted from R. U. Bonnar, M. Dimbat, and F. H. Stross, *Number Average Molecular Weights*, Wiley, New York, 1958.)

* * *

EXAMPLE 3.1 *Molecular Weights from Osmotic Pressure Measurements.* To what molecular weights do the limiting reduced osmotic pressures obtained from Figure 3.5 correspond? The data in Figure 3.5 are presented in such a way that unit problems do not enter the picture. We are not usually so lucky! Consider some possible combinations of units for π, V, and R that are compatible with the units of the ordinate in Figure 3.5.

Solution: According to Equation (35), the molecular weights of the various polymers are simply the reciprocals of the limiting reduced osmotic pressures in the units that have been used for the ordinate in Figure 3.5, namely, mole g^{-1}, Therefore

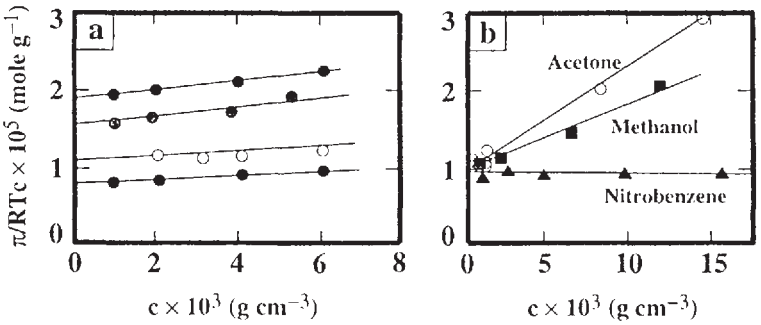


FIG. 3.5. Plots of π/RTc versus concentration: (a) various cellulose acetate fractions in acetone (data from A. Bartovics and H. Mark, *J. Am. Chem. Soc.*, **65**, 1901 (1943)); and (b) nitrocellulose in three different solvents (data from A. Dobry, *J. Chem. Phys.*, **32**, 50 (1935)).

Part in Figure 3.5	$(\pi/RTc)_0 \cdot 10^{-5}$ (mole g^{-1})	M (g mole^{-1})
(a)	1.92	52,000
(a)	1.59	63,000
(a)	1.09	92,000
(a)	0.79	126,000
(b)	0.90	111,000

Osmotic pressures may be expressed in any of the usual pressure units (and then some!) and volumes are often—but not always—expressed in cubic centimeters (milliliters); compatible units of R must be chosen:
If π is in atm and V in cm^3, use $R = 82.05$ atm cm^3 K^{-1} mole^{-1}.
If π is in torr and V in cm^3, use $R = 62{,}360$ torr cm^3 K^{-1} mole^{-1}.
If π is in mm of solution and V in cm^3, use $R = 62{,}360\ (\rho_{Hg}/\rho_{Soln})$ mm cm^3 K^{-1} mole^{-1}, where the ρ's are densities.
If π is in dyne cm^{-2} and V in cm^3, use $R = 8.314 \cdot 10^7$ erg K^{-1} mole^{-1}.
If π is in N m^{-2} and V in m^3, use $R = 8.314$ J K^{-1} mole^{-1}
If the masses (in c) are expressed in g and kg, respectively, in the last two situations, then the units of the ordinate reduce to cm^2 s^{-2} and m^2 s^{-2}, respectively, which do not bear much resemblance to molecular weight units. Note, however, that the acceleration of gravity in appropriate units can be factored out of these latter quantities to leave units of cm and m, respectively, for reduced osmotic pressure. ■

* * *

3.3c Osmometry and Polydispersity of Molecular Weights

As we saw in Section 1.5c, the condition of polydispersity is quite normal with colloidal solutes. Our discussion of osmometry has shown that it is possible to evaluate the molecular weight of a colloidal solute by osmotic pressure measurements. Next we consider the fact that the sample on which such a measurement is made is more than likely polydisperse and that the molecular weight obtained from such an experiment is some average quantity. The objective of this section is to show that it is the *number average molecular weight* that is determined in an osmotic pressure experiment on a polydisperse system. For the purposes of this demonstration, the solution is assumed to be ideal. (Notice, however, that, in the case of surfactant aggregates, i.e., micelles, discussed in Chapter 8, osmometry can lead to weight average molecular weight $\overline{M}_w$, as shown by Puvvada and Blankschtein in 1989.)

Experimental results from a polydisperse system may be related as follows:

$$\pi_{exp} = \frac{c_{exp}RT}{\overline{M}} \tag{37}$$

where π_{exp} and c_{exp} represent the experimental osmotic pressure and concentration, respectively, and $\overline{M}$ is the average molecular weight. It is the method of averaging in Equation (37) that we seek to determine. This relationship applies to the observable quantities. Precisely the same equation may be written, however, for each of the molecular weight fractions of the solute, designated here by the subscript i:

$$\pi_i = \frac{c_iRT}{M_i} \tag{38}$$

Two additional relationships are fairly evident. The experimental osmotic pressure is the sum of the pressure contributions of the individual components:

$$\pi_{exp} = \sum_i \pi_i \tag{39}$$

and the experimental concentration is the sum of the concentrations of the components:

$$c_{exp} = \sum_i c_i \tag{40}$$

Next, Equations (37)–(40) can be combined as follows:

$$\pi_{exp} = \sum_i \pi_i = RT \sum_i \frac{c_i}{M_i} = RT\frac{c_{exp}}{\overline{M}} = RT\frac{\Sigma_i c_i}{\overline{M}} \tag{41}$$

Equation (41) may be simplified to yield

$$\overline{M} = \frac{\Sigma_i c_i}{\Sigma_i (c_i/M_i)} \tag{42}$$

This result still fails to resemble any of the standard averages listed in Table 1.8. However, if we introduce the expression

$$c_i = \frac{n_i M_i}{V} \tag{43}$$

we obtain

$$\overline{M} = \frac{(1/V)\,\Sigma_i n_i M_i}{(1/V)\,\Sigma_i (n_i M_i/M_i)} = \frac{\Sigma_i n_i M_i}{\Sigma_i n_i} \tag{44}$$

Equation (44) shows the average molecular weight determined from osmometry to be the number average molecular weight as defined by Equation (1.14).

This same conclusion may also be reached by the following argument. The product $n_i M_i$ in Equation (43) equals the weight of component i in the solution; the total weight of solute in the solution equals $\Sigma_i n_i M_i$. The experimental osmotic pressure depends on and therefore measures the total number of moles of solute $\Sigma_i n_i$. The ratio of the total weight to the total number of moles of solute defines the number average molecular weight.

Any experiment that "counts" the number of molecules in a given weight of polydisperse material can be interpreted in terms of a number average molecular weight. All of the colligative properties have this as a feature of their shared thermodynamic origin. Another technique, known as end-group analysis, may be used to determine the molecular weight of certain polymers. Example 3.2 makes it clear that this too yields a number average molecular weight for polydisperse samples.

* * *

EXAMPLE 3.2 *Degree of Polymerization and Molecular Weight Distribution.* Polycaprolactam—otherwise known as nylon 6—with a degree of polymerization n has the following molecular structure: $H_2N(CH_2)_5CO[NH(CH_2)_5CO]_{n-2}NH(CH_2)_5COOH$. A 1.06-g sample of this material is dissolved in an appropriate solvent and titrated with an alcoholic KOH solution of normality 0.0250 N. Exactly 5.00 ml of base are required to neutralize the carboxyl groups of the sample. What is the molecular weight of the sample? What is n? Criticize or defend the following proposition: This method can only be used for molecules that contain just one of the analyzed functional groups per molecule.

Solution: This problem is really no different from the molecular weight determinations of unknown acids that are often conducted in general chemistry lab courses. What is important to recognize is that there is one carboxyl group per molecule or one equivalent per mole. Therefore the molecular weight of the polymer is given by

$$\overline{M} = \frac{1.06\text{ g}}{(5.0)(0.025)\text{ mEq}} \cdot \frac{10^3\text{mEq}}{\text{Eq}} \cdot \frac{1\text{ Eq}}{1\text{ mole}} = 8480\text{ g mole}^{-1}$$

This is the same average molecular weight that would be determined by an osmotic pressure experiment on the same sample. Since the molecular weight of the repeat unit is 113 g mole^{-1}, the degree of polymerization $n = 8480/113 \approx 75$.

The proposition recognizes the importance of knowing the number of analyzed groups per molecule, but incorrectly requires that there be only one such group. Some polymers might contain the same functional group at both ends of a linear chain. In this case there are 2 Eq/mole. A Y-shaped molecule could have three such groups, an X-shaped molecule four, and so forth. These last cases would have 3 and 4 Eq/mole, respectively. The number of groups does not matter as long as that number is known. ■

* * *

That end-group analysis and osmometry give number averages for polydisperse samples is simply because the number of solute molecules is counted in each case. We see in Chapter 5 that light scattering effectively weighs the molecules rather than counts them (see, for example, Section 5.4d). Hence light scattering gives a weight average value for M. The fact that osmotic pressure "counts" solute particles has some interesting consequences for charged colloids, as we see in Section 3.5b.

3.4 STATISTICAL FOUNDATIONS OF SOLUTION THERMODYNAMICS

By analogy with Equation (28), we might expect the second virial coefficient to depend on the size and/or the interactions of the molecules in solution. Although this expectation is basically correct, we must not take the form of Equation (28), a *gas* equation, too literally in discussing *solutions*. Furthermore, it must be recalled that, in gases, only interactions between the molecules of the gas are possible. In solution, we may consider solvent-solvent, solute-solute, and solvent-solute interactions. This is a good reminder that the analogy with gases cannot be pushed too far.

By itself, *classical* thermodynamics provides us with no information on the molecular origin of the second virial coefficient, which is merely a phenomenological coefficient from an exclusively thermodynamic viewpoint. *Statistical* thermodynamics undertakes the task of providing molecular interpretations to such quantities. The added complication is that the statistical thermodynamic approach requires the use of models, and models inevitably oversimplify things. Nevertheless, we have arrived at that point at which we add insight only to the extent that we do some modeling. Thus, for example, we can learn something about solute-solute interactions and solute-solvent interactions from experimental B values, but only if we are willing to accept the models on which these interpretations are based. Toward this end, it is important to know what goes into these models. A model that is very plausible in one system may not make any sense for another.

In Section 3.4a we examine a model for the second virial coefficient that is based on the concept of the excluded volume of the solute particles. A solute-solute interaction arising from the spatial extension of particles is the premise of this model. Therefore the potential exists for learning something about this extension (i.e., particle dimension) for systems for which the model is applicable. In Section 3.4b we consider a model that considers the second virial coefficient in terms of solute-solvent interaction. This approach offers a quantitative measure of such interactions through B. In both instances we only outline the pertinent statistical thermodynamics; a somewhat fuller development of these ideas is given in Flory (1953). Finally, we should note that some of the ideas of this section are going to reappear in Chapter 13 in our discussions of polymer-induced forces in colloidal dispersions and of coagulation or "steric" stabilization (Sections 13.6 and 13.7).

3.4a Excluded-Volume Effect

Before considering how the excluded volume affects the second virial coefficient, let us first review what we mean by excluded volume. We alluded to this concept in our model for size-exclusion chromatography in Section 1.6b.2b. The development of Equation (1.27) is based on the idea that the center of a spherical particle cannot approach the walls of a pore any closer than a distance equal to its radius. A zone of this thickness adjacent to the pore walls is a volume from which the particles—described in terms of their centers—are denied entry because of their own spatial extension. The volume of this zone is what we call the *excluded volume* for such a model. The van der Waals constant b in Equation (28) measures the excluded volume of gas molecules; for spherical molecules it equals four times the actual volume of the sphere, as discussed in Section 10.4b, Equation (10.38).

3.4a.1 Statistical Entropy and Entropy of Mixing: Definitions

A point of entry for statistical considerations into thermodynamics is the Boltzmann entropy relationship

$$S = k_B \ln \Omega \tag{45}$$

in which Ω is called the thermodynamic weight or the partition function and k_B is the Boltzmann constant, which is replaced by R when we are working on a "per mole" rather than a "per molecule" basis. The thermodynamic weight is the tricky part of Equation (45); it counts the number of ways a particular state can come about. On a qualitative basis, chemists learn to use this relationship almost intuitively. One associates higher entropy with states that are more disordered, and disordered states can come about in a larger number of ways than states of higher order. The familiar analogy that is often used here is to note that there are more ways for a deck of cards to exist as "shuffled" than as "arranged by suits." We can readily see that Equation (45) encompasses the third law of thermodynamics: There is only one way to organize a perfect crystal at absolute zero, hence $\Omega = 1$ and $S = 0$.

Entropy changes can also be developed in terms of Equation (45):

$$S_2 - S_1 = k_B \ln\left(\frac{\Omega_2}{\Omega_1}\right) \tag{46}$$

Our interest is in the entropy of mixing—that is, ΔS for $1 + 2 \rightarrow$ mixture—which may be written

$$\Delta S_m = k_B \ln\left(\frac{\Omega_{mix}}{\Omega_1 \Omega_2}\right) \tag{47}$$

by virtue of Equation (46). Note that we use the subscripts m for the mixing process and *mix* for the mixture itself. What remains to be done is to count the ways in which N_1 molecules of component 1 and N_2 molecules of component 2 exist when mixed together or, more specifically, the factor by which this number exceeds the number of ways the separate molecules can exist.

The derivation we follow proceeds through four stages:

1. We must devise an expression for Ω_{mix}.
2. From Ω_{mix} a straightforward application of Equation (45) is required to obtain an equation for S_{mix}. It takes a bit of additional argument to convert this to an expression for ΔS_m.
3. A single approximation will enable us to go from ΔS_m to ΔG_m, but the implications of this assumption deserve some comment.
4. Some mathematical manipulations convert ΔG_m into an expression for $\mu_i - \mu_1^0$, which is directly related to π through Equation (21).

3.4a.2 Development of Partition Function

We begin by considering the number of ways a solute molecule of excluded volume u can be placed in an otherwise empty volume V. Suppose we imagine V to be sectioned off a number of sites (say, N_s), each of which can accommodate the volume u. Since V is intended to represent the volume of the solution—a macroscopic quantity—and u is a molecular parameter, the number of such sites is large. The first solute molecule to enter V can occupy any one of these sites, and we represent by ω_1 the number of possible placements for the first solute molecule. Because the total volume is partitioned into these sites, it follows that

$$\omega_1 = KV = N_s \tag{48}$$

where K is an appropriate proportionality constant, equal to $(1/u)$ in our present model. (In general, K may differ from $1/u$ since small amounts of unfilled [unfillable] space may exist even when the N_s units of volume $1/u$ each are closely packed, i.e., $N_s u$ can be less than the total volume V. Therefore, for the sake of generality, we develop the equations here in terms of K also, instead of restricting ourselves to $K = 1/u$ only.)

The second molecule to be placed can go into any one of the remaining sites, hence the number of ways to place the second molecule is given by

$$\omega_2 = K(V - u) = (N_s - 1) \tag{49}$$

For the third molecule, this quantity is given by $\omega_3 = K(V - 2u)$, and for the jth it is

$$\omega_j = K\,[V - (j - 1)\,u] = [N_S - (j - 1)] \tag{50}$$

The number of ways of placing the first, the second, the third, and so on to the ith is

$$\Omega \propto \omega_1\omega_2\omega_3 \ldots \omega_i = N_s\,(N_s - 1)\,(N_s - 2) \ldots [N_s - (i - 1)] = \frac{N_s!}{(N_s - i)!} \tag{51}$$

We have written Equation (51) as a proportion rather than as an equality since the given expression actually overcounts the number of possible placements. If there is a total of N_2 solute molecules, then there are $N_2!$ different permutations of these molecules, and Equation (51) should be divided by $N_2!$ because of the interchangeability of the molecules. Incorporating this idea and Equation (50) into Equation (51) gives

$$\Omega = \frac{1}{N_2!}\prod_{j=1}^{N_2} K\,[V - (j - 1)u] = \frac{N_s!}{N_2!\,(N_s - N_2)!} \tag{52}$$

where Π represents the product of terms. Letting $i = j - 1$ allows Equation (52) to be written more concisely as

$$\Omega = \frac{1}{N_2!}\prod_{i=0}^{N_2-1} K\,(V - iu) = \frac{1}{N_2!}\prod_{i=0}^{N_2-1} KV\left(1 - i\frac{u}{V}\right) \tag{53}$$

After the N_2 solute molecules are placed, all remaining sites are filled with solvent molecules. Since our model is specifically interested in the solute-solute excluded-volume effect, we may say that there is only one way for the solvent molecules to be placed. Although this overlooks details about the solvent, we see presently that such details would eventually be subtracted, so we lose nothing by this simplification. Equation (53) therefore gives the expression for Ω_{mix} we sought as the first step of our derivation.

3.4a.3 Expression for Entropy of Mixing

Substituting Equation (53) into Equation (45) gives a statistical expression for the solute contribution to the configurational entropy of the mixture:

$$\frac{S_{mix}}{k_B} = -\ln N_2! + \sum_{i=0}^{N_2-1} \ln\left|KV\left(1 - i\frac{u}{V}\right)\right| \tag{54}$$

The summation replaces the product in going from Equation (53) to Equation (54) since we are dealing with logarithms in the latter. Note that the configurational entropy refers explicitly to the entropy associated with the mixture itself; the internal entropy of the molecules themselves is clearly not included. Next a series of mathematical manipulations will transform Equation (54) into a more useful form:

1. Divide the summation in Equation (54) into two terms:

$$\sum \ln\left|KV\left(1 - i\frac{u}{v}\right)\right| = \sum \ln(KV) + \sum \ln\left(1 - i\frac{u}{v}\right)$$

2. Note that $\Sigma \ln(KV) = N_2 \ln(KV)$ since there are N_2 identical terms in the summation.
3. Since $u/V << 1$ as already noted, $\ln(1 - iu/V)$ can be expanded as a series (see Appendix A) with only the leading term retained to give

$$\sum \ln\left(1 - i\frac{u}{V}\right) \approx \sum\left(-i\frac{u}{V}\right) \approx -\frac{u}{V}\sum i$$

4. The sum of integers from 0 to y is $y(y + 1)/2$; hence

$$\sum_{i=0}^{N_2-1} i = (N_2 - 1)\frac{N_2}{2} \approx \frac{N_2^2}{2}$$

since N_2 is large.

5. We can change the equation from dealing with numbers of molecules (N) to numbers of moles (n) by dividing both sides by Avogadro's number N_A and writing $k_B N_A = R$.

Applying these manipulations to Equation (54) gives

$$\frac{S_{mix}}{R} = -\frac{1}{N_A}\ln N_2! + n_2 \ln(KV) - \frac{1}{2}N_A n_2^2 \frac{u}{V} \tag{55}$$

Replacing V by the mole-weighted sum of the partial molar volumes, as was done in developing Equation (31), we write

$$\frac{S_{mix}}{R} = -\frac{1}{N_A}\ln N_2! + n_2 \ln[K(n_1\overline{V}_1 + n_2\overline{V}_2)] - \frac{1}{2}\frac{N_A n_2^2 u}{n_1\overline{V}_1 + n_2\overline{V}_2} \tag{56}$$

If we set n_2 equal to zero in Equation (56), the physical system described corresponds to pure solvent, and the equation reduces to $S_1 = 0$. If we set n_1 equal to zero, we are describing pure solute, and Equation (56) becomes

$$\frac{S_2}{R} = n_2 \ln K + \ln(n_2\overline{V}_2) - \frac{1}{N_A}\ln N_2! - \frac{N_A n_2 u}{2\overline{V}_2} \tag{57}$$

Subtracting these expressions for S_1 and S_2 from S_{mix} gives ΔS_m, which was the goal of the second stage of our derivation:

$$\frac{\Delta S_m}{R} = n_2 \ln\left(\frac{n_1\overline{V}_1 + n_2\overline{V}_2}{n_2\overline{V}_2}\right) - \frac{1}{2}N_A u \frac{n_2^2}{n_1\overline{V}_1 + n_2\overline{V}_2} + \frac{1}{2}N_A\frac{n_2 u}{\overline{V}_2} \tag{58}$$

Note that whatever configurational entropy we had attributed to the solvent would have disappeared at this point.

As outlined above, our next objective is to write an expression for ΔG_m. Since $\Delta G = [\Delta H - T\Delta S]$ for a constant temperature process such as this, we can immediately write

$$-\frac{\Delta G_m}{RT} = n_2 \ln\left(\frac{n_1\overline{V}_1 + n_2\overline{V}_2}{n_2\overline{V}_2}\right) - \frac{1}{2}N_A u \frac{n_2^2}{n_1\overline{V}_1 + n_2\overline{V}_2} + \frac{1}{2}N_A\frac{n_2 u}{\overline{V}_2} \tag{59}$$

by assuming $\Delta H_m = 0$. In making this assumption we imply that the energetic interactions between solvent-solvent pairs, solute-solute pairs, and solvent-solute pairs are all the same. Thus, removing a solvent molecule from a bulk sample of the liquid and replacing it with a molecule of the solute does not involve an enthalpy change. It is the introduction of this assumption that leaves the excluded volume as the sole contributor to the second virial coefficient. In the next section we assume a nonzero value for ΔH_m and examine the effect this has on B. For now, it is enough to note that the present model attributes solution nonideality to molecular size and not to the energetics of molecular interactions. The model is clearly inappropriate for any system in which the molecular interactions are important.

3.4a.4 Chemical Potential, Osmotic Pressure, and Second Virial Coefficient

The final stage of our derivation converts Equation (59) to an expression for $(\mu_1 - \mu_2)$ by differentiating Equation (59) with respect to n_1 according to Equation (1):

$$\mu_1 - \mu_1^0 = -RT\left[\frac{n_2\overline{V}_1}{V} + \frac{1}{2}N_A\overline{V}_1\left(\frac{n_2}{V}\right)^2 u\right] \tag{60}$$

Equation (21) shows how to convert Equation (60) into an expression for osmotic pressure:

$$\pi = \frac{RT}{\overline{V}_1}\left[\frac{n_2\overline{V}_1}{V} + \frac{1}{2}N_A u\overline{V}_1\left(\frac{n_2}{V}\right)^2\right] \tag{61}$$

Since $n_2/V = c/M_2$, Equation (61) may be written in practical concentration units:

$$\pi = RT\left(\frac{c}{M} + \frac{1}{2}\frac{N_A u}{M^2}c^2\right) \tag{62}$$

Comparing this result with Equation (34) shows that

$$B = \frac{1}{2}\frac{N_A u}{M^2} \tag{63}$$

according to this model.

The essence of this model for the second virial coefficient is that an excluded volume is defined by surface contact between solute molecules. As such, the model is more appropriate for molecules with a rigid structure than for those with more diffuse structures. For example, protein molecules are held in compact forms by disulfide bridges and intramolecular hydrogen bonds; by contrast, a randomly coiled molecule has a constantly changing outline and imbibes solvent into the domain of the coil to give it a very "soft" surface. The present model, therefore, is much more appropriate for the globular protein than for the latter. Example 3.3 applies the excluded-volume interpretation of B to an aqueous protein solution.

* * *

EXAMPLE 3.3 *Excluded Volume of Bovine Serum Albumin from Osmotic Pressure Measurements.* A plot of π/c versus c for an aqueous solution of the bovine serum albumin molecule at 25°C and pH = 5.37 is shown in Figure 3.6. The molecule is known to be nearly spherical and uncharged at this pH. Evaluate the molecular weight and the excluded volume of this protein from the intercept and slope of this line, 0.268 torr $(\text{g kg}^{-1})^{-1}$ and $1.37 \cdot 10^{-3}$ torr $\text{kg}^2\ \text{g}^{-2}$, respectively. From the particle mass and volume, estimate the partial specific volume of the solute in solution. The specific volume of the unsolvated protein is about 0.75 $\text{cm}^3\ \text{g}^{-1}$; does the solute appear to be solvated?

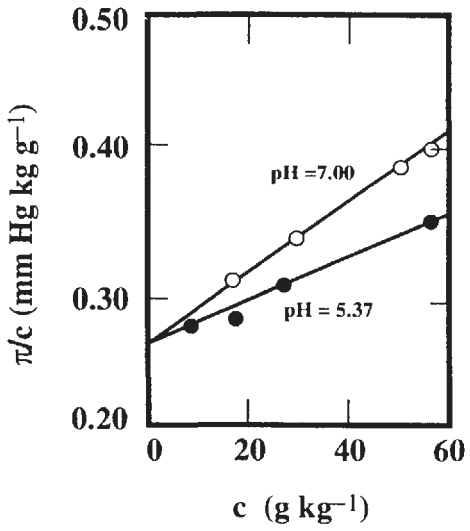


FIG. 3.6 Plot of π/c versus concentration for bovine serum albumin in 0.15 M NaCl at pH 7.00 and 5.37. (Adapted from G. Scatchard, A. C. Batchelder, and A. Brown, *J. Am. Chem. Soc.*, **68**, 2320 (1946).)

Solution: First, convert to SI units, assuming the density of the solution to be 1.0 g cm^{-3}.

For the intercept

$$0.268 \text{ torr (g kg}^{-1})^{-1} \cdot \frac{1.01 \cdot 10^5 \text{ N m}^{-2}}{760 \text{ torr}} \cdot \left(\frac{10^{-3} \text{ m}^3}{\text{kg}}\right) \cdot \left(\frac{10^3 \text{ g}}{\text{kg}}\right) = 35.6 \text{ N m}^{-2} (\text{kg m}^{-3})^{-1}$$

For the slope

$$1.37 \cdot 10^{-3} \text{ torr kg}^2\text{g}^{-2} \cdot \frac{1.01 \cdot 10^5 \text{ N m}^{-2}}{760 \text{ torr}} \cdot \left(\frac{10^{-3} \text{ m}^3}{\text{kg}}\right)^2 \cdot \left(\frac{10^3 \text{ g}}{\text{kg}}\right)^2 = 0.812 \text{ N m}^4 \text{ kg}^{-2}$$

Calculate the molecular weight by Equation (35):

$$M = RT(\pi/c)_0^{-1} = (8.314)(298)/35.6 = 69.6 \text{ kg mole}^{-1} = 69{,}600 \text{ g mole}^{-1}$$

The mass per particle is obtained by dividing by N_A:

$$69.6/6.02 \cdot 10^{23} = 1.16 \cdot 10^{-22} \text{ kg molecule}^{-1}$$

Divide the slope by RT to get B:

$$0.182/(8.314)(298) = 7.35 \cdot 10^{-5} \text{ m}^3 \text{ kg}^{-2} \text{ mole}$$

Use Equation (63) to calculate the excluded volume:

$$u = \frac{2BM^2}{N_A} = \frac{2(7.35 \cdot 10^{-5})(69.6)^2}{6.02 \cdot 10^{23}} = 1.18 \cdot 10^{-24} \text{ m}^3 \text{ molecule}^{-1}$$

Since the actual volume of a spherical particle is one-quarter the excluded volume, the volume of the molecule is $2.95 \cdot 10^{-25}$ m^3 molecule^{-1}.

The partial specific volume is given by the ratio of the molecular volume to the molecular mass:

$$2.95 \cdot 10^{-25}/1.16 \cdot 10^{-22} = 2.54 \cdot 10^{-3} \text{ m}^3 \text{ kg}^{-1} = 2.54 \text{ cm}^3 \text{ g}^{-1}$$

Comparing this with the nonsolvated value of 0.75 cm^3 g^{-1} definitely suggests that the particle is hydrated.

Note that the volume of the spherical molecule may also be converted to a molecular radius, which equals 4.11 nm for this molecule. ■

* * *

The second set of data points in Figure 3.6—measured for the same solute but at a pH = 7.00—shows a steeper slope and therefore a larger B value. It would be incorrect to analyze this slope by the procedure used in the example above, however. The reason for this is that the protein acquires a charge in going from pH 5.37 to 7.00. The charge must be explicitly taken into account to interpret B in this case. We discuss this in Section 3.5b. It is pertinent to note, however, that charged particles require counterions to give them electrical neutrality. These counterions occupy a region in the solution that surrounds the charged colloid; we describe them as setting up an ion atmosphere around the central particle. It makes sense that the particle plus its ion atmosphere should have a larger excluded volume than the uncharged particle, which entrains no ion atmosphere.

3.4b Inclusion of Energetic Interactions: Flory-Huggins Theory

In many colloidal solutions the solute is best described as a random coil in which the domain of the molecule contains both polymer segments and solvent molecules. Vinyl-type synthetic polymers in organic solvents are especially well suited for this representation. Depending on the energetics of the interaction between solvent molecules and polymer chain segments, the solvent may be imbibed into the coil domain to a greater or lesser extent, and the spatial extension of the chain depends on these interactions, as well as on the size of the molecule itself. A solvent that swells the coil dimensions is called a *good solvent*, and one that causes the coil to shrink is called a *poor solvent*. The former situation arises when solvent-solute contacts are favored, and the latter when solute-solute contacts are favored. The "goodness" of a solvent for a particular polymer depends not only on the nature of the species involved,

but also on the temperature. Generally speaking, lowering the temperature causes a decrease in solvent goodness.

In this section we look at a statistical model for a solution that allows for both the random coil geometry of the solute molecule and the variability of interactions between the solute and different solvents. We merely sketch the general outline of the theory; in the next section we examine the implications of this model for interpreting the second virial coefficient as a measure of solvent goodness. The approach we adopt is generally known as the Flory-Huggins theory, after P. J. Flory and M. L. Huggins, whose independent efforts merged in the final result. Flory was awarded the Nobel Prize in 1974 for his numerous contributions to polymer chemistry.

3.4b.1 Entropy of Mixing

The Flory-Huggins theory begins with a model for the polymer solution that visualizes the solution as a three-dimensional lattice of N sites of equal volume. Each lattice site is able to accommodate either one solvent molecule or one polymer segment since both of these are assumed to be of equal volume. The polymer chains are assumed to be monodisperse and to consist of n segments each. Thus, if the solution contains N_1 solvent molecules and N_2 solute (polymer) molecules, the total number of lattice sites is given by

$$N = N_1 + nN_2 \tag{64}$$

Without spelling out the geometry of the lattice explicitly, we further assume that each site in the lattice is surrounded by z neighboring sites, making z the coordination number of the lattice.

On the basis of this model, an expression for ΔS_m can be derived. We will not go through the details of the derivation, but merely note the following similarities and differences between this derivation and the one that leads to Equation (58) for the excluded-volume model:

1. Instead of counting the ways to place solute molecules, this derivation counts the ways to place solute segments.
2. The placement of successive segments is subject to a restriction that did not arise in the previous derivation, namely, segments from the same chain must occupy adjacent lattice sites because they are covalently bonded together.
3. The derivation considers the number of ways ω_i of placing each of the n segments in, say, the ith solute molecule, then goes on to count the number of ways of scaling this up for N_2 solute molecules. The thermodynamic probability of the mixture Ω_{mix} results from these steps.
4. The entropy of the mixture is calculated from this by the Boltzmann entropy equation, Equation (45). By separately letting N_2 and N_1 equal zero, the configurational entropies of the solvent and the solute, respectively, are obtained from S_{mix}. Finally, by subtracting S_1 and S_2 from S_{mix}, an expression is obtained for ΔS_m.

Except for the complication of positioning connected segments in adjacent sites, the general outline of this derivation clearly parallels the previous derivation.

For a mole of solution—i.e., $N_1 + N_2 = N_A$—the theory outlined above results in the following expression for ΔS_m:

$$\frac{\Delta S_m}{R} = -\left[\frac{N_1}{N_A}\ln\left(\frac{N_1}{N}\right) + \frac{N_2}{N_A}\ln\left(\frac{nN_2}{N}\right)\right] \tag{65}$$

The ratios N_1/N_A and N_2/N_A are mole fractions of the two components since N_A is the total number of molecules in a mole of solution. On the other hand, N is the total number of lattice sites and N_1 and nN_2 are the numbers of sites occupied by solvent and solute, respectively. Since all of the lattice sites have equal volumes, the ratios in the logarithms are the volume

fractions occupied by the two components. Letting x_i be the mole fraction of component i and ϕ_i the volume fraction of that component, Equation (65) can be written

$$\Delta S_m = R[x_1 \ln \phi_1 + x_2 \ln \phi_2] \tag{66}$$

Note that $\phi_2 \to x_2$ as $n \to 1$; that is, Equation (66) reduces to the expression for the entropy of mixing for an ideal solution in the event that the solute molecule is no larger than the solvent molecule.

3.4b.2 Energetic Interactions and Enthalpy Change Due to Mixing

At this point we could proceed as in the previous derivation of an expression for B: Evaluate ΔG_m, μ_m and π by setting $\Delta Hm = 0$. Doing this, however, would sacrifice this model's ability to deal with different interaction energies between solvent and solute segments. Hence our next objective is to find an expression for ΔH_m in terms of the lattice model we have developed. Once this is done, we can combine ΔS_m and ΔH_m into an expression for ΔG_m and then proceed as above.

To do this we assign an energy of interaction w_{11} to a pair of solvent molecules and w_{22} to a pair of polymer segments. The latter arises from the intermolecular forces between segments and not from the covalent bonds between them. In the same fashion, we define w_{12} to be the energy of the solvent-segment interaction.

As noted above, a lattice site in the polymer solution has z nearest neighbors that may be occupied by either solvent molecules or solute segments. If ϕ_i is the volume fraction of the solution occupied by species i, then we assume that this fraction also applies to the z sites that adjoin the specific site under consideration. Hence any site is surrounded, on the average, by $z\phi_1$ solvent molecules and $z\phi_2$ chain segments. If the central site is occupied by a solute segment (component 2), then the intermolecular forces between this segment and its neighbors contribute $(z\phi_1 w_{12} + z\phi_2 w_{22})$ to the energy of the system. Since the lattice consists of N sites of which $\phi_2 N$ are occupied by solute segments interacting in the manner just described, we write $[(1/2)N\phi_2(z\phi_1 w_{12} + z\phi_2 w_{22})]$ for the total interaction energy of all such segments. The factor 1/2 enters the expression because each pair of interacting species is counted twice by this procedure. If the lattice is completely filled with polymer segments, then the interaction energy is simply given by $[(1/2)z\phi_2 N w_{22}]$.

Precisely the same series of steps gives the energy contribution by sites occupied by solvent molecules:

1. The contribution of interactions with its neighbors made by each solvent molecule is $(z\phi_1 w_{11} + z\phi_2 w_{12})$.
2. The number of sites occupied by solvent molecules is $\phi_1 N$.
3. The total energy contribution by all solvent molecules is $(1/2)N\phi_1(z\phi_1 w_{11} + z\phi_2 w_{12})$.
4. If the lattice is completely filled with solvent, then the interaction energy is given by $(1/2)z\phi_1 N w_{11}$.

Now consider the change in interaction energies that accompanies the mixing process 1 + 2 → mixture. We must total the interactions of all chain segments and all solvent molecules to obtain the interaction energy of the mixture and then subtract from this the interactions corresponding to pure solute and pure solvent. Assembling the required values from above, we write

$$\Delta H_m = \frac{1}{2} z N [\phi_2(\phi_2 w_{22} + \phi_1 w_{12}) + \phi_1(\phi_1 w_{11} + \phi_2 w_{12}) - \phi_1 w_{11} - \phi_2 w_{22}] \tag{67}$$

Since $\phi_1 + \phi_2 = 1$, Equation (67) can be rearranged as

$$\Delta H_m = -\frac{1}{2} z N \phi_1 \phi_2 (w_{11} + w_{22} - 2w_{12}) \tag{68}$$

In terms of pairwise interactions, the solution process can be represented

$$(1,1) + (2,2) \rightarrow 2(1,2) \tag{69}$$

and, following the usual thermodynamic notation, we can write Δw for the process as

$$\Delta w = 2w_{12} - w_{11} - w_{22} \tag{70}$$

Combining Equations (68) and (70) gives

$$\Delta H_m = \frac{1}{2} z N \phi_1 \phi_2 \Delta w = \frac{1}{2} z N_1 \phi_2 \Delta w \tag{71}$$

The forces between molecules that we measure by these w's are (usually) forces of attraction and, by convention, are represented by negative numbers. Thus when 1,2 attractions are stronger than 1,1 and 2,2 attractions, Δw and ΔH_m are both negative. Since a negative enthalpy of mixing makes a favorable contribution to a negative value for ΔG_m, this sign convention makes sense. Conversely, if the 1,1 and 2,2 attractions are stronger (more negative), then Δw and ΔH_m are positive. The case in which ΔH_m is zero is called *athermal mixing*; Equation (70) shows that this corresponds to a situation in which solute-solute, solvent-solvent, and solute-solvent interactions are all equivalent in energy. This was assumed to be the case in the excluded-volume model for solution nonideality discussed in the last section.

An additional development of Equation (70) can be made by assuming that w_{12} is the geometric mean of w_{11} and w_{22}. It makes sense that the energy of interaction between unlike molecules is somehow related to the homogeneous interactions; this manner of averaging the latter has advantages that will be evident below. By assuming $w_{12} = (w_{11}w_{22})^{1/2}$, Equation (70) becomes

$$\Delta w \propto w_{11} + w_{22} - 2(w_{11}w_{22})^{1/2} \propto (\sqrt{w_{11}} - \sqrt{w_{22}})^2 \tag{72}$$

This development cannot result in a negative value for Δw or ΔH_m and is therefore definitely inapplicable for systems in which the solute and solvent display some specific type of interaction, such as hydrogen bonding. However, when purely physical interactions are involved, Equation (72) has proved to be quite useful.

The utility of this approach lies in the fact that Equation (72) describes the mixing process in terms of homogeneous interactions, which are readily measured for pure liquids. The heat of vaporization, for example, is a liquid property that increases as the strength of intermolecular attractions increases. Rather than working with molar heats of vaporization, it is more convenient to divide the molar heat by the molar volume to define what is known as the *cohesive energy density* (CED) of a material. As the name implies, the CED measures the energy per unit volume that holds the molecules of a liquid together. As such, it is directly proportional to the w's of our discussion and can be evaluated from readily available data. Introducing the concept of cohesive energy density into Equation (72) enables us to write

$$\Delta H_m \propto \Delta w \propto \left| \sqrt{(CED)_1} - \sqrt{(CED)_2} \right|^2 \tag{73}$$

Polymers decompose before they evaporate, so it appears that the concept of CED is not applicable to these materials. However, by finding a solvent with which a particular polymer mixes athermally, we can assign to the polymer by Equation (73) the same CED as that solvent. Thus cohesive energy densities for a number of polymers, as well as low molecular weight solvents, have been determined. Table 3.1 lists some representative examples of such data.

Cohesive energy densities can be used on a limited basis to give quantitative meaning to the chemist's rule of thumb, "like dissolves like." Specifically, the more alike a solvent and a polymer are in CED, the more nearly athermal their mixing will be. The more different the two are in this property, the more endothermic the mixing process will be. Remember that Equation (72) makes no provision for exothermic mixing. In the next section we see how such information might be used. Remember that Equation (71) is not limited to endothermic

TABLE 3.1 Values of the Square Root of the Cohesive Energy Density for Some Polymers and Low Molecular Weight Solvents

Solvent	$(CED)^{1/2}$, $(cal\ cm^{-3})^{1/2}$	Polymer	$(CED)^{1/2}$, $(cal\ cm^{-3})^{1/2}$
n-Decane	6.6	Poly(tetrafluoro ethylene) (Teflon)	6.2
Cyclohexane	8.2	Polyethylene	7.7–8.2
Toluene	8.9	Polystyrene	8.5–9.1
Acetone	9.9	Poly(methyl methacrylate)	9.1–9.5
Cyclohexanol	11.4	Polypropylene	9.2–9.4
Ethanol	12.7	Poly(vinyl chloride)	9.7–9.9
Methanol	14.5	Poly(ethylene terephthalate)	10.7
Water	23.4	Poly(acrylonitrile)	12.3–12.8

Source: H. Burrell in *Polymer Handbook*, 2d ed., (J. Brandrup and E. H. Immergut, Eds.), Wiley, New York, 1975.

situations. In the next section we use the more general form of ΔH_m to consider the Flory-Huggins theory as it applies to the second virial coefficient.

3.4b.3 Chemical Potential, Osmotic Pressure, and Second Virial Coefficient

Equations (65) and (71), respectively, give ΔS_m and ΔH_m according to the Flory-Huggins theory. From these components, ΔG_m can be assembled directly, and by differentiation with respect to N_1 the Flory-Huggins expression for $(\mu_1 - \mu_1^0)$ may be obtained:

$$\mu_1 - \mu_1^0 = RT \ln \phi_1 + RT\left(1 - \frac{1}{n}\right)\phi_2 + \frac{1}{2} z\Delta w \phi_2^2 \tag{74}$$

It is conventional to let $1/2\ z\Delta w = \chi RT$, where χ is called the *Flory-Huggins interaction parameter*. Note that $1/2\ \Delta w$ is the energy change per 1,2 pair according to Equation (69); therefore, with the coordination number z absorbed, the parameter χ measures this in units of RT. Finally, Equation (21) establishes the connection between chemical potential and osmotic pressure; according to this equation,

$$-\pi \overline{V}_1 = RT\left[\ln \phi_1 + \left(1 - \frac{1}{n}\right)\phi_2 + \chi\phi_2^2\right] \tag{75}$$

By expressing all volume fractions in terms of the solute, the first term on the right-hand side of Equation (75) becomes $\ln (1 - \phi_2)$, which may be expanded (see Appendix A) as $(-\phi_2 - \phi_2^2/2)$. With this modification, Equation (75) becomes

$$\frac{\pi \overline{V}_1}{RT} = \phi_2 + \frac{1}{2}\phi_2^2 - \left(1 - \frac{1}{n}\right)\phi_2 - \chi\phi_2^2 \tag{76}$$

All that remains to be done to complete our derivation of the second virial coefficient in terms of the Flory-Huggins theory is convert volume fractions into practical concentration units. First, we can express the volume fraction of the solute in terms of partial molar volumes:

$$\phi_2 = \frac{n_2 \overline{V}_2}{n_1 \overline{V}_1 + n_2 \overline{V}_2} \approx \frac{n_2}{n_1}\frac{\overline{V}_2}{\overline{V}_1} \tag{77}$$

where the approximate form applies to dilute solutions. If we recall Equations (24) and (32), Equation (77) may be written

$$\phi_2 \approx x_2 \frac{\overline{V}_2}{\overline{V}_1} = c_2 \frac{\overline{V}_2}{M_2} \tag{78}$$

Introducing practical concentration units to Equation (78), the Flory-Huggins theory yields

$$\frac{\pi}{RTc} = \frac{\overline{V}_2}{M_2 n \overline{V}_1} - \frac{\frac{1}{2} - \chi}{\overline{V}_1}\left(\frac{\overline{V}_2}{M_2}\right)^2 c = \frac{1}{M_2} + \frac{\frac{1}{2} - \chi}{\overline{V}_1}\left(\frac{\overline{V}_2}{M_2}\right)^2 c; c = c_2 \tag{79}$$

The last simplification is possible because the solute molecule is n times larger than a solvent molecule, and the same size relationship applies to the partial molar volumes. According to Equation (79), the second virial coefficient is given by

$$B = \frac{\frac{1}{2} - \chi}{\overline{V}_1}\left(\frac{\overline{V}_2}{M}\right)^2 \tag{80}$$

Since χ measures Δw for the process described by Equation (69), we see that Equation (80) accomplishes what we set out to do, namely, relate the second virial coefficient to differences in the interaction energies between various pairs of molecules.

3.4b.4 Solvent Goodness and Theta Temperature of Polymer Solutions

One of the first things to observe about Equation (80) is the fact that it allows the second virial coefficient to be positive, negative, or zero, depending on whether χ is less than, greater than, or equal to 1/2, respectively. Figure 3.5b reveals that positive and negative slopes are both observed in plots of reduced osmotic pressure versus concentration. We have more to say about this. For now, it is enough to note that under conditions with $\chi = 1/2$, $B = 0$ and solutions that are not too concentrated behave ideally. This is an advantage from the point of view of molecular weight determination since a polymer-solvent system that meets this requirement satisfies the van't Hoff equation. This means that the molecular weight can be determined from a single solution without the need to do a series of experiments and extrapolate to infinite dilution.

The condition of $B = 0$ also marks the demarcation between good and poor solvent conditions. Positive values of the second virial coefficient characterize good solvents, and negative values characterize poor solvents. This state of affairs—which is usually called the Θ condition—is very important in polymer chemistry. We encounter it again in our discussions of viscosity (Chapter 4) and light scattering (Chapter 5). Before examining the significance of the Θ condition any further, let us first consider the two states on either side of it.

A positive B value indicates a good solvent. According to Equation (80), this is guaranteed for negative (or small positive) values of χ. Since χ is proportional to Δw and ΔH_m, a positive value for the second virial coefficient corresponds to an exothermic (or small endothermic) enthalpy of mixing. Conversely, a negative value for the second virial coefficient corresponds to a positive (or small negative) value of χ and an endothermic (or small exothermic) enthalpy of mixing. Classifying the solvent as good or poor on the basis of the slope of a plot of π/c versus c is therefore consistent with the contribution of ΔH_m to favorable mixing.

The Flory-Huggins expression for ΔG_m can be related to other quantities besides osmotic pressure. One of the things that can be done is to calculate solubility limits, and therefore miscibility diagrams, for various polymer-solvent systems. Although we shall not pursue this in detail, it is of interest to note that this approach leads to the conclusion that $\chi = 1/2$ is a critical value for this parameter; that is, the critical point on a miscibility diagram corresponds to $\chi = 1/2$ for a polymer of infinite molecular weight ($n = \infty$). What is significant about this is the following: Suppose that we could somehow adjust the "goodness" of a particular solvent-polymer system, decreasing this quality from an initially good state. At $\chi = 1/2$ the solvent would go from good to poor for a polymer of infinite molecular weight, and that particular molecular weight fraction would undergo phase separation. A polymer of somewhat lower molecular weight does not undergo phase separation until the value of χ is somewhat

larger than 1/2. The shorter the polymer chain, the more χ must exceed 1/2 to reach the threshold of miscibility. Remember that $\chi > 1/2$ corresponds to "poor" solvent conditions in terms of the second virial coefficient. The extension of these ideas to miscibility limits shows that systems of this sort are on their way to phase separation. It is only because the molecules have finite molecular weights that they remain in solution at all.

In the remarks above we considered adjusting solvent goodness as if it were an imaginary process. In fact, it can be carried out physically in two different ways for a particular polymer solution. One way is to lower the temperature of the system; another way is to dilute the initial system with a poor solvent. An application such as this is one in which the cohesive energy densities listed in Table 3.1 can be put to use. The utility of changes in solvent goodness lies in the possibility of fractionating a specimen with respect to molecular weight by such a variation. Since synthetic polymers are almost always highly polydisperse, the addition of a poor solvent to a solution of the polymer (or lowering its temperature) causes the highest molecular weight fraction to separate out of solution. The separated phase can be physically removed and the process repeated until a series of fractions is obtained. The same thing can be accomplished by temperature variations for a fixed polymer-solvent system.

Since the goodness of a polymer-solvent system can be adjusted by changing the temperature, it is desirable to recast Equation (80) in a way that shows this effect explicitly. Toward this end, we define the following identity to introduce a temperature variable into Equation (80):

$$\left(\frac{1}{2} - \chi\right) = \Psi(1 - \Theta/T) \tag{81}$$

We discuss this more fully below, but one thing to note immediately is that $T = \Theta$ describes the same state as described by $\chi = 1/2$, namely, the condition of $B = 0$. It is apparent that Θ is a temperature—variously known as the *theta temperature* or the *Flory temperature*. Introducing this parameter indicates why the $B = 0$ situation is called the Θ condition. In order to justify the equivalence of the two sides of Equation (81), consider the following steps:

1. The term 1/2 on the left-hand side of Equation (81) enters the Flory-Huggins theory as part of the series expansion of $\ln(1 - \phi_2)$ in the transition between Equations (75) and (76).
2. Thus $(1/2)R$ is an entropy contribution to the second virial coefficient.
3. Our expression for ΔS_m—on which this term is based—was derived by assuming purely random placement of polymer segments and solvent molecules. This may not be fully justified because of intermolecular forces that we did not consider in arriving at Equation (66). We take advantage of this opportunity to allow for some bias in the placement of particles on the lattice and replace $1/2\ R$ with ΔS_ψ as the contribution of entropy to the second virial coefficient.
4. We recall from the transition between Equations (74) and (75) that χRT is the enthalpy contribution to the second virial coefficient. In the present context, we designate this ΔH_ψ.
5. By multiplying the numerator and denominator of the terms by the same factors, the left-hand side of Equation (81) can be transformed as follows:

$$\frac{1}{2} - \chi = \frac{\frac{1}{2}R}{R} - \frac{\chi RT}{RT} = \frac{\Delta S_\psi}{R} - \frac{\Delta H_\psi}{RT}$$

6. If $\Delta S_\psi/R$ is factored out of the last version, we obtain

$$\frac{\Delta S_\psi}{R}\left(1 - \frac{1}{T}\frac{\Delta H_\psi/R}{\Delta S_\psi/R}\right) = \Psi\left(1 - \frac{1}{T}\frac{\Delta H_\psi/R}{\Delta S_\psi/R}\right)$$

where $\Delta S_\psi/R$ has been written Ψ.

7. The ratio of ΔH_Ψ to ΔS_Ψ has kelvin units; this quantity is called the theta temperature. Substituting Θ for this ratio gives the right-hand side of Equation (81).

These manipulations may appear to add little except for needless complication to an interpretation of the second virial coefficient for random coils. Recall, however, that Equation (81) allows the variation of solvent goodness caused by temperature changes to be described quantitatively. Thus the interaction parameter χ is used to describe how B changes when a polymer is dissolved in different solvents. By contrast, Θ is used to describe the variation in B when a given polymer-solvent system is examined at different temperatures. This has been done for the polystyrene-cyclohexane system at three different temperatures; the results are discussed in Example 3.4.

* * *

EXAMPLE 3.4 *Theta Temperature of A Polymer Solution from Second Virial Coefficient Data.* Values of the second virial coefficient along with some pertinent volumes are tabulated below for the polystyrene-cyclohexane system at three temperatures.

T(K)	$B \cdot 10^5$ (cm^3 g^{-2} mole)	$\bar{V}_1$ (cm^3 mole^{-1})	$\bar{V}_2/M_2$ (cm^3 g^{-1})
303	−4.55	109.5	0.930
313	4.45	110.9	0.935
323	9.01	112.3	0.940

Use these data to estimate Θ and Ψ for this system. Does ΔS_Ψ agree with the expected entropy contribution to the second virial coefficient?

Solution: The theta temperature is that value of T at which $B = 0$. It is apparent that B changes sign (i.e., passes through zero) about midway between 303 and 313K. Equations (80) and (81) can be combined to give

$$B = \frac{(\bar{V}_2/M_2)^2}{\bar{V}_1} \Psi \left(1 - \frac{\Theta}{T}\right)$$

which can be solved for Ψ once Θ is known. Using $\Theta = 308\text{K} = 35°\text{C}$, the following values of Ψ can be calculated using the volumes provided:

T(K)	$1 - \Theta/T$	$\bar{V}_1 B(\bar{V}_2/M_2)^{-2}$	Ψ
303	−0.017	$-5.76 \cdot 10^{-3}$	0.339
313	0.016	$5.65 \cdot 10^{-3}$	0.353
323	0.046	$1.15 \cdot 10^{-2}$	0.249

For reasons not discussed, it is correct to use the value of Ψ interpolated to Θ conditions rather than, say, average the divergent values. Thus we estimate $\Psi = 0.34$, or $\Delta S_\Psi = 0.34\ R$. Since the Flory-Huggins theory predicts a contribution of l/2 R to the second virial coefficient, it seems that an additional entropy effect—given by $-0.16\ R$—must be included in order to account for the experimental B value. The fact that this "correction" to ΔS_m is negative implies that the entropy of mixing has been slightly overestimated by assuming random placement of the constituents on the lattice. ■

* * *

We have now looked at two models for the second virial coefficient of uncharged colloidal solutes. In Section 3.5b we see that B depends on the magnitude of the particle charge for polyelectrolyte solutes.

3.5 OSMOTIC EQUILIBRIUM IN CHARGED SYSTEMS

We have had no occasion as yet in this book to note that colloidal solutes may possess an electrical charge just like their low molecular weight counterparts. Portions of Chapter 4 and Chapters 11-13 are concerned with those properties of colloids that are direct consequences of

the charge of the particles. For the present we introduce the idea of charged particles by examining the effect of the charge on the osmotic pressure of the system.

The charge on a colloidal particle may originate either from the dissociation of functional groups that are covalently bonded to the colloid or from the preferential adsorption of ions to the surface. The charge of a colloid cannot be regarded as a fixed quantity like molecular weight, but must be treated as a variable with a value that depends on the nature and concentration of other components of the system. For example, proteins are positively charged at very low pH levels and negatively charged at very high pH levels; the point of electroneutrality varies from one protein to another. In the case of proteins, it is clearly the ionization of acidic and basic functional groups attached to the polypeptide chain that is primarily responsible for the charge characteristics of the molecules.

At this point we shall not concern ourselves any further with the origin of the charge of a colloidal system; rather, our attitude is that charge is one more characteristic that must be measured and understood in order to characterize certain systems fully.

Although it is not particularly difficult to formulate the thermodynamics of charged systems in perfectly general terms, the resulting notation is cumbersome. Instead of the completely general form, therefore, we consider a very specific case. The principal features will emerge clearly from this example; other situations may be readily derived by parallel arguments. The system we are concerned with consists of three components: Component 1 is the solvent, usually water; component 2 is the colloidal electrolyte; and component 3 is a low molecular weight uni-univalent electrolyte MX. We (arbitrarily) designate the colloidal electrolyte PX_z, consisting of a positively charged macroion having a valence number $+z$, paired with $z\,X^-$ ions. It could be the negative ion of the colloidal electrolyte that is the macroion (as in Vignette III) and the low molecular weight solute could have a different stoichiometry, but the essential features would remain the same.

In physical chemistry it is convenient to express concentrations as molalities and to use molality units to express the activity of the components. This is the convention we follow in this section. Accordingly, the standard state for a component consists of a solution in which that component has an activity of 1.0 mole (kg solvent)$^{-1}$.

3.5a Donnan Equilibrium

The specific situation we wish to consider is the osmotic equilibrium that develops in an apparatus that has a semipermeable membrane impermeable to the macroion only. That is, the membrane is assumed to be permeable not only to the solvent but also to both of the ions of the low molecular weight electrolyte, but not to the colloidal ion P^{z+}. At equilibrium the low molecular weight ions will be found on both sides of the membrane, but not in equal concentrations, because of the presence of the macroions on one side of the membrane. We have already come across an example of such a situation in the vignette at the beginning of this chapter on the role of Donnan equilibrium on the so-called resting states of nerve cells.

For the purpose of our discussion, we designate the side of the membrane that contains the macroions as the α phase and the solution from which the macroions are withheld as the β phase. Equation (12) continues to describe the equilibrium condition; applying it to component 3 leads to the following:

$$\mu_{3,\alpha} = \mu_{3,\beta} \tag{82}$$

Substituting Equation (13) for the β phase and Equation (19) for the α phase, which is under an osmotic pressure π, yields

$$\mu_3^0 + RT \ln a_{3,\beta} = \mu_3^0 + RT \ln a_{3,\alpha} + \int_0^{\pi} \overline{V}_3 dp \tag{83}$$

At sufficiently low concentrations of the macroion, the osmotic pressure term will be negligible compared with $RT \ln a_{3,\alpha}$, so Equation (83) becomes

$$a_{3,\alpha} = a_{3,\beta} \tag{84}$$

It may be recalled from physical chemistry that the activity of a 1 : 1 electrolyte is given by the product of the activities of the positive and negative ions of the compounds; therefore

$$(a_{M,\alpha})(a_{X,\alpha}) = (a_{M,\beta})(a_{X,\beta}) \tag{85}$$

This expression describes what is known as the *Donnan equilibrium*. It does *not* say that the activity of M^+ and/or X^- is the same on both sides of the membrane, but that the ion activity product is constant on both sides of the membrane. In the sense that an ion product is involved, the Donnan equilibrium clearly resembles all other ionic equilibria.

Remembering that $a_+ = \gamma_+ m_+$ and $\gamma_\pm^2 = \gamma_+ \gamma_-$, where $\gamma_\pm$ is the mean ionic activity coefficient (appropriate to molality units), enables us to rewrite Equation (85) as

$$(m_{M,\alpha})(m_{X,\alpha})\gamma_{\pm,\alpha}^2 = (m_{M,\beta})(m_{X,\beta})\gamma_{\pm,\beta}^2 \tag{86}$$

Of course, in the limit of infinite dilution $\gamma_\pm \to 1$. For the present we restrict our attention to sufficiently dilute solutions so that activity coefficients may be neglected and molalities may be used instead of activities. It might also be noted that in dilute solutions, for which this simplification is apt to be valid, molality and molarity are almost equal.

Another factor that we have not yet taken into account is the requirement that both sides of the membrane be electrically neutral. For the α phase, which contains the macroion, this condition is expressed by

$$zm_{P,\alpha} + m_{M,\alpha} = m_{X,\alpha} \tag{87}$$

In the β phase, which contains only low molecular weight ions, electroneutrality requires

$$m_{M,\beta} = m_{X,\beta} \tag{88}$$

The significance of the Donnan equilibrium is probably best seen as follows. Combining Equations (86) and (88) yields

$$m_{M,\beta}^2 = m_{X,\beta}^2 = (m_{M,\alpha})(m_{X,\alpha}) \tag{89}$$

Next, we use Equation (87) to substitute for either $m_{M,\alpha}$ or $m_{X,\alpha}$ in Equation (89), obtaining the following quadratic equations for $m_{M,\alpha}$ and $m_{X,\alpha}$:

$$m_{M,\alpha}^2 + zm_P m_{M,\alpha} - m_{M,\beta}^2 = 0 \tag{90}$$

and

$$m_{X,\alpha}^2 - zm_P m_{X,\alpha} - m_{X,\beta}^2 = 0 \tag{91}$$

where we have dropped the subscript α from $m_{P,\alpha}$, for convenience. These expressions permit us to evaluate the concentration of the low molecular weight ions in the compartment with the macroions, the α phase, in terms of z and the concentration of ions in the other compartment.

The situation is most easily understood by considering a numerical example. Table 3.2 lists values of $m_{M,\alpha}$ and $m_{X,\alpha}$ calculated using Equations (90) and (91). These values have been determined for two different values of $m_{M,\beta} = m_{X,\beta}$: 10^{-3} and 10^{-2}. The parameter zm_P has been selected at six evenly spaced intervals between 10^{-3} and 10^{-2}. A solution containing 1 g of colloidal electrolyte of molecular weight 10^5 per 100 g of water, for example, would have a value of $m_P = 10^{-4}$; if the macroion carries a charge of $+10$, the parameter zm_P equals 10^{-3}. It is evident from Table 3.2 that the concentration of low molecular weight positive ions is larger in the β phase than in the α phase (which contains the macroions), and that the situation is reversed for the negative ions. The requirement of electroneutrality brings this about. To show better the uneven distribution of low molecular weight ions on the two sides of the membrane, Table 3.2 also lists the ratio of the concentrations on both sides of the membrane for both the positive and the negative ions.

Two conclusions may be drawn readily from an inspection of the results of Table 3.2. First, the uneven distribution of the simple ions becomes more pronounced as the quantity zm_P increases. The low molecular weight ions are free, after all, to pass through the membrane; it is only electroneutrality that holds them back. The more macroions present or the higher

TABLE 3.2 Values of $m_{M,\alpha}$ and $m_{X,\alpha}$ and the Ratios $(m_\alpha/m_\beta)_M$ and $(m_\alpha/m_\beta)_X$ for Two Values of m_β and a Range of Values of zm_P*

	$m_{M,\beta} = m_{X,\beta} = 10^{-3}$				$m_{M,\beta} = m_{X,\beta} = 10^{-2}$			
zm_P	$m_{M,\alpha}$	$m_{X,\alpha}$	$(m_\alpha/m_\beta)_M$	$(m_\alpha/m_\beta)_X$	$m_{M,\alpha}$	$m_{X,\alpha}$	$(m_\alpha/m_\beta)_M$	$(m_\alpha/m_\beta)_X$
10^{-3}	6.18×10^{-4}	1.62×10^{-3}	0.62	1.62	9.51×10^{-3}	1.05×10^{-2}	0.95	1.05
2×10^{-3}	4.14×10^{-4}	2.41×10^{-3}	0.41	2.41	9.05×10^{-3}	1.11×10^{-2}	0.91	1.11
4×10^{-3}	2.36×10^{-4}	4.24×10^{-3}	0.24	4.24	8.20×10^{-3}	1.22×10^{-2}	0.82	1.22
6×10^{-3}	1.62×10^{-4}	6.16×10^{-3}	0.16	6.16	7.44×10^{-3}	1.34×10^{-2}	0.74	1.34
8×10^{-3}	1.23×10^{-4}	8.12×10^{-3}	0.12	8.12	6.77×10^{-3}	1.48×10^{-2}	0.67	1.48
10^{-2}	9.9×10^{-5}	1.01×10^{-2}	0.10	10.09	6.18×10^{-3}	1.62×10^{-2}	0.62	1.62

*All concentrations are in moles per kilogram of solvent. The α phase contains positive macroions.

charge they carry, the more asymmetrically the simple electrolyte will be distributed. Table 3.2 also shows that the uneven distribution of low molecular weight electrolyte becomes less pronounced as the concentration of this electrolyte is increased. We return to this point in the next section when we discuss the osmotic pressure of charged systems.

The combined effects of electroneutrality and the Donnan equilibrium permits us to evaluate the distribution of simple ions across a semipermeable membrane. If electrodes reversible to either the M^+ or the X^- ions were introduced to both sides of the membrane, there would be no potential difference between them; the system is at equilibrium and the ion activity is the same in both compartments. However, if calomel reference electrodes are also introduced into each compartment in addition to the reversible electrodes, then a potential difference will be observed between the two reference electrodes. This potential, called the *membrane potential*, reflects the fact that the membrane must be polarized because of the macroions on one side. It might be noted that polarized membranes abound in living systems, but the polarization there is thought to be primarily due to differences in ionic mobilities for different solutes rather than the sort of mechanism that we have been discussing. We return to a more detailed discussion of the electrochemistry of colloidal systems in Chapter 11.

3.5b Osmotic Pressure of Charged Colloids

Now let us turn our attention to the osmotic pressure generated by the macroion in this system. Since we have already restricted ourselves to dilute solutions, it is adequate for our purposes to use Equation (35), making allowance for the fact that we have been expressing concentrations as molality in this section. The volume of 1 kg of solvent equals $1000\ V_1^0/M_1$, so Equation (35) becomes

$$\pi \frac{1000\ V_1^0}{M_1} = m\,R\,T \tag{92}$$

where m is the molality of the solute responsible for the osmotic pressure and V_1^0 is the molar volume of the solvent. Since there are solute molecules on both sides of the membrane, the osmotic pressure will be due to the excess solute on the side of the membrane that carries the macroion (the α phase). Therefore we may replace m in Equation (92) by

$$m = m_{P,\alpha} + m_{M,\alpha} + m_{X,\alpha} - m_{M,\beta} - m_{X,\beta} \tag{93}$$

Substituting from the electroneutrality Equations (87) and (88) transforms Equation (93) into

$$m = m_{P,\alpha} + m_{M,\alpha} + zm_{P,\alpha} + m_{M,\alpha} - 2m_{M,\beta} \tag{94}$$

This is further modified by the Donnan relationship, Equation (86), also rewritten to include electroneutrality, to give

$$m = m_{P,\alpha}(1 + z) + 2\,m_{M,\alpha} - 2\,[\mathrm{m}_{\mathrm{M},\alpha}(z\,m_{P,\alpha} + m_{M,\alpha})]^{1/2} \tag{95}$$

Combining Equations (92) and (95) yields the following for the osmotic pressure of the system:

$$\pi = \frac{M_1 RT}{1000\ V_1^0}\left|(1 + z)m_P + 2\,m_M - 2\,m_M\left(1 + \frac{z\,m_P}{m_M}\right)^{1/2}\right| \tag{96}$$

It should be noted that all concentrations have been expressed in terms of the molality of the solute in the compartment that contains the macroion, so the subscript α is no longer necessary.

Although Equation (96) is rather awkward as written, several highly informative variations of it are obtained by considering different limiting situations. These are summarized in Table 3.3 for the cases in which $m_M = 0$, $m_M \gg m_P$, and $m_M > m_P$. In the table, the expressions for π that follow from Equation (96) in each of these cases are written both in terms of the molality of the macroion m_P and with the concentration of the macroion expressed in weight per unit volume c_2, specifically grams of colloid per liter of solution.

TABLE 3.3 Special Cases of Equation (96) for the Osmotic Pressure of a Charged Colloid in the Presence of a Low Molecular Weight Salt

	m_P (mole kg^{-1})	c_2 (g liter^{-1})
$m_M = 0$	$\pi = \dfrac{M_1 RT}{1000 V_1^0}(1 + z)m_P$	$\pi = \dfrac{RT(1 + z)c_2}{M_2}$
$m_M \gg m_P$	$\pi = \dfrac{M_1 RT}{1000 V_1} m_P$	$\pi = \dfrac{RTc_2}{M_2}$
$m_M > m_P$	$\pi = \dfrac{M_1 RT}{1000 V_1^0}\left(m_P + \dfrac{z^2}{4}\dfrac{m_P^2}{m_M}\right)$	$\pi = \dfrac{RTc_2}{M_2} + \dfrac{z^2}{4}\dfrac{1000 V_1^0 RTc_2^2}{M_1 M_2^2 m_M}$
		$\dfrac{\pi}{RTc_2} = \dfrac{1}{M_2} + \dfrac{z^2 1000 c_2}{4\rho_1 M_2^2 m_{MX}}$

We do not present the algebraic manipulations that lead to the various forms presented in Table 3.3; however, two relationships involved in generating these forms might be noted. First, the quantity $(1 + zm_P/m_M)^{1/2}$ in Equation (96) may be approximated by the binomial series expansion (see Appendix A) to give

$$\left(1 + \frac{z m_P}{m_M}\right)^{1/2} = 1 + \frac{1}{2}\frac{z m_P}{m_M} - \frac{1}{8}\left(\frac{z m_P}{m_M}\right)^2 + \cdots \tag{97}$$

When $m_M \gg m_P$, only the first two terms are retained; when $m_M > m_P$, the first three are used. Second, the relationship between molality and grams per liter units is given by the following in dilute solutions:

$$m_P = \frac{1000 V_1^0}{M_1 M_2} c_2 \tag{98}$$

All the forms presented in Table 3.3 are readily obtained from Equation (96) by incorporating Equation (97) or (98) or both into Equation (96). Now, let us consider the physical significance of the resulting special cases.

The two extreme values of m_M, $m_M = 0$ and $m_M \gg m_P$, are especially interesting. The former corresponds to the case of no added salt (since the macroion is positive) and the latter to a large excess or "swamping" amount of added salt. Now suppose an osmotic pressure experiment were conducted on two solutions of the same colloid, assumed to have a fixed charge z, with the objective of determining the molecular weight of the colloid. Further suppose that the two determinations differ from one another in the sense that one corresponds to zero added salt and the second to *swamping electrolyte* conditions. Finally, suppose the results are simply interpreted in terms of Equation (35) to yield the molecular weight of the colloid. Comparing Equation (35) with the results listed in Table 3.3 reveals that the correct molecular weight would be obtained for the charged colloid under swamping electrolyte conditions, but an apparent molecular weight less than the true weight by a factor $z + 1$ is obtained in the absence of salt.

How are we to understand this odd result? The answer is easy when we remember that osmotic pressure counts solute particles. The macroion cannot pass through the semipermeable membrane. In the absence of added salt, its counterions will not pass through the membrane either since the electroneutrality of the solution must be maintained. Therefore the equilibrium pressure is that associated with $(z + 1)$ particles. Failure to consider the presence of the counterions will lead to the interpretation of a low molecular weight for the colloid. As we already saw, the presence of increasing amounts of salt leads to a leveling off of the ion concentrations on the two sides of the membrane. The effect of the charge on the macroion is essentially "swamped out" with increasing electrolyte.

One interesting aspect of the limiting case of swamping electrolyte is the fact that the conclusion is totally independent of the specific nature of the ions. This is a partial justification for an assumption that was implicitly made at an earlier point. In writing Equations (87) and (88), we assume that the only ions present are M^+, X^-, and P^{z+}. In aqueous solutions, however, H^+ and OH^- are always present also, but have clearly been assumed to be negligible in writing Equations (87) and (88). The swamping electrolyte concentration may always be chosen to justify neglecting these contributions.

Table 3.3 also includes an approximation for the case in which the concentration of the salt exceeds that of the colloid, but not to the swamping extent, $m_M > m_P$. Comparison of that case with the result given in Equation (34) suggests that the contribution of charge to the second virial coefficient of the solution is given by

$$B = \frac{1000\, z^2}{4\, M_2^2 \rho_1 m_{MX}} \tag{99}$$

Strictly speaking, this contribution should be added to the excluded volume of the particles. The excluded volume of the particles becomes more important as the concentration of the salt increases, a conclusion that may be seen by considering two facts:

1. Charged colloids behave as if they were uncharged under swamping electrolyte conditions.
2. The electrostatic contribution to B, Equation (99), is inversely proportional to salt concentration.

This effect may be qualitatively understood as follows. In a charged system the colloid consists of the macroion and its low molecular weight counterions, which are, of course, distributed through a portion of the solution in the neighborhood of the macroion. Thus we visualize the colloidal ion as being surrounded by an ion atmosphere, the same sort of model that is invoked in the Debye-Hückel theory of electrolyte nonideality. The "excluded volume" that is required, therefore, includes both the volume of the colloidal particle and the volume of that part of the solution that contains the counterions. The precise distribution of counterions around a charged particle is the subject matter of Chapter 11 and, to a lesser extent, Chapter 12. In those chapters, we see that the extent of the domain over which the ion atmosphere extends increases as the electrolyte concentration decreases. Therefore, it is reasonable to expect the volume of the ion atmosphere to be sufficiently larger than the macroion in dilute solutions so that the volume actually excluded by the colloidal particle can be neglected.

We have already seen that the second virial coefficient may be determined experimentally from a plot of the reduced osmotic pressure versus concentration. Since all other quantities in Equation (99) are measurable, the charge of a macroion may be determined from the second virial coefficient of a solution with a known amount of salt. As an illustration of the use of Equation (99), we consider the data of Figure 3.6 in Example 3.5.

* * *

EXAMPLE 3.5 *Evaluation of Charges of Macroions from Osmotic Pressures.* In Example 3.3, we evaluated M and B for bovine serum albumin at pH = 5.37, at which the molecule is known to be uncharged. Use the data in Figure 3.6 to evaluate B and, from it, the charge of the molecule at pH = 7.00. The data in Figure 3.6 were measured in 0.15 M NaCl.

Solution: The slope of the line at pH = 7.00 is $2.28 \cdot 10^{-3}$ torr kg^2 g^{-2}; we convert to SI units as in Example 3.3:

$$2.28 \cdot 10^{-3}\ \text{torr kg}^2\ \text{g}^{-2} \cdot \frac{1.01 \cdot 10^5\ \text{N m}^{-2}}{760\ \text{torr}} \cdot \left(\frac{10^{-3}\ \text{m}^3}{\text{kg}}\right)^2 \cdot \left(\frac{10^3\ \text{g}}{\text{kg}}\right)^2 = 0.303\ \text{N m}^4\ \text{kg}^{-2}$$

Division by RT gives B:

$$0.303/(8.314)(298) = 12.23 \cdot 10^{-5}\ \text{m}^3\ \text{kg}^{-2}\ \text{mole}$$

As noted above, it is the difference between this value and the B value for the uncharged

molecule (Example 3.3) that is interpreted by Equation (99). Also recall that M was determined to be 69,600 g mole^{-1} in the example cited. Therefore

$$(12.23 - 7.35)10^{-5} = \frac{1000z^2}{4M_2^2\rho_1 m_{MX}}$$

$z^2 = 4(4.88 \cdot 10^{-5}\ \text{m}^3\ \text{kg}^{-2}\ \text{mole})\ (69.6\ \text{kg mole}^{-1})^2 \cdot (1.0\ \text{g cm}^{-3})(0.15\ \text{mole kg}^{-1})(100\ \text{cm m}^{-1})^3 \div 1000\ \text{g kg}^{-1} = 142$

and $z \approx \pm 12$.

Since the pH is higher than that at which the molecule is uncharged, the albumin must be negatively charged. Hence we identify z as -12 in this case. ■

* * *

3.6 SOME APPLICATIONS OF OSMOTIC PHENOMENA

3.6a Dialysis

Substances with particles in the colloidal size range are often obtained in a form that contains low molecular weight impurities. For example, enzymes are separated from homogenized tissue samples by extraction in a buffer solution. The enzyme preparation, therefore, is "contaminated" by the components of the buffer. Likewise, synthetic high polymers generally contain unreacted monomer, initiator, and catalyst. There are many experiments in which traces of low molecular weight impurities would have no effect, as, for example, in sedimentation. This chapter has shown clearly, however, that the presence of low molecular weight solutes may have large effects in an osmotic pressure experiment if these substances are retained by the membrane, either directly or through the electroneutrality condition.

Experiments with semipermeable membranes not only point out the need for purification, but also suggest a means by which this may be accomplished. The procedure of dialysis is one technique for removing low molecular weight salts or nonelectrolytes from a colloid. The method consists merely of enclosing the colloid to be purified in a bag made of some semipermeable material. The sealed bag is then placed in a quantity of the solvent. The membrane must be permeable to the solvent and to any low molecular weight impurities present, but impermeable to the colloid. As a result of the semipermeability of the bag, the impurities will distribute themselves through both compartments. The outer portion is replaced often or even continuously, so the low molecular weight impurities are gradually flushed away.

Since the membrane is also permeable to the solvent, the solvent simultaneously diffuses into the bag, diluting the colloid. Ample air space must be present in the bag at the beginning, otherwise it will rupture owing to the pressure developed by the solvent imbibed. There is also a danger that the porosity of the membrane will increase if the bag is stretched as a result of internal pressure buildup. Cellophane tubing is most commonly used as the membrane material. It is sold in rolls for this purpose and may be cut to length and tied at the ends to make the required bags.

Purification by dialysis is a slow process. Its rate is increased, however, by increasing the surface area of the membrane since that is where the exchange of solute between the two phases takes place. Stirring and frequent replacement of the solvent accelerate the process by maintaining the maximum gradient of concentration across the membrane. With ionic contaminants, the rate of dialysis may be enhanced by placing electrodes in the compartment surrounding the enclosed colloid and taking advantage of the migration of the ions in an electric field. This modification is known as *electrodialysis*.

Finally, it might be noted that colloids may be concentrated by a slight modification of the dialysis procedure. The liquid against which the colloid of interest is being dialyzed may itself be a concentrated colloid. With aqueous dispersions, for example, polyethylene oxide solutions may be used as the second colloid.

The second colloid is prepared at higher activity; therefore the solvent is drawn toward the more concentrated phase. This increases the concentration of the colloid of interest. Alternatively, the concentration increase may be accomplished by allowing the solvent to evaporate from the outer surface of the bag.

Just as with osmotic pressure, the membranes in dialysis must be carefully selected to be compatible with the system under study. Specifically, this amounts to impermeability with respect to the colloid(s) involved and permeability with respect to low molecular weight components.

3.6b Reverse Osmosis

As we discussed in Section 3.2, samples of solution and solvent separated by a semipermeable membrane will be at equilibrium only when the solution is at a greater pressure than the solvent. This is the osmotic pressure. If the solution is under less pressure than the equilibrium osmotic pressure, solvent will flow from the pure phase into the solution. If, on the other hand, the solution is under a pressure greater than the equilibrium osmotic pressure, the pure solvent will flow in the reverse direction, from the solution to the solvent phase. In the last case, the semipermeable membrane functions like a filter that separates solvent from solute molecules. In fact, the process is referred to in the literature by the terms *hyperfiltration* and *ultrafiltration*, as well as *reverse osmosis* (Sourirajan 1970); however, the last term is enjoying common use these days.

As we have discussed already, the property of membrane semipermeability applies to all sorts of systems. Likewise, reverse osmosis may be applied to a wide variety of systems. An application that has attracted a great deal of interest in recent years is the production of potable water from saline water. Since no phase transitions are involved as, for example, in distillation, the method offers some prospect of economic feasibility in coastal regions.

Cellulose acetate seems to be the most thoroughly investigated of many possible membrane materials. Cellulose acetate membranes are capable of yielding 96–98% retention of NaCl, for example, and of delivering about 0.2 $cm^3\ s^{-1}\ atm^{-1}\ m^{-2}$. This amounts to about 50 $gal\ day^{-1}\ ft^{-2}$ at 100 atm. Note that in this application of osmometry, it is not the solute but the membrane that makes up the colloidal system. In this case, the solid portion of the membrane would be the continuous phase and the pores, necessarily small if the membrane is to be effective, the "dispersed" phase. According to this point of view, numerous aspects of membrane technology become part of the interests of colloid and surface chemists. Practical desalination by reverse osmosis depends on a membrane that (a) has enough contact area to process large volumes of solution and (b) is thin enough to do so rapidly, yet (c) is sturdy enough to withstand these pressures. Since some of these points work in opposition to each other, the technical problem is one of optimization.

It is the rate of separation rather than the efficiency of salt retention that is the primary practical issue in the development of reverse osmosis desalination. In addition to a variety of other factors, the rate of reverse osmotic flow depends on the excess pressure across the membrane. Therefore the problem of rapid flow is tied into the technology of developing membranes capable of withstanding high pressures. The osmotic pressure of sea water at 25°C is about 25 atm. This means that no reverse osmosis will occur until the applied pressure exceeds this value. This corresponds to a water column about 840-ft high at this temperature.

Note that a solution more concentrated than the original one also results from the reverse osmosis process. This means that the method of reverse osmosis may also be used as a method for concentrating solutions. Fruit juices and radioactive wastes, for example, have been concentrated by this method.

REVIEW QUESTIONS

1. Describe in simple terms what *osmotic pressure* is. Does a solute have to have charges for it to have an osmotic pressure?
2. What is the relation between the chemical potential of a solute in a solution and its osmotic pressure?
3. Is the presence of a semipermeable membrane separating solutions at two different concentrations needed for osmotic pressure to exist?
4. Describe some examples of phenomena or processes that are influenced by the osmotic pressure.

5. Describe an experiment that can be used to measure osmotic pressures of colloids or macromolecular solutions. Can the osmotic pressure of a colloid or a polymer solution exceed atmospheric pressures? If yes, under what conditions? If not, why?
6. Write the equation of state for the osmotic pressure of a solution that behaves ideally.
7. How does the validity of the *van't Hoff equation* depend on the structure or the nature (e.g., solid particle, flexible coil, etc.) of the solute? Why or why not?
8. How would you correct for deviations from ideality?
9. What is *reduced osmotic pressure*?
10. What is a *virial expansion*?
11. What is the physical significance of the *second virial coefficient*? Of what is it a function?
12. Is the second virial coefficient a function of density? Explain.
13. Sketch the variation of the second virial coefficient with temperature.
14. Is the second virial coefficient positive or negative, or can it be either? What is the physical significance of its sign?
15. How does one use osmotic pressure measurements to determine the molecular weight of a solute? How does polydispersity in molecular weight affect such a measurment?
16. What is the meaning of the term *excluded-volume interaction*?
17. Outline the logic used in deriving expressions for the osmotic pressure and second virial coefficient due to excluded-volume interactions?
18. Outline how the *Flory-Huggins theory* accounts for enthalpy of mixing.
19. What is the physical significance of the χ parameter in the Flory-Huggins theory? How is it related to solute/solute and solute-solvent interactions?
20. What is *theta temperature* (or, the *Flory temperature*)? What are the relative magnitudes of the excluded-volume interactions and the energetic interactions in a dilute polymer solution at its theta temperature?
21. What is *Donnan equilibrium*? Give at least three examples of applications for which the Donnan equilibrium is important.
22. How does added salt affect the osmotic pressure of a charged colloid?
23. Give a few examples of the use of semipermeable membranes in practice.

REFERENCES

General References (with Annotations)

Crow, D. R., *Principles and Applications of Electrochemistry*, 4th ed., Blackie A & P, London, 1994. (Undergraduate level. A treatment of electrochemistry; requires only basic physical chemistry as a prerequisite.)

Flory, P. J., *Principles of Polymer Chemistry*, Cornell University Press, Ithaca, NY, 1953. (Undergraduate level. A classic textbook by a pioneer in the field.)

Hiemenz, P. C., *Polymer Chemistry: The Basic Concepts*, Marcel Dekker, New York, 1984. (Undergraduate level. An introduction to polymer chemistry. Discusses many of the topics relevant to the present chapter.)

Other References

Bergethon, P. R., and Simons, E. R., *Biophysical Chemistry: Molecules to Membranes*, Springer-Verlag, New York, 1990.

Levitan, I. B., and Kaczmarek, L. K., *The Neuron: Cell and Molecular Biology*, Oxford University Press, New York, 1991.

Morawetz, H., *Macromolecules in Solution*, Wiley, New York, 1965.

Nossal, R. J., and Lecar, H., *Molecular and Cell Biophysics*, Addison-Wesley, Redwood City, CA, 1991.

Puvvada, S., and Blankschtein, D., *J. Phys. Chem.*, **93**, 7753 (1989).

Richards, E. G., *An Introduction to the Physical Properties of Large Molecules in Solution*, Cambridge University Press, Cambridge, England, 1980.

Sourirajan, S., *Reverse Osmosis*, Academic Press, New York, 1970.

Tanford, C., *Physical Chemistry of Macromolecules*, Wiley, New York, 1961.

PROBLEMS

1. Criticize or defend the following proposition: As proof that no low molecular weight fractions of polymer have passed through the membrane in an osmotic pressure experiment, the following test may be performed. A quantity of "poor" solvent is added to an aliquot taken from the solvent side of the membrane. The absence of precipitate proves that no low molecular weight polymer passed through the membrane.

2. The osmotic pressure of solutions of a fractionated, atactic poly(isopropylacrylate) solution was measured at 25°C with the following results:*

π (g cm^{-2})	1.39	2.46	4.20	6.52
$c \times 10^2$ (g cm^{-3})	0.47	0.69	1.05	1.36

   Prepare a plot of π/c versus c for these results and evaluate $(\pi/c)_0$. Calculate M and B for this system. Define what is meant by an "atactic" polymer and compare with syndiotactic and isotactic polymers. List reference(s) consulted for these definitions.

3. An important assumption made in truncating Equation (34) is that the third virial coefficient is small. It is known† that the third virial coefficient depends strongly on the second so that it approaches zero in poor solvents even faster than B does. In fact, if Γ_2 is defined as the product BM_2, it is known that Γ_3 is approximately $0.25\Gamma_2^2$ in a good solvent; that is, Equation (34) may be written with one additional term as

$$\frac{\pi}{RTc} = \frac{1}{M}(1 + \Gamma_2 c + 0.25\Gamma_2^2 c^2)$$

   Describe how this result may be used to facilitate the evaluation of M and B in the event that a plot of reduced osmotic pressure versus c still contains too much curvature to permit a meaningful straight line to be drawn.

4. Osmotic pressures for aqueous solutions of *n*-dodecylhexaoxyethylene monoether, $C_{12}H_{25}(OC_2H_4)_5OC_2H_5$, were measured at 25°C. At concentrations below 0.038 g liter^{-1} no osmotic pressure develops, indicating complete membrane permeability. Above this concentration a pressure develops, indicating the presence of impermeable species. In the following data these pressures are reported for various $c - c_0$ values, where $c_0 = 0.038$ g liter^{-1}, the threshold for an osmotic effect:‡

π (cm)	4.90	6.53	7.62	10.58
$c - c_0$ (g liter^{-1})	29.72	38.12	43.90	58.46

   Plot $\pi/(c - c_0)$ versus $c - c_0$ to evaluate the molecular weight and B for the species responsible for the osmotic pressure.

5. The data of the preceding problem may be interpreted by assuming the following model. Above 0.038 g liter^{-1}, solute molecules associate into aggregates. By comparing the M value obtained in Problem 4 with the molecular weight of the original ether, calculate the number of molecules in the aggregate according to this model. Assuming the colloidal particles are spherical, use the second virial coefficient evaluated in Problem 4 to estimate the molar volume and radius of the aggregates. Do the quantities calculated in this problem seem reasonably self-consistent?

6. The cohesive energy density of a low molecular weight liquid is given by the heat of vaporization of that liquid, expressed per unit volume. Verify the value given for the CED of acetone in Table 3.1 from the facts that $\Delta H_v = 30.2$ kJ mole^{-1} and $\rho = 0.792$ g cm^{-3}. Would it be better to use the change in internal energy on vaporization ΔU_v rather than ΔH_v as a measure of intermolecular attraction? Does it make much difference quantitatively whether ΔU_v or ΔH_v is used?

---

*Mark, J. E., Wessling, R. A., and Hughes, R. E., *J. Phys. Chem.*, **70**, 1895 (1966).

†Flory, P. J., *Principles of Polymer Chemistry*, Cornell University Press, Ithaca, NY, 1973.

‡Attwood, D., Elworthy, P. H., and Kayne, S. B., *J. Phys. Chem.*, **74**, 3529 (1970).

7. The osmotic pressure of polystyrene (PS) solutions in toluene and methylethyl ketone (MEK) was measured* at 25°C, and the results were analyzed to give B values of 4.59×10^{-4} and 1.39×10^{-4} cm^3 g^{-2} mole, respectively, for the two solutions. Use these results to criticize or defend the following proposition: According to Table 3.1, $\sqrt{\text{CED}} = 8.5\text{-}9.1$ for polystyrene. For toluene and MEK, $\sqrt{\text{CED}}$ equals 8.9 and 9.04, respectively. Since the MEK solution has a smaller B value than the toluene solution, it appears that the best value to use for the $\sqrt{\text{CED}}$ for polystyrene is at the upper end of the range of values given and close to the value of MEK. In this way the quantity $[(\sqrt{\text{CED}})_{\text{MEK}} - (\sqrt{\text{CED}})_{\text{PS}}]^2$ will be smaller than the same quantity for polystyrene in toluene. This is consistent with the order of the B values.

8. The solvent activity in a solution of polybutadiene in benzene was determined by measuring the vapor pressure p_1 of benzene over solutions containing various concentrations of polymer.† A plot of $\ln(p_1/p_1^0) - \ln \Phi_1 - (1 - 1/n)\Phi_2$ versus Φ_2^2—in which p_1^0 is the vapor pressure of the pure benzene—yields a straight line having an intercept of zero and a slope equal to 0.33. Evaluate the interaction parameter χ from this result. Is the Θ temperature above or below the experimental temperature? Explain.

9. Krigbaum and Geymer (1959)‡ measured the osmotic pressure of polystyrene in cyclohexane at several different temperatures. The following is a sample of their results for a single fraction of polymer:

T = 24°C		T = 34°C		T = 44°C	
c (g cm^{-3})	$(\pi/RTc) \times 10^6$ (mole g^{-1})	c (g cm^{-3})	$(\pi/RTc) \times 10^6$ (mole g^{-1})	c (g cm^{-3})	$(\pi/RTc) \times 10^6$ (mole g^{-1})
0.0976	8.0	0.0081	13.3	0.0959	18.6
0.182	6.0	0.0201	14.2	0.1780	28.1
0.259	8.7	0.0964	14.2	0.2550	40.0
		0.1800	18.7		
		0.2570	26.2		

Plot all of these points on the same graph as π/RTc versus c. Although considerable nonideality exists, estimate the limiting slope at each temperature. Interpretation is assisted by realizing that each set of data approaches the same intercept as $c \to 0$. What is the approximate molecular weight of the polymer? How do the values of B estimated from the limiting slopes compare with the values given for the same system over an overlapping range of temperatures in Example 3.4?

10. Solutions of bovine serum albumin in 0.15 M NaCl were studied at other pH levels in addition to those shown in Figure 3.6. The following data are examples of additional measurements:§

pH	m_P (g protein kg^{-1})	π (mm Hg)
6.19	57.71	21.48
6.64	56.17	21.40

Since the limiting value of π/m_P shown in Figure 3.6 applies to these data also, it is possible to evaluate the second virial coefficient at these pH levels. Evaluate B and z, the effective protein charge, at the pH values shown. (Note that a slight variation in NaCl concentrations at these different pH levels should be taken into account for a more accurate determination of z.)

11. The osmotic pressure of salt-free (electrodialyzed) bovine serum albumin solutions was measured at pH = 5.37 (±?).§ At this pH the net charge of the protein molecules is zero. The following data were obtained in different runs:

*Bawn, C., Freeman, R., and Kamaliddin, A., *Trans. Faraday Soc.*, **46**, 862 (1950).

†Jessup, R. S., *J. Res. Nat. Bur. Stand.*, **60**, 47 (1958).

‡Krigbaum, W. R., and Geymer, D. O., *J. Am. Chem. Soc.*, **81**, 1859 (1959).

§Scatchard, G., Batchelder, A. C., and Brown, A., *J. Am. Chem. Soc.*, **68**, 2320 (1946).

π (mm Hg)	4.26	7.44	8.14	12.97	11.31	19.62
m_P (g protein kg^{-1})	19.56	19.71	40.88	46.05	60.81	63.25

Discuss the reasons why it is so difficult to obtain meaningful osmotic pressure data in salt-free solutions. Consider specifically the reconciliation of the electrolyte-free aspect of the experiment with the accurate control of pH.

12. In reverse osmosis both solvent and solute diffuse because of gradients in their chemical potentials. For the solvent there is no gradient of chemical potential at an osmotic pressure of π; at applied pressures p greater than π, there is such a gradient that is proportional to the difference $p - \pi$. To a first approximation, the gradient of the solute chemical potential is independent of p and depends on the difference between concentrations on opposite sides of the membrane. This leads to the result that the fraction of solute retained varies as $[1 + \text{const.}/(p - \pi)]^{-1}$. Verify that the following data* for a reverse osmosis experiment with 0.1 M NaCl and a cellulose acetate membrane follow this relationship:

Applied p (atm)	10	13	20	38	51	75
Percent salt retained	63	79	88	94	95	97

(π is about 2.6 atm for 0.1 M NaCl.)

*Data of J. E. Breton, Jr., cited by H. K. Lonsdale, in *Desalination by Reverse Osmosis* (U. Merton, Ed.), MIT Press, Cambridge, MA, 1966.

4
The Rheology of Dispersions

Imagine that your Tradesman drags behind his regular and respectable vertex, a parallelogram of twelve or thirteen inches in diagonal: What are you to do with such a monster sticking fast in your house door?

From Abbott's *Flatland*

4.1 INTRODUCTION

4.1a What Is Rheology?

The way liquids flow is one of their most obvious properties. We use a variety of terms in everyday language to describe this aspect of fluidlike substances. Thus we speak of the "thickness" of cream, the "weight" of oil, and the "leveling" of paint to describe the flow behavior and properties of such materials. The science student will probably recognize that all these terms allude in one way or another to a property known as the *viscosity* of the liquid.

However, substituting a technical term does little to make this somewhat elusive subject more intelligible. Moreover, viscosity is just one indicator of the flow characteristics of materials. In order to understand fully the flow behavior of materials, it is important that we understand some of the basic physical laws or rules that materials follow when subjected to external forces so that we can formulate theories to predict the deformation and flow of matter under a given condition. The study of the *flow and deformation of materials* is known as *rheology*. Rheology, a term coined by Professor E. C. Bingham of Lafayette College in Indiana in 1929 in consultation with a colleague from the Department of Classics, comes from the Greek root *rheos* meaning *stream* (Barnes et al. 1989) and is the subject of our focus in this chapter.

4.1b Why Is Rheology Important?

Materials in a colloidal state are frequently preferred in industrial processing operations because their large surface areas per unit volume enhance chemical reactivity, adsorptive capacity, heat transfer rates, and so on. Therefore, one cannot overlook the importance of the flow behavior and properties of colloids since they exert a significant influence on the performance, efficiency, and economy of the process. Note that some examples of this (e.g., ceramic processing, electrophoretic display devices, and food colloids) were mentioned in the vignettes presented in Chapter 1. In addition, one often uses the flow properties and behavior of the products as measures of the microstructure (or, "morphology") of the products* and as a means of quality control (e.g., printing inks, toners, paints, skin creams, blood substitutes,

*Although it is often said that "Morphology from rheology is theology!"

gels used as drug-delivery systems, etc.) and even to ensure consumer appeal and marketability of the products. Vignette IV provides an example of rheology that most people are likely to appreciate.

From a technical standpoint, it is also important to note that colloids display a *wide* range of rheological behavior. Charged dispersions (even at very low volume fractions) and sterically stabilized colloids show *elastic* behavior like solids. When the interparticle interactions are not important, they behave like ordinary liquids (i.e., they flow easily when subjected to even small shear forces); this is known as *viscous behavior*. Very often, the behavior falls somewhere between these two extremes; the dispersion is then said to be *viscoelastic*. Therefore, it becomes important to understand how the interaction forces and fluid mechanics of the dispersions affect the flow behavior of dispersions.

VIGNETTE IV RHEOLOGY IN MANUFACTURING AND QUALITY CONTROL: Rheology of Chocolate

Rheology is of major importance in processing operations in industry. The rheological behavior of process streams and dispersions determines the pumping and transportation costs, the ease of mixing operations in reactors, and the "quality" of the final product in many cases. What we do not often realize is that it is also one of the factors that determine the esthetic and sensual appeal of certain products, especially in the case of food products and cosmetics! Differences in the "feel" of a face lotion or a skin cream or the "consistency" of a ketchup can make or break the market for the product!

Take chocolate, for example. The way chocolate feels in the mouth (in addition to its taste, of course) makes a difference in its appeal to chocolate enthusiasts, and much of it has to do with the rheological changes due to the melting of the fat in the mouth. From a processing standpoint, the intricacies of the molds and coating patterns require that the chocolate be free-flowing at high shear stresses but have negligible flow at low stresses (Dickinson 1992, Chapter 3); that is, it should be *shear-thinning*—a concept we discuss in this chapter.

Why are colloids and the rheology of colloids relevant to chocolate? Melted chocolate is a complex, multiphase fluid consisting of solid nonfat particles (about 70% by volume, mostly sugar granules and the rest from crushed cocoa bean) and cocoa butter. (Milk chocolate includes, in addition, some milk solids and milk fat. In both cases, the particle size range is broad, and the shapes are irregular.) Melted chocolate is a non-Newtonian fluid that is shear-thinning and has a yield stress (see Chapter 4, Section 4.1c). The viscosity could be adjusted by adding more cocoa butter, but that is an expensive proposition. A much less expensive and quite effective way is to add lecithin or a suitable emulsifier made of small molecules. The rheology of melted chocolate is very sensitive to lecithin, which first reduces the viscosity and yield stress but increases them at larger concentrations. The optimum amount needed is less than one-tenth of the amount of cocoa butter that would be needed for similar results. The current theory is that lecithin adsorbs onto the sugar particles and prevents them from forming agglomerates, thereby influencing the rheological behavior. This is the domain of the colloid scientist, who is more likely to excel in the needed task if also a chocolate enthusiast! (The monograph of Dickinson cited in the last paragraph is a very good source of colloidal issues of importance in food science.)

Of course, the processing engineer or the colloid chemist seldom worries about the fundamental aspects of the microstructure of the product and its relation to rheology, but many of the topics we discuss in this chapter are useful for gaining the conceptual basis necessary to deal with the practical aspects of dispersion rheology.

4.1c Focus of This Chapter

Our objective in this chapter is to discuss the flow behavior of fluids and dispersions and introduce some of the basic ideas and theories that are essential for understanding the structure

and flow properties of colloids. We particularly focus on the property "viscosity" in terms of several different models and assumptions for noninteracting as well as interacting dispersions at low and high volume fractions.

1. We begin with a brief discussion of *Newton's law of viscosity* and follow this with a discussion of Newtonian flow (i.e., the flow of liquids that follow Newton's law) in a few standard configurations (e.g., cone-and-plate geometry, concentric cylinders, and capillaries) under certain specific boundary conditions. These configurations are commonly used in viscometers designed to measure viscosity of fluids.

2. Next, we take a closer look at the equation of motion for fluids, known as the *Navier-Stokes equation*. Our intention here is to discuss the general concepts that go into setting up the equation of motion rather than to present a comprehensive treatment of the Navier-Stokes equation.

3. The next topic is Einstein's theory of viscosity of dispersions of rigid, spherical particles. This theory is the starting point for most of the current approaches to flow properties of colloids and plays a practical, pedagogical, as well as historical role.

4. We examine some major conditions under which Einstein's theory breaks down. For our purpose, the two important reasons are (a) the effect of the concentration of the dispersion and (b) the effects of interparticle forces, particularly the electrostatic repulsive forces or polymer additives. This leads us next to the *non-Newtonian* behavior of dispersions.

5. Whether a substance behaves like a liquid or a solid is often a matter of the time during which the behavior is observed and the magnitudes of the applied forces. Even mountains and continents deform and move in geological time scales. Metals and rocks are rigid below a critical stress (force per unit area) known as the *yield stress*, but show significant deformation when stresses exceed this limit. What is interesting about colloids (and what makes them complicated) is that they can behave like solids or liquids in time spans and at stress levels common in daily life. It is therefore important to define some useful time scales and energy scales that relate the physicochemical properties of colloids (e.g., charges, surface potentials, thickness of adsorbed polymer layers, sizes of the particles, etc.) to the flow properties. The section on non-Newtonian behavior reviews, in addition to the basic types of non-Newtonian behavior, the time scales and dimensionless groups of importance in this context. Some sample results for viscosity of charged dispersions and coagulated dispersions are also presented to illustrate the departure of non-Newtonian systems from the simpler systems discussed. It is, however, important to remember that the contents of this chapter barely touch on the vast area of dispersion rheology, which is an area of active research currently.

6. We then conclude the chapter with a brief discussion of the viscosity of polymer solutions.

From the outset, it is helpful to be aware of the differences among the models we consider here. As an example, fluids are treated as continuous matter in discussing experimental viscometry; in contrast, we use a statistical, particulate model in discussing polymer solutions. Another example is the development of different models for lyophobic and lyophilic systems. The *Einstein theory of viscosity* is based on a model of rigid spherical particles. As such, it is most appropriate for two-phase systems or dispersions containing particles with rigid structures (although, as we also point out, they describe the viscosity of dispersions of not-so-rigid particles under certain circumstances). The *Kirkwood-Riseman theory* (discussed in Chapter 4, Section 4.9b), in contrast, applies to flexible chains and is suitable for synthetic polymers in solution. A final example is the assumption—made through the bulk of the chapter—that flow has no effect on the size, shape, or structure of the dispersed units. In Chapter 4, Section 4.8, we relax this constraint and consider the viscous behavior of coagulated or flocculated systems in which the flow may disrupt the flocs.

Both the fluid mechanics and the statistical mechanics on which some of the key theoretical results of the chapter are based are sufficiently complicated that we only sketch the highlights of these topics. We attempt to impart some physical plausibility to these theories, however, by using both force and energy perspectives in discussing the viscous resistance to flow.

4.2 NEWTON'S LAW OF VISCOSITY

The coefficient of viscosity was introduced in Chapter 2, Section 2.3a, but this parameter is elusive enough to warrant further comment. In this section we examine the definition of the coefficient of viscosity—the viscosity, for short—of a fluid. This definition leads directly to a discussion of some experimental techniques for measuring viscosity; these are discussed in the following sections.

Imagine two parallel plates of area A between which is sandwiched a liquid of viscosity η. If a force F parallel to the x direction is applied to one of these plates, it will move in the x direction as shown in Figure 4.1. Our concern is the description of the velocity of the fluid enclosed between the two plates. In order to do this, it is convenient to visualize the fluid as consisting of a set of layers stacked parallel to the boundary plates. At the boundaries, those layers in contact with the plates are assumed to possess the same velocities as the plates themselves; that is, $v = 0$ at the lower plate and equals the velocity of the moving plate at that surface. This is the *nonslip condition* that we described in Chapter 2, Section 2.3. Intervening layers have intermediate velocities. This condition is known as *laminar flow* and is limited to low velocities. At higher velocities, turbulence sets in, but we do not worry about this complication.

We can imagine within the fluid two layers separated by dy, over which distance the velocity changes by an amount dv. Therefore dv/dy defines a velocity gradient; *Newton's law of viscosity* states that the shear stress, $\tau = F/A$, is proportional to dv/dy. The viscosity η of the sandwiched fluid is the factor of proportionality:

$$\tau = \frac{F}{A} = \eta \frac{dv}{dy} = \eta \dot{\gamma} \tag{1}$$

Since the velocity v may be written as dx/dt and dx/dy defines the *shear strain* (i.e., flow deformation per unit length) acting on the sample, the velocity gradient may be written as $d(shear)/dt$ and is called the *rate of shear*, often denoted by $\dot{\gamma}$. Note that (dv/dy), or γ', has as units time^{-1}. Thus a dimensional statement of Equation (1) gives (mass length time^{-2} length^{-2}) = η time^{-1}, which shows that η has dimensions (mass length^{-1} time^{-1}) or (kg m^{-1} s^{-1}) in SI units. This is called a pascal-second (abbreviated as Pa.s). We observed in Section 2.3a that the cgs unit of viscosity, the poise (abbreviated as P), is widely used and that 10 P = 1 kg m^{-1} s^{-1} (i.e., 10 P = 1 Pa.s).

Equation (1) merely hypothesizes that the shear stress is linearly proportional to the rate of strain. Fluids that obey the form predicted by Equation (1) are said to be *Newtonian*. Figure 4.2 is a sketch of F/A versus the velocity gradient for several different modes of behavior. For a Newtonian fluid, this representation gives a straight line of zero intercept and

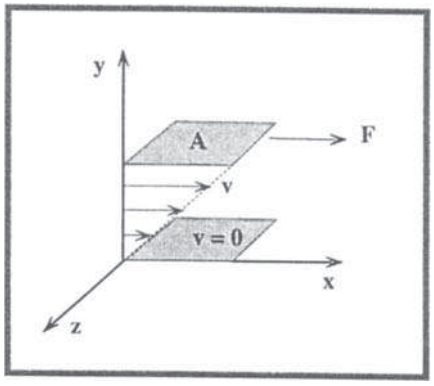


FIG. 4.1 The relationship between applied force per unit area and fluid velocity.

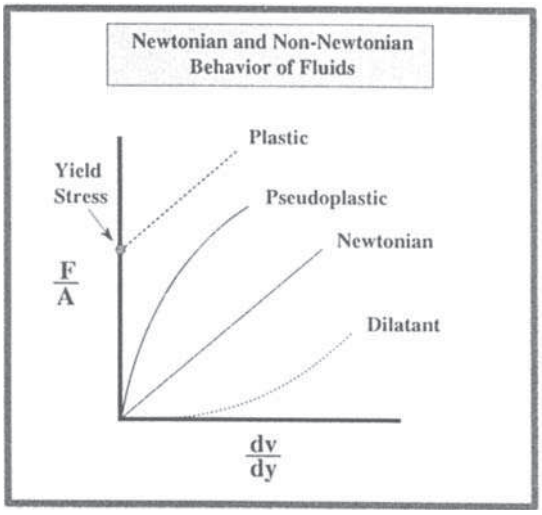


FIG. 4.2 Comparison of Newtonian liquids with several forms of non-Newtonian behavior.

slope equal to η. *Non-Newtonian* fluids generally show nonlinear plots; their "viscosity," the slope of the tangent to the curve at various points, is a function of the rate of shear. Most actual representations of experimental results display the data with the coordinates interchanged from the way they are shown in Figure 4.2. In that case, it is the cotangent of the angle that describes the slope of the line at any point that determines the true (if Newtonian) or apparent (if non-Newtonian) viscosity of the system. For the present, Newtonian behavior is our concern. We discuss some examples of non-Newtonian behavior in Section 4.8.

A second interpretation of η is as valid as Equation (1) and perhaps more illuminating. To arrive at this alternative, we multiply both sides of Equation (1) by dv/dy:

$$\frac{F}{A}\frac{dv}{dy} = \eta\left(\frac{dv}{dy}\right)^2 \tag{2}$$

Now consider the following points to interpret $(F/A)(dv/dy)$:

1. Writing the velocity v as dx/dt, one can regroup the variables on the left-hand side of Equation (2) as $d[Fdx/Ady]/dt$.
2. A force times an increment of distance Fdx equals an increment of energy dE.
3. An area times an increment of distance Ady equals an increment of volume dV.
4. Since the force under consideration measures viscous resistance to flow, the quantity dE/dV measures the energy dissipated per unit volume.
5. Dividing the change in dE/dV, i.e., $d(dE/dV)$, by dt gives the *rate of energy dissipation per unit volume*, which we denote by E_1 for convenience.

Based on these ideas, Equation (2) can be rewritten as

$$\frac{dE_1}{dt} = \eta\left(\frac{dv}{dy}\right)^2 \tag{3}$$

which shows that the volume rate of energy dissipation is proportional to the square of the velocity gradient, with the viscosity of the fluid as the factor of proportionality.

Equations (1) and (3) are equivalent as definitions of the viscosity of a fluid. To convince ourselves that η as defined by these expressions does indeed measure resistance to flow, consider two liquids of widely different viscosity, say, water and molasses:

TABLE 4.1 Viscosities of Some Familiar Materials at Room Temperature

Liquid	Approximate viscosity (Pa.s)
Glass	10^{40}
Molten glass (500°C)	10^{12}
Bitumen	10^{8}
Molten polymers	10^{3}
Golden syrup	10^{2}
Liquid honey	10^{1}
Glycerol	10^{0}
Olive oil	10^{-1}
Bicycle oil	10^{-2}
Water	10^{-3}
Air	10^{-5}

Source: Barnes et al., 1989.

1. Imagine an arrangement like that shown in Figure 4.1 and consider the force that must be applied to sustain a velocity gradient of, say, 1 s^{-1}. The higher the viscosity of the fluid is, the greater will be the force. Molasses would require more force than water.
2. Imagine the rate at which energy is dissipated in maintaining a unit velocity gradient. The higher the viscosity is, the faster will be the energy dissipation per unit volume. Molasses would dissipate energy more rapidly than water.

Table 4.1 presents typical values of viscosities of some common materials. One can see from this table that viscosity varies over several orders of magnitude. One may also note that, although the table lists a viscosity for glasses, the magnitude of the viscosity clearly suggests that the deformation of glasses at room temperatures will be extremely small so glasses are best treated as solids under normal conditions. (Recall the discussion in Section 4.1c of time scales and their relation to whether a substance is defined as a liquid or a solid.) Typical rates of shear for some familiar processes are shown in Table 4.2.

In the next section we consider an experimental approach to viscosity. We generate the apparatus of interest by wrapping—in our imagination—the fluid in Figure 4.1 into a closed ring around the z axis. The two rigid surfaces then describe concentric cylinders, and the instrument is called a *concentric-cylinder viscometer.*

4.3 CONCENTRIC-CYLINDER AND CONE-AND-PLATE VISCOMETERS

4.3a Concentric-Cylinder Viscometers

Imagine a viscous fluid enclosed in a gap between two concentric cylinders as shown in Figure 4.3a. If one of the cylinders is caused to rotate, a viscous resistance to the rotation will be transmitted through the fluid to the nonrotating cylinder and produce a torque on the nonrotating cylinder. An apparatus that is easily visualized—although not widely used—consists of a cup centered on a turntable with a bob concentrically suspended in it (Fig. 4.3a). In addition to the characteristics of the apparatus, the torque on the suspending wire depends on the viscosity of the fluid—it would be greater for molasses than for water, all other things being equal. Such a torque can be measured, and our objective is to see how the viscosity of the fluid can be evaluated from such data.

We define r to be the distance variable in the radial direction and ω to be the velocity of rotation, with ω measured in radians/second. This means the velocity of a cylindrical layer of

TABLE 4.2 Typical Shear Rates for Some Familiar Processes

Situation	Typical range of shear rates (s^{-1})	Application
Sedimentation of fine powders in a suspending liquid	10^{-6}–10^{-4}	Medicines, paints
Leveling due to surface tension	10^{-2}–10^{-1}	Paints, printing inks
Draining under gravity	10^{-1}–10^{1}	Painting and coating, toilet bleaches
Extruders	10^{0}–10^{2}	Polymers
Chewing and swallowing	10^{1}–10^{2}	Foods
Dip coating	10^{1}–10^{2}	Paints, confectionery
Mixing and stirring	10^{1}–10^{3}	Manufacturing liquids
Pipe flow	10^{0}–10^{3}	Pumping, blood flow
Spraying and brushing	10^{3}–10^{4}	Spray drying, painting, fuel atomization
Rubbing	10^{4}–10^{5}	Application of creams and lotions to the skin
Milling pigments in fluid bases	10^{3}–10^{5}	Paints, printing inks
High-speed coating	10^{5}–10^{6}	Paper
Lubrication	10^{3}–10^{7}	Gasoline engines

Source: Barnes et al., 1989.

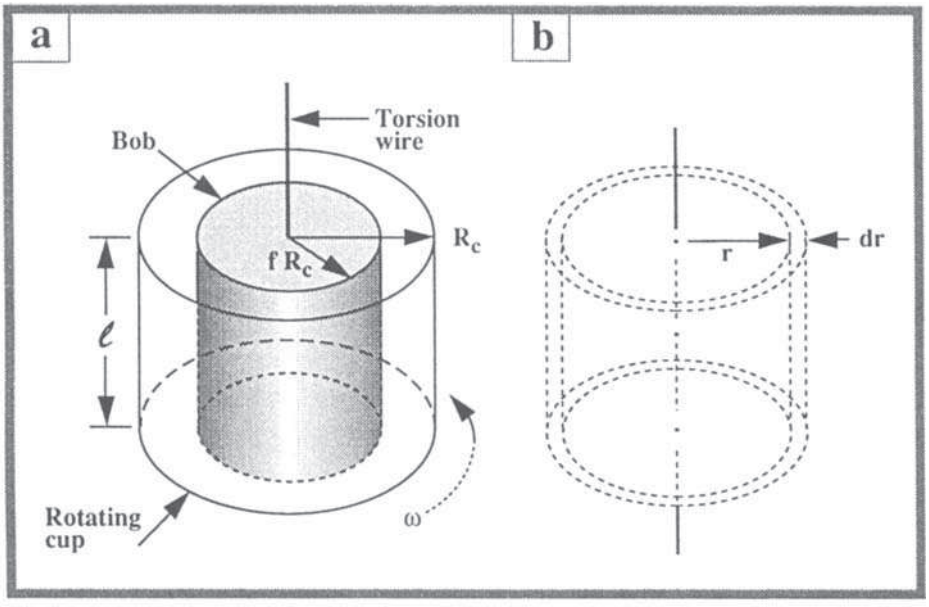


FIG. 4.3 Schematic representation of a concentric-cylinder viscometer: (a) geometry of a cup and bob; and (b) volume element within a liquid gap.

fluid a distance r from the axis of rotation is $r\omega$. There is a velocity gradient across the gap in this apparatus; this can be written $r(d\omega/dr)$. Next we represent a layer of fluid as a cylindrical shell of length ℓ and thickness dr as shown in Figure 4.3b and consider the viscous force acting on such a shell. By Equation (1), this force is given by

$$F = \eta A \frac{dv}{dr} = \eta(2\pi r \ell)\left(r \frac{d\omega}{dr}\right) \tag{4}$$

where $2\pi r\ell$ is the area of the shell. Torque is given by the product of a force and the distance through which it operates, so this element of viscous force must be multiplied by r to give a torque T:

$$T = Fr = 2\pi\eta \ell r^3 \frac{d\omega}{dr} \tag{5}$$

When the apparatus begins to rotate, the fluid experiences an initial acceleration, but a stationary state is rapidly attained in which forces balance and acceleration is zero. Equation (5) gives a generalized expression for torque; under stationary-state conditions it must be independent of r. If this were not the case, forces would be different in different parts of the fluid, and acceleration would occur. Accordingly, we set the torque on this volume element equal to a constant:

$$(2\pi \eta \ell r^3)\left(\frac{d\omega}{dr}\right) = constant \tag{6}$$

Now suppose the outer cylinder has a radius R_c and rotates with an experimental velocity ω_{ex}. Furthermore, assume the inner cylinder has a radius fR_c—where f is a fraction close to unity for small gaps—and is stationary. We can integrate Equation (6) between these limits to obtain a value for the constant torque that characterizes the experiment:

$$2\pi \eta \ell \int_0^{\omega_{ex}} d\omega = constant \int_{fR_c}^{R_c} r^{-3}\, dr = -\frac{const.}{2}\left(\frac{1}{R_c^2} - \frac{1}{(fR_c)^2}\right) \tag{7}$$

or

$$constant = 4\pi \eta \ell R_c^2 \omega_{ex} \frac{f^2}{1 - f^2} = F f R_c \tag{8}$$

where F is the force transmitted to the bob under stationary-state conditions. Since everything in this expression except η is experimentally measurable, Equation (8) can be used to evaluate viscosity. Since ω is the observed angular velocity, the subscript is no longer necessary. As noted above, an apparatus that is based on this geometrical arrangement is called a *concentric-cylinder viscometer*. Just as the geometry of the concentric-cylinder viscometer is based on Figure 4.1, which serves to define η, Equation (8) reduces to Newton's law of viscosity in the appropriate limit. This is examined in Example 4.1.

* * *

EXAMPLE 4.1 *Stress-Strain Relationship for a Concentric-Cylinder Viscometer with Small Gaps.* Examine Equation (8) in the limit $f \to 1$ to show that the relationship reduces to Equation (1) under these conditions.

Solution: The limit $f \to 1$ corresponds to a small gap between the two cylindrical walls. We are therefore justified in saying that the area of contact between the liquid and wall is $2\pi R_c\ell$; if we divide both sides of Equation (8) by this area A, we obtain

$$(F/A) = 2\eta\omega[f^2/(1 - f^2)]$$

Since $1 - f^2 = (1 + f)(1 - f) \approx 2(1 - f)$, this becomes

$$F/A = \eta\omega f^2/(1 - f)$$

Multiplying the numerator and denominator of this expression by R_c and letting $v_{max} = R_c\omega$ gives

$$(F/A) = \eta\omega_{max}f^2/[R_c(1 - f)] = \eta(v_{max} - 0)f^2/[R_c(1 - f)] = \eta f^2 \Delta v/\Delta R_c$$

since the velocity at the stationary inner wall is zero. In the limit we obtain

$$\lim_{f \to 1} (F/A) = \eta\, dv/dr$$

which is equivalent to Equation (1), the defining equation for η. When the separation between the cylinders is negligible compared to their radius of curvature, the concentric cylindrical surfaces approximate the infinite parallel plates of the model used to define the coefficient of viscosity. ■

* * *

One of the appealing features of the concentric-cylinder viscometer is the fact that the rate of shear can be varied by adjusting either the width of the gap or the angular velocity. This is a valuable capability in light of the kinds of non-Newtonian behavior that colloidal systems may display (as shown in Fig. 4.2). Equation (6), however, reminds us that this capability must be viewed with caution. That relationship shows that it is $r^3(d\omega/dr)$ that is constant across the gap, not dv/dr itself. Thus, even though we may change either the speed or the dimensions of the viscometer and thereby achieve different *average* rates of shear, the latter remains an average. The local velocity gradient varies from place to place within the fluid, although this variation is small for narrow gaps.

A variety of commercial viscometers is available that embodies the essential features of the device we have analyzed here. In most commercial instruments, it is the inner cylinder that rotates, even though there is a theoretical advantage to having the outer cylinder rotate rather than the inner one. The reason is that laminar flow is stable when the outer cylinder rotates, whereas centrifugal forces tend to induce turbulent flow in the reverse case. By rotating the inner member, however, interchangeable spindles can be used and the sturdiness and versatility of such a design offsets the theoretical advantage of the rotating cup arrangement. In actual practice, instruments may deviate from the concentric-cylinder arrangement, and their use to evaluate η depends on calibration with known standards rather than mathematical analysis of the instrument geometry.

4.3b Cone-and-Plate Viscometers

We conclude this section with a few remarks about the cone-and-plate type of viscometer, sketched schematically in Figure 4.4. In this viscometer, the fluid is placed between a stationary plate and a cone that touches the plate at its apex. This apparatus also possesses cylindrical symmetry, but this time in order to indicate a location within the fluid we must specify not only r, the distance from the axis of rotation, but also the location within the gap between the cone and the plate, as measured by θ, the angle from the vertical (see Fig. 4.4). Mathematical analysis of this apparatus leads to the result

$$\frac{dv}{dr} \approx \omega \frac{\cos\theta}{\cos\theta_0} \tag{9}$$

where θ_0 is the angle between the vertical and the wetted surface of the cone. This approximation holds for the case in which both θ and θ_0 are close to 90°. This condition is always met in actual practice, for which the angle $(90 - \theta_0)$ is generally less than 5°. Some commercial concentric-cylinder viscometers have interchangeable parts; this means that a conical rotor may be substituted for a cylindrical one. Equation (9) shows that the primary velocity gradient within the fluid is independent of radial position to a very good approximation. This feature distinguishes the cone-and-plate viscometer from the concentric-cylinder viscometer and is a significant advantage in any study that requires accurate knowledge of the rate of shear to which a sample is subjected.

To determine the viscosity of the fluid, the torque T necessary to turn the cone at an angular velocity ω is measured. This torque depends on the viscosity of the fluid according to the equation

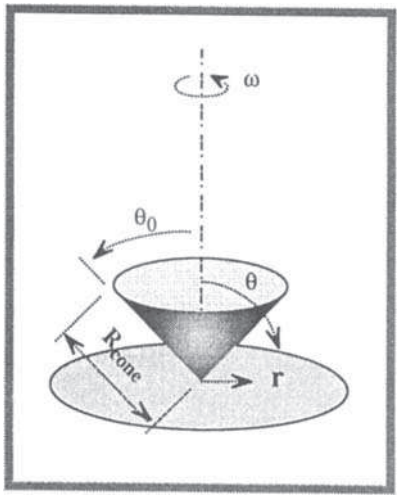


FIG. 4.4 Schematic representation of a cone-and-plate viscometer (the angle is greatly exaggerated).

$$T = \frac{\frac{4}{3}\pi R_{cone}^3 \eta \omega \sin \theta_0}{\cot \theta_0 + \frac{1}{2}\ln [(1 + \cos \theta_0)/(1 - \cos \theta_0)]\sin \theta_0} \tag{10}$$

where R_{cone} is the radius of the cone measured along the wetted surface. Like the concentric-cylinder viscometer, this apparatus can be operated, in principle, at a variety of different angular velocities, and thus η can be studied at different rates of shear. Furthermore, this last quantity is reasonably constant throughout the fluid in the apparatus.

As noted above, dilute colloidal systems display Newtonian behavior; that is, their apparent viscosity is independent of the rate of shear. Accordingly, the capability to measure η under conditions of variable shear is relatively superfluous in these systems. However, non-Newtonian behavior is commonplace in charged colloids and coagulated colloids (see Section 4.8).

Both the concentric-cylinder and the cone-and-plate viscometers are widely used, especially in cases of non-Newtonian behavior. Liquids of low molecular weight and dilute solutions in such solvents generally display Newtonian behavior. In the last case there is no advantage in being able to vary the shear rate, nor is it a disadvantage that this varies locally within the apparatus. Accordingly, a simpler viscometer is often used for such liquids. A capillary viscometer takes advantage of the fact that a velocity gradient exists whenever a fluid flows past a stationary wall. In the next section we examine the capillary viscometer and Poiseuille's law, by which it is analyzed.

4.4 THE POISEUILLE EQUATION AND CAPILLARY VISCOMETERS

4.4a Flow Through Cylinders: The Poiseuille Flow

Figure 4.5 shows a portion of a cylindrical capillary of radius R_c and length ℓ. Our interest is in the flow of a viscous liquid through such a capillary. This arrangement has the same cylindrical symmetry as the viscometers we discussed in the last section and, once again, it is convenient to focus attention on a cylindrical shell of fluid of radius r and thickness dr, as shown in Figure 4.5. Because of the nonslip condition at the wall of the capillary, the liquid shell adjacent to that wall has a velocity equal to zero. The velocity increases for shells of

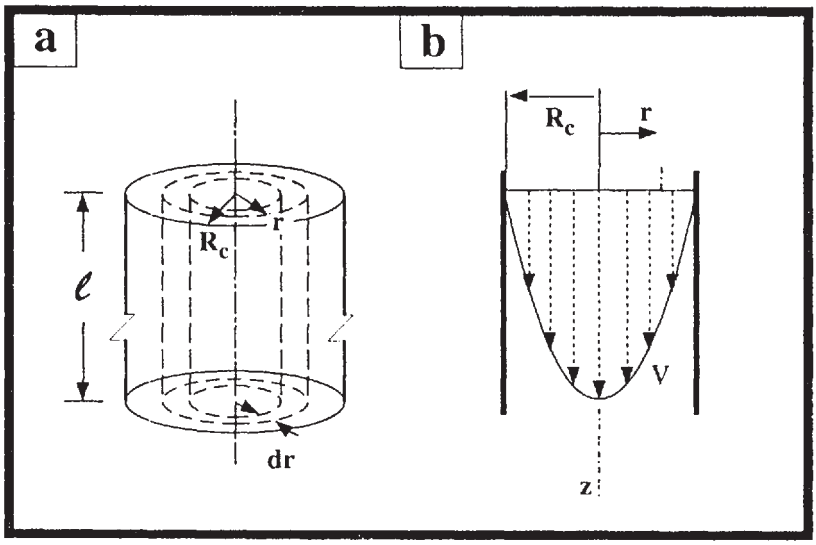


FIG. 4.5 Flow in a cylindrical capillary: (a) a volume element in the flowing liquid; and (b) the parabolic flow profile.

decreasing r so that a velocity gradient exists. Let us attempt to write an expression for v as a function of r, the radius of the cylindrical shell. Since stationary-state conditions hold within the volume element, then any nonviscous forces must exactly balance the viscous force.

The increment in viscous force acting on this element is the difference between the viscous forces on the outer and inner surfaces of the element, with each of these given by Equation (1), with $A = 2\pi r\ell$:

$$\Delta F_{vis} = (F_{vis})_{out} - (F_{vis})_{in} = 2\pi(r + dr)\eta\ell\left(\frac{dv}{dr}\right)_{r+dr} - 2\pi r\eta\ell\left(\frac{dv}{dr}\right)_r \tag{11}$$

Next, we must relate $(dv/dr)_{r+dr}$ to $(dv/dr)_r$. The following expression accomplishes this, provided dr is small:

$$\left(\frac{dv}{dv}\right)_{r+dr} = \left(\frac{dv}{dr}\right)_r + \left(\frac{d^2v}{dr^2}\right)_r dr \tag{12}$$

If this result is substituted into Equation (11), expanded, and only terms linear in dr retained, the expression becomes

$$\Delta F_{vis} = 2\pi\eta\ell\left(r\frac{d^2v}{dr^2}dr + \frac{dv}{dr}dr\right) = 2\pi\eta\ell\frac{d}{dr}\left(r\frac{dv}{dr}\right) \tag{13}$$

This force is counterbalanced by the increments in gravitational and pressure forces:

$$\Delta(F_g + F_{press}) = 2\pi\ell\rho grdr + 2\pi\Delta prdr \tag{14}$$

where the first term equals the weight of the shell and the second is the force on the shell if a pressure difference of Δp exists across the ends of the tube. Note that Δp as defined here is the difference between the pressure at the "exit" of the capillary and the one at the entrance; therefore, Δp is negative since the pressure at the entrance has to be higher for the flow to occur toward the exit. We have assumed (as evident from the figure) that the acceleration due to gravity g acts in the z direction. Setting Equations (13) and (14) equal to each other gives

$$\eta \frac{d}{dr}\left(r\frac{dv}{dr}\right) = \left(\rho g + \frac{\Delta p}{\ell}\right)r \tag{15}$$

Integration of Equation (15) yields

$$\eta r\frac{dv}{dr} = \frac{1}{2}\left(\rho g + \frac{\Delta p}{\ell}\right)r^2 \tag{16}$$

where the condition that $r(dv/dr)$ equals zero at $r = 0$ is used to evaluate the integration constant.

Equation (16) may be integrated again to get

$$\eta \int dv = \frac{1}{2}\left(\rho g + \frac{\Delta p}{\ell}\right)\int r\,dr \tag{17}$$

Using the boundary condition that $v = 0$ at $r = R_c$ to evaluate the constant of integration yields

$$v = \frac{(\rho g + \Delta p/\ell)(r^2 - R_c^2)}{4\eta} \tag{18}$$

This equation describes the velocity of a fluid element as a parabolic function of its radial distance from the center of the tube (see Fig. 4.5b).

The rate of volume flow through the tube V/t equals the summation of the flow rate through each shell, that is, the cross-sectional area of each shell multiplied by the velocity of that shell, where the latter is given by Equation (18):

$$\frac{V}{t} = \int_0^{R_c} \frac{(\rho g \ell + \Delta p)}{4\eta\ell}(r^2 - R_c^2)2\pi r\,dr \tag{19}$$

or

$$\frac{V}{t} = \frac{(\rho g \ell + \Delta p)\pi R_c^4}{8\eta\ell} \tag{20}$$

Equation (20), known as the *Poiseuille equation*, provides the basis for the most common technique for measuring the viscosity of a liquid or a dilute colloidal system, namely, the *capillary viscometer*.

4.4b Capillary Viscometers

The capillary viscometers take advantage of the Poiseuille equation for determining viscosity of fluids. Most capillary viscometers are designed with a relatively large bulb at both ends of the capillary, as shown in Figure 4.6. A constant volume in the upper bulb is designated by two lines etched at either end of the bulb. The viscometer is used by measuring the time required for the liquid level to drop from one line to the other as the fluid drains through the capillary. In such an apparatus, the difference in height of the two liquid columns is relatively constant during the time required for flow. Generally, the only pressure difference across the liquid is due to the weight of the liquid. Under these conditions, Equation (20) can be written

$$\eta = A\rho t \tag{21}$$

in which the constant A incorporates all the parameters that characterize the apparatus. Comparison of the flow times of two substances, one known (subscript 1) and one unknown (subscript 2), through the same apparatus provides an easy way to evaluate η for the unknown. In this case Equation (21) becomes

$$\eta_2 = \frac{\rho_2 t_2}{\rho_1 t_1}\eta_1 \tag{22}$$

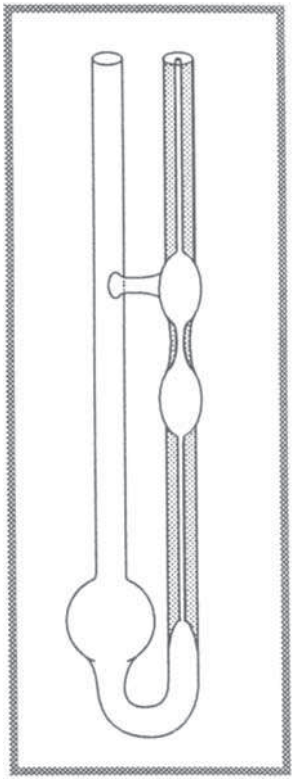


FIG. 4.6 Schematic representation of a capillary viscometer.

For greater accuracy, an additional term may be added to Equation (21) that corrects η for the fact that the Poiseuille equation does not apply exactly at the two ends of the capillary. With this correction for end effects, Equation (21) becomes

$$\eta = A\rho t - B\frac{\rho}{t} \tag{23}$$

where $B = V/8\pi$. With this correction included, both A and B can be regarded as instrument constants and evaluated by calibrating the viscometer with two liquids of known density and viscosity. Commercial capillary viscometers are generally designed to make this end correction small, often negligible.

* * *

EXAMPLE 4.2 *Comparison Between Capillary Viscometers and Concentric-Cylinder Viscometers.* Criticize or defend the following proposition: A set of capillary viscometers with different radii can be used in much the same way as a concentric-cylinder viscometer with variable speed or gap width to conduct studies in which the rate of shear is an independent variable.

Solution: Equation (16) shows that the velocity gradient is not uniform in a capillary viscometer any more than it is in a concentric-cylinder instrument. The rate of shear dv/dr is directly proportional to the radial distance from the axis of the cylinder. At the wall it has its maximum value, which is proportional to R_c; at the center of the tube it equals zero. Some intermediate value, say, the average, might be used to characterize the gradient in a given instrument. This quantity will be different for capillaries of different radii. All of this is similar to the situation in concentric-cylinder viscometers.

The two types of viscometers differ in the following way, however. A much wider range of

shear rates is attainable in concentric-cylinder instruments with adjustable features than in capillaries. Because an average velocity gradient is used in describing these experiments, it is essential that a wide range of averages be spanned; otherwise, there is too much uncertainty in the individual values for meaningful results to be obtained. Since a wider range is possible with the concentric instrument, it is preferable when variable shear rates are to be investigated. As we saw in the last section, cone-and-plate viscometers are better yet. ■

* * *

4.5 THE EQUATION OF MOTION: THE NAVIER-STOKES EQUATION

It is impossible to read much of the literature on viscosity without coming across some reference to the equation of motion. In the area of fluid mechanics, this equation occupies a place like that of the Schrodinger equation in quantum mechanics. Like its counterpart, the equation of motion is a complicated partial differential equation, the analysis of which is a matter for fluid dynamicists. Our purpose in this section is not to solve the equation of motion for any problem, but merely to introduce the physics of the relationship. Actually, both the concentric-cylinder and the capillary viscometers that we have already discussed are analyzed by the equation of motion, so we have already worked with this result without explicitly recognizing it. The equation of motion does in a general way what we did in a concrete way in the discussions above, namely, describe the velocity of a fluid element within a flowing fluid as a function of location in the fluid. The equation of motion allows this to be considered as a function of both location and time and is thus useful in nonstationary-state problems as well.

As is the case with all differential equations, the boundary conditions of the problem are an important consideration since they determine the "fit" of the solution. Many problems are set up to have a high level of symmetry and thereby simplify their boundary descriptions. This is the situation in the viscometers that we discussed above and that could be described by cylindrical symmetry. Note that the cone-and-plate viscometer—in which the angle from the axis of rotation had to be considered—is a case for which we skipped the analysis and went straight for the final result, a complicated result at that. Because it is often solved for problems with symmetrical geometry, the equation of motion is frequently encountered in cylindrical and spherical coordinates, which complicates its appearance but simplifies its solution. We base the following discussion on rectangular coordinates, which may not be particularly convenient for problems of interest but are easily visualized.

4.5a The Navier-Stokes Equation: General Considerations

Figure 4.7 shows a small rectangular volume element $\Delta x\, \Delta y\, \Delta z$ located within a flowing fluid. The fluid passing through this element has a velocity $\mathbf{v}$, which may be resolved into x, y, and z components—represented by v_x, v_y, and v_z, respectively—since, in general, $\mathbf{v}$ bears no special relationship to the coordinates. Note that $\mathbf{v}$ was in the axial direction in the cylindrical problems we discussed above (i.e., $\mathbf{v} = \mathbf{i}v_x$, where $\mathbf{i}$ is the unit vector in the x direction, and $v_y = v_z = 0$); no such restriction operates here. Now let us examine the rate at which momentum accumulates in this volume element. To begin, we recognize that the rate of change of momentum equals a force $\mathbf{F}$:

$$\frac{d(m\mathbf{v})}{dt} = m\left(\frac{d\mathbf{v}}{dt}\right) = m\mathbf{a} = \mathbf{F} \tag{24}$$

where $\mathbf{a}$ is the acceleration of the fluid element. Furthermore, the net rate of change of momentum per unit volume is given by $d(\rho\mathbf{v})/dt$. Therefore the net rate of change in momentum per unit volume equals the sum of all forces acting per unit volume. This last quantity is made up of three contributions, which include external forces (such as gravity) and pressure forces in addition to viscous forces; therefore we write

$$\frac{d(\rho\mathbf{v})}{dt} = \mathbf{F}_{ext} + \mathbf{F}_{press} + \mathbf{F}_{vis} \tag{25}$$

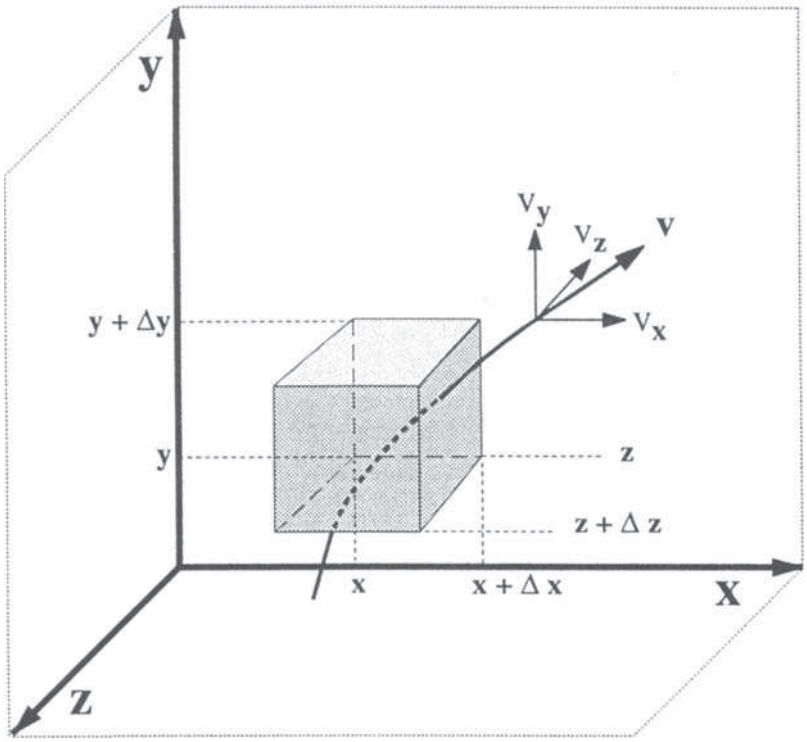


FIG. 4.7 A volume element in a liquid flowing with velocity **v**.

Treating density as a constant, the case for an incompressible fluid, enables us to write

$$\frac{d(\rho \mathbf{v})}{dt} = \rho\left(\frac{\partial \mathbf{v}}{\partial x}\frac{dx}{dt} + \frac{\partial \mathbf{v}}{\partial y}\frac{dy}{dt} + \frac{\partial \mathbf{v}}{\partial z}\frac{dz}{dt} + \frac{\partial \mathbf{v}}{\partial t}\right) \tag{26}$$

where we have used the chain rule for differentiation to expand $d\mathbf{v}/dt$ into the form shown.

The external, pressure, and viscous forces acting on the volume element can be developed more fully also. We do not go through the details of this development, but instead cite the final result followed by some explanatory remarks:

$$\rho\left(\frac{\partial \mathbf{v}}{\partial x}v_x + \frac{\partial \mathbf{v}}{\partial y}v_y + \frac{\partial \mathbf{v}}{\partial z}v_z + \frac{\partial \mathbf{v}}{\partial t}\right) = \rho\,\mathbf{g} - \mathbf{i}\frac{\partial p}{\partial x} - \mathbf{j}\frac{\partial p}{\partial y} - \mathbf{k}\frac{\partial p}{\partial z} + \eta\left(\frac{\partial^2 \mathbf{v}}{\partial x^2} + \frac{\partial^2 \mathbf{v}}{\partial y^2} + \frac{\partial^2 \mathbf{v}}{\partial z^2}\right) \tag{27}$$

where the bold **g** now indicates the *vectorial* acceleration due to gravity and **i**, **j**, and **k** are unit vectors in the x, y, and z directions, respectively. Equation (27) is the equation of motion for an *incompressible* fluid and is a vectorial equation (that is, quantities on both the left-hand side and the right-hand side are vectors and have *directions* as well as magnitudes). Remember, the solution to this differential equation for a set of specified boundary conditions gives a general expression for **v** as a function of x, y, z, and t.

Now let us briefly consider the various terms in Equation (27) with the idea of establishing their plausibility. Rewriting Equation (27) in vector notation condenses it by combining terms with similar significance; in this notation Equation (27) becomes

$$\rho(\mathbf{v}\cdot\nabla\mathbf{v}) + \rho\frac{\partial \mathbf{v}}{\partial t} = \rho\,\mathbf{g} - \nabla p + \eta\nabla^2\mathbf{v} \tag{28a}$$

where $\nabla = \mathbf{i}\partial/\partial x + \mathbf{j}\partial/\partial y + \mathbf{k}\partial/\partial z$, $\nabla^2 = \partial^2/\partial x^2 + \partial^2/\partial y^2 + \partial^2/\partial z^2$ and the term in parentheses is a dot product. Equation (28a) is the well-known *Navier-Stokes equation* that is the centerpiece of fluid dynamics. The above vectorial equation can be written individually for each direction and is more convenient for explaining in an elementary fashion the physical significance of the viscous terms:

$$x \text{ direction:} \quad \rho\left(\frac{\partial v_x}{\partial x} v_x + \frac{\partial v_x}{\partial y} v_y + \frac{\partial v_x}{\partial z} v_z + \frac{\partial v_x}{\partial t}\right) = \rho g_x - \frac{\partial p}{\partial x} + \eta\left(\frac{\partial^2 v_x}{\partial x^2} + \frac{\partial^2 v_x}{\partial y^2} + \frac{\partial^2 v_x}{\partial z^2}\right) \tag{28b}$$

$$y \text{ direction:} \quad \rho\left(\frac{\partial v_y}{\partial x} v_x + \frac{\partial v_y}{\partial y} v_y + \frac{\partial v_y}{\partial z} v_z + \frac{\partial v_y}{\partial t}\right) = \rho g_y - \frac{\partial p}{\partial y} + \eta\left(\frac{\partial^2 v_y}{\partial x^2} + \frac{\partial^2 v_y}{\partial y^2} + \frac{\partial^2 v_y}{\partial z^2}\right) \tag{28c}$$

$$z \text{ direction:} \quad \rho\left(\frac{\partial v_z}{\partial x} v_x + \frac{\partial v_z}{\partial y} v_y + \frac{\partial v_z}{\partial z} v_z + \frac{\partial v_z}{\partial t}\right) = \rho g_z - \frac{\partial p}{\partial z} + \eta\left(\frac{\partial^2 v_z}{\partial x^2} + \frac{\partial^2 v_z}{\partial y^2} + \frac{\partial^2 v_z}{\partial z^2}\right) \tag{28d}$$

Notice that if the gravity acts in the z direction only, $g_x = g_y = 0$.

4.5b Significance of the Terms and Some Simplifications

Starting on the right-hand sides of these equations, we note the following:

1. The first term on the right in each case describes the force of gravity acting on the volume element. In this development, gravity is the only external force we consider. The gravitational force per unit volume is given by $\rho\mathbf{g}$. (Notice that when the acceleration due to gravity acts in the z direction, $\mathbf{g} = \mathbf{i}0 + \mathbf{j}0 + \mathbf{k}g_z = \mathbf{k}g$, where $g_z = g$ is the magnitude of the vector $\mathbf{g}$.)
2. The second term on the right in each case describes the force on the volume element due to pressure. The net force in the x direction due to pressure is given by $(p_{x+\Delta x} - p_x)\Delta y\Delta z$. Dividing through by $\Delta x\Delta y\Delta z$ gives the x component of force per unit volume as $[(p_{x+\Delta x} - p_x)/\Delta x]$ or $\partial p/\partial x$ in the limit $\Delta x \to 0$. Similar expressions apply to the y and z components of F_{press}. (Note that here p_x stands for the pressure p at x, i.e., the subscript x denotes the position at which the pressure p is evaluated. The pressure is a scalar quantity. In contrast, $\mathbf{v}$ and $\mathbf{g}$ are vectors, and the subscripts in those cases denote the components in the appropriate directions.)
3. The third term on the right describes *viscous forces*. The viscous force acting on, say, the face of area $\Delta x\Delta y$ of the volume element in the y direction is given by $\eta\Delta x\Delta y(\partial v_y/\partial z)_{z+\Delta z}$ on one side of the volume element and by $\eta\Delta x\Delta y(\partial v_y/\partial z)_z$ on the other side. Dividing the difference between these two terms by the volume of the element $\Delta x\Delta y\Delta z$ and taking the limit $\Delta z \to 0$ gives the contribution to F_{vis} from the velocity gradient of v_y in the z direction, namely, $\eta(\partial^2 v_y/\partial z^2)$, in Equation (28c). Similar considerations apply to the other viscous terms in Equations (28b)–(28d).

Next, we turn our attention to the terms on the left-hand side of Equation (28a):

1. The first term on the left is called the *inertial* term, and the second arises from the temporal variation in the velocity at any given position. For low velocities, the former may be neglected.
2. Under conditions for which the inertial term can be neglected compared to the other terms in the equation, Equation (28a) becomes

$$\rho\frac{\partial \mathbf{v}}{\partial t} = \rho\mathbf{g} - \nabla p + \eta\nabla^2\mathbf{v}. \tag{29a}$$

This is called the *Stokes approximation* to the equation of motion. At steady-state conditions, this may be written as

$$-\rho\mathbf{g} + \nabla p = \eta\nabla^2\mathbf{v} \tag{29b}$$

This equation, with or without the body-force term $\rho\mathbf{g}$, is called the *Stokes equation*.*

3. Under stationary-state conditions, the velocity is independent of time, and Equation (28a) becomes

$$\rho\mathbf{g} - \nabla p + \eta\nabla^2\mathbf{v} = \rho(\mathbf{v}\cdot\nabla\mathbf{v}) \tag{30}$$

This is called the steady-state *Navier-Stokes equation*.

We remarked at the beginning of this section that the equation of motion is the cornerstone of any discussion of fluid dynamics. When one considers the various coordinate systems in which it may be expressed, the vector identities that may transform it, or the approximations that may be used to simplify it, Equation (28a) takes on many forms, some of which are scarcely recognizable as the same relationship. The purpose of this section is to illustrate that—despite its complexity and variations—the equation of motion is really nothing more than a statement of *Newton's second law of motion*!

Until now, we have been primarily concerned with the definition and measurement of viscosity without regard to the nature of the system under consideration. Next we turn our attention to systems containing dispersed particles with dimensions in the colloidal size range. Viscosity measurements can be used to characterize both lyophobic and lyophilic systems; we discuss both in the order cited.

4.6 EINSTEIN'S THEORY OF VISCOSITY OF DISPERSIONS

In 1906 Albert Einstein (Nobel Prize, 1921) published his first derivation of an expression for the viscosity of a dilute dispersion of solid spheres. The initial theory contained errors that were corrected in a subsequent paper that appeared in 1911. It would be no mistake to infer from the historical existence of this error that the theory is complex. Therefore we restrict our discussion to an abbreviated description of the assumptions of the theory and some highlights of the derivation. Before examining the Einstein theory, let us qualitatively consider what effect the presence of dispersed particles is expected to have on the viscosity of a fluid.

4.6a The Effect of Particles on the Viscosity of Dispersions

The presence of a colloidal size particle in the liquid increases the viscosity because of the effect it has on the flow pattern. Two effects that readily come to mind are illustrated in Figure 4.8. In Figure 4.8a, the velocity profile near a wall is shown for a pure liquid. The variation of velocity among layers is indicated by the arrows of different lengths. In Figure 4.8b, a nonrotating particle is pictured cutting across several layers in the flowing liquid. Since the particle does not rotate (by hypothesis), it must slow down the fluid so that the layers on opposite sides of the particle have the same velocity, that of the particle itself. The overall velocity gradient is thus reduced. Since the applied force is presumably the same in both Figures 4.8a and 4.8b, the reduced velocity gradient must be offset by an increase in η.

Alternatively, we might consider a particle that is induced to rotate by its position in the velocity gradient. Such a situation is shown in Figure 4.8c. In this case some of the energy that would otherwise keep the liquid flowing is deposited in the particle, causing it to rotate. In both cases the presence of the particle increases the viscosity of the fluid.

The increase in viscosity due to dispersed particles is expected to increase with the concen-

*This is different from "Stokes's law" we discussed in Equation (2.7), which results from the solution of Equation (29b) for flow over spheres (Bird et al. 1960). Equation (2.7) is for the *frictional force* on a sphere and is also known as the Stokes equation.

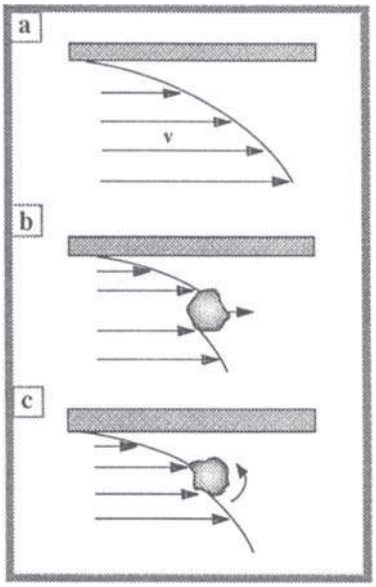


FIG. 4.8 Schematic illustration of the flow pattern near a stationary wall for a pure liquid and for a dispersion: (a) a pure liquid near a stationary wall; (b) a dispersion containing nonrotating particles; and (c) a dispersion containing rotating particles.

tration of the particles, a dependence we may tentatively describe in terms of a power series in concentration c:

$$\eta = A + Bc + Cc^2 + \ldots \tag{31}$$

In this equation A, B, C, . . . are constants with values to be determined. This much is evident: As the concentration of a dispersion goes to zero, its viscosity must go to that of the continuous phase. Therefore, $A = \eta_0$, the viscosity of the medium. Furthermore, the constants B, C, . . . might reasonably be expected to depend on the size, shape, orientation, and so on of the dispersed units.

Einstein's derivation is based on the hydrodynamic equations of the preceding section. As such, it is limited to those cases in which ρ and η for the fluid are constant and in which the flow velocity is low. Furthermore, the theory postulates an extremely dilute dispersion of rigid spheres with *no slippage of the liquid at the surface of the spheres*. Finally, the spheres are assumed to be large enough compared to the solvent molecules to permit us to regard the solvent as a continuum, but small enough compared to the dimensions of the viscometer to permit us to ignore wall effects. These size restrictions make this result applicable to particles in the colloidal size range.

Einstein considered a fluid in laminar flow through a dilute, random array of spherical particles. An obvious difficulty in applying these equations to the case in question is the enormous number of surfaces at which boundary conditions must be specified. This leads to the first reason why an infinitely dilute dispersion must be considered. If the particles are sufficiently far apart, each will modify the flow pattern in its environment as if it alone were present. This introduces two boundary conditions: no slippage at the surface and unperturbed flow at larger distances from the surface. Note the resemblance between this model and that of the Stokes law, Equation (2.7), for the drag force on a spherical particle. As in that derivation, two different expressions for the rate of energy dissipation in the dispersion are equated to give the final result. It is important to remember Equation (3) in this context since it establishes the coefficient of viscosity as the pertinent parameter in any discussion of the rate of energy dissipation per unit volume dE_v/dt.

Einstein was able to solve the equation of motion, first for the case of a single sphere present, to give dE_V/dt for a spherical volume ("cell") of dispersion of radius r centered at the spherical particle of radius R_s. The distance r is so much larger than the radius of the spherical particle that the disturbance of flow due to the particle has vanished at the surface of the hypothetical enclosing sphere. Einstein has shown the result of this integration to be

$$\left(\frac{dE_V}{dt}\right)_{Solv.\ +\ Sph.} = K\eta_0\left(\frac{8}{3}\pi r^3 + \frac{4}{3}\pi R_s^3\right) \tag{32}$$

where η_0 is the viscosity of the solvent and K is a constant with a significance that does not concern us since it cancels out. The subscripts *Solv.* and *Sph.* refer to the solvent in the cell and the sphere of radius R_s, respectively. Since $(4/3)\pi r^3$ represents the volume V of the system and $(4/3)\pi R_s^3$ represents the volume occupied by the spherical particle, we can rewrite this result as

$$\left(\frac{dE_V}{dt}\right)_{Solv.\ +\ Sph.} = K\eta_0 V(2 + \phi) \tag{33}$$

where ϕ is the volume fraction occupied by the sphere. Examining the limit of Equation (32) as $R_s \to 0$ permits us to evaluate the rate of dissipation of energy in the same volume element in the absence of the sphere; that is, if solvent alone is present,

$$\left(\frac{dE_V}{dt}\right)_{Solv.} = 2K\eta_0 V \tag{34}$$

Next, we can use Equations (33) and (34) to consider the increment per sphere in the rate of energy dissipation. Subtracting Equation (34) from Equation (33) and rearranging gives

$$\left(\frac{dE_V}{dt}\right)_{Solv.\ +\ Sph.} = \left(1 + \frac{\phi}{2}\right)\left(\frac{dE_V}{dt}\right)_{Solv.} \tag{35}$$

If the dispersion contains N spheres sufficiently far part for the effects of each to be independent of the others, we may write for the whole dispersion (subscript *Disp.*)

$$\begin{aligned}\left(\frac{dE_V}{dt}\right)_{Disp.} &= N\left(\frac{dE_V}{dt}\right)_{Solv.\ +\ Sph.} \\ &= \left(1 + \frac{\phi}{2}\right)N\left(\frac{dE_V}{dt}\right)_{Solv.} \\ &= \left(1 + \frac{\phi}{2}\right)\left(\frac{dE_V}{dt}\right)_{Total\ Solv.}\end{aligned} \tag{36}$$

Here, the subscript *Total Solv.* stands for all the solvent (of volume $V_{Total} = NV$, i.e., N times the solvent "cells" of volume V each). Of course, the volume fraction ϕ remains unchanged in Equation (36) since both the volume of spheres and the volume of solvent cells are multiplied by the amount N. Thus, one has from Equation (36)

$$\left(\frac{dE_V}{dt}\right)_{Disp.} = 2K\eta_0 V_{Total}\left(1 + \frac{\phi}{2}\right) \tag{37}$$

Einstein was able to derive another equation describing the rate of energy dissipation per unit volume for a dispersion of spheres:

$$\left(\frac{dE_V}{dt}\right)_{Disp.} = 2K\eta(1 - \phi)^2 V_{Total} \tag{38}$$

in which η (without a subscript) is the measured viscosity *of the dispersion* and K has the same significance as in Equation (37). Again in the derivation of this result, the particles are assumed to be far apart. Therefore, Equations (37) and (38), two expressions for the same quantity derived with comparable assumptions, may be equated:

$$2K\eta(1 - \phi)^2 V_{Total} = 2K\eta_0 V_{Total}\left(1 + \frac{\phi}{2}\right) \tag{39}$$

Rearranging this equation yields

$$\frac{\eta}{\eta_0} = \frac{(1 + \phi/2)}{(1 - \phi)^2} = \left(1 + \frac{\phi}{2}\right)(1 + \phi + \phi^2 + \cdots)^2 \tag{40}$$

where only the leading terms in ϕ have been retained since ϕ must be small to satisfy the requirement of large distances between spheres. If no term higher than first order in ϕ is retained, Equation (40) becomes

$$\eta/\eta_0 = 1 + 2.5\phi + \cdots \tag{41}$$

a result known as *Einstein's equation of viscosity of dispersions*. This derivation was not only an important accomplishment in its own right, but also served as a model for many subsequent derivations. Before considering these, however, let us examine Einstein's law from an experimental point of view.

4.6b Einstein's Theory: Experimental Tests

The Einstein equation is one of those pleasant surprises that occasionally emerges from complex theories: a remarkably simple relationship between variables, in this case the viscosity of a dispersion and the volume fraction of the dispersed spheres. A great many restrictive assumptions are made in the course of the derivation of this result, but the major ones are (a) that the particles are solid spheres and (b) that their concentration is small.

These conditions are relatively easy to meet experimentally, so Equation (41) has been tested in numerous studies. Figure 4.9 is an example of the sort of verification that has been

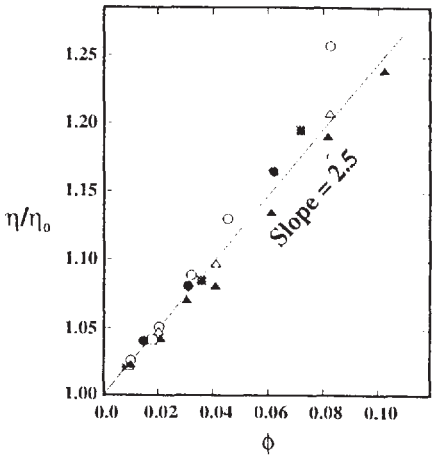


FIG. 4.9 Experimental verification of Einstein's law of viscosity for spherical particles of several different sizes (Squares are yeast particles, $R_s = 2.5$ μm; circles are fungus spores, $R_s = 4.0$ μm; triangles are glass spheres, $R_s = 80$ μm). Open symbols represent measurements in concentric-cylinder viscometers, and closed symbols represent measurements in capillary viscometers. (Data from F. Eirich, M. Bunzl, and H. Margaretha, *Kolloid Z.*, **74**, 276 (1936).)

obtained. In the work summarized in this figure, a variety of model systems were investigated using both capillary and concentric-cylinder viscometers. The solid line in the figure is the Einstein prediction; the agreement between theory and experiment is seen to be very good.

Note that the applicability of the Einstein equation seems equally good regardless of the size of the spheres used. Although the range of particle radii in Figure 4.9 is relatively narrow, this conclusion has been verified for particles as small as individual molecules and as large as grains of sand. It might also be noted that experiments of this sort are carried out in mixed solvents or electrolyte solutions with densities equal to that of the dispersed particles (i.e., in density-matched liquids) so that the spheres do not settle under the influence of gravity.

The simplicity of the Einstein equation makes it relatively easy to test, but also limits its usefulness rather sharply. With so few variables involved, the quantities we may evaluate by Equation (41) are few. Viscosity is a measurable quantity from which we try to extract information about the dispersion. All that Equation (41) offers directly in this area is the evaluation of ϕ from viscosity measurements, again provided ϕ is small and the particles are spheres.

The data in Figure 4.9 are limited to concentrations below $\phi = 0.10$. We might ask, What are the consequences of increasing the concentration to higher volume fractions of spherical particles? Equation (40) may appear to suggest that the range of applicability of Einstein's equation might be extended by retaining terms higher than first order in the power series; that is,

$$\eta/\eta_0 = 1 + 2.5\phi + k_1\phi^2 + \cdots \tag{42}$$

might give a better fit to the data from dispersions with concentrations more than infinitely dilute. There are several points about this result that should be noted:

1. Equations (31) and (42) are identical in form. Equation (42) shows that the volume fraction is the theoretically preferred unit of concentration as far as viscosity is concerned.
2. The theoretical evaluation of k_1 and the higher order coefficients in Equation (42) requires more than merely retaining additional terms in Equation (40). Einstein's derivation of Equations (32) and (38) is based on the restriction that the dispersion is very dilute, and therefore simply retaining higher order coefficients in Equation (40) will not eliminate the above restriction.
3. A number of theoretical attempts have been made to evaluate k_1 by going back through the Einstein derivation and superimposing the effects of neighboring particles rather than treating them as independent. A few of the theoretical values that have been obtained for k_1 from different models are 14.1, 12.6, and 7.35.

Rather than attempting to choose among these theoretical approaches, let us examine an empirical approach to the problem of deviations from the Einstein equation at high concentrations. Toward this end, it is convenient to rearrange Equation (42) as follows:

$$\frac{(\eta/\eta_0 - 1)}{\phi} = 2.5 + k_1\phi + \cdots \tag{43}$$

This has the effect of reducing the order of each term on the right-hand side of the equation. The term on the left-hand side of the equation is defined as the *reduced viscosity* (see Table 4.3). Now if we plot the reduced viscosity against ϕ, the result should be a straight line of slope k_1 and intercept 2.5, at least at low concentrations before still higher order terms become important.

Figure 4.10 shows the viscosity of dispersions of glass spheres of radius 65 μm plotted in the manner suggested by Equation (43). Several conclusions are evident from the data replotted in this way:

TABLE 4.3 Symbol, Common and International Union for Pure and Applied Chemistry (IUPAC) Names, and Limiting Values for a Variety of Forms Commonly Used to Present Viscosity Data

Functional form	Symbol	Common name	IUPAC name	$\lim_{c \to 0}$
—	η	Viscosity	—	η_0
η/η_0	η_r	Relative viscosity	Viscosity ratio	1
$\eta/\eta_0 - 1$	η_{sp}	Specific viscosity	—	0
$(\eta/\eta_0 - 1)/c$	η_{red}	Reduced viscosity	Viscosity number	$[\eta]$
$c^{-1} \ln(\eta/\eta_0)$	η_{inh}	Inherent viscosity	Logarithmic viscosity number	$[\eta]$
$\lim_{c \to 0} \eta_{red}$ or $\lim_{c \to 0} \eta_{inh}$	$[\eta]$	Intrinsic viscosity	Limiting viscosity number	—

1. The intercept of the curve is clearly the Einstein coefficient, 2.5, as suggested by Equation (43).
2. The initial slope of Figure 4.10 suggests k_1 is about 10.0 for these data, a reasonable value in the light of theoretical predictions.
3. At still higher volume fractions, η increases even more steeply than predicted by the two-term version of Equation (43).

The results of experiments in viscometry are routinely reported in a variety of forms; the ones we have used in Figures 4.9 and 4.10 are only two of the possibilities. Table 4.3 lists some of the functional forms in which data are often presented. Also listed are the symbols, common and IUPAC names, and the limiting values for these quantities as the concentration of the colloid goes to zero. In Table 4.3 the symbol c is used to represent the concentration of the dispersed phase. For the present, we continue to let ϕ be the unit of concentration.

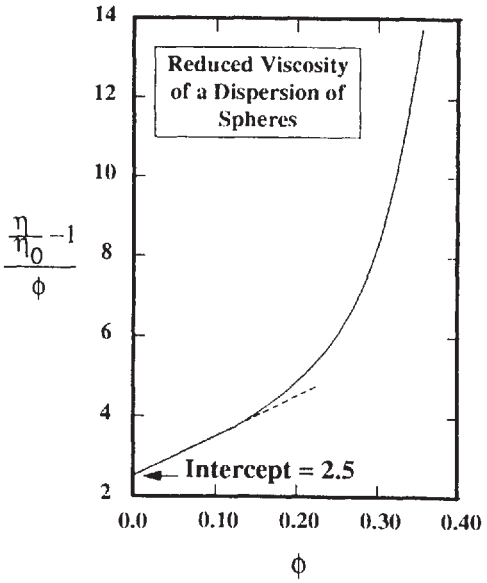


FIG. 4.10 Reduced viscosity versus volume fraction for a dispersion of glass spheres up to $\phi \approx 0.40$. ($R_s = 6.5 \cdot 10^{-3}$ cm. (Data from V. Vand, *J. Phys. Colloid Chem.*, **52**, 300 (1948).)

The definition of the *reduced viscosity* (see Table 4.3) is somewhat analogous to the reduced osmotic pressure given in Equation (3.34), and the *intrinsic viscosity* $[\eta]$ is the limiting value of this quantity. Note that the reduced viscosity gives the relative increase in the viscosity of the solution over the solvent per unit of concentration. Since $[\eta]$ equals the limiting value of the reduced viscosity, $[\eta]$ has the significance of measuring the first increment of viscosity due to the dispersed particles. The intrinsic viscosity, therefore, is a characteristic of the dispersed particles. The Einstein equation predicts that it should have a value of 2.5 for spherical particles. Note that $[\eta]$ may also be evaluated by extrapolating inherent viscosities to $\phi = 0$. Example 4.3 demonstrates the equivalence of these two procedures for evaluating $[\eta]$.

* * *

EXAMPLE 4.3 *The Limiting Behavior of the Inherent Velocity at Low Volume Fractions.* Show that the inherent viscosity $(1/\phi)\ln(\eta/\eta_0)$ reduces to $[\eta]$—defined as the *limiting reduced viscosity*—as $\phi \to 0$.

Solution: By using volume fraction for c in Equation (31), one has $\eta = A + B\phi + C\phi^2 + \ldots$. Since $A = \eta_0$, one can rewrite this as $\eta/\eta_0 = 1 + B'\phi + C'\phi^2 + \ldots$, where $B' = (B/\eta_0)$, $C' = (C/\eta_0)$, etc. The intrinsic viscosity is the coefficient of the first-order concentration term; therefore this may be written as $\eta/\eta_0 = 1 + [\eta]\phi + C'\phi^2 + \ldots$. Replacing the argument of the logarithm in the definition of inherent viscosity by this expansion gives

$$\eta_{inh} = (1/\phi)\ln(1 + [\eta]\phi + C'\phi^2 + \ldots) = (1/\phi)([\eta]\phi + C'\phi^2 + \ldots)$$

where the second form is obtained by using the series expansion (see Appendix A) for the logarithm. Further simplification leads to

$$\eta_{inh} = [\eta] + C'\phi + \ldots$$

In the limit of $\phi \to 0$, this reduces to $[\eta]$, the limiting reduced viscosity. ■

* * *

The Einstein theory shows that volume fraction is the theoretically favored concentration unit in the expansion for viscosity, even though it is not a practical unit for unknown solutes. As was the case in the Flory-Huggins theory in Chapter 3, Section 3.4b, it is convenient to convert volume fractions into "mass/volume" concentration units for the colloidal solute. According to Equation (3.78), $\phi = c(\bar{V}_2/M_2)$, where c has units "mass/volume" and $\bar{V}_2$ and M_2 are the partial molar volume and molecular weight, respectively, of the solute. In viscosity work, volumes are often expressed in deciliters—a testimonial to the convenience of the 100-ml volumetric flask! In this case, $\bar{V}_2$ must be expressed in these units also. The reader is advised to be particularly attentive to the units of concentration in an actual problem since the units of intrinsic viscosity are concentration ^{1} when the reduced viscosity is written as an expansion of powers of concentration c. (The intrinsic viscosity is *dimensionless* when the reduced viscosity is written as an expansion of powers of volume fraction ϕ.) With the substitution of Equation (3.78), Equation (42) becomes

$$\frac{\eta}{\eta_0} = 1 + 2.5\frac{\bar{V}_2}{M_2}c + k_1\left(\frac{\bar{V}_2}{M_2}\right)^2 c^2 + \cdots \tag{44}$$

or

$$\frac{1}{c}\left(\frac{\eta}{\eta_0} - 1\right) = 2.5\frac{\bar{V}_2}{M_2} + k_1\left(\frac{\bar{V}_2}{M_2}\right)^2 c + \cdots \tag{45}$$

Equation (45) shows that as long as balances, volumetric flasks, and viscometers are available, $[\eta]$ can be determined. All that is required is to measure viscosity at a series of concentrations, work up the data as $(1/c)[(\eta/\eta_0) - 1]$, and extrapolate to $c = 0$. If the experimental value of $[\eta]$ turns out to be 2.5 $(\bar{V}_2/M_2)$, then the particles are shown to be unsolvated spheres. If $[\eta]$ differs from this value, the dispersed units deviate from the requirements of the Einstein model. In the next section we examine how such deviations can be interpreted for lyophobic colloids.

4.7 DEVIATIONS FROM THE EINSTEIN MODEL

The Einstein theory is based on a model of dilute, unsolvated spheres. In this section we examine the consequences on intrinsic viscosity of deviations from the Einstein model in each of the following areas:

1. Concentrations that are not very low
2. Particle swelling due to solvation
3. Nonspherical particles approximated as ellipsoids of revolution

In addition, we introduce one other major source of deviation discussed in more detail in Section 4.8, namely, effects of interparticle interactions due to charges on the particles and polymer additives.

We consider each of these items as potential causes of deviation in the following subsections.

4.7a Effect of High Volume Fractions

We have already indicated that the coefficient k_1 in Equation (42) has been calculated for spheres by various theoretical models. While this coefficient is a measure of concentration effects, we do not pursue its derivation. Instead, we qualitatively examine the effect of particle crowding as the origin of the positive deviations from the Einstein theory that inevitably set in at higher concentrations, as seen in Figures 4.10 and 4.11.

To do this, we consider a dispersion of volume fraction ϕ and examine the increment in viscosity $d\eta$ as a small amount of particles is added to the dispersion. If we take ϕ to be small enough that the Einstein equation, Equation (41), holds, the increment $d\eta$ that accompanies the addition of particles is then given by

$$d\eta = 2.5\eta d\phi^* \tag{46}$$

where η is the viscosity of the dispersion prior to the addition of the new particles. The change in volume fraction denoted by $d\phi^*$ represents the volume of added particles divided by the volume of space available for those particles. It can be written in terms of the *total* volume fraction ϕ of all spheres in the system as $d\phi/(1 - \phi)$. Therefore, Equation (46) becomes

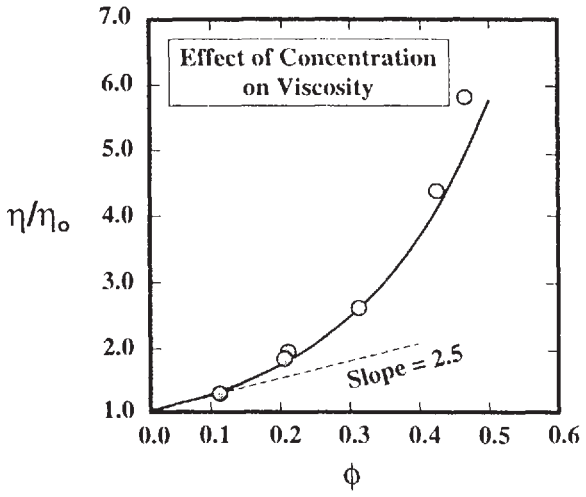


FIG. 4.11 The effect of particle crowding on viscosity. The solid line is drawn according to Equation (48); the points are experimental results. (Data from R. Roscoe, *Br. J. Appl. Phys.*, **3**, 267 (1952).)

$$d\eta = 2.5\,\eta \frac{d\phi}{1 - \phi} \tag{47}$$

One can now integrate this result, recalling that $\eta = \eta_0$ at $\phi = 0$, to give

$$\eta/\eta_0 = (1 - \phi)^{-2.5} \tag{48}$$

This result reduces to the Einstein equation as $\phi \to 0$. The solid line in Figure 4.11 was drawn according to Equation (48). This result shows that the upward curvature displayed in Figures 4.10 and 4.11 can be explained on the basis of particle crowding; unfortunately, this approach produces no new parameters to quantify the effect. It might be noted that the original derivation of this result was undertaken to explore the effects of polydispersity. Specifically, the different categories of spheres were assumed to be of different sizes. Since the Einstein result is independent of particle size, it is really irrelevant to the effect shown in Figure 4.11.

4.7a.1 The Krieger-Dougherty Equation

Equation (48) has been derived under the assumption that the volume fraction can reach unity as more and more particles are added to the dispersion. This is clearly physically impossible, and in practice one has an upper limit for ϕ, which we denote by ϕ_{max}. This limit is approximately 0.64 for random close packing and roughly 0.71 for the closest possible arrangement of spheres (face-centered cubic packing or hexagonal close packing). In this case, $d\phi^*$ in Equation (46) is replaced by $d\phi/[1 - (\phi/\phi_{max})]$. Equation (47) then becomes, when written in terms of the intrinsic viscosity $[\eta]$,

$$d\eta = [\eta]\eta \frac{d\phi}{[1 - (\phi/\phi_{max})]} \tag{49}$$

which, on integration, leads to

$$\frac{\eta}{\eta_0} = [1 - (\phi/\phi_{max})]^{-[\eta]\phi_{max}} \tag{50}$$

As noted above, we have replaced the Einstein coefficient of 2.5 by the more general intrinsic viscosity $[\eta]$ in the above equations. Equation (50) is known as the *Krieger-Dougherty equation* and has been found to be highly useful for relating viscosity data for both low shear rates ($\dot{\gamma} \to 0$) and high shear rates ($\dot{\gamma} \to \infty$). For instance, the experimental results of Krieger (1972) and de Kruif et al. (1985) lead to

$$\phi_{max} = 0.632; \qquad [\eta] = 3.13 \text{ for } \dot{\gamma} \to 0 \tag{51}$$

and

$$\phi_{max} = 0.708; \qquad [\eta] = 2.71 \text{ for } \dot{\gamma} \to \infty \tag{52}$$

Equations (51) and (52) imply an interesting feature. The respective values of ϕ_{max} suggest that at low shear the particles pack themselves randomly as the volume fraction increases (i.e., $\phi_{max} \approx 0.64$), whereas they approach ordered packing at high shear rates.

It turns out that the Krieger-Dougherty equation can also be used for intermediate shear rates with suitable modifications. Experimental data also suggest that $[\eta]$ and $\dot{\gamma}$ are independent of particle size, although they are stress dependent. Therefore viscosities of monodispersed colloids of different particle sizes can be represented by a single equation by suitably defining the variables. A discussion of these and other extensions may be found in Barnes et al. (1989).

4.7b Effects of Solvation and Shapes

Figure 2.4 in Chapter 2 suggested how either solvation or ellipticity could increase the effective size of a particle as far as its friction factor is concerned. A moment's reflection will convince us that this same conclusion is qualitatively true with respect to intrinsic viscosity also.

The solvation of a sphere swells its volume above the "dry" volume, which is presumably what is used to evaluate ϕ. Therefore whatever factor describes the increase in particle volume due to solvation is absorbed into $[\eta]$. This is easily seen as follows. Suppose the mass of colloidal solute in a solution is converted to the volume of unsolvated material using the "dry" density. In this way, mass/volume units are converted to "volume/volume" units, which we label ϕ_{dry} since the unsolvated density was used in evaluating this quantity. If the solvation is uniform throughout the particle—as would be the case for, say, aqueous proteins—then the solvated particle exceeds the unsolvated particle in volume by the factor $[1 + (m_{1,b}/m_2)(\rho_2/\rho_1)]$ as shown by Equation (2.38). Recall that in this expression, $m_{1,b}$ is the mass of bound solvent, m_2 is the mass of the solute particle, and ρ_i is the density of solvent or solute, as appropriate. Still assuming ϕ_{dry} has been used to evaluate $[\eta]$, we see

$$\frac{1}{\phi_{dry}}\left(\frac{\eta}{\eta_0} - 1\right) = 2.5\left(1 + \frac{m_{1,b}}{m_2}\left|\frac{\rho_2}{\rho_1}\right|\right) + \cdots \tag{53}$$

$$\frac{1}{c}\left(\frac{\eta}{\eta_0} - 1\right) = 2.5\left(1 + \frac{m_{1,b}}{m_2}\left|\frac{\rho_2}{\rho_1}\right|\right)\left(\frac{1}{\rho_2}\right) + \cdots \tag{54}$$

in practical units. Example 4.4 considers how intrinsic viscosity measurements can be used to reach conclusions about the state of solvation of colloidal particles.

* * *

EXAMPLE 4.4 *Extent of Hydration of a Protein Molecule from Intrinsic Viscosity Measurements.* Suppose an aqueous solution of a spherical protein molecule shows an intrinsic viscosity of 3.36 $cm^3\ g^{-1}$. Taking $\rho_2 = 1.34\ g\ cm^{-3}$ for the dry protein, estimate the extent of hydration of the protein.

Solution: It is apparent from the units of $[\eta]$ that solute concentration has been expressed in g/cm^3. Dividing this concentration by the density of the unsolvated protein converts the concentration to "dry" volume fraction units. Since the concentration appears as a reciprocal in the definition of $[\eta]$, we must multiply $[\eta]$ by ρ_2 to obtain $(1/\phi_{dry})[(\eta/\eta_0) - 1]$. For this protein the latter is given by (3.36)(1.34) = 4.50. If the particles were unsolvated, this quantity would equal 2.5 since the molecules are stated to be spherical. Hence the ratio 4.50/2.50 = 1.80 gives the volume expansion factor, which equals $[1 + (m_{1,b}/m_2)(\rho_2/\rho_1)]$. Therefore $(m_{1,b}/m_2) = 0.80$ (1.00/1.34) = 0.60. The intrinsic viscosity reveals the solvation of these particles to be 0.60 g H_2O per gram of protein. ■

* * *

We noted above that either solvation or ellipticity could cause the intrinsic viscosity to exceed the Einstein value. Simha and others have derived extensions of the Einstein equation for the case of ellipsoids of revolution. As we saw in Section 1.5a, such particles are characterized by their axial ratio. If the particles are too large, they will adopt a preferred orientation in the flowing liquid. However, if they are small enough to be swept through all orientations by Brownian motion, then they will increase $[\eta]$ more than a spherical particle of the same mass would. Again, this is very reminiscent of the situation shown in Figure 2.4.

Figure 4.12a shows plots of the intrinsic viscosity—in volume fraction units—as a function of axial ratio according to the *Simha equation*. Figure 4.12b shows some experimental results obtained for tobacco mosaic virus particles. These particles—an electron micrograph of which is shown in Figure 1.12a—can be approximated as prolate ellipsoids. Intrinsic viscosities are given by the slopes of Figure 4.12b, and the parameters on the curves are axial ratios determined by the Simha equation. Thus we see that particle asymmetry can also be quantified from intrinsic viscosity measurements for unsolvated particles.

Finally, we note that both solvation and ellipticity can occur together. The contours shown in Figure 4.13a illustrate how various combinations of solvation and ellipticity are compatible with an experimental intrinsic viscosity. The particle considered in Example 4.4 has an intrinsic viscosity of 4.50 and was calculated to be hydrated to the extent of 0.60 g H_2O per gram of protein. The same value of $[\eta]$ is also compatible with nonsolvated ellipsoids of

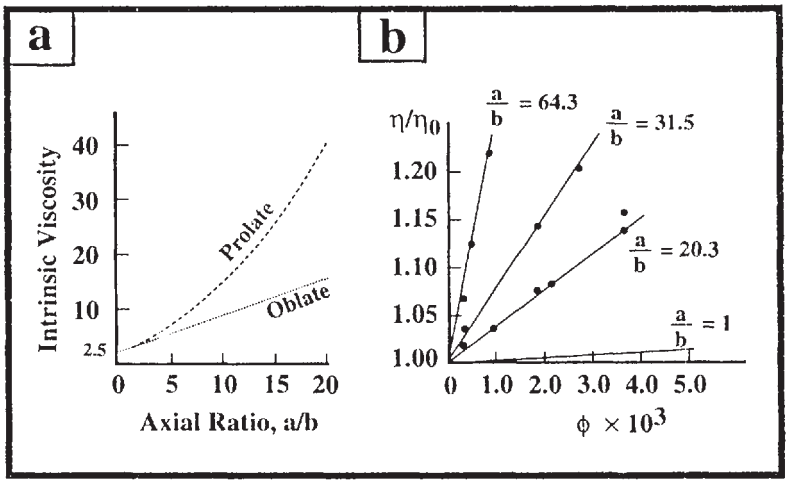


FIG. 4.12 Viscosity of dispersions of some nonspherical particles: (a) intrinsic viscosity as a function of the axial ratio a/b for oblate and prolate ellipsoids of revolution according to the Simha theory (redrawn with permission of Hiemenz 1984); (b) experimental values of relative viscosity versus volume fraction for tobacco mosaic virus particles of different a/b ratios (data from M. A. Lauffer, *J. Am. Chem. Soc.*, **66**, 1188 (1944)).

axial ratios 4.0 and 0.22. Intermediate combinations of solvation and ellipticity can also be determined from Figure 4.13a.

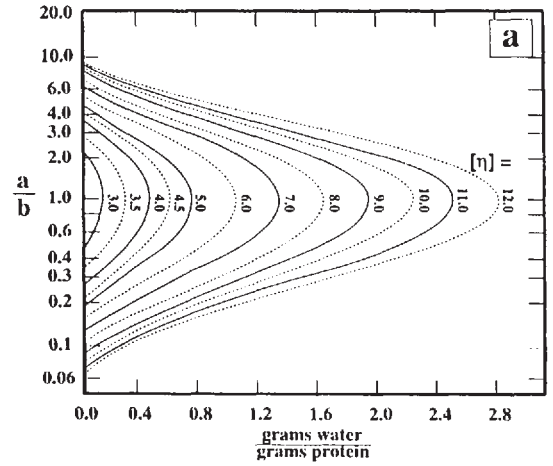
Contours like this are qualitatively the same sort of thing we obtain from sedimentation-diffusion experiments as shown in Figure 2.9. Therefore let us consider the relationship between the two types of data. In general, exactly the same factors affect both the intrinsic viscosity and the friction factor ratio, but the functional dependencies are somewhat different. Figure 4.13b shows how a contour of f/f_0 selected from a sedimentation-diffusion study and an intrinsic viscosity contour selected on the basis of viscosity experiments might overlap. In this case the solvation-ellipticity combination is characterized unambiguously: $a/b = 2.5$ and $(m_{1,b}/m_2) = 1.0$. Figure 4.13b shows the complementarity of viscosity and sedimentation-diffusion data.

The attentive reader will realize that we have strayed rather far from the hard spheres of the Einstein theory to find applications for it. It should also be appreciated, however, that the molecules we are discussing are proteins that—through disulfide bridges and hydrogen bonding—have fairly rigid structures. Therefore the application of the theory—amended to allow for solvation and ellipticity—is justified. This would not be the case for synthetic polymers, which are best described as random coils and for which a different formalism is employed. This is the topic of Section 4.9.

4.7c Electroviscous Effects and Viscoelectric Effects

In Chapter 3 we saw that electrical effects complicate the osmotic equilibrium for charged particles. Not surprisingly, particle charge complicates analysis of viscosity as well. Like osmotic phenomena, the viscosity of dispersions of charged particles is highly sensitive to the concentration of the electrolytes. Chapters 11–13 examine the behavior of simple ions near charged surfaces. For now, we review some of the concepts, variables, and definitions of those chapters as background for our discussion of charge effects on viscosity (and, more generally, rheology).

For simplicity, consider a lyophobic colloidal particle with a surface that carries a charge. Electroneutrality requires that an equal amount of opposite charges (counterions) be present



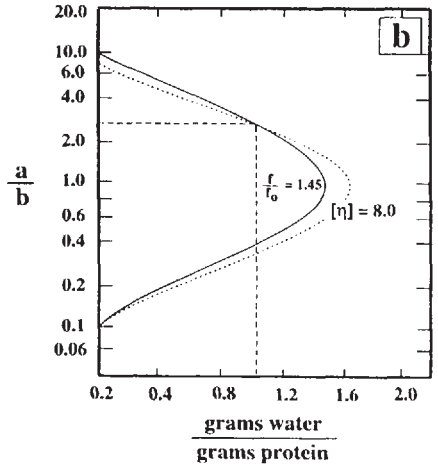


FIG. 4.13 Intrinsic viscosity of a protein solution: (a) variation of the intrinsic viscosity of aqueous protein solutions with axial ratio a/b and extent of hydration $m_{1,b}/m_2$ (redrawn from L. Oncley, *Ann. NY Acad. Sci.*, **41**, 121 (1941)); (b) superposition of the $[\eta] = 8.0$ contour from Fig. 4.13a and the $f/f_0 = 1.45$ contour from Figure 2.9. The crossover unambiguously characterizes particles with respect to hydration and axial ratio.

not too far from the charged surface. The charged surface of the particle and the counterions are referred to as an *electrical double layer*. The counterions are distributed over a region that extends some distance from the surface of the particles because of the thermal (diffusive) motion of the counterions. This region is known as the *diffuse* part of the double layer; its range—represented by the reciprocal of the so-called Debye-Hückel parameter κ—depends on the concentration and charge of the counterions and the electrical potential at the surface of the particles.

It is convenient to think of the diffuse part of the double layer as an ionic atmosphere surrounding the particle. Any movement of the particle affects the particle's ionic atmosphere, which can be thought of as being "dragged along" through bulk motion and diffusional motion of the ions. The resulting electrical contribution to the resistance to particle motion manifests itself as an additional viscous effect, known as the *electroviscous* effect. Further,

when the particle moves, the surface where slippage occurs lies within the diffuse part of the double layer at ionic dimensions from the surface. The potential at that location can be measured using electrophoresis experiments, among others (Chapter 12), and is called the *zeta potential* ζ.

Two charged particles approaching each other sense the presence of each other through the overlap of their electrical double layers. This double-layer overlap results in a repulsive force between similarly charged particles.

The dispersed particles also attract each other through *van der Waals* forces (Chapter 10), which decrease with increasing interparticle separation. A material property known as the *Hamaker constant*, A, can be used as a measure of the strength of the van der Waals attraction.

The net outcome of the encounter between two particles—that is, whether attraction or repulsion prevails—depends on magnitudes of attractive and repulsive forces as the particles approach each other. Table 4.4 summarizes some pertinent relationships from Chapters 10–12 that will be useful in the following discussion of the effects of electrostatic and van der Waals forces on the rheology of dispersions.

The electroviscous effects are usually classified in three categories depending on the origin of the underlying mechanism.

1. If the fluid in capillary flows considered in Section 4.4a is an electrolyte solution, the flow will give rise to a *streaming potential*, as discussed in Chapter 12. The streaming potential causes a backflow because of electroosmotic effects (also discussed in Chapter 12), thereby causing a reduction in the net flow in the forward direction. Since the flow rate for an identical pressure drop along the length of the capillary will be larger in the absence of the above double-layer effect, the presence of the electrical double layer makes the liquid appear to have an enhanced viscosity. In the case of dilute, charged dispersions, a similar effect occurs as a consequence of the additional energy dissipation caused by the distortion of the double layer under shear. The net effect is an increase in the viscosity of the dispersions. This effect is called the *primary electroviscous effect*.
2. The *secondary electroviscous effect* refers to the change in the rheological behavior of a charged dispersion arising from interparticle interactions, i.e., the interactions between the electrical double layers around the particles.
3. The term *tertiary electroviscous effect* is applied to the changes in the conformation of polyelectrolytes that are caused by *intra*molecular double-layer interactions. It is customary to extend this definition to include all effects in which the geometry of the system is altered as a result of double-layer interactions.

TABLE 4.4 Summary of Some Colloidal Interaction Energies and Parameters from Chapters 10–12

Quantity described	Equation
Debye-Hückel Parameter κ (unit, 1/length)	$\kappa = [(1000\, e^2 N_A/\varepsilon k_B T)\sum_i z_i^2 M_i]^{1/2}$ Equation (11.41)
Van der Waals attraction between two flat surfaces (unit, energy)	$\Phi_A = -(A/12\pi)d^{-2}$ Equation (10.63)
Electrical double-layer repulsion between two flat surfaces (unit, energy)	$\Phi_R = 64\, k_B T\, n_\infty\, \kappa^{-1} \Upsilon_0^2\, exp\,(-d/\kappa^{-1})$ Equation (11.86)
Zeta potential ζ (unit, volts)	$\zeta = 3\,\eta\, u/2\varepsilon$ From Equation (12.27)

Note: e = electron charge; N_A = Avogadro's number; z_i = charge of ion of type i; M_i = molar concentration of ions in the bulk; ε = dielectric constant of the medium; Φ_A = energy of attraction; A = Hamaker constant; d = distance between the surfaces; Φ_R = energy of repulsion; n_∞ = ionic concentration (in number/volume); $\Upsilon_0 \approx 1$; η = viscosity of the liquid; u = electrophoretic mobility

In studying and interpreting the effects of electrical double layers on the viscosities of dispersions, one usually assumes that the viscosity of the suspending medium itself remains unchanged. However, situations do exist in which the electrical fields at a fluid/particle interface exert a significant influence on the structure of the fluid itself in the vicinity of the interface. This can modify the viscosity of the fluid. This class of electrically induced effects is known as *viscoelectric effects*. We consider this briefly in Chapter 12 in the context of its influence on the migration of charged particles in an electric field. In this chapter, however, we are concerned only with electroviscous effects and not viscoelectric effects. A good introduction to the viscoelectric effects (and a broader treatment of electroviscous effects) may be found in Hunter (1981).

Our objective in this chapter is modest, namely, to provide a general discussion of the electroviscous effects and to present a few equations that serve as guidelines for understanding the effects of colloidal forces on the viscosity of dispersions. The underlying theories are rather complicated and fall outside our scope.

The electroviscous effects and the other effects discussed in Sections 4.7a–c lead to what is called non-Newtonian behavior in the flow of dispersions. In the next section, we begin with a brief review of the basic concepts concerning deviations from Newtonian flow behavior and then move on to consider how high particle concentrations and electroviscous effects affect the flow and viscosity.

4.8 NON-NEWTONIAN BEHAVIOR

We have already devoted a considerable amount of space to our discussion of viscosity without ever venturing beyond Newtonian systems. At least as much—probably more—could be said about non-Newtonian systems.

Colloids (even those consisting of only spherical particles) do not necessarily behave like Newtonian fluids. Deviations from Newtonian behavior occur for a number of reasons, but we are concerned mainly with two effects: (a) those due to interparticle hydrodynamic interactions and (b) those due to colloidal forces such as electrostatic effects, effects due to polymer layers adsorbed on the particles, etc. The hydrodynamic effects exist even for neutral particles although they become important only for moderate to large concentrations of the particles. The colloidal forces can exert a significant influence even at low volume fractions if the ranges of colloidal attractions and repulsions are significant. The formal theoretical developments that account for the effects of colloidal and hydrodynamic forces are rather complex and are not sufficiently well established. We shall not go into any of the details of these since they require a more advanced background in fluid mechanics than introduced in the previous sections of this chapter; some introduction to the relevant ideas are available in advanced books or reviews such as Russel (1980), Tadros (1984, Chapter 6), Hunter (1990, Chapter 18), and Russel et al. (1989, Chapter 14). However, a discussion of the rheology of dispersions without at least some indication of what happens in the case of interacting particles is not only incomplete, but is also misleading. Therefore, our goal in this section is to provide a *general idea* of what happens to the flow behavior of dispersions when the above forces become important, and what physicochemical factors or features of the phenomena one should pay attention to in attempting to understand the rheology of dispersions.

4.8a Examples of Non-Newtonian Features

Non-Newtonian fluids are generally those for which the viscosity is not constant even at constant temperature and pressure. The viscosity depends on the shear rate or, more accurately, on the previous kinematic history of the fluid. The linear relationship between the shear stress and the shear rate, noted in Equation (1), is no longer sufficient. Strictly speaking, the coefficient of viscosity is meaningful only for Newtonian fluids, in which case it is the slope of a plot of stress versus rate of shear, as shown in Figure 4.2. For non-Newtonian fluids, such a plot is generally nonlinear, so the slope varies from point to point. In actual practice, the data

are traditionally represented with the rate of shear (dv/dy) as the ordinate and the stress (F/A) as the abscissa, as shown in Figure 4.14. In this case, the apparent viscosity at some particular point is given by the cotangent of the angle that defines the slope at that particular point.

The non-Newtonian fluids in shear flows may be classified broadly into three types:

1. Fluids with shear stresses that at any point depend on the shear rates only and are *independent* of time. These include (a) what are known as *Bingham plastics*, materials that require a minimum amount of stress known as yield stress before deformation, (b) pseudoplastic (or shear-thinning) fluids, namely, those in which the shear stress decreases with the shear rate (these are usually described by "power-law" expressions for the shear stress; i.e., the rate of strain on the right-hand-side of Equation (1) is raised to a suitable power), and (c) dilatant (or shear-thickening) fluids, in which the stress increases with the shear rate (see Fig. 4.2).

2. Fluids in which the shear stress/shear rate behavior depends on time, i.e., the kinematic history of the fluid. These fall in one of two categories: (a) *thixotropic* or (b) *rheopectic*. Many polymer gels display time-dependent stress/shear rate behavior. In thixotropic fluids, the microstructure of the fluid progressively breaks down and the viscosity decreases. In rheopectic fluids, the applied shear promotes gradual formation of local structure and the apparent viscosity increases. As we discuss below, thixotropic behavior is associated with flocs of asymmetrical particles. In certain instances, quicksand apparently operates through a thixotropic mechanism. The struggles of the victim merely decrease the viscosity of the trap and worsen the victim's plight. By contrast, wet beach sand is an example of a dilatant system in which the apparent viscosity increases with shear. Anyone who has wiggled their toes in the wet sand knows that under these conditions (low shear) the wet sand is very fluid. However, the same sand is hard to a firm footstep (high shear). In contrast to thixotropy, dilatancy is favored by symmetrical, nonflocculated particles. It is almost always observed at concentrations in the neighborhood of 40% dispersed particles. At this concentration and in the absence

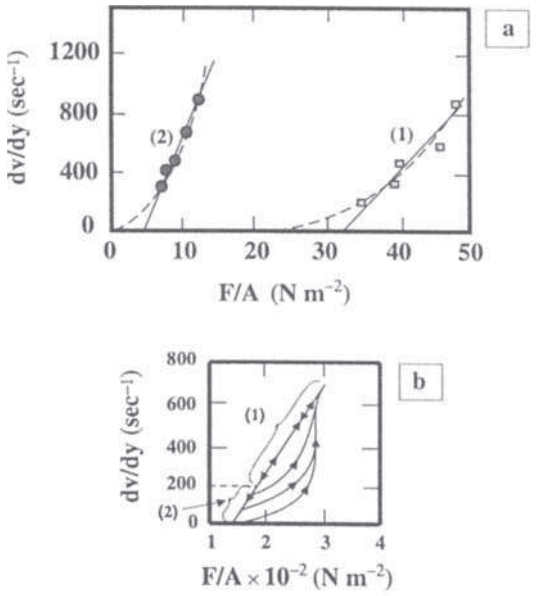


FIG. 4.14 Rate of shear versus shearing stress: (a) for 7% aqueous carbon black dispersions (data from A. I. Medalia and E. Hagopian, *Ind. Eng. Chem. Prod. Res. Dev.*, **3**, 120 (1964)); and (b) for an 11% aqueous bentonite dispersion (pH = 8.7) for which the time of the cycle is 70 sec (based on data of W. F. Gabrysh, H. Eyring, P. Lin-Sen, and A. F. Gabrysh, *J. Am. Ceram. Soc.*, **46**, 523 (1963)).

of disruptable flocs, the only way such a system can flow is by the gradual rolling of particles past one another. If the rate of shear is too great, this deformation is impossible.

3. Viscoelastic materials, as mentioned in Section 4.1, display behavior somewhere between a solid and a liquid.

Colloidal dispersions, polymer solutions, and both colloidal and polymer gels can in principle behave like any of the above-mentioned materials.

4.8b Time Scales and Dimensionless Parameters of Importance

While in the case of noninteracting dispersions one needed to consider only the effect of the particle concentration, in interacting dispersions one needs to consider the time over which the flow behavior is observed and its magnitude relative to the time scales over which either shear or colloidal forces alter the local structure of the dispersions. What the flow behavior is, which interaction effects dominate the behavior, and how they do depend on the competing influences of the applied shear and interaction effects. In this section, we outline some of the important parameters one can formulate to judge the relative effects of various colloidal interactions and the physical significance of those parameters.

4.8b.1 The Deborah Number

We mentioned in Section 4.1 that whether a material deforms under applied stress is a matter of the magnitudes of the shear force exerted and the time of observation. It is common to use silicone putty (known as *Silly Putty*) to illustrate the above statement. If you have enough patience, you will notice that silly putty is a highly viscous material (although you may not think of it that way) that will find its own level when placed in a container. In this sense, it behaves like a "liquid." On the other hand, as its name is meant to suggest, a ball of Silly Putty will also bounce when dropped to the floor. That is, under severe and sudden deformation, it behaves like a solid. The behavior of the Silly Putty thus brings to our attention the importance of time scales and deformation rates in classifying the "flow behavior" of materials.

The dimensionless number in rheology that compares relative importance of the time scale of the deformation process t_D over the observation time t_O is called the *Deborah number* (*De*):

$$De = t_D/t_O \tag{55}$$

The time t_D is typically very small for liquids and is very large for solids. Large Deborah numbers signify solidlike behavior, and small values typify liquidlike behavior. Correspondingly, a colloid or a gel may behave like a solid because either its characteristic time is very large (we discuss this below) or the deformation process is very fast (this is the reason we compare the time for deformation t_D with the time for observation t_O). The name *Deborah* comes from the fifth chapter of the book of Judges in the Old Testament, in which Deborah is reported to have said that "the mountains flowed before the Lord." The idea is that everything flows if one has a (very) long life to observe the motion!

4.8b.2 The Peclet Number

While the Deborah number is often used to compare the time for deformation with the time of observation in experiments, it also inspires us to identify and formulate other dimensionless groups that compare the various characteristic times and forces relevant in colloidal phenomena. We discuss some of the important ones.

The Peclet number compares the effect of imposed shear (known as the *convective* effect) with the effect of diffusion of the particles. The imposed shear has the effect of altering the local distribution of the particles, whereas the diffusion (or Brownian motion) of the particles tries to restore the equilibrium structure. In a quiescent colloidal dispersion the particles move continuously in a random manner due to Brownian motion. The thermal motion establishes an equilibrium statistical distribution that depends on the volume fraction and interparticle potentials. Using the Einstein-Smoluchowski relation for the time scale of the motion, with the Stokes-Einstein equation for the diffusion coefficient, one can write the time taken for a particle to diffuse a distance equal to its radius R_s as

$$t_{Diff} \sim R_s^2/D = (6\pi\eta_0 R_s^3/k_B T) \tag{56}$$

where we have used the Stokes-Einstein equation for the diffusion coefficient D (see Equations (2.32) and (2.8)),

$$D = k_B T/6\pi\eta_0 R_S \tag{57}$$

This characterizes the time taken for the restoration of the equilibrium microstructure after a disturbance caused, for example, by convective motion, i.e., this is the relaxation time of the microstructure. The time scale of shear flow is given by the reciprocal of the shear rate, $\dot{\gamma}$. The dimensionless group formed by the ratio (t_{Diff}/t_{Shear}) is the Peclet number

$$Pe = (6\pi\eta_0 R_s^3\dot{\gamma}/k_B T) = R_s^2\dot{\gamma}/D \tag{58}$$

which specifies the relative importance of convection over diffusion. For example, if $Pe \ll 1$, the distribution of particles is only slightly altered by the flow, and the behavior is dominated by diffusional relaxation of the particles. When $Pe \gg 1$, convective (hydrodynamic) effects dominate the behavior, and experiments suggest that a dispersion of spherical particles behaves like a shear-thinning fluid.

Before leaving this discussion, it is important to note that other forms of Peclet numbers are also possible and may be more appropriate depending on the type of convective influence studied. For example, in the case of oscillatory flows (as in oscillatory viscometers), it is more useful to define the Peclet number as $(R_s^2\omega/D)$, where ω is the frequency of oscillation. Regardless of the particular definition, the general significance of the Peclet number remains the same, i.e., it compares the effect of convection relative to diffusion.

4.8b.3 Relative Energies or Lengths

In addition to the Peclet number, one can also define other dimensionless groups that compare either relevant time scales or energies of interaction. Using some of the concepts previewed in Section 4.7c and Table 4.4, one can define an electrostatic group (in terms of the zeta potential ζ and relative permittivity ε_r of the liquid) as

$$N_{El} = \varepsilon_r\varepsilon_0\zeta^2 R_s/k_B T \tag{59}$$

which is a measure of the electrostatic energy relative to the thermal energy, represented by $k_B T$. Similarly, the range of electrostatic repulsion κ^{-1} (the Debye length) relative to the particle radius R_s is represented by

$$N_{DL} = \kappa R_s \tag{60}$$

The strength of the van der Waals attraction relative to thermal energy is represented by

$$N_{vdW} = A/k_B T \tag{61}$$

where A is the effective Hamaker constant between the particles through the intervening fluid.

Other dimensionless groups that compare the thickness of the adsorbed polymer layer to the radius of the particle or the radius of gyration of the polymer to the particle radius in polymer/colloid mixtures can also be easily defined. We are mostly concerned with the volume fraction ϕ and the Peclet number Pe in our discussions in this chapter. However, the other dimensionless groups may appear in the equations for intrinsic viscosity of dispersions when the dominant effects are electroviscous or sterically induced.

4.8c Charged Particles

In Section 4.7c we outlined the types of effects one can expect in the response of charged dispersions to deformation. In this section, we present some results for the viscosity of charged colloids for which electroviscous effects could be important. As mentioned above, we shall not go into the theoretical details behind the equations since they require a fairly advanced knowledge of fluid dynamics and, in some cases, statistical mechanics. Moreover, some of

the concepts, such as electrical double layers, double-layer thickness (i.e., the Debye or the Debye-Hückel length) κ^{-1} and zeta potential ζ, used in the following discussion require material from Chapters 11 and 12, and it will help to return to this section after reviewing those chapters.

4.8c.1 The Primary Electroviscous Effect

The first analysis of the primary electroviscous effect dates to 1916 when Smoluchowski presented the following equation for the intrinsic viscosity:

$$[\eta] = 2.5\left\{1 + \frac{4(\varepsilon_r\varepsilon_0\zeta)^2}{k\,\eta_0 R_s^2}\right\} \tag{62}$$

where ε_r is the relative permittivity, ε_0 is the permittivity of free space, k is the specific conductivity of the continuous phase, and ζ is the zeta potential, which is used in place of the potential at the surface of the particle (a sphere of radius R_s). The physical significance and measurement of the zeta potential are discussed in Chapter 12. Smoluchowski's result is supposed to apply for thin electrical double layers (i.e., large κR_s), a range for which the primary electroviscous effects are smallest. A subsequent analysis by Krasny-Ergen in 1936 for the same conditions led to a correction for the second term in the brackets:

$$[\eta] = 2.5\left\{1 + \frac{6(\varepsilon_r\varepsilon_0\zeta)^2}{k\,\eta_0 R_s^2}\right\} \tag{63}$$

These equations (and others that follow) are based on the solutions of equations of motion for the particles, as well as the electrolyte, that we use in Chapter 12 in the context of electrophoresis. It turns out that the analysis of Krasny-Ergen fails to satisfy one of the boundary conditions and does not take into account energy dissipations caused by the electric currents arising from the motion of the electrolyte. It is more appropriate for $\kappa R_s \to \infty$.

A corrected and more general analysis of the primary electroviscous effect for weak flows, i.e., for low Pe numbers (for small distortions of the diffuse double layer), and for small zeta potentials, i.e., $\zeta < 25$ mV, was carried out by Booth in 1950. The result of the analysis leads to the following result for the intrinsic viscosity $[\eta]$ for charged particles in a 1 : 1 electrolyte:

$$[\eta] = 2.5\left\{1 + 4\pi(\varepsilon_r\varepsilon_0\zeta)^2\frac{(\kappa R_s)^2(1 + \kappa R_s)^2 Z(\kappa R_s)}{k\,\eta_0 R_s^2}\right\} \tag{64}$$

The function $Z(\kappa R_s)$ is a power series with the two limiting forms, for extended diffuse layers:

1. Thick double layers, i.e., small κR_s:

$$Z(\kappa R_s) \approx (200\pi\kappa R_s)^{-1} + (11\kappa R_s/3200\pi) \tag{65a}$$

2. Thin double layers, i.e., large κR_s:

$$Z(\kappa R_s) = (3/2\pi)(\kappa R_s)^{-4} \tag{65b}$$

The substitution of Equation (65b) in Equation (64) leads to the Krasny-Ergen equation for $\kappa R_s \to \infty$, as one would expect.

Although Booth's result already indicates the influence of the charges on the particles and the electrolyte in the dispersion on the viscosity of the dispersion, one sees more complex behavior only when the effect of larger distortions of the double layer is included in the analysis. An extension for larger distortions (represented by the Peclet number Pe) of the double layer is available (Russel 1978a) and can be written as

$$[\eta] = 2.5\left\{1 + \frac{6(\varepsilon_r\varepsilon_0\zeta)^2}{k\,\eta_0 R_s^2}\frac{1}{1 + Pe^2}\right\} \tag{66}$$

for small zeta potentials (less than about 25 mV) and for $Pe \ll \kappa R_s$ and $\kappa R_s \gg 1$. Russel's result indicates that the primary electroviscous effect leads to a shear-thinning (pseudoplastic)

behavior even for moderate *Pe* and thin double layers with low potentials. Moreover, variations in normal stresses also occur, although they cannot be deduced from the above equation alone.

The results presented here provide a glimpse of the complications in the flow behavior that may arise as a result of the distortions of the electrical double layer even at low volume fractions of the dispersions. Extensions of Booth's result and others are available in the literature (see Hunter 1981), but we shall not go into those here.

The variations in the intrinsic viscosity predicted by the primary electroviscous effect are often small, and it is difficult to attribute variations in the experimentally observed $[\eta]$ from the Einstein value of 2.5 to the above effect since such variations can be caused easily by small amounts of aggregation. Nevertheless, Booth's equation has been found to be adequate in most cases. Further discussions of this and related issues are available in advanced books (Hunter 1981).

4.8c.2 The Secondary Electroviscous Effect

The secondary electroviscous effect is often interpreted in terms of an increase in the effective "collision diameter" of the particles due to electrostatic repulsive forces (i.e., the particles begin to "feel" the presence of other particles even at larger interparticle separations because of electrical double layer). A consequence of this is that the excluded volume is greater than that for uncharged particles, and the electrostatic particle-particle interactions in a flowing dispersion give an additional source of energy dissipation.

Let us consider situations for which the double-layer thickness is large enough (i.e., κR_s is small) and interparticle distances are large (Russel 1978b). Then, at typical interparticle distances in a quiescent dispersion, the van der Waals attraction (see Chapter 10) is insignificant. We can then use the electrostatic energy expression presented in Chapter 11 to get a dimensionless parameter α that represents the ratio of electrostatic energy to thermal energy (i.e., k_BT):

$$\alpha = (4\pi\varepsilon_r\varepsilon_0\zeta^2R_s^2\kappa/k_BT)\exp(2\kappa R_s) = 4\pi N_{El}N_{DL}\exp(2N_{DL}) \tag{67}$$

where N_{El} and N_{DL} are defined in Equations (59) and (60), respectively. The parameter α is another way of expressing the strength of electrostatic energy relative to the thermal energy. For values of $\alpha \gg 1$ (i.e., the electrostatic energy is dominant), the effective collision diameter L for the particles in the absence of convective forces can be shown to be

$$L = \kappa^{-1}\ln[\alpha/\ln(\alpha/\ln\alpha)] \tag{68}$$

Under shear flow, the minimum center-to-center separation, r, between the particles will be in the interval $\{2R_s \le r \le L\}$; i.e., at the high-shear limit $r = 2R_s$ and at quiescent conditions $r = L$. An analysis of the flow behavior in this case leads to the following expression for the zero shear rate (i.e., $Pe \ll 1$) limit of the relative viscosity:

$$\frac{\eta(0)}{\eta_0} = 1 + [\eta]\phi + \frac{2}{5}([\eta]\phi)^2 + \frac{3}{40}\ln\left(\frac{\alpha}{\ln\alpha}\right)\left|\ln\left(\frac{\alpha}{\ln(\alpha/\ln\alpha)}\right)\right|^4\frac{\phi^2}{(\kappa R_s)^5}\cdots \tag{69}$$

The intrinsic viscosity $[\eta]$ in the above expression *includes* the primary electroviscous effect. The experimental data of Stone-Masui and Watillon (1968) for polymer latices seem to be consistent with the above equation (Hunter 1981). Corrections for α for large values of κR_s are possible, and the above equation can be extended to larger Peclet numbers. However, because of the sensitivity of the coefficients to κR_s and the complications introduced by multiparticle and cooperative effects, the theoretical formulations are difficult and the experimental measurements are uncertain. For our purpose here, the above outline is sufficient to illustrate how secondary electroviscous effects affect the viscosity of charged dispersions.

4.8c.3 The Tertiary Electroviscous Effect

As mentioned in Section 4.7c, the tertiary electroviscous effect is at least partly due to the expansion and contraction of particles arising from the conformational changes of the polyelectrolytes (adsorbed or chemically bound to the surface of the particles) with changes in

the electrolyte concentration or pH of the medium. For particles stabilized by a layer of polyelectrolyte, one can calculate an effective volume fraction ϕ' (if the thickness δ of the polyelectrolyte layer is known) using

$$\phi' = \phi[1 + (\delta/R_s)]^3 \tag{70}$$

where ϕ is the actual volume fraction of the particles (of radius R_s). One can then use an equation such as Equation (50), the Krieger-Dougherty equation, to estimate the viscosity. The analysis of the tertiary electroviscous effect is complicated by the fact that factors such as the effect of the particle surface on the conformation of the polyelectrolyte chains, the structure of the chains in the layer, and the extent of the penetration of the liquid into the layer make it difficult to estimate the thickness and the hydrodynamic effects of the stabilizer layer. This area is still in its infancy. Some additional information may be obtained from Hunter (1981).

4.8d Dispersions of Aggregates

So far we have focused our attention on dispersions in which interparticle repulsions dominate. Attractive forces (see Chapter 10) have been assumed to be negligible, and as a consequence the dispersions are stable. When the van der Waals attraction begins to play a significant role (i.e., the net effect of repulsion and attraction is such that we have coagulation of the dispersion), more complicated flow behavior occurs. Such dispersions are more common in practice and deserve our attention.

Aggregated colloids are common examples of systems in which the apparent viscosity depends on the rate of shear. The existence of velocity gradients means that differences in flow velocity may extend over the dimensions of colloidal particles. If the dispersed "particle" is an aggregate of primary particles, a *floc*, then the shearing forces associated with viscosity that operate across the floc may disrupt or rearrange the aggregate. This corresponds to a dissipation of translational energy and hence contributes to the viscosity. It is clear that the extent to which this occurs may vary with the velocity gradient; this is one way in which the apparent viscosity may depend on the rate of shear.

Let us see how thixotropic behavior relates to the phenomena of coagulation. The data shown in Figure 4.14a were obtained for a 7% slurry of carbon black in water. Curve 1 shows the results obtained immediately after the dispersion was prepared in a high-speed blender. After additional mild agitation, the results shown in curve 2 are obtained; these are independent of further agitation. With shorter periods of mild agitation, a family of curves lying between 1 and 2 would be obtained.

The data presented in Figure 4.14a are consistent with the following mechanism. The dispersion that emerges from the blender is fundamentally unstable with respect to coagulation and coagulates rapidly to form a volume-filling network throughout the continuous phase. Except for the size and structure of the "chains," the situation is comparable to a cross-linked polymer swollen by solvent. In both, the liquid is essentially immobilized by the network of chains, and the system behaves as an elastic solid under low stress. The term *gel* is used to describe such systems whether the dispersed particles are lyophilic or lyophobic.

As the force applied to the surface of the gel is increased, however, a point is ultimately reached—the yield value—at which the network begins to break apart and the system begins to flow (curve 1 in Fig. 4.4a). Increasing the rate of shear may result in further deflocculation, in which case the apparent viscosity would decrease further with increased shear. Highly asymmetrical particles can form volume-filling networks at low concentrations and are thus especially well suited to display these phenomena.

As the system is subjected to ongoing, low-level mechanical agitation, the network structure is rearranged to a dispersion of more compact flocs that display both a lower yield value and a lower apparent viscosity than the initial dispersion (curve 2). A certain amount of time is required for the dispersed units to acquire a size and structure compatible with the prevailing low level of agitation. This is why intermediate cases (not shown in Fig. 4.14a) are observed before the actual stationary-state condition is obtained.

If the time for measurement of a stress/shear curve is short compared to the time required for rearrangement of the structure of the dispersed particles, then different results are obtained depending on whether the rate of shear is increasing or decreasing. Figure 4.14b is an example of such hysteresis for 11% dispersions of bentonite (a montmorillonite clay with plate-shaped primary particles) in which the entire cycle is measured in 70 sec. These data show the sensitivity of such experiments to the level of shear and to the time of observation. If the direction of the cycle is reversed along the descending branch of the curve, different results are obtained depending on whether the reversal is done at a rate of shear above (region (1) in Fig. 4.14b) or below (region (2)) about 200 s^{-1}. This shows that the structure within the colloid builds up rapidly (compared to the cycle time of 70 sec) at rates of shear below 200 s^{-1}, with negligible buildup at greater rates of shear.

These complicated observations are difficult to interpret in terms of fundamental interactions between particles; nevertheless, they have tremendous practical significance. For example, carbon black strengthens rubber against deformation, and for this reason 3–4 kg of carbon black is introduced into every tire made. This accounts for an annual worldwide consumption of carbon black of over 4 billion kg for automobile tires alone. Likewise, dispersions of clay in oils are used as lubricating greases. Obviously, the viscosity of these substances under various conditions of shear is an important consideration. Printing inks, drilling muds (circulated around the shaft in well-drilling operations to cool the bit and flush away cuttings), paper coatings, paints, and innumerable industrial slurries may all be considered examples of areas for which these considerations are vitally important.

In a paint, for example, a controlled level of aggregation is important in both the actual application of the paint and its storage in the container during application. In the former, thixotropy (the word means "changing with touch") permits the paint to "thin" under the shearing influence of the paintbrush or spraygun. Once applied, it thickens, preventing the drip or sag of the paint on the surface. In addition, the time required for this yield value to develop should be sufficient to allow for the leveling of brush marks. Thixotropy is an important property of paint in the bucket as well as on the wall. The buildup of a yield value interferes with the sedimentation of the pigment and eliminates the need to stir the paint continuously to assure uniformity. The fact that these requirements are well met by commercial paints indicates the success of paint chemists in regulating thixotropy.

One of the major difficulties in developing theories of the rheology of coagulated or flocculated dispersions is that the microstructures of the aggregates are nonequilibrium structures under shear. Understandably, the rheology of such dispersions is history dependent, as we have seen above, and requires computer simulations and nonequilibrium statistical mechanics for proper study.

4.9 VISCOSITY OF POLYMER SOLUTIONS

The viscosity of a polymer solution is one of its most distinctive properties. The spatial extension of the molecules is great enough so that the solute particles cut across velocity gradients and increase the viscosity in the manner suggested by Figure 4.8. In this regard they are no different from the rigid spheres of the Einstein model. What is different for these molecules is the internal structure of the dispersed units, which are flexible and swollen with solvent. The viscosity of a polymer solution depends, therefore, on the polymer-solvent interactions, as well as on the properties of the polymer itself.

4.9a The Staudinger-Mark-Houwink Equation

An important empirical generalization about the intrinsic viscosity of polymer solutions is given by

$$[\eta] = kM^a \tag{71}$$

in which k and a are experimentally determined constants. We shall call Equation (71) the *Staudinger-Mark-Houwink equation*. Staudinger (Nobel Prize, 1953), one of the founders of

modern polymer chemistry, originally proposed this relationship with $a = 1$. It was subsequently recognized that more commonly $0.5 < a < 0.8$. The constants k and a are called the Mark-Houwink coefficients; they depend on the temperature as well as the polymer-solvent system. Values of k and a are determined for a particular system by measuring the intrinsic viscosity of polymer fractions of known molecular weight. Example 4.5 illustrates how this is done.

* * *

EXAMPLE 4.5 *Empirical Determination of the Mark-Houwink Coefficients for a Polymer Solution.* The molecular weights of various polycaprolactam preparations were determined by end-group analysis (see Example 3.2); intrinsic viscosities of the various fractions in *m*-cresol were measured at 25°C. The following values are representative of the results obtained (Reimschussel and Dege 1971):

$M \cdot 10^{-3}$	3.50	4.46	7.69	13.0	17.6	21.6	30.8
$[\eta]$(dl/g)	0.36	0.43	0.61	0.87	1.10	1.25	1.59

Determine the values of k and a that fit these data.

Solution: By taking the logarithms of Equation (71), a linear form is obtained:

$$\ln [\eta] = \ln k + a \ln M.$$

This could be plotted with a and $\ln k$ evaluated from the slope and intercept, respectively. Alternatively, a least squares analysis of the data can be performed. When this is carried out using the logarithms of the above results, it is found that $a = 0.683$ and $\ln k = -6.593$, or $k = 1.37 \cdot 10^{-3}$. The units of k are consistent with the concentration units "dl/g." This result can be inverted to give M directly: $M = 1.51 \cdot 10^{4}[\eta]^{1.46}$. ■

* * *

The practical significance of the result of this example lies in the great ease with which viscosity measurements can be made. Once the k and a values for an experimental system have been established by an appropriate calibration, molecular weights may readily be determined for unknowns measured under the same conditions. Extensive tabulations of Mark-Houwink coefficients are available, so the calibration is often unnecessary for well-characterized polymers (see Table 4.5).

If intrinsic viscosity is used to evaluate the molecular weight of a polydisperse sample, the molecular weight so obtained is an average value. Equation (1.20) defined the viscosity average, which is the kind of average obtained. We are now in a better position to see how this comes about.

TABLE 4.5 Mark-Houwink Coefficients for Some Typical Polymer-Solvent Systems at the Indicated Temperatures

Polymer	Solvent	Temperature (°C)	$k \times 10^{3}$ ($cm^{3}g^{-1}$)	a
Polypropylene	Cyclohexanone	92 = Θ	172	0.50
Poly(vinyl alcohol)	Water	80	94	0.56
Poly(oxyethylene)	CCl_4	25	62	0.64
Poly(methyl methacrylate)	Acetone	30	7.7	0.70
Polystyrene	Toluene	34	9.7	0.73
Natural rubber	Benzene	30	18.5	0.74
Poly(acrylonitrile)	Dimethyl formamide	20	17.7	0.78
Poly(vinyl chloride)	Tetrahydrofuran	20	3.63	0.92
Poly(ethylene terephthalate)	*m*-Cresol	25	0.77	0.95

Source: M. Kurata, M. Iwama, and K. Kamada, in *Polymer Handbook*, 2d ed. (J. Srandrup and E. H. Immergut, Eds.), Wiley, New York, 1975.

It is apparent that the experimental (subscript *ex*) intrinsic viscosity depends on an average molecular weight $\overline{M}$; it is the *nature of this average* that we want to establish. According to Equation (71), we write

$$[\eta]_{ex} = k\overline{M}^{a} \tag{72}$$

The intrinsic viscosity contains a contribution of concentration that we note by writing $[\eta]_{ex} = (\eta_{sp})_{ex}/c_{ex}$ for which it is understood that this is a limiting value. If the sample is polydisperse, both $(\eta_{sp})_{ex}$ and c_{ex} are made up of contributions from molecules in different molecular weight classes, which we designate by the subscript *i*; that is, we can write

$$(\eta_{sp})_{ex} = \sum_{i}(\eta_{sp})_{i} \tag{73}$$

and

$$c_{ex} = \sum_{i} c_{i} = \frac{1}{V}\sum_{i} m_{i} \tag{74}$$

where the rightmost expression for c_{ex} recognizes that practical units are used in these experiments.

Each of the molecular weight fractions is expected to obey Equation (71) independently; therefore

$$|\eta_{sp}|_{i} = c_{i}[\eta]_{i} = c_{i}kM_{i}^{a} \tag{75}$$

Substituting Equations (73)–(75) into Equation (72) yields

$$\frac{k\Sigma_{i} c_{i} M_{i}^{a}}{\Sigma_{i} c_{i}} = \frac{(k/V)\Sigma_{i} m_{i} M_{i}^{a}}{(1/V)\Sigma_{i} m_{i}} = k\overline{M}^{a} \tag{76}$$

which shows that

$$\overline{M}^{a} = \frac{\Sigma_{i} m_{i} M_{i}^{a}}{\Sigma_{i} m_{i}} \tag{77}$$

Since $m_{i} = n_{i}M_{i}$, Equation (77) may be written as

$$\overline{M} = \overline{M}_{v} = \left(\frac{\Sigma_{i} n_{i} M_{i}^{(a+1)}}{\Sigma_{i} n_{i} M_{i}}\right)^{1/a} \tag{78}$$

which is the form given in Chapter 1 in Equation (1.20) and Table 1.8.

In this section we have looked at the Staudinger-Mark-Houwink equation as a purely empirical relationship useful for determining the molecular weight of unknown polymeric solutes. A considerable amount of work has been directed toward understanding the theoretical basis for this result. Although a detailed discussion would take us too far afield, we examine certain special cases of Mark-Houwink *a* values in the next section.

4.9b Polymer Chain Extension and Viscosity

Two special cases of the Staudinger-Mark-Houwink equation can be justified without much difficulty: $a = 1.0$ and $a = 0.5$. In our discussion we ignore all numerical coefficients and concentrate on the variables, particularly with respect to molecular weight dependence. This is sufficient to arrive at an understanding of the significance of the Mark-Houwink exponent. We examine two models for polymer chains in solution. These models picture the polymer chain as being either unwound so that each segment experiences the flow streamlines or tightly coiled so that the flow leaves interior segments unaffected. The two models we discuss are known, respectively, as the free-draining and nondraining models. We see presently that these models—which clearly represent extremes of behavior—correspond to exponents of 1.0 and 0.5, respectively, in the Staudinger-Mark-Houwink equation. It is apparent that many poly-

mer-solvent-temperature situations will result in chain coils of intermediate conformation; for these a is expected to take on values between the extremes we consider.

4.9b.1 Free-Draining Model

We begin by examining the free-draining model. Figure 4.15a shows a portion of a polymer chain situated in a velocity gradient. This will cause the molecule to rotate, thus converting some translational kinetic energy to rotational kinetic energy. This amounts to a dissipation of energy and thus connects with viscosity. Although not self-evident, it can be shown that the angular velocity induced in the molecule in this situation is directly proportional to the velocity gradient. Furthermore, a chain segment at a distance $r = r_i$ from the center of mass of the molecule, say, segment i, acquires a velocity $v_i = r_i\omega$ from the rotation. Figure 4.15b defines the location of the ith segment relative to the center of mass.

Next, we consider the force of viscous resistance experienced by this segment as a result of moving through its surroundings with a velocity v. Equation (2.2) relates the force of viscous resistance to the velocity through the expression $F_v = fv$, where f is the friction factor. The friction factor of a chain segment plays the same role but is given the symbol ξ to emphasize that it applies to a segment rather than the entire molecule. Therefore, for the ith segment, we write

$$F_{v,i} = \xi v_i \tag{79}$$

Both $F_{v,i}$ and v_i can be resolved into x and y components. Using the geometry defined by Figure 4.15b, we obtain

$$F^x_{v,i} = \xi r_i \omega \sin\theta \tag{80}$$

and

$$F^y_{v,i} = \xi r_i \omega \cos\theta \tag{81}$$

Since force times distance equals energy, the rate at which energy is dissipated by viscous forces on segment i is given by $F^x_{vi}v_x + F^y_{vi}v_y$, which we identify as $(dE/dt)_i$. Substituting Equations (80) and (81) gives

$$\left(\frac{dE}{dt}\right)_i = \xi r_i^2 \omega^2 (\sin^2\theta + \cos^2\theta) \propto \xi r_i^2 \left(\frac{dv}{dy}\right)^2 \tag{82}$$

For the entire polymer chain (subscript p), this result must be totaled for the n segments of the chain to produce $(dE/dt)_p$:

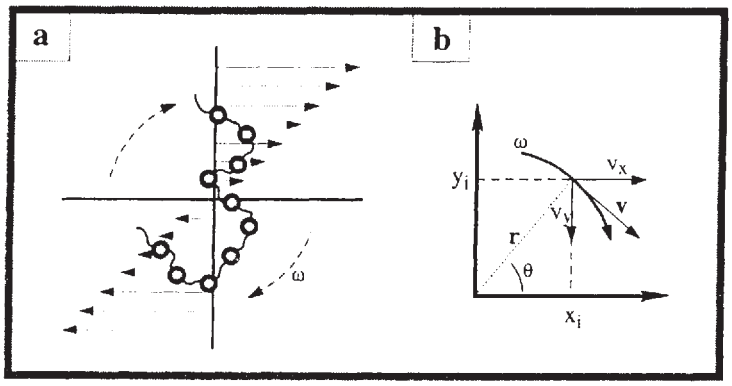


FIG. 4.15 Flow of polymer solutions: (a) velocity gradient through the center of mass of a polymer chain; (b) definition of coordinates for ith segment of rotating chain. The distance $r = r_i$ is the distance of the ith segment from the center of mass. (Redrawn with permission from Hiemenz 1984.)

$$\left(\frac{dE}{dt}\right)_p \propto \xi \sum_{i=1}^{n} r_i^2 \left(\frac{dv}{dy}\right)^2 \tag{83}$$

Since r_i is the distance from the center of mass, we use Equation (2.75) to replace $\Sigma_i r_i^2$ with $n\overline{R_g^2}$, where R_g is the radius of gyration of the molecule. Thus, Equation (83) becomes

$$\left(\frac{dE}{dt}\right)_p \propto \xi n \overline{R_g^2} \left(\frac{dv}{dy}\right)^2 \tag{84}$$

To scale up this last result from a single molecule to all of the molecules in a volume of solution V, we multiply Equation (84) by the particle concentration, given by c/M:

$$\frac{dE_V}{dt} \propto \xi n \overline{R_g^2} \frac{c}{M} \left(\frac{dv}{dy}\right)^2 \tag{85}$$

Notice that the left-hand side is now energy dissipation per unit volume. Comparing this result with Equation (3) shows that the increment in viscosity caused by the free-draining chain is given by the coefficient of $(dv/dy)^2$ and may be written

$$\eta - \eta_0 \propto \xi n \overline{R_g^2} \frac{c}{M} \propto \xi n \overline{R_g^2} \frac{c}{n} \tag{86}$$

where the second form recognizes that M is proportional to the number of segments in the chain. Finally, Equation (86) can be cast in the form of the intrinsic viscosity to yield

$$[\eta] = \frac{1}{c} \frac{(\eta - \eta_0)}{\eta_0} \propto \frac{\xi \overline{R_g^2}}{\eta_0} \tag{87}$$

which shows that the intrinsic viscosity of a solution of freely draining chains is proportional to the square of the radius of gyration of those chains. The segmental friction factor has dimensions "mass/time," and η_0 has dimensions "mass/(length time)"; therefore, the right-hand side of Equation (87) has dimensions of (length^3/mass), which are the reciprocal concentration units appropriate for intrinsic viscosity.

Independent studies of the segmental friction factor reveal it to be essentially independent of the molecular weight of the polymer. Therefore for this model the molecular weight dependence of the intrinsic viscosity is the same as that of $\overline{R_g^2}$. According to Equation (2.77), $\overline{R_g^2} \propto n \propto M$; therefore the only molecular weight dependence that survives out of all of this is a first-power dependence. We are thus led to conclude that the Mark-Houwink coefficient a equals unity for the case of free-draining chains. Since polymer chains are generally jumbled into coils, for which this may not be a good model, it is not surprising that experimental a values are usually less than this.

We discussed solvent goodness in Chapter 3, Section 3.4b.4. There are several aspects of that discussion that are pertinent here:

1. Solvent is squeezed out of the coil domain more and more as the solvent becomes poorer.
2. A change in solvent goodness can come about either from addition of a poorer solvent or from decreases in temperature.
3. Theta conditions correspond to a solvent so poor that precipitation would occur for a polymer of infinite molecular weight.
4. Theta conditions are identified experimentally as the situation in which the second virial coefficient of the osmotic pressure is zero.

In view of these considerations, it is not surprising that experimental a values vary systematically with decreasing solvent goodness. As the solvent goodness decreases, the chain becomes more tightly coiled so that flow streamlines penetrate the coil to a lesser extent. In an extreme situation we can imagine the coil so impermeable to the solvent flow that it behaves as

a rigid sphere. Such a coil is said to be *nondraining* since any solvent imbibed by the coil is essentially immobilized. There are two very important things to realize about this state of affairs. First, a coil in which polymer-polymer contacts are this highly favored sounds like an alternate description of theta conditions. Second, if the coil approaches a rigid sphere in behavior, Einstein's equation for viscosity becomes appropriate again.

4.9b.1 The Kirkwood-Riseman Theory

Kirkwood and Riseman have developed a theory that allows for variable degrees of solvent drainage through the coil domain. We shall not go into this theory in any detail, except to note that it should reduce to Equation (87) in the free-draining limit and to the Einstein equation in the nondraining limit. The Kirkwood-Riseman theory can be written in the form

$$[\eta] \propto \frac{(\overline{R_{g,0}^2})^{3/2}}{n} f(X) \tag{88}$$

where the subscript 0 indicates Θ conditions and $X = [n\xi/\eta_0(\overline{R_{g,0}^2})^{1/2}]$. The function $f(X)$ in Equation (88) is our concern. For our purposes, it is enough to note that $f(X)$ approaches a value that is proportional to $[n\xi/\eta_0 (\overline{R_{g,0}^2})^{1/2}]$ in the free-draining limit (we continue to ignore numerical constants) and approaches a constant value in the nondraining limit. If we substitute the free-draining limit into Equation (88), we obtain

$$[\eta] \propto \frac{(\overline{R_{g,0}^2})^{3/2}}{n} \frac{n\xi}{\eta_0(\overline{R_{g,0}^2})^{1/2}} \propto \frac{\xi}{\eta_0} \overline{R_{g,0}^2} \tag{89}$$

The radius of gyration is expected to be different under theta and nontheta conditions since the extent of coil swelling due to imbibed solvent changes with solvent goodness. We define a coil expansion factor α* as follows:

$$\alpha = \frac{R_g}{R_{g,0}} \tag{90}$$

Although α is ordinarily greater than unity, fractional values are also possible. The range of fractional values is more limited, however, since the polymer tends to precipitate rather than squeeze out much more solvent under conditions poorer than theta conditions. Incorporating α into Equation (89) gives

$$[\eta] \propto \frac{\xi}{\eta_0} \frac{\overline{R_g^2}}{\alpha^2} \tag{91}$$

which is equivalent to Equation (87) as required.

Using the nondraining limit of the Kirkwood-Riseman theory gives

$$[\eta]_\Theta \propto \frac{(\overline{R_{g,0}^2})^{3/2}}{n} \tag{92}$$

Since Equation (2.77) shows that $\overline{R_{g,0}^2} \propto n$, Equation (92) becomes $[\eta] \propto n^{1/2} \propto M^{1/2}$. This important result shows that $a = 0.5$ is expected under Θ conditions. This expectation has been repeatedly verified by viscosity experiments under independently determined conditions. The subscript Θ has been attached to $[\eta]$ in Equation (92) in recognition of this.

For high molecular weight polymers in good solvents, $[\eta]$ exceeds $[\eta]_\Theta$ because of coil expansion under nondraining conditions; that is, as more solvent enters the coil domain than would be present under Θ conditions, Equation (92) continues to apply, with $\overline{R_g^2}$ replacing $\overline{R_{g,0}^2}$. Using Equation (90) to quantify this expansion effect, we obtain

$$[\eta] \propto \frac{\alpha^3(\overline{R_{g,0}^2})^{3/2}}{n} \tag{93}$$

*Note that this α is different from the one defined in Eq. (67).

Taking the ratio of Equation (93) to Equation (92) gives

$$\frac{[\eta]}{[\eta]_\Theta} = \alpha^3 \tag{94}$$

which shows how the effect of solvent uptake on the spatial extension of polymer coils can be quantitatively determined from intrinsic viscosity experiments under theta and nontheta conditions.

We remarked above that the Einstein equation might be pertinent in the case of a nondraining coil; Equation (93) does not seem to bear this out. The Einstein theory in the form of Equation (41) predicts that $(\eta/\eta_0 - 1)$ is proportional to the volume fraction of the dispersed particles. For the nondraining coils, the volume fraction is proportional to the volume of each coil domain times the number concentration of the particles. The first of these factors is proportional to R_g^3 and the second to c/M. This leads to the prediction

$$\left(\frac{\eta}{\eta_0} - 1\right) \propto \phi \propto R_g^3 \frac{c}{M} \propto \alpha^3 (\overline{R_{0,0}^2})^{3/2} \frac{c}{M} \tag{95}$$

which is equivalent to Equation (92).

We have omitted a great deal of detail in this discussion of polymer viscosity. The interested reader will find some of the missing information supplied in Flory (1953). In particular, we have omitted all numerical coefficients, which limits us to ratios as far as computational capability is concerned. Numerical coefficients are available for Equation (92), for example, and this allows coil dimensions to be evaluated from viscosity measurements. A general conclusion that unifies all of this section is that any factor that causes a polymer chain to be more extended in space—whether by coil unfolding or swelling by solvent—tends to increase $[\eta]$. This is exactly what we expect in terms of the purely qualitative picture provided by Figure 4.8. Example 4.6 illustrates this for some actual polymers.

* * *

EXAMPLE 4.6 *Variation of Viscosity with Polymer Configuration.* For cellulose triacetate in acetone at 25°C, $k = 8.97 \cdot 10^{-5}$ dl/g and $a = 0.9$. For polyisobutene in benzene at 24°C, $k = 1.07 \cdot 10^{-3}$ dl/g and $a = 0.5$. Calculate $[\eta]$ for these systems if each of the respective polymers has $M = 10^5$. Comment on the correlation of the a values with the nature of the system in each case.

Solution: Intrinsic viscosities are calculated by direct substitution into the Staudinger-Mark-Houwink equation. For cellulose triacetate,

$$[\eta] = 8.97 \cdot 10^{-5}(10^5)^{0.9} = 2.84 \text{ dl/g}$$

For polyisobutene,

$$[\eta] = 1.07 \cdot 10^{-3}(10^5)^{0.5} = 0.34 \text{ dl/g}$$

Even though the two polymers have the same molecular weight, the cellulose triacetate has an intrinsic viscosity more than eight times greater than the polyisobutene. Note that the Mark-Houwink coefficient a is primarily responsible for this; the intrinsic viscosities would be ranked oppositely if k were responsible.

These results are fully consistent with the nature of the systems involved. Cellulose triacetate repeat units are six-member rings that carry three bulky acetate groups each. Internally, the rings are inflexible with respect to rotation and, because of the bulk of the acetate groups, are expected to be severely hindered in their rotation with respect to each other. As a consequence, such a molecule exists in solution in a highly extended form that approaches the predictions of the free-draining model in intrinsic viscosity. Polyisobutene, by contrast, is an aliphatic hydrocarbon for which low molecular weight aliphatic molecules are expected to be better solvents. Although benzene dissolves this polymer, it is a poorer solvent because of the polarizable pi electrons. It is therefore plausible that this system corresponds to theta conditions with an a value of 0.5. ■

* * *

REVIEW QUESTIONS

1. What is meant by *rheology*? What role does viscosity of a fluid play in rheology?
2. Describe the physical significance of *Newton's law of viscosity*. Is Newton's law always applicable?
3. What is meant by a *fluid*, and how does it differ from (a) a solid and (b) a gel in terms of the flow behavior?
4. Why is rheology important in colloid science? Give as many examples as you can.
5. Explain what a *non-Newtonian* fluid is, and list the different types of non-Newtonian flow behavior.
6. Why is the shear history often important in the case of dispersions and polymer solutions?
7. Why are time scales and ratios of forces or energies important in studying flow behavior of dispersions?
8. What is the *Deborah number*? What is its physical significance? What is the *Peclet number*? Describe at least two ways of defining the Peclet number for flow of dispersions.
9. What is the *Navier-Stokes equation*? What is the physical significance of each of the terms appearing in the Navier-Stokes equation? How does the Navier-Stokes equation differ from the Stokes equation? Can you use the Navier-Stokes equation for a non-Newtonian fluid?
10. Why does the viscosity of a dispersion change, and how does it change, with the addition of colloidal particles?
11. List some of the conditions under which the Einstein equation for viscosity of dispersions fails and how one can correct the situation.
12. What is *inherent viscosity*, and what are its units?
13. What is *reduced viscosity*?
14. How do charges on particles change the viscosity of a dispersion? What are electroviscous effects? How do they differ from the *viscoelectric* effect?
15. How does the viscosity of a polymer solution differ from the viscosity of dispersions? What factors are important in the case of the former?
16. What is meant by "free-draining model" and "nondraining model" in the case of viscosities of polymer solutions?
17. Describe a way to determine the molecular weight of a polymer through viscosity measurements.

REFERENCES

General References (with Annotations)

Barnes, H. A., Hutton, J. F., and Walters, K., *An Introduction to Rheology*, Elsevier, Amsterdam, Netherlands, 1989. (Undergraduate and graduate levels. A practical introduction to the rheology of suspensions and polymer solutions. Chapter 2 presents introductory concepts and measurement techniques at the undergraduate level. Effects of colloidal interactions are not described.)

Bird, R. B., Stewart, W. E., and Lightfoot, E. N., *Transport Phenomena*, Wiley, New York, 1960. (Undergraduate and graduate levels. A classic on transport phenomena. Roughly one-third of the book is on fluid mechanics.)

Dickinson, E., *An Introduction to Food Colloids*, Oxford University Press, Oxford, England, 1992. (Undergraduate level. Although devoted to food colloids, the discussion of rheology in Chapter 3 of this volume is generally applicable to dispersions.)

Hiemenz, P. C., *Polymer Chemistry: The Basic Concepts*, Marcel Dekker, New York, 1984. (Undergraduate level. An introduction to fundamental concepts in polymer chemistry.)

Hunter, R. J., *Zeta Potentials in Colloid Science: Principles and Applications*, Academic Press, London, 1981. (Advanced. A research-level monograph on electrokinetic phenomena and electroviscous and viscoelectric effects.)

Hunter, R., *Foundations of Colloid Science*, Vol. 2, Oxford University Press, Oxford, England, 1990. (Mostly graduate level. Chapter 18 presents a graduate-level introduction to rheology of dispersions.)

Russel, W. B., Saville, D., and Schowalter, W., *Colloidal Dispersions*, Cambridge University Press, Cambridge, England, 1989. (This graduate-level treatment of colloidal dispersions devotes Chapter 14 to a brief overview of rheology and discusses the physical significance of the dimensionless groups relevant in rheology. Table 14.2 summarizes the relevant dimensionless groups.)

Other References

Booth, F., *Proc. Roy. Soc.*, **A205**, 533 (1950).
Cohn, E. J., and Edsall, J. T., *Proteins, Amino Acids and Peptides*, ACS Monograph, Hafner, New York, 1965.
de Kruif, C. G., van Iersel, E. M. F., Vrij, A., and Russel, W. B., *J. Chem. Phys.*, **83**, 4717 (1985).
Einstein, A., *Investigations on the Theory of the Brownian Movement*, Dover, New York, 1956.
Flory, P. J., *Principles of Polymer Chemistry*, Cornell University Press, Ithaca, NY, 1953.
Frisch, H. L., and Simha, R., The Viscosity of Colloidal Suspensions and Macromolecular Solutions. In *Rheology*, Vol. 1 (F. R Eirich, Ed.), Academic Press, New York, 1956.
Hermans, J. J., *Flow Properties of Disperse Systems*, North-Holland, Amsterdam, Netherlands, 1953.
Krasny-Ergen, B., *Kolloid Z.*, **74**, 172 (1936).
Krieger, I. M., *Adv. Colloid Interface Sci.*, **3**, 111 (1972).
Reimschussle, H. K., and Dege, G. J., *J. Polym. Sci.*, **9**, 2343 (1971).
Richards, E. G., *An Introduction to the Physical Properties of Large Molecules in Solution*, Cambridge University Press, Cambridge, England, 1980.
Russel, W. B., *J. Fluid Mech.*, **85**, 673 (1978a).
Russel, W. B., *J. Fluid Mech.*, **85**, 209 (1978b).
Russel, W. B., *Adv. Colloid Interface Sci.*, **24**, 287 (1980).
Stone-Masui, J., and Watillon, A., *J. Colloid and Interface Sci.*, **28**, 187 (1968).
Tadros, Th. F. (Ed.), *Surfactants*, Academic Press, London, 1984.
Tanford, C., *Physical Chemistry of Macromolecules*, Wiley, New York, 1961.

PROBLEMS

1. Gillespie and Wiley* used a cone-and-plate viscometer to measure F/A versus dv/dx for dispersions of silica and cross-linked polystyrene in dioctyl phthalate. At a volume fraction of 0.35 for both solids, the following results were obtained:

Silica	$F/A \times 10^{-3}$ (dyne cm^{-2})	2.2	1.4	1.0	0.50	0.25
	dv/dx (s^{-1})	500	325	235	125	60
Polystyrene	$F/A \times 10^{-3}$ (dyne cm^{-2})	1.6	0.80	0.55	0.25	
	dv/dx (s^{-1})	500	235	160	100	

   Use these data to determine either η or the yield value for these dispersions depending on whether or not the system is Newtonian. Are these results consistent with the fact that the axial ratio was nearer unity and the particle size distribution narrower for the polystyrene than the silica? Explain.

2. An aqueous polybutyl methacrylate latex (average radius = 200 Å) has a viscosity of 50,500 cP at $\phi = 0.25$ and a viscosity of 36.7 cP at the same rate of shear when 1.71×10^{-5} g NaCl is added per gram of polymer.† Assuming that these charged particles must be surrounded by a layer of dissolved ions in solution, what conclusions can you draw about the dependence of the thickness of this layer of ions on the electrolyte content of the continuous phase?

3. A copolymer of vinylpyridine and methacrylic acid (62 and 38 mole%, respectively, in polymer) was studied in 90% methanol–10% water solution. The specific viscosity was measured‡ as a function of added NaOH or HCl, and the following results were obtained:

η_{sp}	0.1	0.25	0.28	0.24	0.21	0.21	0.20
mEq added per gram	0	2	4	6	2	4	6
		HCl			NaOH		

*Gillespie, T., and Wiley, R., *J. Phys. Chem.*, **66**, 1077 (1962).
†Brodnyan, J. G., and Kelley, E. L., *J. Colloid Sci.*, **19**, 488 (1964).
‡Alfrey, T., Jr., and Morawitz, H., *J. Am. Chem. Soc.*, **74**, 436 (1952).

Discuss these results in terms of the apparent effect of acid and base on the configuration of the polymer chain. Be sure your explanation is consistent with the chemical nature of the copolymer.

4. A dispersion of polydisperse spheres shows a relative viscosity of about 2.6 at a volume fraction of about 0.31.* Compare this result with the predictions of Equation (48).

5. The viscosity of cross-linked polymethyl methacrylate spheres in benzene was measured† and found to be

ϕ	0.050	0.035	0.028	0.019	0.014	0.010
η/η_0	2.15	1.61	1.41	1.19	1.18	1.12

Calculate $[\eta]$ for these spheres. Is there evidence in these results that the polymer particles may be swollen by the solvent? Explain.

6. The viscosity of uniform, cross-linked polystyrene spheres of two different diameters was measured in benzyl alcohol at 30°C:‡

$d = 0.382\ \mu$m	ϕ	0.013	0.030	0.059	0.075
	η_{sp}	0.036	0.086	0.178	0.233
$d = 0.433\ \mu$m	ϕ	0.02	0.04	0.08	
	η_{sp}	0.056	0.116	0.251	

Evaluate the intrinsic viscosity for each size of spherical particle and comment on the results in terms of the Einstein prediction that $[\eta]$ should be independent of particle size. Is the fact that benzyl alcohol is only a moderately good solvent for linear polystyrene consistent with the observed deviation between the experimental and theoretical values for $[\eta]$? Explain.

7. Criticize or defend the following proposition using the data of Problem 6 and the fact that the intrinsic viscosities of cross-linked polystyrene spheres in solvents such as benzene and CCl_4 (good solvents for linear polystyrene) lie in the range 5.8 to 7.5: Cross-linked polystyrene spheres swell by imbibing solvent, the effect being more extensive the better the solvent properties of the continuous phase for the non-cross-linked polymer.

8. Use the following data to evaluate the Mark-Houwink a and k constants for cellulose acetate in acetone at 25°C:§

M_n	c(g dl^{-1})	η_{sp}
130,000	0.094	0.289
	0.273	0.990
	0.546	2.770
86,000	0.114	0.286
	0.351	1.100
	0.703	3.120
76,000	0.118	0.247
	0.353	0.890
	0.775	2.700
61,000	0.138	0.239
	0.275	0.520
	0.428	0.880
48,000	0.152	0.209
	0.303	0.450
	0.684	1.230

*Eilers, H., *Kolloid Z.*, **97**, 313 (1941).

†Kose, A., and Hachisu, S., *J. Colloid Interface Sci.*, **46**, 460 (1974).

‡Papir, Y. S., and Krieger, I. M., *J. Colloid Interface Sci.*, **34**, 126 (1970).

§Sookne, A. M., and Harris, M., *Ind. Eng. Chem.*, **37**, 475 (1945).

9. Various molecular weight fractions of cellulose nitrate were dissolved in acetone, and the intrinsic viscosity was measured at 25°C:*

$M \times 10^{-3}$ (g mole^{-1})	77	89	273	360	400	640	846	1550	2510	2640
$[\eta]$(dl g^{-1})	1.23	1.45	3.54	5.50	6.50	10.6	14.9	30.3	31.0	36.3

Use these data to evaluate the constants k and a in the Staudinger-Mark-Houwink equation for this system.

10. The relative viscosity of solutions of cellulose nitrate in acetone was measured and extrapolated to zero rate of shear:*

η/η_0	1.45	1.53	1.67	1.89	2.31	3.41
c(g dl^{-1})	0.0151	0.0176	0.0212	0.0264	0.0352	0.0528

Use these data to evaluate the intrinsic viscosity for the polymer and use the a and k values from Problem 9 to calculate the molecular weight from the value of $[\eta]$.

11. The following intrinsic viscosity values of some high molecular weight polystyrene fractions have been reported†:

$\overline{M} \times 10^{-6}$ (g mole^{-1})	$[\eta]$ dl g^{-1})	
43.8	5.5	
27.4	4.4	Cyclohexane at 35.4°C
43.5	67.7	
26.8	36.5	Benzene at 40°C

Use these data to evaluate the constants in the Staudinger-Mark-Houwink equation. Are the values obtained consistent with the known facts that 35.4°C is the Flory (Θ) temperature for polystyrene in cyclohexane while benzene is a good solvent for polystyrene at 40°C.

12. Solutions of nylon-6,6 were studied in 90% formic acid solutions at 25°C, and the following data were obtained for two different molecular weight fractions:‡

c(g dl^{-1})	0.744	0.527	0.368	0.164
η_{sp}/c (dl g^{-1})	0.485	0.477	0.478	0.450

and

c (g dl^{-1})	0.742	0.640	0.537	0.436	0.332	0.225	0.132	0.058
η_{sp}/c (dl g^{-1})	0.897	0.892	0.886	0.876	0.864	0.847	0.805	0.778

Calculate the intrinsic viscosity for these two polymers. The Mark-Houwink constants for this system are known to be $a = 0.72$ and $k = 11 \times 10^{-4}$ dl g^{-1}; calculate the molecular weights of the two nylon fractions.

13. The radius of gyration of polymer coils can be determined independently from light scattering. Fox and Flory§ measured both R_g and $[\eta]$ for various molecular weight fractions of polystyrene in various solvents at several temperatures. The following results were obtained:

*Holtzer, A. M., Benoit, H., and Doty, P., *J. Phys. Chem.*, **58**, 624 (1954).
†McIntyre, D., Fetlers, L. J., and Slagowski, E., *Science*, **176**, 1041 (1972).
‡Taylor, G. B., *J. Am. Chem. Soc.*, **69**, 635 (1947).
§Fox, T. G., Jr., and Flory, P. J., *J. Am. Chem. Soc.*, **73**, 1915 (1951).

Solvent	T (°C)	$M \times 10^{-3}$ (g mole^{-1})	$[\eta]$ (dl g^{-1})	R_g(Å)
Methylethyl ketone	22	1760	1.65	437
	22	1620	1.61	414
	67	1620	1.50	400
	22	1320	1.40	367
	25	980	1.21	343
	22	940	1.17	306
	22	520	0.77	222
	25	318	0.60	194
	22	230	0.53	163
Dichloroethane	22	1780	2.60	576
	22	1620	2.78	545
	67	1620	2.83	529
	22	562	1.42	310
	22	520	1.38	278
Toluene	22	1620	3.45	527
	67	1620	3.42	523

Use these data to evaluate the factor of proportionality in Equation (93). Does this factor seem reasonably constant?

14. The intrinsic viscosity of a polystyrene solution in cyclohexane was measured* under theta conditions and found to be 0.078 dl g^{-1} for a polymer of $M = 8370$ g mole^{-1}. Use the average value of the factor of proportionality from Problem 13 to evaluate the radius of gyration of the polystyrene molecules in this solution. Calculate what this quantity is expected to be in terms of Equation (2.77), using twice the degree of polymerization as the number of steps in the random walk (since there are two C–C bonds per segment) and taking $0.154(2)^{1/2} = 0.218$ nm as the value of ℓ_0 (the C–C bond length corrected for tetrahedral bond angles). Briefly explain why the experimental value for R_g is larger than that calculated by Equation (2.77).

*Krigbaum, W. R., Mandelkern, L., and Flory, P. J., *J. Polym. Sci.*, **9**, 381 (1952).

5

Static and Dynamic Light Scattering and Other Radiation Scattering

An interesting and oft-investigated question, "What is the origin of light?" and the solution of it, has been repeatedly attempted, with no other result than to crowd our lunatic asylums with the would-be solvers. Hence, after fruitless attempts to suppress such investigations by making them liable to a heavy tax, the Legislature . . . absolutely prohibited them.

From Abbott's *Flatland*

5.1 INTRODUCTION

5.1a What Is Radiation Scattering?

In Chapter 1 we described dark-field microscopy in which particles too small for direct microscopic observation could be detected against a dark field by horizontal illumination. Airborne dust or smoke particles show in a beam of light in an otherwise dark room in the same way. In both cases, the particles interact with the light that strikes them and deflect some of that light from its original direction. We speak of this light as being "scattered." A whole assortment of optical phenomena related to this are generally known as *light scattering* effects.

It turns out that the intensity of scattered light at any angle depends on the wavelength of the incident light, the size and shape of the scattering particles, and the optical properties of the scatterers, as well as the angle of observation. Furthermore, the functional relationship among these variables is known, at least for spherical particles and other geometries, under certain circumstances. By applying these relationships to light scattering experiments, information about the particles responsible for the scattering can be deduced. A similar statement can be made about experiments based on other forms of radiation (e.g., x-ray or neutron scattering), although the mechanism of scattering depends on the particular type of radiation.

5.1b What Are Static and Dynamic Scattering and Why Are They Important?

Let us focus on light scattering. Light scattering can be classified as *static* or *dynamic* depending on how the intensity is measured. In *static light scattering* the time-averaged total intensity is measured as a function of scattering angle. We see in this chapter how the weight and a characteristic linear dimension of the particle may be determined for some systems from static light scattering. Moreover, information on the internal structure and shapes of the particles as well as *inter*particle structure can be deduced by measuring the angle dependence of the intensity.

In contrast, in *dynamic light scattering* (DLS) the temporal variation of the intensity is measured and is represented usually through what is known as the intensity autocorrelation function. The diffusion coefficients of the particles, particle size, and size distribution can be deduced from such measurements. There are many variations of dynamic light scattering, and