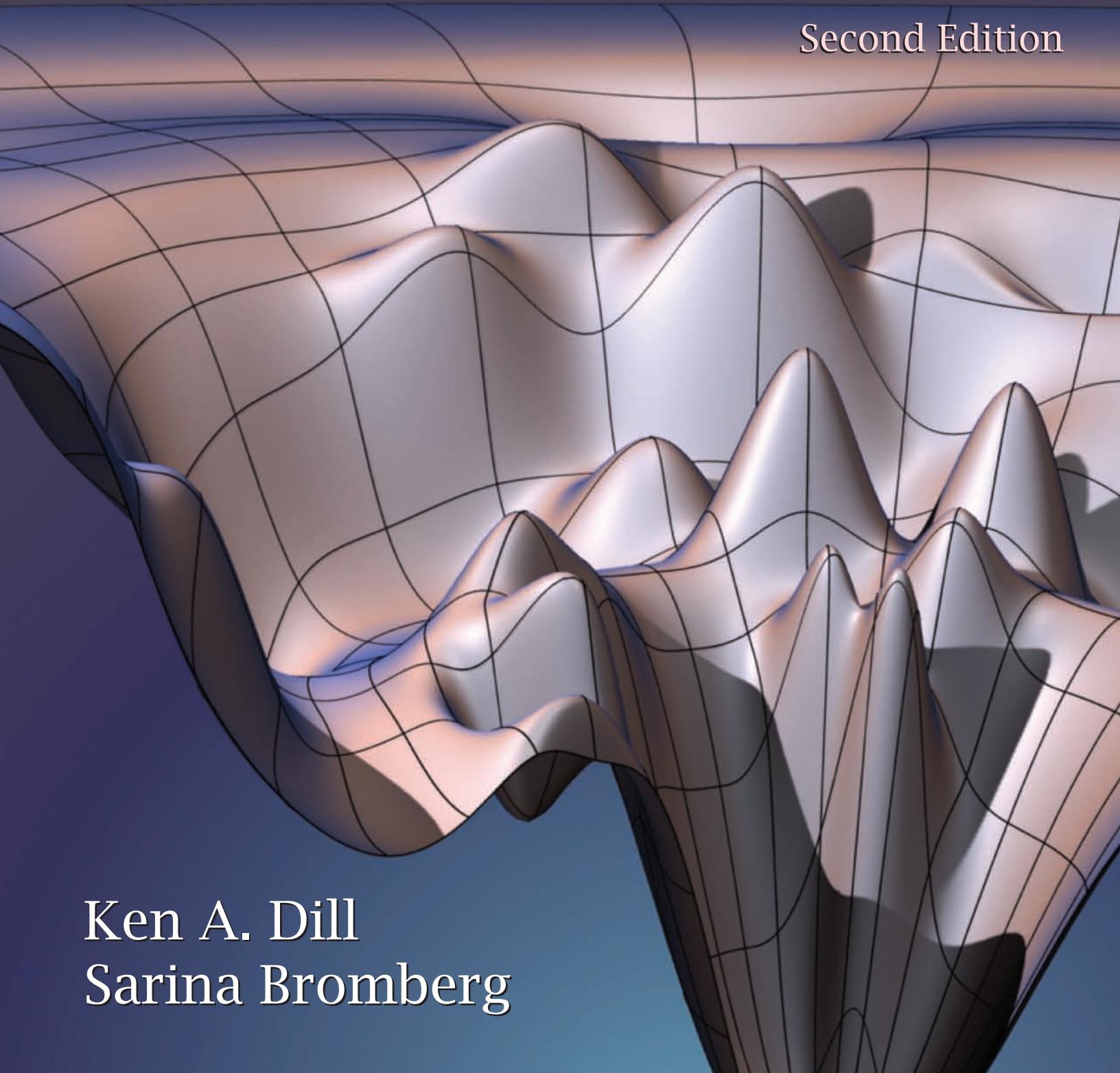


Molecular Driving Forces

Statistical Thermodynamics in Biology,
Chemistry, Physics, and Nanoscience

Second Edition



Ken A. Dill
Sarina Bromberg

This page is intentionally left blank.

Molecular Driving Forces

**Statistical Thermodynamics in
Biology, Chemistry, Physics,
and Nanoscience**

Second Edition

This page is intentionally left blank.

Molecular Driving Forces

**Statistical Thermodynamics in
Biology, Chemistry, Physics,
and Nanoscience**

Second Edition

Ken A. Dill Sarina Bromberg

With the assistance of Dirk Stigter on the Electrostatics chapters

 **Garland Science**
Taylor & Francis Group
LONDON AND NEW YORK

Garland Science

Vice President: Denise Schanck
Senior Editor: Michael Morales
Assistant Editor: Monica Toledo
Production Editor: H.M. (Mac) Clarke
Illustrator & Cover Design: Xavier Studio
Copyeditor: H.M. (Mac) Clarke
Typesetting: Techset Composition Limited
Proofreader: Sally Huish
Indexer: Merrall-Ross International, Ltd.

Ken A. Dill is Professor and Director of the Laufer Center for Physical and Quantitative Biology at Stony Brook University. He received his undergraduate training at MIT, his PhD from the University of California, San Diego, did postdoctoral work at Stanford, and was Professor at the University of California, San Francisco until 2010. A researcher in statistical mechanics and protein folding, he has been the President of the Biophysical Society and is a member of the US National Academy of Sciences.

Sarina Bromberg received her BFA at the Cooper Union for the Advancement of Science and Art, her PhD in molecular biophysics from Wesleyan University, and her postdoctoral training at the University of California, San Francisco. She writes, edits, and illustrates scientific textbooks.

Dirk Stigter (1924–2010) received his PhD from the State University of Utrecht, The Netherlands, and did postdoctoral work at the University of Oregon, Eugene.

©2011 by Garland Science, Taylor & Francis Group, LLC

This book contains information obtained from authentic and highly regarded sources. Reprinted material is quoted with permission, and sources are indicated. A wide variety of references are listed. Reasonable efforts have been made to publish reliable data and information, but the author and the publisher cannot assume responsibility for the validity of all materials or for the consequences of their use. All rights reserved. No part of this publication may be reproduced, stored in a retrieval system, or transmitted in any form or by any means—graphic, electronic, or mechanical, including photocopying, recording, taping, or information storage and retrieval systems—with permission of the copyright holder.

ISBN 978-0-8153-4430-8

Library of Congress Cataloging-in-Publication Data

Dill, Ken A.

Molecular driving forces : statistical thermodynamics in chemistry, physics, biology, and nanoscience / Ken Dill, Sarina Bromberg. -- 2nd ed.

p. cm.

Includes index.

ISBN 978-0-8153-4430-8 (pbk.)

1. Statistical thermodynamics. I. Bromberg, Sarina. II. Title.

QC311.5.D55 2010

536'.7--dc22

2010034322



Taylor & Francis Group, an informa business

Published by Garland Science, Taylor & Francis Group, LLC, an informa business,
270 Madison Avenue, New York, NY 10016, USA, and
2 Park Square, Milton Park, Abingdon, OX14 4RN, UK

Printed in the United States of America

15 14 13 12 11 10 9 8 7 6 5 4 3 2 1

Visit our web site at <http://www.garlandscience.com>

Dedicated to Austin, Peggy, Jim, Jolanda, Tyler, and Ryan
and
In memory of Dirk Stigter

This page is intentionally left blank.

Preface

What forces drive atoms and molecules to bind, to adsorb, to dissolve, to permeate membranes, to undergo chemical reactions, and to undergo conformational changes? This is a textbook on statistical thermodynamics. It describes the forces that govern molecular behavior. Statistical thermodynamics uses physical models, mathematical approximations, and empirical laws that are rooted in the language of entropy, distribution function, energy, heat capacity, free energy, and partition function, to predict the behaviors of molecules in physical, chemical, and biological systems.

This text is intended for graduate students and advanced undergraduates in physical chemistry, biochemistry, biophysics, bioengineering, polymer and materials science, pharmaceutical chemistry, chemical engineering, and environmental science.

We had three goals in mind as we wrote this book. First, we tried to make extensive connections with experiments and familiar contexts, to show the practical importance of this subject. We have included many applications in biology and polymer science, in addition to applications in more traditional areas of chemistry and physics. Second, we tried to make this book accessible to students with a variety of backgrounds. So, for example, we have included material on probabilities, approximations, partial derivatives, vector calculus, and the historical basis of thermodynamics. Third, we strove to find a vantage point from which the concepts are revealed in their simplest and most comprehensible forms. For this reason, we follow the axiomatic approach to thermodynamics developed by HB Callen, rather than the more traditional inductive approach; and the Maximum Entropy approach of Jaynes, Skilling, and Livesay, in preference to the Gibbs ensemble method. We have drawn from many excellent texts, particularly those by Callen, Hill, Atkins, Chandler, Kubo, Kittel & Kroemer, Carrington, Adkins, Weiss, Doi, Flory, and Berry, Rice, & Ross.

Our focus here is on molecular driving forces, which overlaps with—but is not identical to—the subject of thermodynamics. While the power of thermodynamics is its generality, the power of statistical thermodynamics is the insight it gives into microscopic interactions through the enterprise of model-making. A central theme of this book is that making models, even very simple ones, is a route to insight and to understanding how molecules work. A good theory, no matter how complex its mathematics, is usually rooted in some very simple physical idea.

Models are mental toys to guide our thinking. The most important ingredients in a good model are predictive power and insight into the causes of the predicted behavior. The more rigorous a model, the less room for ambiguity. But models don't need to be complicated to be useful. Many of the key insights in statistical mechanics have come from simplifications that may seem unrealistic at first glance: particles represented as perfect spheres with atomic detail left out, neglecting the presence of other particles, using crystal-like lattices of particles in liquids and polymers, and modeling polymer chains as random flights, etc. To borrow a quote, statistical thermodynamics has a history of what might be called the unreasonable effectiveness of unrealistic simplifications. Perhaps the classic example is the two-dimensional Ising model of magnets as two types of arrows, up spins or down spins, on square lattices. Lars Onsager's

famous solution to this highly simplified model was a major contribution to the modern revolution in our understanding of phase transitions and critical phenomena.

We begin with entropy. Chapter 1 gives the underpinnings in terms of probabilities and combinatorics. Simple models are used in Chapters 2 and 3 to show how entropy is a driving force. This motivates more detailed treatments throughout the text illustrating the Second Law of Thermodynamics and the concept of equilibrium. Chapter 4 lays out the mathematical foundations in multivariate calculus that are needed for the following chapters.

These threads culminate in Chapter 5, which defines the entropy and gives the Boltzmann distribution law, the lynch-pin of statistical thermodynamics. The key expressions, $S = k \ln W$ and $S = -k \sum_i p_i \ln p_i$, are shown to provide the mathematics for an extremum principle that underpins thermodynamics.

The principles of thermodynamics are described in Chapters 6–9. The statistical mechanics of simple systems follows in Chapters 10 and 11. While temperature and heat capacity are often regarded as needing no explanation (perhaps because they are so readily measured), Chapter 12 uses simple models to shed light on the physical basis of those properties. Chapter 13 applies the principles of statistical thermodynamics to chemical equilibria.

Chapters 14–16 develop simple models of liquids and solutions. We use lattice models here, rather than ideal solution theories, because such models give more microscopic insight into real molecules and into the solvation processes that are central to computational chemistry, biology, and materials science. For example, theories of mixtures often begin from the premise that Raoult's and Henry's laws are experimental facts. Our approach, instead, is to show why molecules are driven to obey these laws. An equally important reason for introducing lattice models here is as background. Lattice models are standard tools for treating complex systems: phase transitions and critical phenomena in Chapters 25 and 26, and polymer conformations in Chapters 32–34.

We explore the dynamic processes of diffusion, transport, and physical and chemical kinetics in Chapters 17–19, through Fick's law, the random-walk model, Poisson and waiting-time distributions, the Langevin model, Onsager relations, Brownian ratchets, time correlation functions and transition-state theory.

We treat electrostatics in Chapters 20–23. Our treatment is more extensive than in other physical chemistry texts because of the importance, in our view, of electrostatics in understanding the structures of proteins, nucleic acids, micelles, and membranes; for predicting protein and nucleic acid ligand interactions and the behaviors of ion channels; as well as for the classical areas of electrochemistry and colloid science. We explore acid-base and oxidation-reduction equilibria and develop the Nernst and Poisson-Boltzmann equations and the Born model, modern workhorses of quantitative biology. Chapter 24 describes intermolecular forces.

We describe simple models of complex systems, including polymers, colloids, surfaces, and catalysts. Chapters 25 and 26 focus on cooperativity: phase equilibria, solubilities, critical phenomena, and conformational transitions, described through mean-field theories, the Ising model, helix-coil model, and Landau theory. Chapters 27–29 describe binding polynomials and how proteins

and other biomolecules can function as molecular machines, essential to modern pharmaceutical science. Chapters 30 and 31 describe water, the hydrophobic effect, and ion solvation. And Chapters 32–34 focus on the conformations of polymers and biomolecules that give rise to the elasticity of rubber, the viscoelasticities of solutions, the immiscibilities of polymers, reptational motion, and the folding of proteins and RNA molecules.

For those familiar with the First Edition, this Second Edition includes three new chapters: (1) “Microscopic Dynamics” describes the microscopic basis for dynamical processes and single-molecule processes; (2) “Molecular Machines” describes how nanoscale machines and engines work, and how energy is transduced into actions; and (3) “The Logic of Thermodynamics” was added in response to requests from users of the First Edition to describe more of classical thermodynamics, including more details on the distinction between reversible and irreversible processes. New practical applications, examples, illustrations, and end-of-chapter questions are integrated throughout the revised and updated text, exploring topics in biology, environmental and energy science, and nanotechnology.

Instructor’s Resources

The following teaching supplements are available to qualified instructors:

- The Art of *Molecular Driving Forces*, Second Edition

The images and tables from the book are available in two convenient formats: PowerPoint® and JPEG. The PowerPoint slides can be customized for lectures.

- Solution’s Manual

The solutions to all problems are available in PDF format.

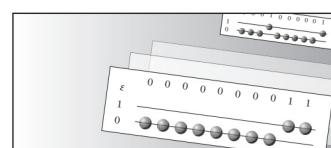
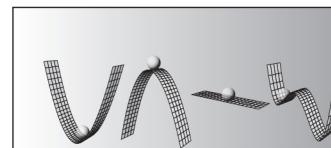
To request these resources, please email science@garland.com, and include full contact and course information.

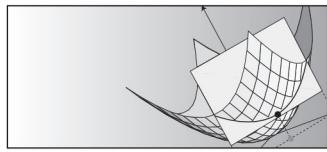
Acknowledgements

We owe a great debt to Neeraj Agrawal, Jose Luis Alonso, Brad Anderson, Jayanth Banavar, Victor Bloomfield, Eric B Brauns, Joseph M Brom, Kevin Cahill, Robert Cantor, Silvia Cavagnero and her classes, Hue Sun Chan, John Chodera, Pietro Cicuta, Roger Coziol, Margaret Daugherty, Siavash Dejgosha, Jim Dill, Roland Dunbrack, Isi Dunietz, Pablo Echenique, Burak Erman, Joao Paulo Ferreira, Kings Ghosh, Guido Guidotti, Steve Harvey, Tony Haymet, Jenny Hsiao, Wonmuk Hwang, Sune Jespersen, Suckjoon Jun, Morteza Khabiry, Peter Kollman, ID Kuntz, Michael Laskowski, Bernard J Leyh, Yong Liu, Cynthia Lo, Jodie Lutkenhaus, Alenka Luzar, Jeffrey D Madura, Amos Maritan, Andy McCammon, Chris Miller, Alex Mogilner, Jerome Nilmeier, Kira Novakofski, Terry Oas and Eric Toone and their students, Folusho Oyerokun, Jose Onuchic, Matteo Palassini, Andrea Peterson, George Peterson, Rob Phillips, Jed Pitera, Tom Powers, Steve Presse, Ravi Radhakrishnan, Miranda Robertson, Stuart Rothstein, Jonathan N Sachs, Samuel Sia, Kimberly Hamad-Schifferli, Kim Sharp, Krista Shipley, Noel Southall, Sue Stanton, Peter Tielemans, Tom Truskett, Vojko Vlachy, John Van Drie, Peter von Hippel, Hong Qian, Ben Sellers, David Spellmeyer, Bill Swope, Roger Tsien, Arieh Warshel, Ben Widom, Dave Wu, Ron Zhang, and Bruno Zimm, for very helpful comments on this text. We owe special thanks to Jan WH Schreurs, Eugene Stanley, John Schellman, and Judy Herzfeld, for careful reading and detailed criticism of large parts of various drafts. We are grateful to Richard Shafer and Ron Siegel, who in co-teaching this course have contributed considerable improvements, and to the UCSF graduate students of this course over the past several years, who have helped form this material and correct mistakes. We thank Claudia Johnson, who created the original course notes manuscript—with the good spirits and patience of a saint. We are deeply grateful to the talented people who worked so hard on book production: Larry Schweitzer, Danny Heap, and Patricia Monohon, who keyboarded, coded, extended, and converted the manuscript into this book format for the First Edition; to Jolanda Schreurs, who worked tirelessly on the graphics for the First Edition; to Matt Hasselmann and Ken Probst of Xavier Studio for work on graphics for the Second Edition; to Mac Clarke, Georgina Lucas, and Denise Schanck, who patiently and professionally saw us through to bound books on the Second Edition; and to Mike Morales and Adam Sendroff, who have managed over many years to help us with all matters of publication with great patience and exceptional commitment. The preparation of the First Edition of this text was partially funded by a grant from the Polymer Education Committee of the Divisions of Polymer Chemistry and Polymeric Materials of the American Chemical Society.

Contents

Preface	vii
Acknowledgements	x
1 Principles of Probability	1
The Principles of Probability Are the Foundations of Entropy	1
What Is Probability?	2
The Rules of Probability Are Recipes for Drawing	
Consistent Inferences	3
Correlated Events Are Described by Conditional Probabilities	7
Combinatorics Describes How to Count Events	10
Collections of Probabilities Are Described by Distribution	
Functions	14
Distribution Functions Have Average Values and	
Standard Deviations	18
Summary	22
Problems	25
References	27
Suggested Reading	28
2 Extremum Principles Predict Equilibria	29
What Are Extremum Principles?	29
What Is a State of Equilibrium?	30
An Extremum Principle: Maximizing Multiplicity Predicts the	
Most Probable Outcomes	32
Simple Models Show How Systems Tend Toward Disorder	33
Summary	37
Problems	38
Suggested Reading	38
3 Heat, Work, & Energy	39
Energy Is a Capacity to Store and Move	
Work and Heat	39
Energy Is Ubiquitous and Important	39
Some Physical Quantities Obey Conservation Laws	45
Heat Was Thought to Be a Fluid	47
The First Law of Thermodynamics: Heat Plus Work Is Conserved	49
Atoms and Molecules Have Energies	50
Why Does Heat Flow?	52
The Second Law of Thermodynamics Is an Extremum Principle	55
Summary	55
Problems	56
References	56
Suggested Reading	56

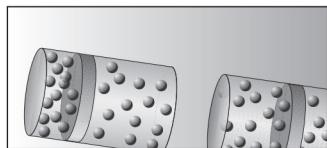




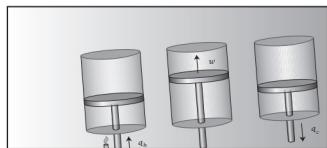
4 Math Tools: Multivariate Calculus	59
Multivariate Calculus Applies to Functions of Multiple Variables	59
Partial Derivatives Are the Slopes of the Tangents to Multivariate Functions	60
The Extrema of Multivariate Functions Occur Where the Partial Derivatives Are Zero	63
How Can You Find the Extrema of Multivariate Functions That Are Subject to Constraints?	66
Integrating Multivariate Functions Sometimes Depends on the Pathway of Integration, and Sometimes It Doesn't	72
The Chain Rule Applies to Functions of Several Variables	75
Summary	77
Problems	78
References	79
Suggested Reading	79



5 Entropy & the Boltzmann Law	81
What Is Entropy?	81
Maximum Entropy Predicts Flat Distributions When There Are No Constraints	85
Maximum Entropy Predicts Exponential Distributions When There Are Constraints	86
What Is the Reason for Maximizing the Entropy?	90
Summary	90
Problems	91
References	92
Suggested Reading	92

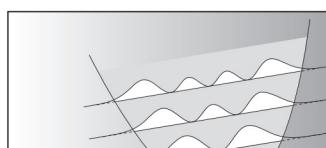
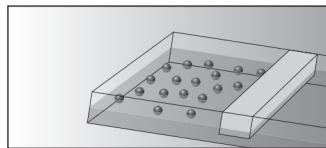
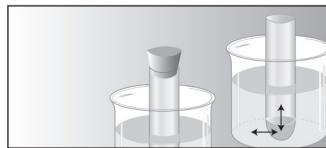


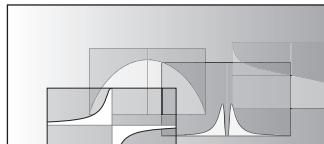
6 Thermodynamic Driving Forces	93
Thermodynamics Is Two Laws and a Little Calculus	93
The Fundamental Thermodynamic Equations, $S = S(U, V, N)$ or $U = U(S, V, N)$, Predict Equilibria	95
The Fundamental Equations Define the Thermodynamic Driving Forces	96
The Second Law Predicts Thermal, Mechanical, and Chemical Equilibria	99
Summary	106
Problems	107
Suggested Reading	107



7 The Logic of Thermodynamics	109
Thermodynamics Helps You Reason About Heat, Work, Processes, and Cycles	109
We Start with a Few Basic Premises	110
Four Different Processes Can Change an Ideal Gas	114
<i>Pathways</i> Are Combinations of Processes	117
A <i>Cycle</i> is a Set of Outbound and Returning Pathways	118
The Maximum Possible Work Is Achieved via Reversible Processes . . .	122
Why Is There an <i>Absolute</i> Temperature Scale?	128
Summary	128

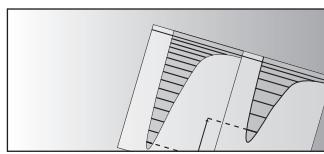
Problems	129
Suggested Reading	130
8 Laboratory Conditions & Free Energies	131
Switch from Maximum Entropy to Minimum Free Energy	131
Free Energy Defines Another Extremum Principle	132
Thermochemistry: The Enthalpy of a Molecule Is the Sum of the Enthalpies of Its Covalent Bonds	144
Summary	145
Problems	146
Suggested Reading	147
9 Maxwell's Relations & Mixtures	149
Thermodynamics Applies to Various Types of Forces	149
Maxwell's Relations Interrelate Partial Derivatives	152
Susceptibilities Are Measurable and Useful	154
Multicomponent Systems Have Partial Molar Properties	160
Summary	165
Problems	166
Suggested Reading	168
10 The Boltzmann Distribution Law	169
Statistical Mechanics Gives Probability Distributions for Atoms and Molecules	169
The Boltzmann Distribution Law Describes the Equilibria Among Atoms and Molecules	171
What Does a Partition Function Tell You?	176
Thermodynamic Properties Can Be Predicted from Partition Functions	182
Summary	189
Problems	190
References	192
Suggested Reading	192
11 The Statistical Mechanics of Simple Gases & Solids	193
Statistical Mechanics Predicts Macroscopic Properties from Atomic Structures	193
The Particle in a Box Is the Quantum Mechanical Model for Translational Motion	195
Vibrations Can Be Treated by the Harmonic Oscillator Model	201
Rotations Can Be Treated by the Rigid Rotor Model	203
The Electronic Partition Function	205
The Total Partition Function Predicts the Properties of Molecules	207
Ideal Gas Properties Are Predicted by Quantum Mechanics	207
The Equipartition Theorem Says That Energies Are Uniformly Distributed to Each Degree of Freedom	212
Summary	216
Problems	218
References	220
Suggested Reading	220



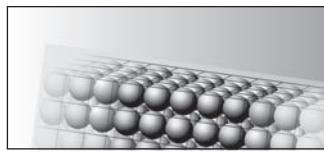


12 What Is Temperature? What Is Heat Capacity?	221
A Microscopic Perspective on Temperature and Heat Capacity	221
A Graphical Procedure Shows the Steps from Fundamental Functions to Experimental Measurables	225

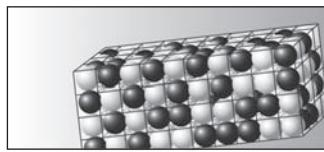
What Drives Heat Exchange?	227
The Heat Capacity Is Proportional to the Energy Fluctuations in a System	228
Summary	232
Problems	233
References	234
Suggested Reading	234



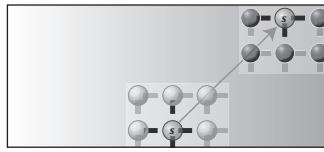
13 Chemical Equilibria	235
Simple Chemical Equilibria Can Be Predicted from Atomic Structures	235
Le Chatelier's Principle Describes the Response to Perturbation from Equilibrium	243
The Temperature Dependence of Equilibrium Is Described by the van't Hoff Equation	244
Summary	249
Problems	250
Suggested Reading	251



14 Equilibria Between Liquids, Solids, & Gases	253
Phase Equilibria Are Described by the Chemical Potential	253
The Clapeyron Equation Describes $p(T)$ at Phase Equilibrium	257
How Do Refrigerators and Heat Pumps Work?	262
Surface Tension Describes the Equilibrium Between Molecules at the Surface and in the Bulk	265
Summary	267
Problems	268
Suggested Reading	268

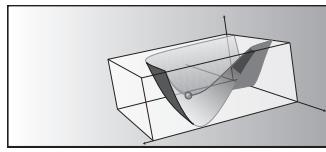
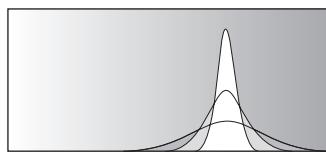


15 Solutions & Mixtures	269
A Lattice Model Describes Mixtures	269
Interfacial Tension Describes the Free Energy of Creating Surface Area	276
What Have We Left Out?	278
Summary	279
Problems	280
References	280
Suggested Reading	280



16 Solvation & the Transfer of Molecules Between Phases	283
The Chemical Potential Describes the Tendency of Molecules to Exchange and Partition	283
Solvation Is the Transfer of Molecules Between Vapor and Liquid Phases	284
A Thermodynamic Model Defines the Activity and Activity Coefficient	287

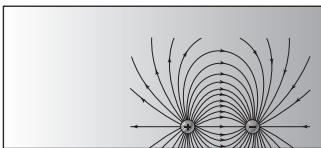
Adding Solute Can Raise the Boiling Temperature of the Solvent	289
Adding Solute Can Lower the Freezing Temperature of a Solvent	291
Adding Solute on One Side of a Semipermeable Membrane Causes an Osmotic Pressure	293
Solutes Can Transfer and Partition from One Medium to Another	295
Dimerization in Solution Involves Desolvation	299
Summary	303
Problems	304
References	306
Suggested Reading	306
17 Physical Kinetics: Diffusion, Permeation, & Flow	307
Applied Forces Cause Liquids to Flow	307
Fick's Law and Fourier's Law: Gradients of Particles or Heat Cause Flows	308
The Diffusion Equation Describes How Concentration Gradients Change Over Time and Space	310
The Smoluchowski Equation Describes Particles Driven by Both Applied Forces and Diffusion	317
The Einstein-Smoluchowski Equation Relates Diffusion and Friction .	318
Physical and Chemical States Evolve Toward Equilibrium	320
Sources and Sinks Also Contribute to Flows: Examples from Population Biology	321
Onsager Reciprocal Relations Describe Coupled Flows	324
Summary	325
Problems	326
References	327
Suggested Reading	327
18 Microscopic Dynamics	329
The Dynamical Laws Arise from Stochastic Processes	329
Diffusing Particles Can be Modeled as Random Flights	329
Inter-Arrival Rates Are Described by Poisson and Waiting-Time Distributions	340
Single-Particle Kinetics is Described by Trajectories	343
Master Equations Are Differential Equations for Markov Dynamics .	345
Brownian Ratchets Convert Binding and Release Events into Directed Motion	348
The Fluctuation-Dissipation Theorem Relates Equilibrium Fluctuations to the Rate of Approach to Equilibrium	351
Summary	353
Problems	354
References	354
Suggested Reading	354
19 Chemical Kinetics & Transition States	357
Chemical Reaction Rates Depend on Temperature	357
The Mass Action Laws Describe Mechanisms in Chemical Kinetics .	357
At Equilibrium, Rates Obey Detailed Balance	358
Chemical Reactions Depend Strongly on Temperature	360



Activated Processes Can Be Modeled by Transition-State Theory	363
Catalysts Speed Up Chemical Reactions	372
The Brønsted Law of Acid and Base Catalysis: The Stronger the Acid, the Faster the Reaction it Catalyzes	375
Funnel Landscapes Describe Diffusion and Polymer Folding Processes	379
Summary	380
Problems	381
References	383
Suggested Reading	384

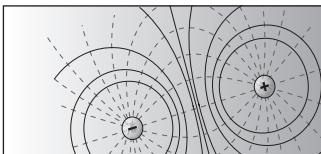
20 Coulomb's Law of Electrostatic Forces 385

Charge Interactions Are Described by Coulomb's Law	385
Charge Interactions Are Long-Ranged	387
Charge Interactions Are Weaker in Media	389
Electrostatic Forces Add Like Vectors	392
What Is an Electrostatic Field?	393
Electric Fields Have Fluxes	395
Summary	401
Problems	402
References	402
Suggested Reading	402



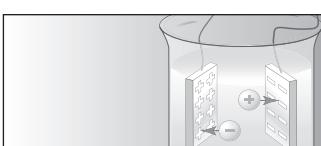
21 The Electrostatic Potential 403

We Switch from Fields (Forces) to Electrostatic Potentials (Energies)	403
Dipoles Are Equal and Opposite Charges Separated by a Distance	409
Voltage-Gated Ion Channels May Be Modeled as Orienting Dipoles	411
Poisson's Equation Is Used to Compute the Potential for Any Constellation of Charges	414
You Can Model Ions Near Interfaces Using Image Charges	418
Summary	421
Problems	422
References	423
Suggested Reading	423

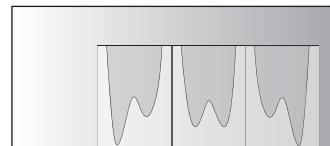
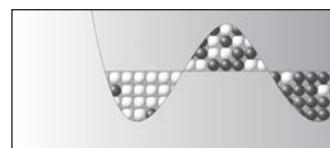
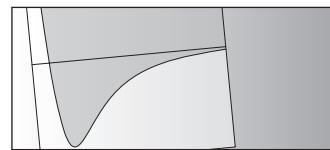
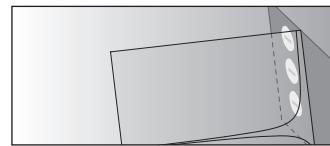


22 Electrochemical Equilibria 425

Electrochemical Potentials Describe Equilibria in Solutions of Charged Species	425
The Nernst Equation: Electrostatic Fields Drive Ion Distributions	426
In Redox Reactions, Electrons Hop from One Molecule to Another	429
Selective Ion Flow Through Membranes Can Create a Voltage	437
Acid-Base Equilibria: Chemical Reactions Are Affected by Electrostatic Interactions	438
Acid-Base Equilibria Are Shifted by the Presence of Charged Surfaces	439
Gradients of Electrostatic Potential Cause Ion Flows	441
Creating a Charge Distribution Costs Free Energy	444
Summary	451
Problems	452



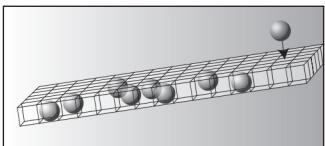
References	453
Suggested Reading	453
23 Salt Ions Shield Charged Objects in Solution	455
Salts Dissociate into Mobile Ions and Shield Charged Objects	
In Water	455
Electrolytes Are Strong or Weak, Depending on Whether	
They Fully Dissociate in Water	462
Summary	467
Problems	468
References	469
Suggested Reading	469
24 Intermolecular Interactions	471
Molecules Repel Each Other at Very Short Range	
and Attract at Longer Distances	471
Short-Ranged Attractions Can Be Explained as	
Electrostatic Interactions	472
The van der Waals Gas Model Accounts for Intermolecular	
Interactions	479
Radial Distribution Functions Describe Structural	
Correlations in Liquids	483
The Lattice Model Contact Energy w	
Approximates Intermolecular Interactions	485
Summary	486
Problems	487
References	488
Suggested Reading	488
25 Phase Transitions	489
Two States Can Be Stable at the Same Time	489
Phase Diagrams Describe What Phases Are Stable Under	
Which Conditions	492
Peaks on Energy Landscapes Are Points of Instability	497
Thermal Phase Equilibria Arise from a Balance between Energies	
and Entropies	500
The Spinodal Curve Describes the Limit of Metastability	502
The Critical Point Is Where Coexisting Phases Merge	504
The Principles of Boiling Are Related to the Principles	
of Immiscibility	505
Boiling a Liquid Mixture Involves Two Types of Phase Transition	510
Summary	512
Problems	514
Suggested Reading	517
26 Cooperativity: the Helix-Coil, Ising, & Landau Models	519
Abrupt Transitions Are Common in Nature	519
Transitions and Critical Points Are Universal	519
The Landau Model Is a General Description of Phase	
Transitions and Critical Exponents	522



The Ising Model Describes Magnetization	525
Polymers Undergo Helix-Coil Transitions	527
The Kinetics of Phase Transitions Can Be Controlled by Nucleation	535
Summary	536
Problems	537
References	538
Suggested Reading	539

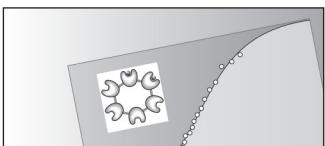
27 Adsorption, Binding, & Catalysis 541

Binding and Adsorption Processes Are Saturable	541
The Langmuir Model Describes Adsorption of Gas Molecules on a Surface	541
The Langmuir Model Also Treats Binding and Saturation in Solution	546
The Independent-Site Model Describes the Principle of Adsorption Chromatography	548
The Michaelis-Menten Model Describes Saturation in Rate Processes	549
Sabatier's Principle: Catalysts Should Bind Neither Too Tightly Nor Too Weakly	552
Summary	554
Problems	555
References	556
Suggested Reading	556



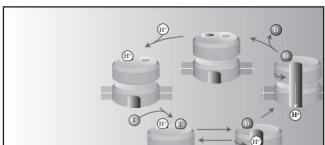
28 Multi-site & Cooperative Ligand Binding 559

Binding Polynomials Are Used to Compute Binding Curves	560
The Simplest Model of Binding Cooperativity Involves Two Binding Sites	562
'Cooperativity' Describes the Depletion of Intermediate States	565
Binding Polynomials Can Be Constructed Using the Addition and Multiplication Rules of Probability	567
Aggregation and Micellization Are Cooperative Processes	570
The Model of McGhee and von Hippel Treats Ligands that Crowd Out Other Ligands	575
The Grand Canonical Ensemble Gives the Basis for Binding Polynomials	578
Summary	581
Problems	582
References	583
Suggested Reading	584

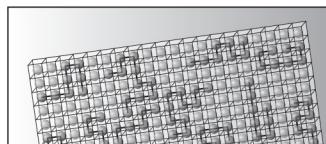
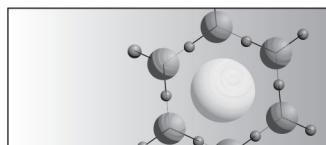
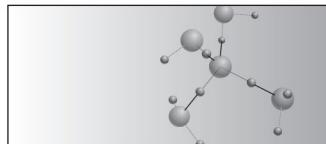


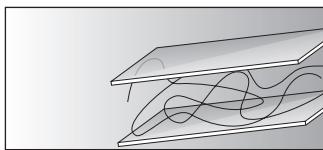
29 Bio & Nano Machines 585

Biochemical Machines Are Powered by Coupled Binding Processes	585
Oxygen Binding to Hemoglobin Is a Cooperative Process	586
Binding Polynomials Can Treat Multiple Types of Ligand	590
Rates Can Often Be Treated by Using Binding Polynomials	594
The Molecular Logic of Biology Is Encoded in Coupled-Binding Actions	595
Biochemical Engines Harness Downhill Processes to Drive Uphill Processes	598
Summary	608



Problems	609
References	612
Suggested Reading	612
30 Water	615
Water Has Anomalous Properties	615
Hydrogen Bonds Are a Defining Feature of Water Molecules	615
Pure Water Has Anomalous Properties	620
Summary	627
Problems	628
References	628
Suggested Reading	628
31 Water as a Solvent	629
Oil and Water Don't Mix: the Hydrophobic Effect	629
The Signature of Hydrophobicity Is Its Temperature Dependence	630
Alcohols Constrict the Volumes in Mixtures with Water	636
Ions Can Make or Break Water Structure	636
Ion Pairing Preferences in Water Depend on Charge Densities	639
Summary	640
References	641
Suggested Reading	642
32 Polymer Solutions	643
Polymer Properties Are Governed by Distribution Functions	643
Polymers Have Distributions of Conformations	643
Polymer Solutions Differ from Small-Molecule Solutions	644
The Flory-Huggins Model Describes Polymer Solution Thermodynamics	646
Flory-Huggins Theory Predicts Nonideal Colligative Properties for Polymer Solutions	651
The Phase Behavior of Polymers Differs from that of Small Molecules	651
The Flory Theorem Says that the Intrinsic Conformations of a Chain Molecule Are Not Perturbed by Surrounding Chains	655
Summary	656
Problems	657
Suggested Reading	657
33 Polymer Elasticity & Collapse	659
Polymeric Materials Are Often Elastic	659
Real Chain Molecules Have Bending Correlations and Stiffness	662
Random-Flight Chain Conformations Are Described by the Gaussian Distribution Function	665
Polymer Elasticity Follows Hooke's Law	666
The Elasticity of Rubbery Materials Results from the Sum of Chain Entropies	670
Polymers Expand in Good Solvents, Are Random Flights in θ Solvents, and Collapse In Poor Solvents	673
Polymer Solutions Exert Osmotic Pressures	678





Polymer Radius Depends on Polymer Concentration: the Semi-Dilute Cross-Over Regime	681
Summary	683
Problems	684
References	684
Suggested Reading	684

34 Polymers Resist Confinement & Deformation 685

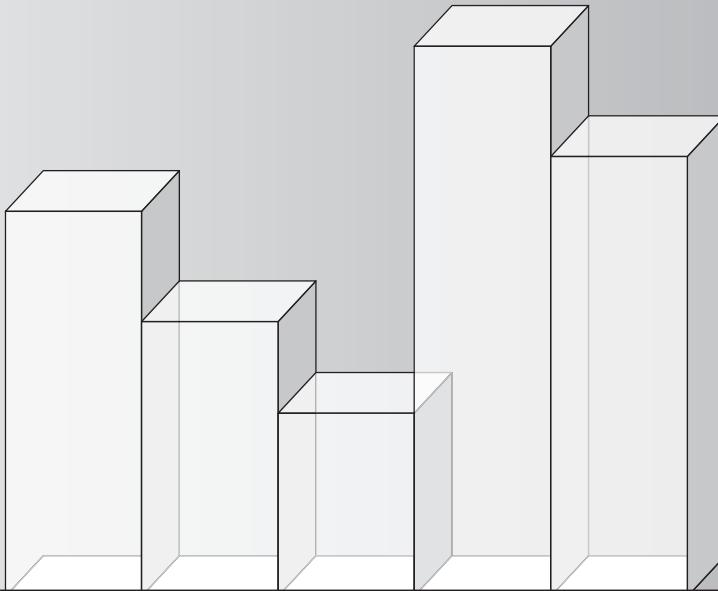
'Excluded Volume' Describes the Large Volume Inside a Polymer Conformation that Is Inaccessible to Other Chains	685
Chain Conformations Are Perturbed Near Surfaces	686
Polymer Conformations Can Be Described by the Diffusion Equation	689
Polymers Tend to Avoid Confined Spaces	691
The Rouse-Zimm Model Describes the Dynamics of Viscoelastic Fluids	693
Reptation Describes the Snakelike Motions of Chains in Melts and Gels	696
Summary	697
Problems	698
References	698
Suggested Reading	698

Appendices

A: How to Make Approximations	699
B: Stirling's Approximation	703
C: A Math Trick for Getting Averages from Derivatives	705
D: Proof of the Euler Reciprocal Relationship	707
E: A Principle of Fair Apportionment Leads to the Function $S = k \ln W$	709
F: Legendre Transforms	717
G: Vectors Describe Forces & Flows	719
H: Table of Constants	731
I: Table of Units	733
J: Useful Taylor Series Expansions	735
K: Useful Integrals	737
L: Multiples of Units, Their Names, & Symbols	739

Index 741

1 Principles of Probability



The Principles of Probability Are the Foundations of Entropy

Fluids flow, boil, freeze, and evaporate. Solids melt and deform. Oil and water don't mix. Metals and semiconductors conduct electricity. Crystals grow. Chemicals react and rearrange, take up heat, and give it off. Rubber stretches and retracts. Proteins catalyze biological reactions. What forces drive these processes? This question is addressed by statistical thermodynamics, a set of tools for modeling molecular forces and behavior, and a language for interpreting experiments.

The challenge in understanding these behaviors is that the properties that can be measured and controlled, such as density, temperature, pressure, heat capacity, molecular radius, or equilibrium constants, do not predict the tendencies and equilibria of systems in a simple and direct way. To predict equilibria, we must step into a different world, where we use the language of *energy*, *entropy*, *enthalpy*, and *free energy*. Measuring the density of liquid water just below its boiling temperature does not hint at the surprise that, just a few degrees higher, above the boiling temperature, the density suddenly drops more than a thousandfold. To predict density changes and other measurable properties, you need to know about the driving forces, the entropies and energies. We begin with entropy.

Entropy is one of the most fundamental concepts in statistical thermodynamics. It describes the tendency of matter toward disorder. Entropy explains

how expanding gases drive car engines, liquids mix, rubber bands retract, heat flows from hot objects to cold objects, and protein molecules tangle together in some disease states. The concepts that we introduce in this chapter, *probability*, *multiplicity*, *combinatorics*, *averages*, and *distribution functions*, provide a foundation for describing entropy.

What Is Probability?

Here are two statements of probability. In 1990, the probability that a person in the United States was a scientist or an engineer was 1/250. That is, there were about a million scientists and engineers out of a total of about 250 million people. In 1992, the probability that a child under 13 years old in the United States ate a fast-food hamburger on any given day was 1/30 [1].

Let's generalize. Suppose that the possible outcomes fall into categories *A*, *B*, or *C*. 'Event' and 'outcome' are generic terms. An event might be the flipping of a coin, resulting in heads or tails. Alternatively, it might be one of the possible conformations of a molecule. Suppose that outcome *A* occurs 20% of the time, *B* 50% of the time, and *C* 30% of the time. Then the probability of *A* is 0.20, the probability of *B* is 0.50, and the probability of *C* is 0.30.

The **definition of probability** is as follows: If *N* is the total number of possible outcomes, and n_A of the outcomes fall into category *A*, then p_A , the probability of outcome *A*, is

$$p_A = \left(\frac{n_A}{N} \right). \quad (1.1)$$

Probabilities are quantities in the range from zero to one. If only one outcome is possible, the process is *deterministic*—the outcome has a probability of one. An outcome that never occurs has a probability of zero.

Probabilities can be computed for different combinations of events. Consider one roll of a six-sided die, for example (die, unfortunately, is the singular of dice). The probability that a 4 appears face up is 1/6 because there are *N* = 6 possible outcomes and only $n_4 = 1$ of them is a 4. But suppose you roll a six-sided die three times. You may ask for the probability that you will observe the sequence of two 3's followed by one 4. Or you may ask for the probability of rolling two 2's and one 6 in any order. The rules of probability and combinatorics provide the machinery for calculating such probabilities. Here we define the relationships among events that we need to formulate these rules.

Definitions: Relationships Among Events

MUTUALLY EXCLUSIVE. Outcomes A_1, A_2, \dots, A_t are mutually exclusive if the occurrence of each one of them precludes the occurrence of all the others. If *A* and *B* are mutually exclusive, then if *A* occurs, *B* does not. If *B* occurs, *A* does not. For example, on a single die roll, 1 and 3 are mutually exclusive because only one number can appear face up each time the die is rolled.

COLLECTIVELY EXHAUSTIVE. Outcomes A_1, A_2, \dots, A_t are collectively exhaustive if they constitute the entire set of possibilities, and no other outcomes are possible. For example, [heads, tails] is a collectively exhaustive set of outcomes for a coin toss, provided that you don't count the occasions when the coin lands on its edge.

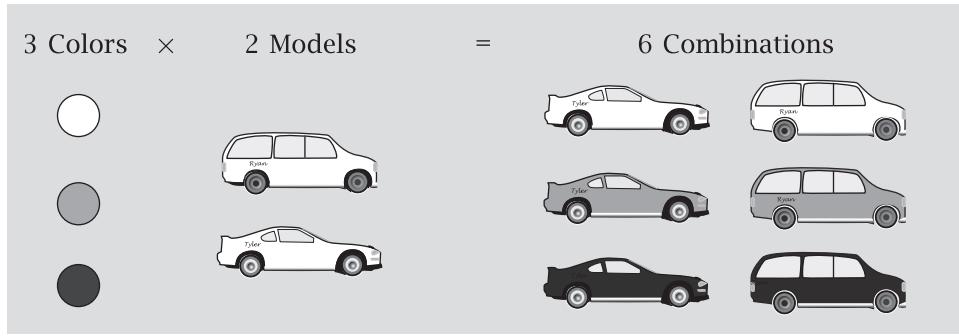


Figure 1.1 If there are three car colors for each of two car models, there are six different combinations of color and model, so the multiplicity is six.

INDEPENDENT. Events A_1, A_2, \dots, A_t are independent if the outcome of each one is unrelated to (or not *correlated* with) the outcome of any other. The score on one die roll is independent of the score on the next, unless there is trickery.

MULTIPLICITY. The multiplicity of events is the total number of ways in which different outcomes can possibly occur. If the number of outcomes of type A is n_A , the number of outcomes of type B is n_B , and the number of outcomes of type C is n_C , the total number of possible combinations of outcomes is the multiplicity W :

$$W = n_A n_B n_C. \quad (1.2)$$

Figure 1.1 shows an example of multiplicity.

The Rules of Probability Are Recipes for Drawing Consistent Inferences

The addition and multiplication rules permit you to calculate the probabilities of certain combinations of events.

ADDITION RULE. If outcomes A, B, \dots, E are mutually exclusive, and occur with probabilities $p_A = n_A/N, p_B = n_B/N, \dots, p_E = n_E/N$, then the probability of observing either A OR B OR ... OR E (the union of outcomes, expressed as $A \cup B \cup \dots \cup E$) is the sum of the probabilities:

$$\begin{aligned} p(A \text{ OR } B \text{ OR } \dots \text{ OR } E) &= \frac{n_A + n_B + \dots + n_E}{N} \\ &= p_A + p_B + \dots + p_E. \end{aligned} \quad (1.3)$$

The addition rule holds only if two criteria are met: the outcomes are *mutually exclusive* and we seek the probability of one outcome OR another outcome.

When they are not divided by N , the broader term for the quantities $n_i (i = A, B, \dots, E)$ is *statistical weights*. If outcomes A, B, \dots, E are both collectively exhaustive and mutually exclusive, then

$$n_A + n_B + \dots + n_E = N, \quad (1.4)$$

and dividing both sides of Equation (1.4) by N , the total number of trials, gives

$$p_A + p_B + \dots + p_E = 1. \quad (1.5)$$

MULTIPLICATION RULE. If outcomes A, B, \dots, E are independent, then the probability of observing A AND B AND \dots AND E (the intersection of outcomes, expressed as $A \cap B \cap \dots \cap E$) is the product of the probabilities,

$$\begin{aligned} p(A \text{ AND } B \text{ AND } \dots \text{ AND } E) &= \left(\frac{n_A}{N}\right)\left(\frac{n_B}{N}\right)\dots\left(\frac{n_E}{N}\right) \\ &= p_A p_B \dots p_E. \end{aligned} \quad (1.6)$$

The multiplication rule applies when the outcomes are *independent* and we seek the probability of one outcome AND another outcome AND possibly other outcomes. A more general multiplication rule, described on page 8, applies even when outcomes are not independent.

Here are a few examples using the addition and multiplication rules.

EXAMPLE 1.1 Rolling a die. What is the probability that either a 1 or a 4 appears on a single roll of a die? The probability of a 1 is $1/6$. The probability of a 4 is also $1/6$. The probability of either a 1 OR a 4 is $1/6 + 1/6 = 1/3$, because the outcomes are mutually exclusive (1 and 4 can't occur on the same roll) and the question is of the OR type.

EXAMPLE 1.2 Rolling twice. What is the probability of a 1 on the first roll of a die and a 4 on the second? It is $(1/6)(1/6) = 1/36$, because this is an AND question, and the two events are independent. This probability can also be computed in terms of the multiplicity. There are six possible outcomes on each of the two rolls of the die, giving a product of $W = 36$ possible combinations, one of which is 1 on the first roll and 4 on the second.

EXAMPLE 1.3 A sequence of coin flips. What is the probability of getting five heads on five successive flips of an unbiased coin? It is $(1/2)^5 = 1/32$, because the coin flips are independent of each other, this is an AND question, and the probability of heads on each flip is $1/2$. In terms of the multiplicity of outcomes, there are two possible outcomes on each flip, giving a product of $W = 32$ total outcomes, and only one of them is five successive heads.

EXAMPLE 1.4 Another sequence of coin flips. What is the probability of two heads, then one tail, then two more heads on five successive coin flips? It is $p_H^2 p_T p_H^2 = (1/2)^5 = 1/32$. You get the same result as in Example 1.3 because p_H , the probability of heads, and p_T , the probability of tails, are both $1/2$. There are a total of $W = 32$ possible outcomes and only one is the given sequence. The probability $p(n_H, N)$ of observing one particular sequence of N coin flips having exactly n_H heads is

$$p(n_H, N) = p_H^{n_H} p_T^{N-n_H}. \quad (1.7)$$

If $p_H = p_T = 1/2$, then $p(n_H, N) = (1/2)^N$.

EXAMPLE 1.5 Combining events—both, either/or, or neither. If independent events A and B have probabilities p_A and p_B , the probability that *both* events happen is $p_A p_B$. What is the probability that A happens AND B does not? The probability that B does not happen is $(1 - p_B)$. If A and B are independent events, then the probability that A happens and B does not is $p_A(1 - p_B) = p_A - p_A p_B$. What is the probability that *neither* event happens? It is

$$p(\text{not } A \text{ AND not } B) = (1 - p_A)(1 - p_B), \quad (1.8)$$

where $p(\text{not } A \text{ AND not } B)$ is the probability that A does not happen AND B does not happen.

EXAMPLE 1.6 Combining events—something happens. What is the probability that *something* happens, that is, A OR B OR both happen? This is an OR question, but the events are independent and not mutually exclusive, so you cannot use either the addition or multiplication rules. You can use a simple trick instead. The trick is to consider the probabilities that events *do not* happen, rather than that events *do* happen. The probability that something happens is $1 - p(\text{nothing happens})$:

$$1 - p(\text{not } A \text{ AND not } B) = 1 - (1 - p_A)(1 - p_B) = p_A + p_B - p_A p_B. \quad (1.9)$$

Multiple events can occur as ordered sequences in *time*, such as die rolls, or as ordered sequences in *space*, such as the strings of characters in words. Sometimes it is more useful to focus on collections of events rather than the individual events themselves.

Elementary and Composite Events

Some problems in probability cannot be solved directly by applying the addition or multiplication rules. Such questions can usually be reformulated in terms of *composite events* to which the rules of probability can be applied. Example 1.7 shows how to do this. Then on page 14 we'll use reformulation to construct probability distribution functions.

EXAMPLE 1.7 Elementary and composite events. What is the probability of a 1 on the first roll of a die OR a 4 on the second roll? If this were an AND question, the probability would be $(1/6)(1/6) = 1/36$, since the two rolls are independent, but the question is of the OR type, so it cannot be answered by direct application of either the addition or multiplication rules. But by redefining the problem in terms of composite events, you can use those rules. An individual coin toss, a single die roll, etc. could be called an elementary event. A composite event is just some set of elementary events, collected together in a convenient way. In this example it's convenient to define each composite event to be a pair of first and second rolls of the die. The advantage is that the complete list of composite events is mutually exclusive. That allows us to frame the problem in terms of an OR question and use the multiplication and addition rules. The composite events are:

[1, 1]*	[1, 2]*	[1, 3]*	[1, 4]*	[1, 5]*	[1, 6]*
[2, 1]	[2, 2]	[2, 3]	[2, 4]*	[2, 5]	[2, 6]
[3, 1]	[3, 2]	[3, 3]	[3, 4]*	[3, 5]	[3, 6]
[4, 1]	[4, 2]	[4, 3]	[4, 4]*	[4, 5]	[4, 6]
[5, 1]	[5, 2]	[5, 3]	[5, 4]*	[5, 5]	[5, 6]
[6, 1]	[6, 2]	[6, 3]	[6, 4]*	[6, 5]	[6, 6]

The first and second numbers in the brackets indicate the outcome of the first and second rolls respectively, and * indicates a composite event that satisfies the criterion for ‘success’ (1 on the first roll OR 4 on the second roll). There are 36 composite events, of which 11 are successful, so the probability we seek is 11/36.

Since many of the problems of interest in statistical thermodynamics involve huge systems (say, 10^{20} molecules), we need a more systematic way to compute composite probabilities than enumerating them all.

To compute this probability systematically, collect the composite events into three mutually exclusive classes, A , B , and C , about which you can ask an OR question. Class A includes all composite events with a 1 on the first roll AND anything but a 4 on the second. Class B includes all events with anything but a 1 on the first roll AND a 4 on the second. Class C includes the one event in which we get a 1 on the first roll AND a 4 on the second. A , B , and C are mutually exclusive categories. This is an OR question, so add p_A , p_B , and p_C to find the answer:

$$\begin{aligned} p(\text{1 first OR 4 second}) &= p_A(\text{1 first AND not 4 second}) \\ &\quad + p_B(\text{not 1 first AND 4 second}) \\ &\quad + p_C(\text{1 first AND 4 second}). \end{aligned} \quad (1.10)$$

The same probability rules that apply to elementary events also apply to composite events. Moreover, p_A , p_B , and p_C are each products of elementary event probabilities because the first and second rolls of the die are independent:

$$\begin{aligned} p_A &= \left(\frac{1}{6}\right) \left(\frac{5}{6}\right), \\ p_B &= \left(\frac{5}{6}\right) \left(\frac{1}{6}\right), \\ p_C &= \left(\frac{1}{6}\right) \left(\frac{1}{6}\right). \end{aligned}$$

Add p_A , p_B , and p_C : $p(\text{1 first OR 4 second}) = 5/36 + 5/36 + 1/36 = 11/36$. This example shows how elementary events can be grouped together into composite events to take advantage of the addition and multiplication rules. Reformulation is powerful because virtually any question can be framed in terms of combinations of AND and OR operations. With these two rules of probability, you can draw inferences about a wide range of probabilistic events.

EXAMPLE 1.8 A different way to solve it. Often, there are different ways to collect up events for solving probability problems. Let’s solve Example 1.7 differently. This time, use $p(\text{success}) = 1 - p(\text{fail})$. Because $p(\text{fail}) = [p(\text{not 1 first}) \text{ AND } p(\text{not 4 second})] = (5/6)(5/6) = 25/36$, you have $p(\text{success}) = 11/36$.

Two events can have a more complex relationship than we have considered so far. They are not restricted to being either independent or mutually exclusive. More broadly, events can be *correlated*.

Correlated Events Are Described by Conditional Probabilities

Events are correlated if the outcome of one depends on the outcome of the other. For example, if it rains on 36 days a year, the probability of rain is $36/365 \approx 0.1$. But if it rains on 50% of the days when you see dark clouds, then the probability of observing rain (event B) depends upon, or is conditional upon, the appearance of dark clouds (event A). Example 1.9 and Table 1.1 demonstrate the correlation of events when balls are taken out of a barrel.

EXAMPLE 1.9 **Balls taken from a barrel with replacement.** Suppose a barrel contains one red ball, R , and two green balls, G . The probability of drawing a green ball on the first try is $2/3$, and the probability of drawing a red ball on the first try is $1/3$. What is the probability of drawing a green ball on the second draw? That depends on whether or not you put the first ball back into the barrel before the second draw. If you replace each ball before drawing another, then the probabilities of different draws are uncorrelated with each other. Each draw is an independent event.

However, if you draw a green ball first, and don't put it back in the barrel, then 1 R and 1 G remain after the first draw, and the probability of getting a green ball on the second draw is now $1/2$. The probability of drawing a green ball on the second try is different from the probability of drawing a green ball on the first try. It is *conditional* on the outcome of the first draw.

Table 1.1 All of the probabilities for the three draws without replacement described in Examples 1.9 and 1.10.

1st Draw	2nd Draw	3rd Draw
$p(R_1) = \frac{1}{3}$	$\begin{cases} p(R_2 R_1)p(R_1) \\ 0 \cdot \left(\frac{1}{3}\right) = 0 \end{cases}$ $\begin{cases} p(G_2 R_1)p(R_1) \\ 1 \cdot \left(\frac{1}{3}\right) = \frac{1}{3} \end{cases}$	$p(G_3 G_2R_1)p(G_2R_1)$ $1 \cdot \left(\frac{1}{3}\right) = \frac{1}{3}$
$p(G_1) = \frac{2}{3}$	$\begin{cases} p(R_2 G_1)p(G_1) \\ \left(\frac{1}{2}\right) \cdot \left(\frac{2}{3}\right) = \frac{1}{3} \end{cases}$ $\begin{cases} p(G_2 G_1)p(G_1) \\ \left(\frac{1}{2}\right) \cdot \left(\frac{2}{3}\right) = \frac{1}{3} \end{cases}$	$p(G_3 R_2G_1)p(R_2G_1)$ $1 \cdot \left(\frac{1}{3}\right) = \frac{1}{3}$ $p(R_3 G_2G_1)p(G_2G_1)$ $1 \cdot \left(\frac{1}{3}\right) = \frac{1}{3}$

Here are some definitions and examples describing the conditional probabilities of correlated events.

CONDITIONAL PROBABILITY. The conditional probability $p(B | A)$ is the probability of event B , *given that* some other event A has occurred. Event A is the *condition* upon which we evaluate the probability of event B . In Example 1.9, event B is getting a green ball on the second draw, event A is getting a green ball on the first draw, and $p(G_2 | G_1)$ is the probability of getting a green ball on the second draw, given a green ball on the first draw.

JOINT PROBABILITY. The joint probability of events A and B is the probability that both events A AND B occur. The joint probability is expressed by the notation $p(A \text{ AND } B)$, or more concisely, $p(AB)$.

GENERAL MULTIPLICATION RULE (BAYES' RULE). If outcomes A and B occur with probabilities $p(A)$ and $p(B)$, the joint probability of events A AND B is

$$p(AB) = p(B | A)p(A) = p(A | B)p(B). \quad (1.11)$$

If events A and B happen to be independent, the pre-condition A has no influence on the probability of B . Then $p(B | A) = p(B)$, and Equation (1.11) reduces to $p(AB) = p(B)p(A)$, the multiplication rule for independent events. A probability $p(B)$ that is not conditional is called an *a priori* probability. The conditional quantity $p(B | A)$ is called an *a posteriori* probability. The general multiplication rule is general because independence is not required. It defines the probability of the *intersection* of events, $p(AB) = p(A \cap B)$.

GENERAL ADDITION RULE. A general rule can also be formulated for the union of events, $p(A \cup B) = p(A) + p(B) - p(A \cap B)$, when we seek the probability of A OR B for events that are not mutually exclusive. When A and B are mutually exclusive, $p(A \cap B) = 0$, and the general addition rule reduces to the simpler addition rule on page 3. When A and B are independent, $p(A \cap B) = p(A)p(B)$, and the general addition rule gives the result in Example 1.6.

DEGREE OF CORRELATION. The degree of correlation g between events A and B can be expressed as the ratio of the conditional probability of B , given A , to the unconditional probability of B alone. This indicates the degree to which A influences B :

$$g = \frac{p(B | A)}{p(B)} = \frac{p(AB)}{p(A)p(B)}. \quad (1.12)$$

The second equality in Equation (1.12) follows from the general multiplication rule, Equation (1.11). If $g = 1$, events A and B are independent and not correlated. If $g > 1$, events A and B are positively correlated. If $g < 1$, events A and B are negatively correlated. If $g = 0$ and A occurs, then B will not. If the *a priori* probability of rain is $p(B) = 0.1$, and if the conditional probability of rain, given that there are dark clouds, A , is $p(B | A) = 0.5$, then the degree of correlation of rain with dark clouds is $g = 5$. Correlations are important in statistical thermodynamics. For example, attractions and repulsions among molecules in liquids can cause correlations among their positions and orientations.

EXAMPLE 1.10 Balls taken from that barrel again. As before, start with three balls in a barrel, one red and two green. The probability of getting a red ball on the first draw is $p(R_1) = 1/3$, where the notation R_1 refers to a red ball on the first draw. The probability of getting a green ball on the first draw is $p(G_1) = 2/3$. If balls are not replaced after each draw, the joint probability for getting a red ball first and a green ball second is $p(R_1G_2)$:

$$p(R_1G_2) = p(G_2 | R_1)p(R_1) = (1)\left(\frac{1}{3}\right) = \frac{1}{3}. \quad (1.13)$$

So, getting a green ball second is correlated with getting a red ball first:

$$g = \frac{p(R_1G_2)}{p(R_1)p(G_2)} = \frac{\frac{1}{3}}{\left(\frac{1}{3}\right)\left(\frac{1}{3} + \frac{1}{3}\right)} = \frac{3}{2}. \quad (1.14)$$

RULE FOR ADDING JOINT PROBABILITIES. The following is a useful way to compute a probability $p(B)$ if you know joint or conditional probabilities:

$$p(B) = p(BA) + p(BA') = p(B|A)p(A) + p(B|A')p(A'), \quad (1.15)$$

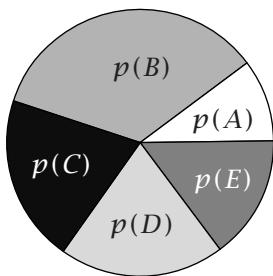
where A' means that event A does not occur. If the event B is rain, and if the event A is that you see clouds and A' is that you see no clouds, then the probability of rain is the sum of joint probabilities of (rain, you see clouds) plus (rain, you see no clouds). By summing over the mutually exclusive conditions, A and A' , you are accounting for all the ways that B can happen.

EXAMPLE 1.11 Applying Bayes' rule: Predicting protein properties. *Bayes' rule*, a combination of Equations (1.11) and (1.15), can help you compute hard-to-get probabilities from ones that are easier to get. Here's a toy example. Let's figure out a protein's structure from its amino acid sequence. From modern genomics, it is easy to learn protein sequences. It's harder to learn protein structures. Suppose you discover a new type of protein structure, call it a *helicoil h*. It's rare; you've searched 5000 proteins and found only 20 helicoils, so $p(h) = 0.004$. If you could discover some special amino acid *sequence feature*, call it sf, that predicts the *h* structure, you could search other genomes to find other helicoil proteins in nature. It's easier to turn this around. Rather than looking through 5000 sequences for patterns, you want to look at the 20 helicoil proteins and find the fraction of them that have your sequence feature. If your sequence feature (say alternating glycine and lysine amino acids) appears in 19 out of the 20 helicoils, you have $p(sf | h) = 0.95$. You also need $p(sf | \bar{h})$, the fraction of non-helicoil proteins (let's call those \bar{h}) that have your sequence feature. Suppose you find $p(sf | \bar{h}) = 0.001$. Combining Equations (1.11) and (1.15) gives Bayes' rule for the probability you want:

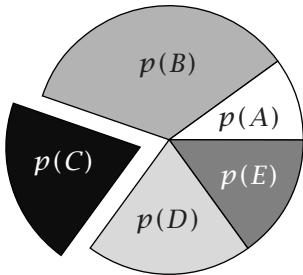
$$\begin{aligned} p(h | sf) &= \frac{p(sf | h)p(h)}{p(sf)} = \frac{p(sf | h)p(h)}{p(sf | h)p(h) + p(sf | \bar{h})p(\bar{h})} \\ &= \frac{(0.95)(0.004)}{(0.95)(0.004) + (0.001)(0.996)} = 0.79. \end{aligned} \quad (1.16)$$

In short, if a protein has the sf sequence, it will have the *h* structure about 80% of the time.

(a) Who will win?



(b) Given that C won...



(c) Who will place second?

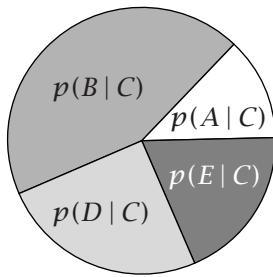


Figure 1.2 (a) *A priori* probabilities of outcomes A to E , such as a horse race. (b) To determine the *a posteriori* probabilities of events A , B , D , and E , given that C has occurred, remove C and keep the relative proportions of the rest the same. (c) *A posteriori* probabilities that horses A , B , D , and E will come in second, given that C won.

Conditional probabilities are useful in a variety of situations, including card games and horse races, as the following example shows.

EXAMPLE 1.12 A gambling equation. Suppose you have a collection of mutually exclusive and collectively exhaustive events A, B, \dots, E , with probabilities p_A, p_B, \dots, p_E . These could be the probabilities that horses A, B, \dots, E will win a race (based on some theory, model, or prediction scheme), or that card types A to E will appear on a given play in a card game. Let's look at a horse race [2]. If horse A wins, then horses B or C don't win, so these are mutually exclusive.

Suppose you have some information, such as the track records of the horses, that predicts the *a priori* probabilities that each horse will win. Figure 1.2 gives an example. Now, as the race proceeds, the events occur in order, one at a time: one horse wins, then another comes in second, and another comes in third. Our aim is to compute the conditional probability that a particular horse will come in second, given that some other horse has won. The *a priori* probability that horse C will win is $p(C)$. Now assume that horse C has won, and you want to know the probability that horse A will be second, $p(A \text{ is second} | C \text{ is first})$. From Figure 1.2, you can see that this conditional probability can be determined by eliminating region C , and finding the fraction of the remaining area occupied by region A :

$$\begin{aligned} p(A \text{ is second} | C \text{ is first}) &= \frac{p(A)}{p(A) + p(B) + p(D) + p(E)} \\ &= \frac{p(A)}{1 - p(C)}. \end{aligned} \quad (1.17)$$

$1 - p(C) = p(A) + p(B) + p(D) + p(E)$ follows from the mutually exclusive addition rule.

The probability that event i is first is $p(i)$. Then the conditional probability that event j is second is $p(j)/[1 - p(i)]$. The joint probability that i is first, j is second, and k is third is

$$p(i \text{ is first}, j \text{ is second}, k \text{ is third}) = \frac{p(i)p(j)p(k)}{[1 - p(i)][1 - p(i) - p(j)]}. \quad (1.18)$$

Equations (1.17) and (1.18) are useful for computing the probability of drawing the queen of hearts in a card game, once you have seen the seven of clubs and the ace of spades. They are also useful for describing the statistical thermodynamics of liquid crystals, and ligand binding to DNA (see pages 575–578).

Combinatorics Describes How to Count Events

Combinatorics, or counting events, is central to statistical thermodynamics. It is the basis for entropy, and the concepts of *order* and *disorder*, which are defined by the numbers of ways in which a system can be configured. Combinatorics is concerned with the *composition* of events rather than the *sequence* of events. For example, compare the following two questions. The first is a question of sequence: What is the probability of the specific sequence of four coin flips, $HTHH$? The second is a question of composition: What is the probability of observing three H 's and one T in *any* order? The sequence question is answered

by using Equation (1.7): this probability is $1/16$. However, to answer the composition question you must count the number of different possible sequences with the specified composition: $HHHT$, $HHTH$, $HTHH$, and $THHH$. The probability of getting three H 's and one T in any order is $4/16 = 1/4$. When you seek the probability of a certain *composition* of events, you count the possible sequences that have the correct composition.

EXAMPLE 1.13 Permutations of ordered sequences. How many permutations, or different sequences, of the letters w , x , y , and z are possible? There are 24:

$wxyz$	$wxzy$	$wyxz$	$wyzx$	$wzxy$	$wzyx$
$xwyz$	$xwzy$	$xywz$	$xyzw$	$xzwy$	$xzyw$
$ywxz$	$ywzx$	$yxwz$	$yxzw$	$yzwx$	$yzxw$
$zwyx$	$zwxy$	$zxwy$	$zxyw$	$zywx$	$zyxw$

How can you compute the number of different sequences without having to list them all? You can use the strategy developed for drawing letters from a barrel without replacement. The first letter of a sequence can be any one of the four. After drawing one, the second letter of the sequence can be any of the remaining three letters. The third letter can be any of the remaining two letters, and the fourth must be the one remaining letter. Use the definition of multiplicity W (Equation (1.2)) to combine the numbers of outcomes n_i , where i represents the position 1, 2, 3, or 4 in the sequence of draws. We have $n_1 = 4$, $n_2 = 3$, $n_3 = 2$, and $n_4 = 1$, so the number of permutations is $W = n_1 n_2 n_3 n_4 = 4 \cdot 3 \cdot 2 \cdot 1 = 24$.

In general, for a sequence of N *distinguishable* objects, the number of different permutations W can be expressed in factorial notation

$$\begin{aligned} W &= N(N-1)(N-2) \cdots 3 \cdot 2 \cdot 1 = N! \\ &= 4 \cdot 3 \cdot 2 \cdot 1 = 24. \end{aligned} \tag{1.19}$$

The Factorial Notation

The notation $N!$, called N factorial, denotes the product of the integers from one to N :

$$N! = 1 \cdot 2 \cdot 3 \cdots (N-2)(N-1)N.$$

$0!$ is defined to equal 1.

EXAMPLE 1.14 Letters of the alphabet. Consider a barrel containing one each of the 26 letters of the alphabet. What is the probability of drawing the letters out in exactly the order of the alphabet, A to Z? The probability of drawing the A first is $1/26$. If you replace each letter after it is drawn, the probability of drawing the B on the second try would be $1/26$, and the probability of drawing the alphabet in order would be $(1/26)^{26}$. But if each letter were *not* replaced in the barrel, the probability of drawing the B on the second trial would be $1/25$.

The probability of drawing the **C** on the third trial would be $1/24$. Without replacement, the probability of drawing the exact sequence of the alphabet is

$$p(\mathbf{ABC\dots XYZ}) = \frac{1}{26 \cdot 25 \cdot 24 \cdots 2 \cdot 1} = \frac{1}{N!}, \quad (1.20)$$

where $N = 26$ is the number of letters in the alphabet. $N!$ is the number of permutations, or different sequences in which the letters could be drawn. $1/N!$ is the probability of drawing one particular sequence.

In Examples 1.13 and 1.14, all the letters are distinguishable from each other: **w**, **x**, **y**, and **z** are all different. But what happens if some of the objects are indistinguishable from each other?

EXAMPLE 1.15 Counting sequences of distinguishable and indistinguishable objects. How many different arrangements are there of the letters **A**, **H**, and **A**? That depends on whether or not you can tell the **A**'s apart. Suppose first that one **A** has a subscript 1 and the other has a subscript 2: **A**₁, **H**, and **A**₂. Then all the characters are distinguishable, as in Examples 1.13 and 1.14, and there are $W = N! = 3! = 6$ different arrangements of these three distinguishable characters:

$$\mathbf{HA}_1\mathbf{A}_2 \quad \mathbf{A}_1\mathbf{H}\mathbf{A}_2 \quad \mathbf{A}_1\mathbf{A}_2\mathbf{H} \quad \mathbf{H}\mathbf{A}_2\mathbf{A}_1 \quad \mathbf{A}_2\mathbf{H}\mathbf{A}_1 \quad \mathbf{A}_2\mathbf{A}_1\mathbf{H}.$$

However, now suppose that the two **A**'s are indistinguishable from each other: they have no subscripts. There are now only $W = 3$ distinguishable sequences of letters: **HAA**, **AHA**, and **AAH**. (Distinguishable is a term that applies either to the letters or to the sequences. We have three distinguishable sequences, each containing two distinguishable letters, **A** and **H**.) The previous expression $W = N!$ overcounts by a factor of two when the two **A**'s are indistinguishable. This is because we have counted each sequence of letters, say **AAH**, twice—**A**₁**A**₂**H** and **A**₂**A**₁**H**. Written in a more general way, the number of distinguishable sequences is $W = N!/N_A! = 3!/2! = 3$. The $N!$ in the numerator comes from the number of permutations as if all the characters were distinguishable from each other, and the $N_A!$ in the denominator corrects for overcounting. The overcounting correction $2!$ is simply the count of all the permutations of the indistinguishable characters, the number of ways in which the **A**'s can be arranged among themselves.

EXAMPLE 1.16 Permutations of mixed sequences. Consider the word **cheese** as **che**₁**e**₂**se**₃, in which the **e**'s are distinguished from each other by a subscript. Then $N = 6$ and there are $6! = 720$ distinguishable ways of arranging the characters. By counting in this way, we have reckoned that **che**₁**e**₂**se**₃ is different from **che**₂**e**₁**se**₃. This correct spelling is counted exactly six times because there are six permutations of the subscripted **e**'s. There are also exactly six permutations of the **e**'s in every other specific sequence. For example:

se ₁ e ₂ che ₃	se ₁ e ₃ che ₂	se ₂ e ₁ che ₃
se ₂ e ₃ che ₁	se ₃ e ₁ che ₂	se ₃ e ₂ che ₁

There are $3! = 6$ permutations of \mathbf{e}_1 , \mathbf{e}_2 , and \mathbf{e}_3 for every sequence of the other characters. So when the \mathbf{e} 's are indistinguishable, there are $6!/3!$ permutations of the letters in the word cheese. The $3!$ in the denominator corrects for the indistinguishability of the \mathbf{e} 's. In general, the denominator needs a factor to account for the indistinguishability of each type of character, so $W = N!/(n_c!n_h!n_e!n_s!) = 6!/(1!1!3!1!) = 120$ is the number of different sequences if the \mathbf{e} 's are indistinguishable from each other.

EXAMPLE 1.17 Another mixed sequence. For the word **freezer**, you have three indistinguishable \mathbf{e} 's and two indistinguishable \mathbf{r} 's. There are $7!/(3!2!)$ permutations of the letters that spell **freezer**.

In general, for a collection of N objects with t categories, of which n_i objects in each category are *indistinguishable* from one another, but distinguishable from the objects in the other $t - 1$ categories, the number of permutations W is

$$W = \frac{N!}{n_1!n_2!\cdots n_t!}. \quad (1.21)$$

When there are only two categories (success/failure, or heads/tails, ...), $t = 2$, so $W(n, N)$, the number of sequences with n successes out of N trials, is

$$W(n, N) = \binom{N}{n} = \frac{N!}{n!(N-n)!}, \quad (1.22)$$

where the shorthand notation $\binom{N}{n}$ for combinations is pronounced ‘N choose n .’ Use $W = N!$ if you can distinguish every single sequence from every other, or $W = N!/n!$ if only heads are indistinguishable from each other, and tails are distinguishable. Or use Equation (1.22) if tails are indistinguishable from other tails, and heads are indistinguishable from other heads, the case we’ll be most interested in subsequently. Example 1.18 applies Equation (1.22) to coin flips and die rolls.

EXAMPLE 1.18 Counting sequences of coin flips and die rolls. You flip a coin $N = 4$ times. How many different sequences have three heads? According to Equation (1.22),

$$W(n_H, N) = \frac{N!}{n_H!n_T!} = \frac{4!}{3!1!} = 4.$$

They are *THHH*, *HTHH*, *HHTH*, and *HHHT*. How many different sequences have two heads?

$$W(2, 4) = \frac{4!}{2!2!} = 6.$$

They are *TTHH*, *HHTT*, *THTH*, *HTHT*, *THHT*, and *HTTH*.

You flip a coin 117 times. How many different sequences have 36 heads?

$$W(36, 117) = \frac{117!}{36!81!} \approx 1.84 \times 10^{30}.$$

We won’t write the sequences out.

You roll a die 15 times. How many different sequences have three 1's, one 2, one 3, five 4's, two 5's, and three 6's? According to Equation (1.21),

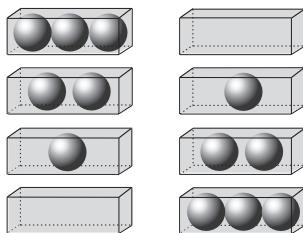
$$W = \frac{15!}{3!1!1!5!2!3!} = 151,351,200.$$

EXAMPLE 1.19 What is the probability of drawing a royal flush in poker?

There are four different ways to draw a royal flush in poker: an ace, king, jack, queen, and ten, all from any one of the four suits. To compute the probability, you need to know how many five-card hands there are in a deck of 52 cards. Use the barrel metaphor: put the 52 cards in the barrel. On the first draw, there are 52 possibilities. On the second draw, there are 51 possibilities, etc. In five draws, there are

$$\frac{52 \cdot 51 \cdot 50 \cdot 49 \cdot 48}{5!} = \frac{52!}{5!(52-5)!} = 2,598,960$$

(a) Balls in Boxes



(b) Movable Walls

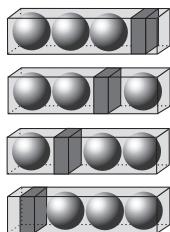


Figure 1.3 How many ways can you put $n = 3$ balls into $M = 2$ boxes (Example 1.20)? (a) There are four ways to partition $n = 3$ balls into $M = 2$ boxes when each box can hold any number of balls. (b) Look at this as four ways to partition three balls and one movable wall.

possible poker hands. The $5!$ in the denominator corrects for all the possible permutations of each sequence (you don't care whether you draw the king or the ace first, for example). The probability is $1/(2,598,960)$ of drawing a royal flush in one suit or $4/(2,598,960) = 1.5 \times 10^{-6}$ that you will draw a royal flush in any of the four suits.

Here's an example of a type of counting problem in statistical thermodynamics.

EXAMPLE 1.20 Bose–Einstein statistics. How many ways can n indistinguishable particles be put into M boxes, with any number of particles per box? This type of counting is needed to predict the properties of particles called *bosons*, such as photons and ${}^4\text{He}$ atoms. Bose–Einstein statistics counts the ways that n particles can be distributed in M different energy levels, when several particles can occupy the same quantum mechanical energy levels. For now, our interest is not in the physics, but just in the counting problem. Figure 1.3 shows that one way to count the number of arrangements is to think of the system as a linear array of n particles interspersed with $M - 1$ movable walls that partition the system into M boxes (spaces between walls). There are $M + n - 1$ objects, counting walls plus particles. If the objects were all distinguishable, there would be $(M + n - 1)!$ arrangements. However, because the n particles are indistinguishable from each other and each of the $M - 1$ walls is indistinguishable from the other walls, and because the walls are distinguishable from the particles, the number of arrangements is

$$W(n, M) = \frac{(M+n-1)!}{(M-1)!n!}. \quad (1.23)$$

Collections of Probabilities Are Described by Distribution Functions

The probabilities of events can be described by *probability distribution functions*. For t mutually exclusive outcomes, $i = 1, 2, 3, \dots, t$, the distribution

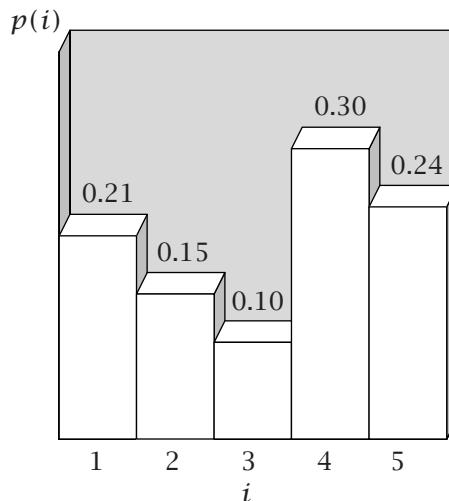


Figure 1.4 A probability distribution function. The possible outcomes are indexed on the horizontal axis. The probability of each outcome is shown on the vertical axis. In this example, outcome **4** is the most probable and outcome **3** is the least probable.

function is $p(i)$, the set of probabilities of all the outcomes. Figure 1.4 shows a probability distribution function for a system with $t = 5$ outcomes.

A property of probability distribution functions is that the sum of the probabilities equals 1. Because the outcomes are mutually exclusive and collectively exhaustive, Equations (1.3) and (1.5) apply and

$$\sum_{i=1}^t p(i) = 1. \quad (1.24)$$

For some types of events, the order of the outcomes $i = 1, 2, 3, \dots, t$ has meaning. For others, it does not. For statistical thermodynamics, the order usually has meaning, and i represents the value of some physical quantity. On the other hand, the index i may be just a label. The index $i = 1, 2, 3$ can represent the colors of socks, [red, green, blue], or [green, red, blue], where the order is irrelevant. Probability distributions can describe either case.

Summations

The sigma notation means to sum terms. For example,

$$\sum_{i=1}^6 ip_i = p_1 + 2p_2 + 3p_3 + 4p_4 + 5p_5 + 6p_6 \quad (1.25)$$

means ‘sum the quantity ip_i from $i = 1$ up to $i = 6$.’ Sometimes the index i above and/or below the sigma is omitted in concise shorthand expressions.

Continuous Probability Distribution Functions

In some situations, the outcomes of an event are best represented by a continuous variable x rather than by a discrete variable. Think of a bell curve. Or, for example, a particle might have some probability $p(x) dx$ of being between position $x = 1.62$ cm and $x + dx = 1.63$ cm or $p(\theta) d\theta$ of having an orientation angle between $\theta = 25.6^\circ$ and $\theta + d\theta = 25.8^\circ$. If x is continuous, $p(x)$ is

called a *probability density*, because $p(x) dx$ is the probability of finding a value between x and $x+dx$. If x ranges from $x=a$ to $x=b$, Equation (1.24) becomes

$$\int_a^b p(x) dx = 1. \quad (1.26)$$

Some distribution functions aren't *normalized*: the statistical weights do not sum to 1. Then you first need to normalize them. For a continuous distribution function $g(x)$, where x ranges from a to b , you can normalize to form a proper probability distribution function. Find the normalization constant g_0 by integrating over x :

$$g_0 = \int_a^b g(x) dx. \quad (1.27)$$

The normalized probability density is

$$p(x) = \frac{g(x)}{g_0} = \frac{g(x)}{\int_a^b g(x) dx}. \quad (1.28)$$

The Binomial and Multinomial Distribution Functions

Some probability distribution functions occur frequently in nature, and have simple mathematical expressions. Two of the most useful functions are the binomial and multinomial distributions. These will be the basis for our development of the concept of entropy in Chapter 2. The binomial distribution describes processes in which each independent elementary event has two mutually exclusive outcomes, such as heads/tails, yes/no, up/down, or occupied/vacant. Independent trials with two such possible outcomes are called *Bernoulli trials*. Let's label the two possible outcomes \bullet and \circ . Let the probability of \bullet be p . Then the probability of \circ is $1-p$. We choose composite events that are pairs of Bernoulli trials. The probability of \bullet followed by \circ is $P_{\bullet\circ} = p(1-p)$. The probabilities of the four possible composite events are

$$\begin{aligned} P_{\bullet\bullet} &= p^2, & P_{\bullet\circ} &= p(1-p), \\ P_{\circ\bullet} &= (1-p)p, & P_{\circ\circ} &= (1-p)^2. \end{aligned} \quad (1.29)$$

This set of composite events is mutually exclusive and collectively exhaustive.

The same probability rules apply to the composite events that apply to elementary events. For example, Equation (1.24) for the normalization of discrete distributions requires that the probabilities must sum to 1:

$$\begin{aligned} P_{\bullet\bullet} + P_{\bullet\circ} + P_{\circ\bullet} + P_{\circ\circ} &= p^2 + 2p(1-p) + (1-p)^2 \\ &= [p + (1-p)]^2 = 1. \end{aligned} \quad (1.30)$$

In Example 1.7, we defined composite events as pairs of elementary events. More generally, a composite event is a sequence of N repetitions of independent elementary events. The probability of a *specific sequence* of n \bullet 's and $N-n$ \circ 's is given by Equation (1.7). What is the probability that a series of N trials has

n ●'s and $N-n$ ○'s in any order? Equation (1.22) gives the total number of sequences that have n ●'s and $N-n$ ○'s. The product of Equations (1.7) and (1.22) gives the probability of n ●'s and $N-n$ ○'s irrespective of their sequence. This is the **binomial distribution**:

$$P(n, N) = p^n(1-p)^{N-n} \frac{N!}{n!(N-n)!}. \quad (1.31)$$

Because the set of all possible sequences of N trials is mutually exclusive and collectively exhaustive, the composite probabilities sum to one, $\sum_{n=0}^N P(n, N) = 1$, as illustrated below.

A simple visualization of the combinatoric terms in the binomial distribution is *Pascal's triangle*. Make a triangle in which the lines are numbered $N = 0, 1, 2, \dots$. Compute $N!/[n!(N-n)!]$ at each position:

$N = 0$	1
$N = 1$	1 1
$N = 2$	1 2 1
$N = 3$	1 3 3 1
$N = 4$	1 4 6 4 1
$N = 5$	1 5 10 10 5 1

Each term in Pascal's triangle is the sum of the two terms to the left and right from the line above it. Pascal's triangle gives the coefficients in the expansion of $(x+y)^N$. For example, for $N = 4$, using $x = p$ and $y = 1-p$, Equation (1.31) is

$$\begin{aligned} [p + (1-p)]^4 &= p^4 + 4p^3(1-p) + 6p^2(1-p)^2 \\ &\quad + 4p(1-p)^3 + (1-p)^4. \end{aligned} \quad (1.32)$$

This sums to one, $\sum_{n=0}^N P(n, N) = 1$, because $[p + (1-p)]^N = 1$.

EXAMPLE 1.21 Distribution of coin flips. Figure 1.5 shows a distribution function, the probability $p(n_H, N)$ of observing n_H heads in $N = 4$ coin flips, given by Equation (1.31) with $p = 0.5$. This shows that in four coin flips, the most probable number of heads is two. It is least probable that all four will be heads or all four will be tails.

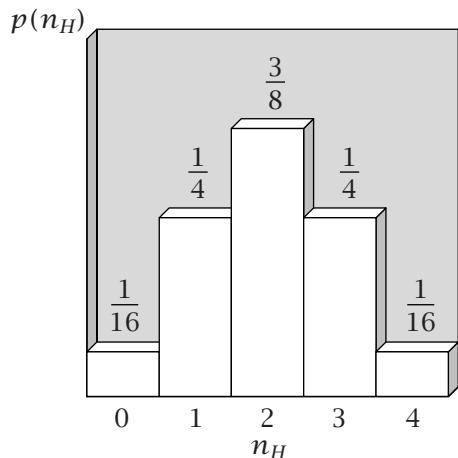


Figure 1.5 The probability distribution for the numbers of heads in four coin flips in Example 1.21.

The multinomial probability distribution is a generalization of the binomial probability distribution. A binomial distribution describes two-outcome events such as coin flips. A multinomial probability distribution applies to t -outcome events where n_i is the number of times that outcome $i = 1, 2, 3, \dots, t$ appears. For example, $t = 6$ for die rolls. For the multinomial distribution, the number of distinguishable outcomes is given by Equation (1.21): $W = N!/(n_1!n_2!n_3!\cdots n_t!)$. The **multinomial probability distribution** is

$$P(n_1, n_2, \dots, n_t, N) = p_1^{n_1} p_2^{n_2} p_3^{n_3} \cdots p_t^{n_t} \left(\frac{N!}{n_1!n_2!\cdots n_t!} \right), \quad (1.33)$$

where each factor $n_i!$ accounts for the indistinguishability of objects in category i . The n_i are constrained by the condition $\sum_{i=1}^t n_i = N$.

Distribution Functions Have Average Values and Standard Deviations

Averages

A probability distribution function contains all the information that can be known about a probabilistic system. A full distribution function, however, is rarely accessible from experiments. Generally, experiments can measure only certain *moments* of the distribution. The **n th moment** of a probability distribution function $p(x)$ is

$$\langle x^n \rangle = \int_a^b x^n p(x) dx = \frac{\int_a^b x^n g(x) dx}{\int_a^b g(x) dx}, \quad (1.34)$$

where the second expression is appropriate for a non-normalized distribution function $g(x)$. Angle brackets $\langle \rangle$ are used to indicate the moments, also called the expectation values or averages, of a distribution function. For a *probability* distribution the zeroth moment always equals one, because the sum of the probabilities equals one. The first moment of a distribution function ($n = 1$ in Equation (1.34)) is called the mean, average, or expected value. For discrete functions,

$$\langle i \rangle = \sum_{i=1}^t i p(i), \quad (1.35)$$

and for continuous functions,

$$\langle x \rangle = \int_a^b x p(x) dx. \quad (1.36)$$

For distributions over t discrete values, the mean of a function $f(i)$ is

$$\langle f(i) \rangle = \sum_{i=1}^t f(i) p(i). \quad (1.37)$$

For distributions over continuous values, the mean of a function $f(x)$ is

$$\langle f(x) \rangle = \int_a^b f(x)p(x) dx = \frac{\int_a^b f(x)g(x) dx}{\int_a^b g(x) dx}. \quad (1.38)$$

Equations (1.35)–(1.38) quantify the familiar notion of average, as Example 1.22 shows.

EXAMPLE 1.22 Taking an average. The average of the set of numbers [3, 3, 2, 2, 2, 1, 1] is 2. The average may be computed by the usual procedure of summing the numbers and dividing by the number of entries. Let's compute the average using Equation (1.35) instead. Since two of the seven outcomes are 3's, the probability of a 3 is $p(3) = 2/7$. Similarly, three of the seven outcomes are 2's, so $p(2) = 3/7$, and two of the seven outcomes are 1's, so $p(1) = 2/7$. The average $\langle i \rangle$ is then

$$\begin{aligned} \langle i \rangle &= \sum_{i=1}^3 ip(i) = 1p(1) + 2p(2) + 3p(3) \\ &= 1\left(\frac{2}{7}\right) + 2\left(\frac{3}{7}\right) + 3\left(\frac{2}{7}\right) = 2. \end{aligned} \quad (1.39)$$

Here are two useful and general properties of averages, derived from the definition given in Equation (1.38):

$$\begin{aligned} \langle af(x) \rangle &= \int af(x)p(x) dx = a \int f(x)p(x) dx \\ &= a\langle f(x) \rangle, \quad \text{where } a \text{ is a constant.} \end{aligned} \quad (1.40)$$

$$\begin{aligned} \langle f(x) + g(x) \rangle &= \int [f(x) + g(x)] p(x) dx \\ &= \int f(x)p(x) dx + \int g(x)p(x) dx \\ &= \langle f(x) \rangle + \langle g(x) \rangle. \end{aligned} \quad (1.41)$$

Variance

The *variance* σ^2 is a measure of the width of a distribution. A broad, flat distribution has a large variance, while a narrow, peaked distribution has a small variance. The variance σ^2 is defined as the average square deviation from the mean,

$$\sigma^2 = \langle (x - \mu)^2 \rangle = \langle x^2 - 2\mu x + \mu^2 \rangle, \quad (1.42)$$

where $\mu = \langle x \rangle$ is the mean value, or first moment. We use μ instead of $\langle x \rangle$ as a reminder here that this quantity is just a constant, not a variable. Using

Equation (1.41), Equation (1.42) becomes

$$\sigma^2 = \langle x^2 \rangle - \langle 2ax \rangle + \langle a^2 \rangle.$$

Using Equation (1.40),

$$\sigma^2 = \langle x^2 \rangle - 2a\langle x \rangle + a^2 = \langle x^2 \rangle - \langle x \rangle^2. \quad (1.43)$$

Second-moment quantities are important for understanding heat capacities (Chapter 12), random walks (Chapter 18), diffusion (Chapter 17), and polymer chain conformations (Chapters 32–34). The square root of the variance is σ , which is also called the *standard deviation*. Moments higher than the second describe asymmetries in the shape of the distribution. Examples 1.23–1.26 show calculations of means and variances for discrete and continuous probability distributions.

EXAMPLE 1.23 Coin flips: mean and variance. Compute the average number of heads $\langle n_H \rangle$ in $N = 4$ coin flips by using the distribution in Example 1.21:

$$\begin{aligned} \langle n_H \rangle &= \sum_{n_H=0}^4 n_H p(n_H, N) \\ &= 0\left(\frac{1}{16}\right) + 1\left(\frac{4}{16}\right) + 2\left(\frac{6}{16}\right) + 3\left(\frac{4}{16}\right) + 4\left(\frac{1}{16}\right) = 2, \end{aligned}$$

and

$$\begin{aligned} \langle n_H^2 \rangle &= \sum_{n_H=0}^4 n_H^2 p(n_H, N) \\ &= 0\left(\frac{1}{16}\right) + 1\left(\frac{4}{16}\right) + 4\left(\frac{6}{16}\right) + 9\left(\frac{4}{16}\right) + 16\left(\frac{1}{16}\right) = 5. \end{aligned}$$

According to Equation (1.43), the variance σ^2 is

$$\sigma^2 = \langle n_H^2 \rangle - \langle n_H \rangle^2 = 5 - 2^2 = 1.$$

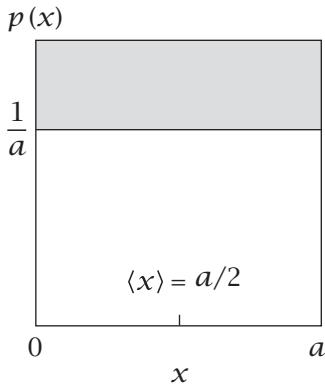


Figure 1.6 Flat distribution function, $0 \leq x \leq a$. The average value is $\langle x \rangle = a/2$ (see Example 1.24).

EXAMPLE 1.24 The average and variance of a continuous function. Suppose you have a flat probability distribution $p(x) = 1/a$ (shown in Figure 1.6) for a variable $0 \leq x \leq a$. To compute $\langle x \rangle$, use Equation (1.36):

$$\langle x \rangle = \int_0^a x p(x) dx = \frac{1}{a} \int_0^a x dx = \left(\frac{1}{a}\right) \frac{x^2}{2} \Big|_0^a = \frac{a}{2}.$$

Equation (1.34) gives the second moment $\langle x^2 \rangle$:

$$\langle x^2 \rangle = \int_0^a x^2 p(x) dx = \frac{1}{a} \int_0^a x^2 dx = \left(\frac{1}{a}\right) \frac{x^3}{3} \Big|_0^a = \frac{a^2}{3}.$$

The variance is $\langle x^2 \rangle - \langle x \rangle^2 = a^2/3 - a^2/4 = a^2/12$.

EXAMPLE 1.25 The average of an exponential distribution. Figure 1.7 shows a distribution function $g(x) = e^{-ax}$ over the range $0 \leq x \leq \infty$. First normalize $g(x)$ to make it a probability distribution. According to Equation (1.28), $p(x) = g(x)/g_0$. Integrate $g(x)$ to determine g_0 :

$$g_0 = \int_0^\infty e^{-ax} dx = -\left(\frac{1}{a}\right)e^{-ax} \Big|_0^\infty = \frac{1}{a} \quad \text{for } a > 0.$$

The normalized distribution function is $p(x) = g(x)/g_0 = ae^{-ax}$. Now, to compute $\langle x \rangle$ for this distribution, use Equation (1.34):

$$\begin{aligned} \langle x \rangle &= \int_0^\infty x p(x) dx = a \int_0^\infty x e^{-ax} dx \\ &= -\left[x e^{-ax} \left(x + \frac{1}{a}\right)\right]_0^\infty = \frac{1}{a}. \end{aligned}$$

EXAMPLE 1.26 Averaging the orientations of a vector. For predicting the conformations of a polymer or spectroscopic properties, you may have a vector that is free to orient uniformly over all possible angles θ . If you want to compute its average projection on an axis, using quantities such as $\langle \cos \theta \rangle$ or $\langle \cos^2 \theta \rangle$, put the beginning of the vector at the center of a sphere. If the vector orients uniformly, it points to any given patch on the surface of the sphere in proportion to the area of that patch.

The strip of area shown in Figure 1.8 has an angle θ with respect to the z -axis. The area of the strip is $(r d\theta)(2\pi\ell)$. Since $\ell = r \sin \theta$, the area of the strip is $2\pi r^2 \sin \theta d\theta$. A strip has less area if θ is small than if θ approaches 90° . The fraction of vectors $p(\theta)$ that point to, or end in, this strip is

$$p(\theta) = \frac{2\pi r^2 \sin \theta d\theta}{\int_0^\pi 2\pi r^2 \sin \theta d\theta} = \frac{\sin \theta d\theta}{\int_0^\pi \sin \theta d\theta}. \quad (1.44)$$

The average $\langle \cos \theta \rangle$ over all vectors is

$$\langle \cos \theta \rangle = \int_0^\pi \cos \theta p(\theta) d\theta = \frac{\int_0^\pi \cos \theta \sin \theta d\theta}{\int_0^\pi \sin \theta d\theta}. \quad (1.45)$$

Figure 1.8 A vector that can orient in all directions can be represented as starting at the origin and ending on the surface of a sphere. The area $2\pi r \ell d\theta$ represents the relative proportion of all the vectors that land in the strip at an angle between θ and $\theta + d\theta$ relative to the z axis.

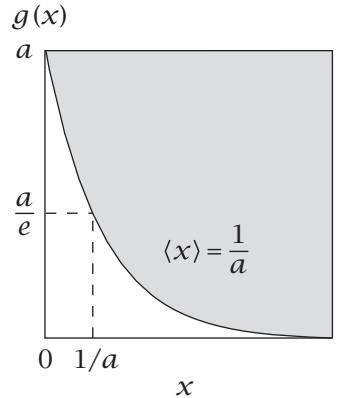
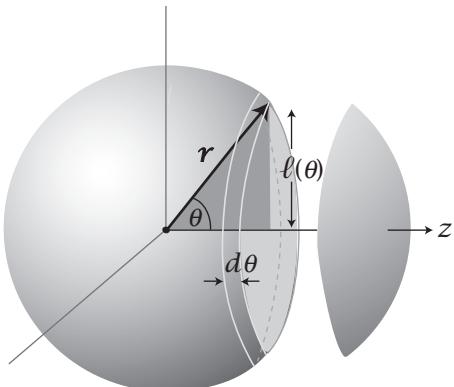


Figure 1.7 Exponential distribution function, $0 \leq x \leq \infty$. The average of $p(x) = ae^{-ax}$ is $\langle x \rangle = 1/a$ (see Example 1.25).

This integration is simplified by noticing that $\sin \theta d\theta = -d \cos \theta$, by letting $x = \cos \theta$, and by replacing the limits 0 and π by 1 and -1 . Then Equation (1.45) becomes

$$\langle \cos \theta \rangle = \frac{\int_1^{-1} x dx}{\int_1^{-1} dx} = \frac{\frac{1}{2}x^2 \Big|_1^{-1}}{x \Big|_1^{-1}} = 0. \quad (1.46)$$

Physically, this says that the average projection on the z axis of uniformly distributed vectors is zero. You can also see this by symmetry: just as many vectors point forward ($0^\circ < \theta \leq 90^\circ$) as backward ($90^\circ < \theta \leq 180^\circ$), so the average is zero.

Later we will find the quantity $\langle \cos^2 \theta \rangle$ to be useful. Following the same logic, you have

$$\langle \cos^2 \theta \rangle = \frac{\int_0^\pi \cos^2 \theta \sin \theta d\theta}{\int_0^\pi \sin \theta d\theta} = \frac{\int_1^{-1} x^2 dx}{\int_1^{-1} dx} = \frac{\frac{1}{3}x^3 \Big|_1^{-1}}{x \Big|_1^{-1}} = \frac{1}{3}. \quad (1.47)$$

Summary

Probabilities describe frequencies or incomplete knowledge. The addition and multiplication rules allow you to draw consistent inferences about probabilities of multiple events. Distribution functions describe collections of probabilities. Such functions have mean values and variances. Combined with combinatorics—the counting of arrangements of systems—probabilities provide the basis for reasoning about entropy, and about driving forces among molecules, described in the next chapter.

Examples of Distributions

Here are some probability distribution functions that commonly appear in statistical mechanics.

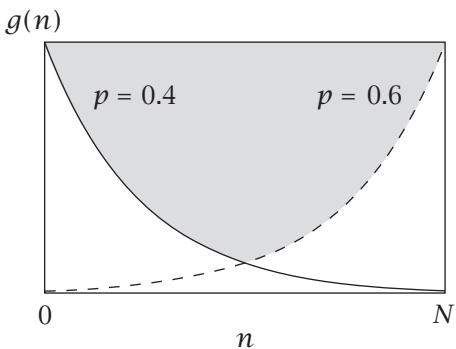


Figure 1.9 Bernoulli

$$g(n) = p^n(1-p)^{N-n}, \quad n = 0, 1, 2, \dots, N. \quad (1.48)$$

The Bernoulli distribution describes independent trials with two possible outcomes (see page 16). $g(n)$ is a distribution function, not a probability, because it is not normalized to sum to one.

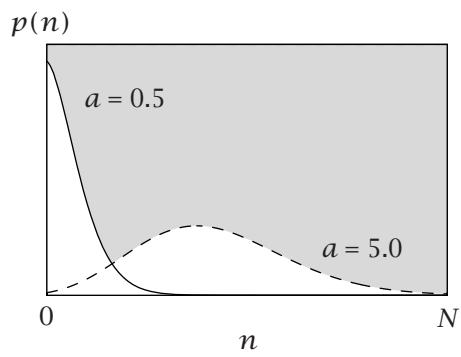


Figure 1.10 Poisson

$$p(n) = \frac{a^n e^{-a}}{n!}, \quad n = 0, 1, 2, \dots, N. \quad (1.49)$$

The Poisson distribution approximates the binomial distribution when the number of trials is large and the probability of each one is small [3]. It is useful for describing radioactive decay, the number of vacancies in the Supreme Court each year [4], the numbers of dye molecules taken up by small particles, or the sizes of colloidal particles.

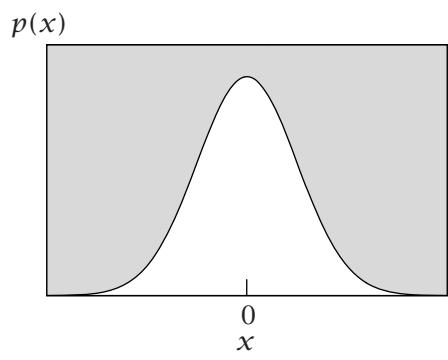


Figure 1.11 Gaussian

$$p(x) = \frac{1}{\sigma \sqrt{2\pi}} e^{-x^2/2\sigma^2}, \quad -\infty \leq x \leq \infty. \quad (1.50)$$

The Gaussian distribution is derived from the binomial distribution for large N [5]. It is important for statistics, error analysis, diffusion, conformations of polymer chains, and the Maxwell–Boltzmann distribution law of gas velocities.

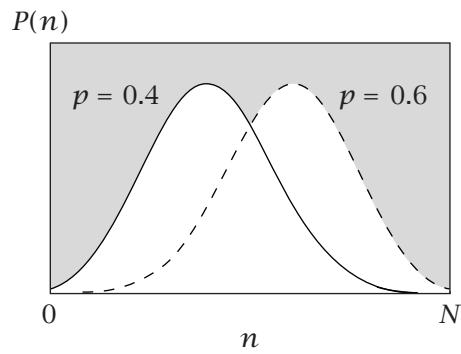


Figure 1.12 Binomial

$$P(n) = p^n (1-p)^{N-n} \times \left(\frac{N!}{n!(N-n)!} \right), \quad n = 0, 1, 2, \dots, N. \quad (1.51)$$

The binomial distribution for collections of Bernoulli trials is derived on pages 16–17.

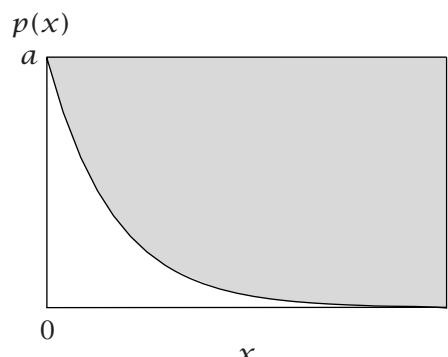


Figure 1.13 Exponential (Boltzmann)

$$p(x) = ae^{-ax}, \quad 0 \leq x \leq \infty. \quad (1.52)$$

The exponential, or Boltzmann distribution, is central to statistical thermodynamics (see Chapters 5 and 10).

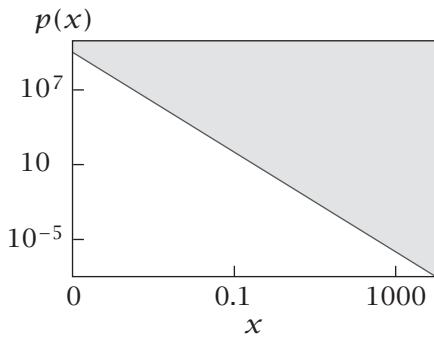


Figure 1.14 Power law

$$p(x) = 1/x^q, \quad (1.53)$$

where q is a constant called the *power law exponent*. Power laws describe the frequencies of earthquakes, the numbers of links to World Wide Web sites, the distribution of incomes ('the rich get richer'), and the noise spectrum in some electronic devices.

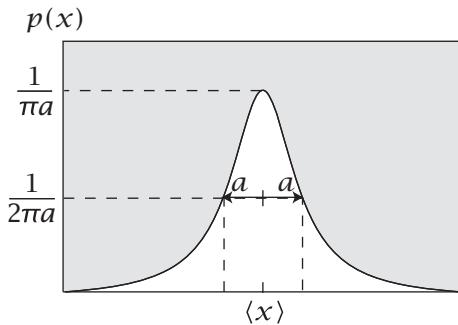


Figure 1.15 Lorentzian

$$p(x) = \frac{1}{\pi} \frac{a}{(x - \langle x \rangle)^2 + a^2}, \quad -\infty \leq x \leq \infty. \quad (1.54)$$

$2a$ is the width of the Lorentzian curve at the level of half the maximum probability. Lorentzian distributions are useful in spectroscopy [3].

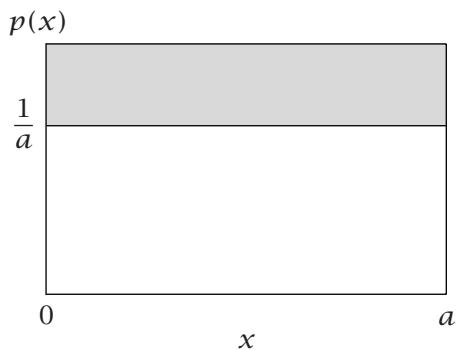


Figure 1.16 Flat

$$p(x) = 1/a, \quad (1.55)$$

where a is a constant independent of x (see Example 1.24).

Problems

1. Combining independent probabilities. You have applied to three schools: University of California at San Francisco (UCSF), Duluth School of Mines (DSM), and Harvard (H). You guess that the probabilities you'll be accepted are $p(\text{UCSF}) = 0.10$, $p(\text{DSM}) = 0.30$, and $p(\text{H}) = 0.50$. Assume that the acceptance events are independent.

- (a) What is the probability that you get in somewhere (at least one acceptance)?
- (b) What is the probability that you will be accepted by both Harvard and Duluth?

2. Probabilities of sequences. Assume that the four bases A, C, T, and G occur with equal likelihood in a DNA sequence of nine monomers.

- (a) What is the probability of finding the sequence AAATCGAGT through random chance?
- (b) What is the probability of finding the sequence AAAAAAAA through random chance?
- (c) What is the probability of finding any sequence that has four A's, two T's, two G's, and one C, such as that in (a)?

3. The probability of a sequence (given a composition). A scientist has constructed a secret peptide to carry a message. You know only the composition of the peptide, which is six amino acids long. It contains one serine S, one threonine T, one cysteine C, one arginine R, and two glutamates E. What is the probability that the sequence SECRET will occur by chance?

4. Combining independent probabilities. You have a fair six-sided die. You want to roll it enough times to ensure that a 2 occurs at least once. What number of rolls k is required to ensure that the probability is at least $2/3$ that at least one 2 will appear?

5. Predicting compositions of independent events. Suppose you roll a fair six-sided die three times.

- (a) What is the probability of getting a 5 twice from all three rolls of the dice?
- (b) What is the probability of getting a total of *at least* two 5's from all three rolls of the die?

6. Computing a mean and variance. Consider the probability distribution $p(x) = ax^n$, $0 \leq x \leq 1$, for a positive integer n .

- (a) Derive an expression for the constant a , to normalize $p(x)$.
- (b) Compute the average $\langle x \rangle$ as a function of n .
- (c) Compute $\sigma^2 = \langle x^2 \rangle - \langle x \rangle^2$ as a function of n .

7. Computing the average of a probability distribution. Compute the average $\langle i \rangle$ for the probability distribution function shown in Figure 1.17.

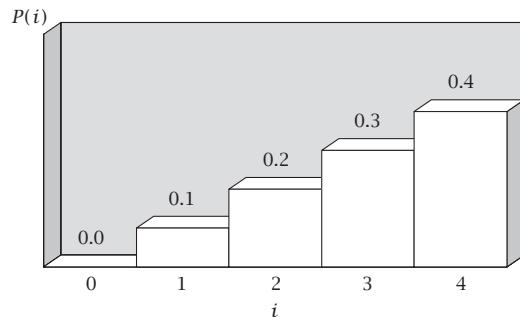


Figure 1.17 A simple probability distribution.

8. Predicting coincidence. Your statistical mechanics class has 25 students. What is the probability that at least two classmates have the same birthday?

9. The distribution of scores on dice. Suppose that you have n dice, each a different color, all unbiased and six-sided.

- (a) If you roll them all at once, how many distinguishable outcomes are there?
- (b) Given two distinguishable dice, what is the most probable sum of their face values on a given throw of the pair? (That is, which sum between 2 and 12 has the greatest number of different ways of occurring?)
- (c) What is the probability of the most probable sum?

10. The probabilities of identical sequences of amino acids. You are comparing protein amino acid sequences for homology. You have a 20-letter alphabet (20 different amino acids). Each sequence is a string n letters in length. You have one test sequence and s different database sequences. You may find any one of the 20 different amino acids at any position in the sequence, independent of what you find at any other position. Let p represent the probability that there will be a 'match' at a given position in the two sequences.

- (a) In terms of s , p , and n , how many of the s sequences will be perfect matches (identical residues at every position)?
- (b) How many of the s comparisons (of the test sequence against each database sequence) will have exactly one mismatch at any position in the sequences?

11. The combinatorics of disulfide bond formation. A protein may contain several cysteines, which may pair together to form disulfide bonds as shown in Figure 1.18. If there is an even number n of cysteines, $n/2$ disulfide bonds can form. How many different disulfide pairing arrangements are possible?

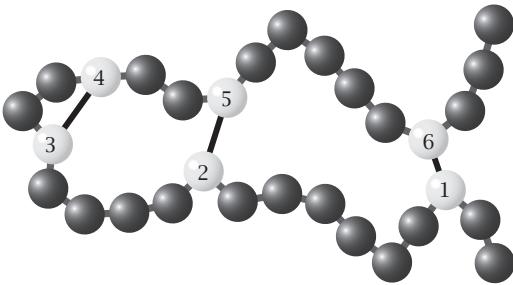


Figure 1.18 This disulfide bonding configuration with pairs 1–6, 2–5, and 3–4 is one of the many possible pairings. Count all the possible pairing arrangements.

12. Predicting combinations of independent events. If you flip an unbiased green coin and an unbiased red coin five times each, what is the probability of getting four red heads and two green tails?

13. A pair of aces. What is the probability of drawing two aces in two random draws without replacement from a full deck of cards?

14. Average of a linear function. What is the average value of x , given a distribution function $q(x) = cx$, where x ranges from zero to one, and $q(x)$ is normalized?

15. The Maxwell-Boltzmann probability distribution function. According to the kinetic theory of gases, the energies of molecules moving along the x direction are given by $\varepsilon_x = (1/2)mv_x^2$, where m is mass and v_x is the velocity in the x direction. The distribution of particles over velocities is given by the Boltzmann law, $p(v_x) = e^{-mv_x^2/2kT}$. This is the Maxwell-Boltzmann distribution (velocities may range from $-\infty$ to $+\infty$).

- (a) Write the probability distribution $p(v_x)$, so that the Maxwell-Boltzmann distribution is correctly normalized.
- (b) Compute the average energy $\langle \frac{1}{2}mv_x^2 \rangle$.
- (c) What is the average velocity $\langle v_x \rangle$?
- (d) What is the average momentum $\langle mv_x \rangle$?

16. Predicting the rate of mutation based on the Poisson probability distribution function. The evolutionary process of amino acid substitution in proteins is sometimes described by the Poisson probability distribution function. The probability $p_s(t)$ that exactly s substitutions at a given amino acid position occur over an evolutionary time t is

$$p_s(t) = \frac{e^{-\lambda t} (\lambda t)^s}{s!},$$

where λ is the rate of amino acid substitution per site per unit time. Fibrinopeptides evolve rapidly: $\lambda_F = 9.0$ substitutions per site per 10^9 years. Lysozyme is intermediate: $\lambda_L \approx 1.0$. Histones evolve slowly: $\lambda_H = 0.010$ substitutions per site per 10^9 years.

- (a) What is the probability that a fibrinopeptide has no mutations at a given site in $t = 1$ billion years?

(b) What is the probability that lysozyme has three mutations per site in 100 million years?

(c) We want to determine the expected number of mutations $\langle s \rangle$ that will occur in time t . We will do this in two steps. First, using the fact that probabilities must sum to one, write $\alpha = \sum_{s=0}^{\infty} (\lambda t)^s / s!$ in a simpler form.

(d) Now write an expression for $\langle s \rangle$. Note that

$$\sum_{s=0}^{\infty} \frac{s(\lambda t)^s}{s!} = (\lambda t) \sum_{s=1}^{\infty} \frac{(\lambda t)^{s-1}}{(s-1)!} = \lambda t \alpha.$$

(e) Using your answer to part (d), determine the ratio of the expected number of mutations in a fibrinopeptide to the expected number of mutations in histone protein, $\langle s \rangle_{\text{fib}} / \langle s \rangle_{\text{his}}$ [6].

17. Probability in court. In forensic science, DNA fragments found at the scene of a crime can be compared with DNA fragments from a suspected criminal to determine the probability that a match occurs by chance. Suppose that DNA fragment A is found in 1% of the population, fragment B is found in 4% of the population, and fragment C is found in 2.5% of the population.

- (a) If the three fragments contain independent information, what is the probability that a suspect's DNA will match all three of these fragment characteristics by chance?
- (b) Some people believe such a fragment analysis is flawed because different DNA fragments do not represent independent properties. As before, suppose that fragment A occurs in 1% of the population. But now suppose the conditional probability of B , given A , is $p(B|A) = 0.40$ rather than 0.040, and $p(C|A) = 0.25$ rather than 0.025. There is no additional information about any relationship between B and C . What is the probability of a match now?

18. Flat distribution. Given a flat distribution, from $x = -a$ to $x = a$, with probability distribution $p(x) = 1/(2a)$:

- (a) Compute $\langle x \rangle$.
- (b) Compute $\langle x^2 \rangle$.
- (c) Compute $\langle x^3 \rangle$.
- (d) Compute $\langle x^4 \rangle$.

19. Family probabilities. Given that there are three children in a family, what is the probability that:

- (a) two are boys and one is a girl?
- (b) all three are girls?

20. Evolutionary fitness. Suppose that the probability of having the dominant allele (D) in a gene is p and the probability of the recessive allele (R) is $q = 1 - p$. You have two alleles, one from each parent.

- (a) Write the probabilities of all the possibilities: DD, DR, and RR.

- (b) If the fitness of **DD** is f_{DD} , the fitness of **DR** is f_{DR} , and the fitness of **RR** is f_{RR} , write the average fitness in terms of p .

21. Ion-channel events. A biological membrane contains N ion-channel proteins. The fraction of time that any one protein is open to allow ions to flow through is q . Express the probability $P(m, N)$ that m of the channels will be open at any given time.

22. Joint probabilities: balls in a barrel. For Example 1.10, two green balls and one red ball drawn from a barrel without replacement:

- (a) Compute the probability $p(RG)$ of drawing one red and one green ball in either order.
- (b) Compute the probability $p(GG)$ of drawing two green balls.

23. Sports and weather. The San Francisco football team plays better in fair weather. They have a 70% chance of winning in good weather, but only a 20% chance of winning in bad weather.

- (a) If they play in the Super Bowl in Wisconsin and the weatherman predicts a 60% chance of snow that day, what is the probability that San Francisco will win?
- (b) Given that San Francisco lost, what is the probability that the weather was bad?

24. Monty Hall's dilemma: a game show problem. You are a contestant on a game show. There are three closed doors: one hides a car and two hide goats. You point to one door, call it *C*. The gameshow host, knowing what's behind each door, now opens either door *A* or *B*, to show you a goat; say it's door *A*. To win a car, you now get to make your final choice: should you stick with your original choice *C*, or should you now switch and choose door *B*? (*New York Times*, July 21, 1991; *Scientific American*, August 1998.)

25. Probabilities of picking cards and rolling dice.

- (a) What is the probability of drawing either a queen or a heart in a normal deck of 52 cards?
- (b) What is the probability P of getting three 7's and two 4's on five independent rolls of a die?

26. Probability and translation-start codons. In prokaryotes, translation of mRNA messages into proteins is most often initiated at start codons on the mRNA having the sequence AUG. Assume that the mRNA is single-stranded and consists of a sequence of bases, each described by a single letter A, C, U, or G.

Consider the set of all random pieces of bacterial mRNA of length six bases.

- (a) What is the probability of having either no A's or no U's in the mRNA sequence of six base pairs long?
- (b) What is the probability of a random piece of mRNA having exactly one A, one U, and one G?
- (c) What is the probability of a random piece of mRNA of length six base pairs having an A directly

followed by a U directly followed by a G; in other words, having an AUG in the sequence?

- (d) What is the total number of random pieces of mRNA of length six base pairs that have exactly one A, exactly one U, and exactly one G, with A appearing first, then the U, then the G? (e.g., AXUXG)

27. DNA synthesis. Suppose that upon synthesizing a molecule of DNA, you introduce a wrong base pair, on average, every 1000 base pairs. Suppose you synthesize a DNA molecule that is 1000 bases long.

- (a) Calculate and draw a bar graph indicating the yield (probability) of each product DNA, containing 0, 1, 2, and 3 mutations (wrong base pairs).
- (b) Calculate how many combinations of DNA sequences of 1000 base pairs contain exactly 2 mutant base pairs.
- (c) What is the probability of having specifically the 500th base pair and the 888th base pair mutated in the pool of DNA that has only two mutations?
- (d) What is the probability of having two mutations side-by-side in the pool of DNA that has only two mutations?

28. Presidential election. Two candidates are running for president. Candidate *A* has already received 80 electoral votes and only needs 35 more to win. Candidate *B* already has 50 votes, and needs 65 more to win.

Five states remain to be counted. Winning a state gives a candidate 20 votes; losing gives the candidate zero votes. Assume both candidates otherwise have equal chances to win in those five states.

- (a) Write an expression for $W_{A,\text{total}}$, the number of ways *A* can succeed at winning 40 more electoral votes.
- (b) Write the corresponding expression for $W_{B,\text{total}}$.
- (c) What is the probability candidate *A* beats candidate *B*?

References

- [1] L Krantz, *What the Odds Are*. Harper Perennial, New York, 1992.
- [2] WT Ziemba and DB Hansch, *Dr Z's Beat the Racetrack*. W Morrow, New York, 1987.
- [3] P Bevington and DK Robinson, *Reduction and Error Analysis for the Physical Sciences*, 3rd edition. McGraw-Hill, New York, 2002.
- [4] S Ross, *A First Course in Probability*, 8th edition. Pearson, Upper Saddle River, NJ, 2010.
- [5] F Reif, *Fundamentals of Statistical and Thermal Physics*. McGraw-Hill, New York, 1965.
- [6] M Nei, *Molecular Evolutionary Genetics*. Columbia Press, New York, 1987.

Suggested Reading

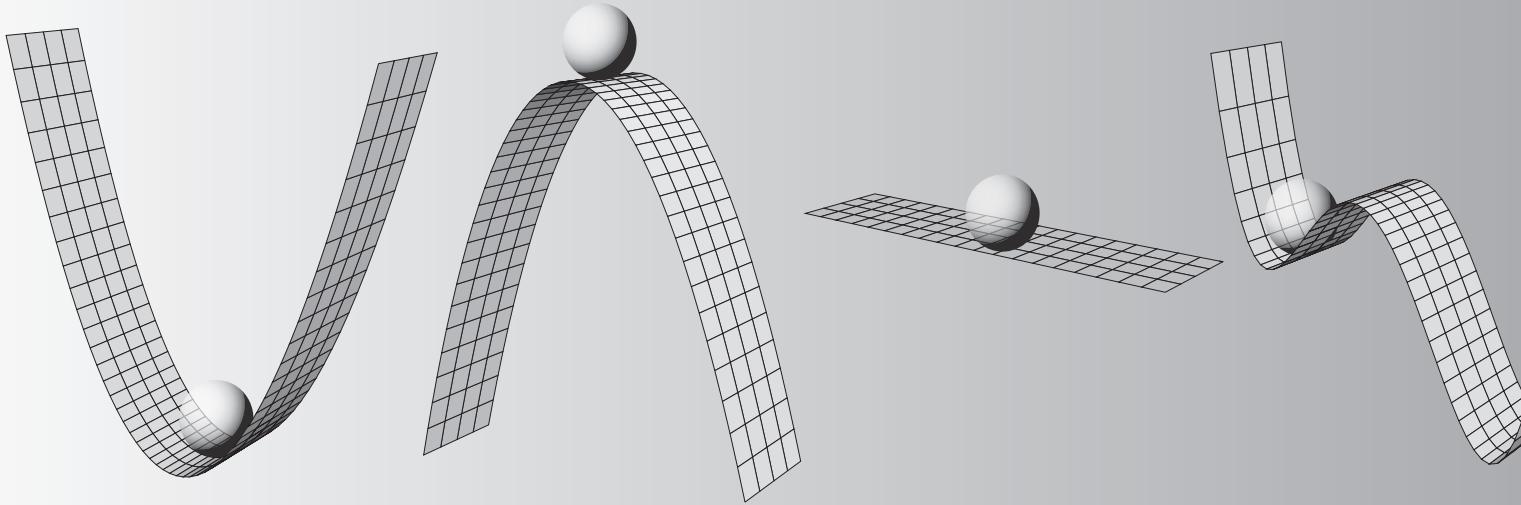
- W Feller, *An Introduction to Probability Theory and its Applications*, Volume 1, 3rd edition and Volume 2, 2nd edition, Wiley, New York, 1968 and 1971. Advanced treatise on principles and applications of probability theory.
- PG Hoel, *Introduction to Mathematical Statistics*, 5th edition, Wiley, New York, 1984. Excellent review of uses of probabilities in statistics: statistical inference, hypothesis testing, and estimation.

E Nagel, *Principles of Probability Theory*, University of Chicago, Chicago, 1939. Philosophical foundations of probability theory.

S Ross, *A First Course in Probability*, 8th edition, Pearson, Upper Saddle River, NJ, 2010. An excellent introduction to probability theory.

RA Rozanov, *Probability Theory: A Concise Course*, Dover, New York, 1977. A concise summary of probabilities and distribution functions.

2 Extremum Principles Predict Equilibria



What Are Extremum Principles?

The forces on atoms and molecules can be described in terms of two tendencies called **energy** and **entropy**. Molecules react, change conformations, bind, and undergo other chemical or physical changes in ways that **cause a quantity** called the **energy to reach its minimum** possible value and a quantity called the entropy to reach **its maximum possible value**. We can predict the tendencies of matter by computing the minima or maxima of certain mathematical functions. These are called extremum (or variational) principles.

To illustrate, we first show that balls rolling downhill can be predicted by the minimization of energy. We describe the various types of extrema, called *states of equilibrium*: stable, unstable, neutral, and metastable states. Then, as a first step toward introducing the entropy and the Second Law of Thermodynamics, we use the multiplicity concept from Chapter 1, and illustrate how gases exerting pressure, the mixing and diffusion of molecules, and rubber elasticity can be explained by a maximization principle.

EXAMPLE 2.1 Mechanical equilibrium is the state of minimum potential energy. For a ball of mass m on a hillside at height z , the gravitational potential energy $V(z)$ is

$$V(z) = mgz,$$

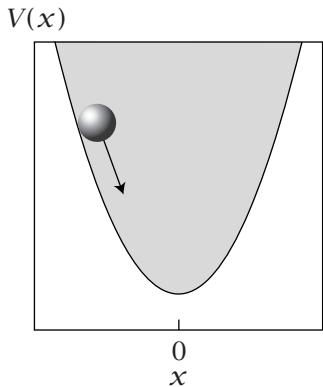


Figure 2.1 The equilibrium position $x^* = 0$ for a ball in a quadratic valley has the minimum potential energy $V(x^*)$.

where g is the gravitational acceleration constant (see Appendix H). Figure 2.1 shows a valley in which the height $z(x)$ depends quadratically on the lateral position x :

$$z = x^2.$$

Substitute $z = x^2$ into $V(z) = mgz$ to find the potential energy of a ball in this quadratic valley:

$$V(x) = mgx^2.$$

The extremum principle says that the ball will roll to wherever the potential energy is a minimum.¹ To find the minimum, determine what value, $x = x^*$, causes the derivative to be zero:

$$\frac{dV(x)}{dx} \Big|_{x^*} = 0 \implies 2mgx^* = 0. \quad (2.1)$$

Equation (2.1) is satisfied when $x^* = 0$. Because the second derivative of V with respect to x is positive, this extremum is a minimum (the mathematics of extrema are discussed in Chapter 4).

Degrees of Freedom and Constraints

A quantity such as the position x of the ball in the valley is called a *degree of freedom* of the system, because the system is free to change that quantity. The alternative to a degree of freedom is a *constraint*. If the ball were fixed at position $x = x_c$, it would be *constrained*. A system is not able to change a constraint. Constraints are imposed on the system from outside. A system governed by an extremum principle will change its degree of freedom until it reaches the maximum or minimum of some function allowed by the constraints. Systems can have multiple degrees of freedom and constraints.

What Is a State of Equilibrium?

A ball rolls downhill until it reaches the bottom, and then stays there. The bottom is its *equilibrium* position. Equilibrium defines where a system tends to go and stay; the net force is zero, $f(x^*) = 0$. The *force* f on the ball defines the strength of its tendency toward equilibrium. The potential energy $V(x) = -\int f(x) dx$ is defined as the integral of the force, or

$$f = -\left(\frac{dV}{dx}\right). \quad (2.2)$$

Therefore the state of equilibrium can be defined either as an extremum in the energy or as the point where the net force is zero.

¹For systems called ‘conservative,’ the sum of the potential energy and the kinetic energy is a constant. But systems that give off heat due to friction, such as balls rolling down hills, are not conservative. In those cases, the kinetic energy decreases over time and the potential energy tends toward a minimum.

Stable, Unstable, Neutral, and Metastable States

Zero net force can be achieved in different ways. The potential energy is a function of the degrees of freedom, $V(x)$ in Example 2.1. Potential energy functions can take any shape. They may be perfectly flat, or smooth and curved, or rough and bumpy. An equilibrium state is defined as *stable* where a potential is at a minimum, *neutral* where a potential is flat, or *metastable* where a potential energy is at a minimum but other minima are deeper. Where a potential energy is at a maximum, a state is *unstable*. The stability of a system depends not just on whether there is zero force, but also on how the system responds to perturbations. Think about a ping-pong ball in a slight breeze. The breeze simply represents a small random fluctuating force.

The force on the ball due to gravity is $f(x) = -dV/dx$. Our sign convention specifies that as we move the ball in a direction to increase its potential energy (up), the force acts in the opposite direction (down) to reduce the energy.

STABLE EQUILIBRIUM. For a system in a *stable* equilibrium at $x = x^*$, the energy is at a minimum, as shown in Figure 2.2. The potential energy increases as the system moves away from a stable equilibrium:

$$V(x) > V(x^*) \quad \text{for all } x \neq x^*.$$

A stable equilibrium can be identified by tests of the first and second derivatives:

$$\frac{dV}{dx} = 0 \quad \text{and} \quad \frac{d^2V}{dx^2} > 0 \quad \text{at } x = x^*.$$

To move a system incrementally *toward* a stable equilibrium requires a downward step, $dV \leq 0$. If a small fluctuation displaces the system away from $x = x^*$, a stable system tends to return to $x = x^*$. If you put a ping-pong ball in a hole in your backyard, it will still be there later, even with the breeze blowing.

NEUTRAL EQUILIBRIUM. In a *neutral* equilibrium, the potential energy surface is flat, as shown in Figure 2.3:

$$\frac{dV}{dx} = 0 \quad \text{for all } x.$$

If you put a ping-pong ball on a flat surface, when you return the breeze may have blown it elsewhere on the surface.

METASTABLE EQUILIBRIUM. A system in *metastable* equilibrium is stable to small displacements, but unstable to larger displacements, as shown in Figure 2.4:

$$V(x) > V(x^*) \quad \text{for small } |x - x^*|,$$

$$\frac{dV}{dx} = 0 \quad \text{and} \quad \frac{d^2V}{dx^2} > 0 \quad \text{at } x = x^*,$$

$$V(x) < V(x^*) \quad \text{for large } |x - x^*|.$$

In metastable equilibrium, a system is locally but not globally stable. If you put a ping-pong ball in a small hole on the side of a deeper ditch, and if the breeze

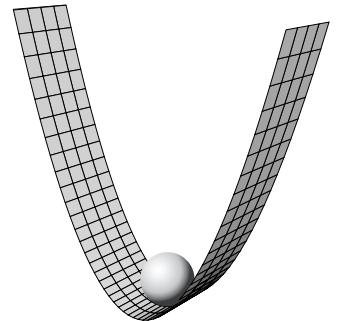


Figure 2.2 Stable.

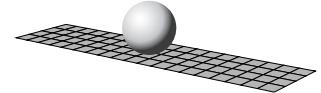


Figure 2.3 Neutral.

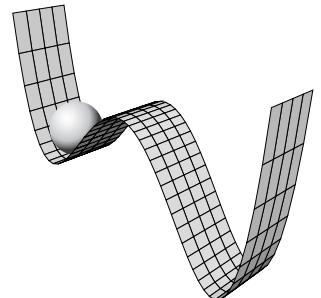


Figure 2.4 Metastable.

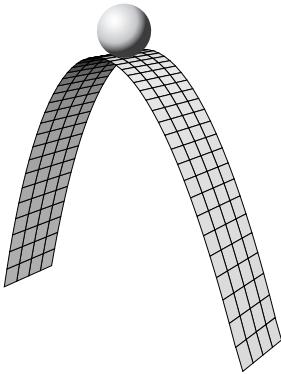


Figure 2.5 Unstable.

Table 2.1 $N = 4$

n	W	$\ln W$
4	$\frac{4!}{0!4!} = 1$	0
3	$\frac{4!}{1!3!} = 4$	1.386
2	$\frac{4!}{2!2!} = 6$	1.792
1	$\frac{4!}{3!1!} = 4$	1.386
0	$\frac{4!}{4!0!} = 1$	0
Total $W = 16$		

Table 2.2 $N = 10$

n	W	$\ln W$
10	1	0
9	10	2.303
8	45	3.807
7	120	4.787
6	210	5.347
5	252	5.529
4	210	5.347
3	120	4.787
2	45	3.807
1	10	2.303
0	1	0
Total $W = 1024$		

is small, it will stay in the small hole. However, in a stiffer wind, you might later find the ball in the deeper ditch.

UNSTABLE STATES. An *unstable* state is the top of a hill as shown in Figure 2.5. The system can reduce its energy by moving away from the point of zero force x^* in either the $+x$ or $-x$ direction:

$$V(x) < V(x^*) \quad \text{for } x \neq x^*.$$

To determine whether a state is stable or unstable requires a second-derivative test to see whether the curvature of $V(x)$ is up or down. For an unstable state,

$$\frac{dV}{dx} = 0 \quad \text{and} \quad \frac{d^2V}{dx^2} < 0 \quad \text{at } x = x^*.$$

Moving toward an unstable state is described by $dV \geq 0$. A ping-pong ball placed on top of a hill won't be there long. If the slightest breeze is blowing, the ball will roll off the hill.

The cover of this book shows an energy surface involving multiple degrees of freedom, and various stable, metastable, and unstable states.

An Extremum Principle: Maximizing Multiplicity Predicts the Most Probable Outcomes

Now we introduce a different extremum principle, one that predicts the distributions of outcomes in statistical systems, such as coin flips or die rolls. This will lead to the concept of entropy and the Second Law of Thermodynamics.

Let's consider the difference between two questions of probability, one based on the *sequence* and the other on the *composition* of outcomes. First, in a series of 4 coin tosses, which *sequence*, HHHH or HTHH, is more likely? The answer, given in Examples 1.3 and 1.4, is that all sequences are equally probable. Each sequence occurs with a probability of $(1/2)^4$. Contrast this with a question of *composition*. Which *composition* is more probable: 4 H's and 0 T's, or 3 H's and 1 T? Although each *sequence* is equally probable, each *composition* is not. 3 H's and 1 T is a more probable composition than 4 H's and 0 T's. Only 1 of the 16 possible sequences is composed of 4 H's and 0 T's, but 4 of the 16 possible sequences are composed of 3 H's and 1 T (HHHT, HHTH, HTHH, and THHH).

Predicting Heads and Tails by a Principle of Maximum Multiplicity

A simple maximization principle predicts the most probable composition: 2 heads and 2 tails in this case. Table 2.1 shows the number of sequences $W(n, N) = N!/[n!(N-n)!]$, as a function of n , the number of heads in $N = 4$ coin flips. W is a maximum (six sequences) when there are 2 heads, $n = 2$. The most probable composition of 4 coin flips is 2H's and 2T's, because more sequences (6 out of 16) have this composition than any other. Table 2.1 shows that the value of n ($n^* = 2$) that maximizes W also maximizes $\ln W$. Later, we will use $\ln W$ to define the entropy.

What happens if you have $N = 10$ coin flips instead of $N = 4$? Table 2.2 shows that the maximum multiplicity occurs again when the number of heads equals

the number of tails. You would observe the $5H, 5T$ composition 24.6% of the time (252 arrangements out of 1024 possible). Any other composition is less probable. Nearly 90% of the time, the composition will be within ± 2 heads or tails of $5H, 5T$.

The multiplicity function $W(n)$ becomes increasingly peaked as you increase the number of coin flips N (see Figure 2.6). Suppose you have $N = 100$ coin flips. The number of sequences of coin flips that have 50 heads and 50 tails is nearly a millionfold greater than the number of sequences that have 25 heads and 75 tails:

$$W(50, 100) = \frac{100!}{50!50!} = 1.01 \times 10^{29}$$

and

$$W(25, 100) = \frac{100!}{25!75!} = 2.43 \times 10^{23}.$$

If you observed 25 heads and 75 tails more often than 50 heads and 50 tails in many sets of $N = 100$ coin flips, you would have evidence of bias. Furthermore, because there is only one sequence that is all heads, the probability of observing 100 heads in 100 flips of an unbiased coin is 9.9×10^{-30} , virtually zero.

The implication of these results is remarkable. While the random process of flipping coins can result in any composition of heads and tails, the composition has a strong tendency toward 50% H and 50% T . Indeed, even though this is a random process, if the number of trials is large enough, the composition of heads and tails **becomes predictable with great precision**.

Like the energy minimization principle, here we have a function $W(n)$, the maximum of which identifies the value of n that we are most likely to observe. To maximize W or $\ln W$ for a series of coin flips, we compute $dW(n)/dn$ for fixed N , and find the value $n = n^*$ that causes the derivative to equal zero. The result is that $n^* = N/2$. In this case, the number of heads n acts like a degree of freedom of the system. n can take on any value from zero to N . Any individual sequence is just as likely as any other. The reason that the 50% heads, 50% tails composition is so strongly favored is that there are more distinguishable sequences of coin flips with that composition than any other. No individual coin flip has any particular tendency, or ‘feels’ any force. As we’ll see, this is **the nature of entropy as a driving force**. Systems tend to be found in the states with the **highest multiplicity**. The maximization of W or $\ln W$ serves the same predictive role for the composition of coin flips as the **minimization of energy** serves for the position of the ball in the valley.

Now we will use very simple models to illustrate how this principle of **maximum multiplicity** also predicts the tendencies of gases to expand and exert pressure, the tendencies of molecules to mix and diffuse, and the tendency of rubber to retract when it is stretched.

Simple Models Show How Systems Tend Toward Disorder

In the examples below and throughout the text, we will rely on one of the major tools of physics, simplified models. Simple models are caricatures. They seek

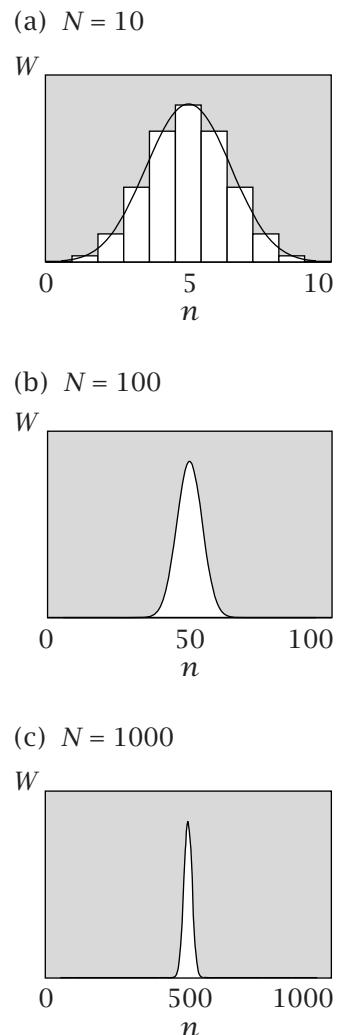


Figure 2.6 The multiplicity function W of the number of heads n narrows as the total number of trials N increases.

to keep **what is essential**, and omit less important details. Model-makers are storytellers. Stories can be involved and detailed, or they can describe just the essentials. Simple models do the latter. The art of model building is in recognizing what is essential. Sometimes a given phenomenon can be described by different models. For example, we will find that both lattice models and quantum mechanics can give the ideal gas law. The best models are the least arbitrary—they have the fewest assumptions or adjustable parameters. A model is useful if it makes testable predictions (right or wrong), and if it helps **encapsulate or unify our knowledge**.

Here and in later chapters we use lattice models. In lattice models, atoms or parts of molecules are represented as hard spherical beads. Space is divided up into bead-sized boxes, called lattice sites, which are artificial, mutually exclusive, and collectively exhaustive units of space. Each lattice site holds either zero or one bead. Two or more beads cannot occupy the same site. So, *occupied* or *not* are mutually exclusive options for a given lattice site. For now, the occupancy or not of one lattice site is independent of that of another site (later, we'll consider correlations). The lattice model just captures the idea that particles can be located at different positions in space, and that no two particles can be in the same place at the same time. This is really all we need for some problems, such as the following.

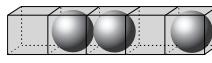
EXAMPLE 2.2 Why do gases exert pressure? A lattice model. Imagine a gas of N spherical particles that are free to distribute throughout either a large volume or a small one. Why does the gas spread out into a large volume?

The tendency to spread out, called pressure, can be explained either in terms of mechanics or maximum multiplicity. These are two different perspectives on the same process. According to the mechanical interpretation, pressure results from particles banging against the container walls. In general, few problems can be solved by the mechanical approach. The multiplicity perspective is more powerful, particularly for complex problems. We illustrate the multiplicity perspective here.

Let's miniaturize the problem. Our aim is to devise a model that we can visualize easily and solve exactly, and that will capture the essence of the problem without mathematical complexity. Figure 2.7 shows $N = 3$ spherical particles spread out in five boxes (in *A*), contained in a smaller volume (in *B*), and bunched up against the left wall of a one-dimensional lattice container (in *C*). We take the volume M (the number of lattice sites in which the particles are distributed) as the degree of freedom: $M_A = 5$, $M_B = 4$, and $M_C = 3$.

What value of M maximizes the multiplicity? You can compute the multiplicity $W(N, M)$ of N particles in M lattice sites the same way that you count the

Figure 2.7 In Example 2.2, Case *A*, three particles are distributed throughout $M_A = 5$ sites. In Case *B*, the three particles are in $M_B = 4$ sites. In Case *C*, the three particles fill $M_C = 3$ sites.

Case	Configuration	Volume
<i>A</i>		5
<i>B</i>		4
<i>C</i>		3

number of distinguishable sequences of N heads in M coin flips. That is, the sequence [vacant, occupied, occupied, vacant, occupied] is just a set of binary outcomes like the coin flip sequence [THHTH]. Our premise is that every sequence is equally likely, no matter how its vacancies and occupancies are arranged. The number of distinguishable arrangements of vacancies and occupancies of N particles in M sites is

$$W(N, M) = \frac{M!}{N!(M-N)!}, \quad (2.3)$$

so there are $W_A(3, 5) = 5!/(3!2!) = 10$ possible arrangements for $M_A = 5$, $W_B(3, 4) = 4!/(3!1!) = 4$ arrangements for $M_B = 4$, and $W_C(3, 3) = 3!/(3!0!) = 1$ arrangement for $M_C = 3$. The multiplicity increases as the volume increases. If the system has only these three possible volumes available, then the probability is $p_C = 1/15$ [i.e., $1/(1+4+10)$] that the particles will be bunched up on the left side of the container, $p_B = 4/15$ that the gas will expand to an intermediate degree, and $p_A = 2/3$ that the particles will be fully spread out.

If the degree of freedom of the system is its volume, the particles will spread out into the largest possible volume to maximize the multiplicity of the system. This is the basis for the force called *pressure*. In Chapters 6 and 24, after we have developed the thermodynamics that we need to define the pressure, we will show that, despite the simplicity of this model, it accurately gives the ideal and van der Waals gas laws.

Our states A , B , and C are called *macrostates* because they are macroscopic, observable, or controllable: we could perform an experiment in which we vary or measure the volume V or the density N/V of the system. Each macrostate is made up of one or more *microstates*, which are the individual snapshots that we counted above. The principle that we assert here is that each microstate is equally probable. So, the probability of observing any particular macrostate will be proportional to how many microstates it has.

While the *pressure* is a force for expansion, the *chemical potential* is a force for *mixing*, as when a dye diffuses in water. The chemical potential describes the tendencies of substances to *dissolve*, the tendency to partition between different phases of matter, or the tendency to transform into different chemical species. Here we use the lattice model to show why particles tend to mix.

EXAMPLE 2.3 Why do materials diffuse? A lattice model for the chemical potential. Suppose that you have four black particles  and four white particles  in eight lattice sites. There are four lattice sites on the left and four lattice sites on the right, separated by a permeable wall. The total volume is fixed. All the lattice sites are occupied by either a  or a  particle. Now the degrees of freedom are the numbers of  and  particles on each side of the permeable wall, not the volume. Again we consider three cases. As in Example 2.2, Figure 2.8 shows one particular spatial configuration for each case, but several different configurations might be possible for a given composition. What is the most probable composition of  and  particles on each side of the wall? (We predict compositions rather than particle sequences because this is what experiments can measure.)

Once again, the statistical mechanical approach is to assume that each spatial configuration (sequence) is equally probable. Find the most probable mixture by

Case	Left	Right
A		
B		
C		
Permeable Barrier →		

Figure 2.8 In Example 2.3, Case A, the composition is two ● and two ○ particles on the left, and two ● and two ○ on the right. In Case B, the composition is three ● and one ○ on the left and one ● and three ○ on the right. In Case C, the composition is four ● and zero ○ on the left, and zero ● and four ○ on the right.

maximizing the multiplicity of arrangements. For each given value of left and right compositions, the total multiplicity is the product of the multiplicities for the left and the right sides:

$$\text{Case A: } W = W(\text{left}) \cdot W(\text{right}) = \frac{4!}{2!2!} \frac{4!}{2!2!} = 36.$$

$$\text{Case B: } W = W(\text{left}) \cdot W(\text{right}) = \frac{4!}{1!3!} \frac{4!}{3!1!} = 16.$$

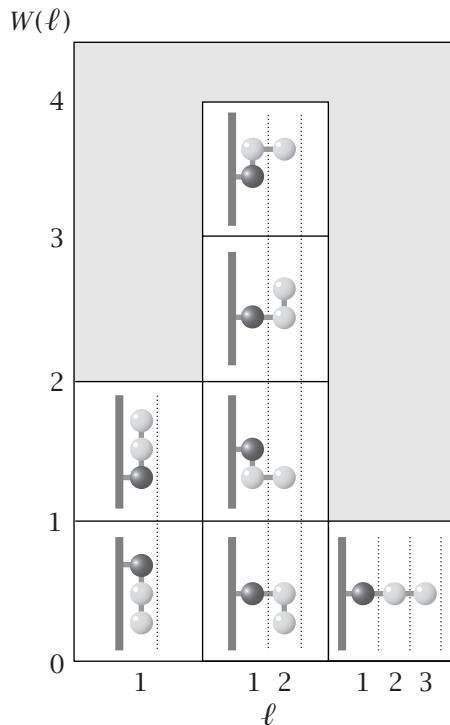
$$\text{Case C: } W = W(\text{left}) \cdot W(\text{right}) = \frac{4!}{0!4!} \frac{4!}{4!0!} = 1.$$

Case A, which has the most uniform particle distribution, has the highest multiplicity and therefore is the most probable. Case C, which has the greatest particle segregation, is the least probable. If the degree of freedom is the extent of particle exchange, then the multiplicity will be greatest when the particles are distributed most uniformly. You can predict the tendencies of particles to mix (i.e., diffuse) with the principle that the system tends toward the distribution that maximizes its multiplicity.

EXAMPLE 2.4 Why is rubber elastic? When you stretch a rubber band, it pulls back. This retractive force is due to the tendency of polymers to adopt conformations that maximize their multiplicities. Polymers have fewer *conformational* states when fully stretched (see Chapters 32–34). Figure 2.9 illustrates the idea. Fix the first monomer of a chain to a wall. The degree of freedom is the position of the other end, at distance $\ell = 1, 2$, or 3 from the wall. $W(\ell)$ is the multiplicity of configurations that have the second chain end in layer ℓ .

This simple model shows that a polymer will tend to be found in a partly *retracted* state $\ell = 2$, rather than stretched to length $\ell = 3$, because of the greater number of conformations ($W = 4$) for $\ell = 2$ than for $\ell = 3$ ($W = 1$). For the same reason, the polymer also tends not to be *flattened* on the surface: $W = 2$ when $\ell = 1$. If you stretch a polymer molecule, it will retract on the basis of the principle that it will tend toward its state of maximum multiplicity. This is the nature of entropy as a force.

Figure 2.9 The simple two-dimensional lattice model of Example 2.4 for a polymer chain in different conformations. The chain monomers are shown as beads. Each chain has three monomers connected by two bonds. Bead number one is attached to a wall. Starting at the wall, the layers of the lattice are numbered $\ell = 1, 2, 3$. To measure the degree of chain stretching, we use ℓ , the lattice layer number where monomer three resides. ℓ is a degree of freedom relevant to this problem because the chain is free to alter its conformations and change its degree of stretching. Count the number of conformations $W(\ell)$ as a function of ℓ . W is maximal when the chain ends in layer $\ell = 2$. This is a more retracted state than $\ell = 3$. To maximize W , a polymer molecule retracts when stretched.



In Chapter 5, we will define the quantity called *entropy*, $S = \text{constant} \times \ln W$, which has a maximum for the values of the degrees of freedom that maximize W . The tendencies of systems toward states of maximum multiplicity are also tendencies toward states of maximum entropy.

Summary

Forces act on atoms and molecules. The degrees of freedom of a system will change until the system reaches a state in which the net forces are zero. An alternative description is in terms of extremum principles: systems tend toward states that have minimal energies or maximal multiplicities, or, as we will see in Chapter 8, a combination of the two tendencies. The tendency toward maximum multiplicity or maximum entropy accounts for the pressures of gases, the mixing of fluids, the retraction of rubber, and, as we'll see in the next chapter, the flow of heat from hot objects to cold ones.

Problems

1. A lattice gas. How many arrangements are there of 15 indistinguishable lattice gas particles distributed on:

- (a) $V = 20$ sites?
- (b) $V = 16$ sites?
- (c) $V = 15$ sites?

2. Maximum of binomial distribution. Find the value $n = n^*$ that causes the function

$$W = \frac{N!}{n!(N-n)!} p^n (1-p)^{N-n}$$

to be at a maximum, for constants p and N . Use Stirling's approximation, $x! \approx (x/e)^x$ (see Appendix B). Note that it is easier to find the value of n that maximizes $\ln W$ than the value that maximizes W . The value of n^* will be the same.

3. Finding extrema. For the function

$$V(x) = \frac{1}{3}x^3 + \frac{5}{2}x^2 - 24x$$

- (a) Where is the maximum?
- (b) Where is the minimum?

4. The binomial distribution narrows as N increases. Flip a coin $4N$ times. The most probable number of heads

is $2N$, and its probability is $p(2N)$. If the probability of observing N heads is $p(N)$, show that the ratio $p(N)/p(2N)$ decreases as N increases.

5. De-mixing is improbable. Using the diffusion model of Example 2.3, with $2V$ lattice sites on each side of a permeable wall and a total of $2V$ white particles and $2V$ black particles, show that perfect de-mixing (all white on one side, all black on the other) becomes increasingly improbable as V increases.

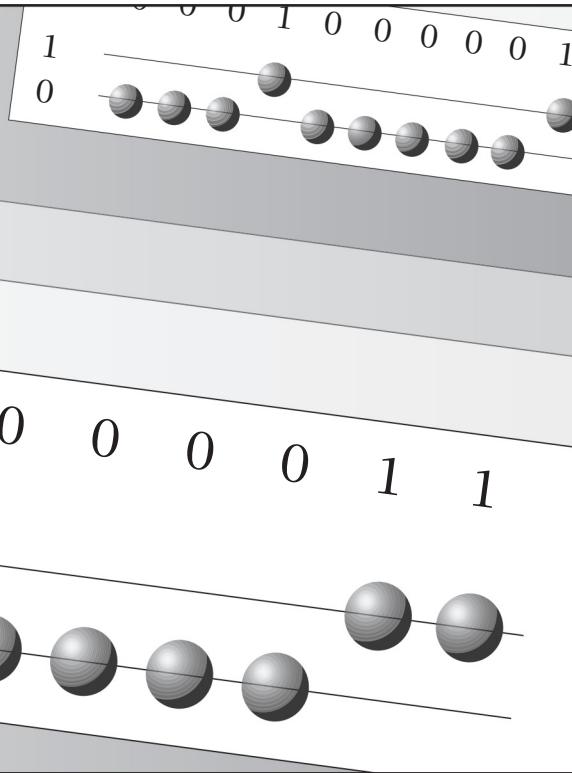
6. Stable states. For the energy function $V(\theta) = \cos \theta$ for $0 \leq \theta \leq 2\pi$, find the values $\theta = \theta_s$ that identify stable equilibria, and the values $\theta = \theta_u$ that identify unstable equilibria.

7. One-dimensional lattice. You have a one-dimensional lattice that contains N_A particles of type A and N_B particles of type B . They completely fill the lattice, so the number of sites is $N_A + N_B$. Write an expression for the multiplicity $W(N_A, N_B)$, the number of distinguishable arrangements of the particles on the lattice.

Suggested Reading

HB Callen, *Thermodynamics and an Introduction to Thermostatistics*, 2nd edition, Wiley, New York, 1985.
In-depth discussion of stability and equilibrium.

3 Heat, Work, & Energy



Energy Is a Capacity to Store and Move Work and Heat

In this chapter, we introduce energy, work, and heat. Energy and related quantities describe the capacity of a system to perform work. Energy is conserved, which means that it flows, so this capacity to perform work can be moved from one place to another. We introduce two fundamentally important ideas: the First and Second Laws of Thermodynamics. The First Law describes the conservation of energy and the interchangeability of heat and different types of work. The Second Law describes the tendencies of systems toward equilibrium.

Energy commonly has units of joules (J), calories (cal), watt-hours (Wh), or BTUs (British Thermal Units). The rate of energy transfer or conversion *per unit time* from one form to another is called *power*, and is typically measured in watts (W) or J s^{-1} .

Energy Is Ubiquitous and Important

Energy is important in commerce, technology, the environment, climate, and life. For example, a country's standard of living, reflected approximately by its gross domestic product (GDP) per person, is correlated with its energy utilization per person. A person in one of the poorest countries adds \$1000 to that country's annual GDP and uses around 200 W [1]. A person in one of the

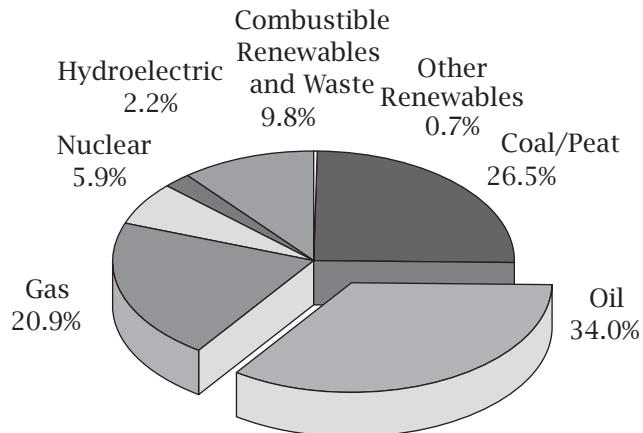


Figure 3.1 Distribution of world energy sources, 2007.
Source: *Key World Energy Statistics 2009*, ©OECD/IEA, 2009, p. 6.

richest countries adds \$35,000 dollars to its GDP and uses about 11 kW. The GDP, which is a measure of a nation's income and output, depends on building, manufacturing, transportation, and services that rely on motors, pumps, engines, generators, batteries, heating, cooling, and energy interconversion of all types. In 2005, the worldwide human 'consumption' of energy was 487 exajoules ($1\text{ EJ} = 10^{18}\text{ J}$) or 139 petawatt-hours ($1\text{ PWh} = 10^{15}\text{ Wh}$). Because energy is conserved, the term *consumption* or *energy usage* here just means that energy is converted from one form that can be readily captured and stored to another form that cannot. For example, car engines convert the chemical covalent bond energy in hydrocarbons to the work of moving the car and to heating its surroundings.

About 80% of the planet's energy currently comes from the covalent bonds of hydrocarbons—oil, gas, and coal; see Figure 3.1. These are called *nonrenewable* sources because they are not readily replaceable. Because the world's oil supplies are finite (estimated in 2008 to be about 1 yottajoule ($1\text{ YJ} = 10^{24}\text{ J}$), and being depleted at a substantial rate, there are efforts to find renewable sources of energy and to use nonrenewable sources wisely and efficiently. Systematic advances in thermodynamics have led to increases in efficiencies, from 5% in the steam engines of T Newcomen and J Watt 200 years ago to more than 50% in today's fuel cells.

EXAMPLE 3.1 You are energetically equivalent to a light bulb. A typical adult eats about 2000 kcal per day (1 'food Calorie' equals 1 kilocalorie). Converting units, you see that this would power a 100 W light bulb:

$$\frac{2000\text{ kcal}}{\text{day}} = \left(\frac{2 \times 10^6\text{ cal}}{\text{day}}\right) \left(\frac{4.184\text{ J}}{\text{cal}}\right) \left(\frac{1\text{ W}}{\text{J s}^{-1}}\right) \left(\frac{1\text{ day}}{3600 \cdot 24\text{ s}}\right) = 97\text{ W} \quad (3.1)$$

This 100 W light bulb energy expenditure per person is a useful reference quantity. People in poor nations contribute about twice their body's energy consumption to their country's GDP, while people in richer nations contribute about 100 times their consumption.

Energy is also important for understanding the Earth's environment and climate. The Earth has been approximately in thermal balance for most of its

history, neither cooling nor warming very much, on average. (Ice ages and warming periods are small fluctuations relative to the total energy flows.) So, the energy flow to the Earth equals the energy flow away from the Earth. A major source of energy influx to the Earth is *electromagnetic radiation*. The Earth receives almost all of its energy from sunlight, about 3.83 YJ per year, or 174 petawatts (PW).

EXAMPLE 3.2 How much solar energy hits your backyard? How much solar energy arrives on the Earth per unit time per square meter? Divide the total energy rate, $1.74 \times 10^{17} \text{ W}$, arriving on the Earth by the Earth's surface area, $4\pi R^2 = 5.1 \times 10^{14} \text{ m}^2$ (the Earth's radius is $R = 6.37 \times 10^6 \text{ m}$), to get 341 W m^{-2} , the solar energy per unit time arriving on an average patch of the Earth. A typical solar panel in 2009 is about 20% efficient, which means that 1 W of sunlight coming into the cell converts to 0.2 W of electricity out. So a solar panel of 1 m^2 in size could run a 70 W light bulb ($0.2 \times 341 \text{ W m}^{-2} \times 1 \text{ m}^2 \approx 70 \text{ W}$).

EXAMPLE 3.3 How much of the Earth's solar energy budget do humans consume? Divide the 487 EJ annual consumption rate above by the total energy from the Sun, 3.83 YJ per year, to get

$$\frac{487 \times 10^{18} \text{ J}}{3.83 \times 10^{24} \text{ J}} = 0.012\% \quad (3.2)$$

We consume only a small fraction of the Earth's total solar energy.

Figure 3.2 shows a peculiar feature of the Earth's electromagnetic radiation balance. Solar energy arrives on the Earth at high frequencies (i.e., short wavelengths, in the visible region of the spectrum, around $0.4\text{--}0.7 \mu\text{m}$). That radiation is then converted to lower frequencies (i.e., longer wavelengths, in the infrared part of the spectrum, around $5\text{--}15 \mu\text{m}$) by physical and biological processes on the Earth, and is then re-radiated away from the Earth as heat.

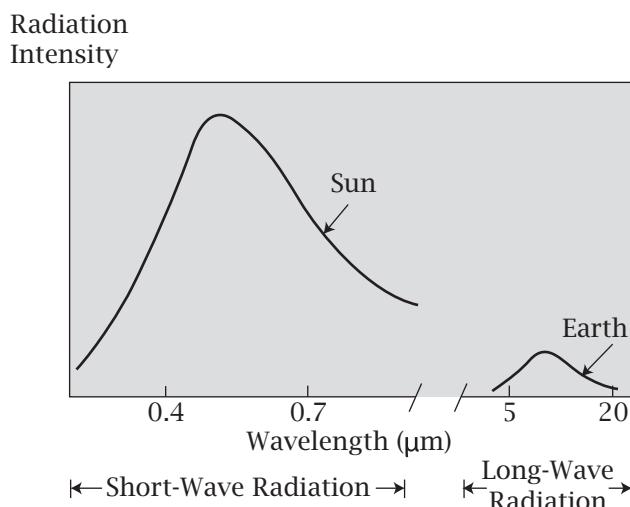


Figure 3.2 The Sun's energy is maximal in the visible part of the spectrum. This is what arrives on the Earth. The Earth radiates away electromagnetic energy too, mostly in the infrared part of the spectrum, called heat radiation.

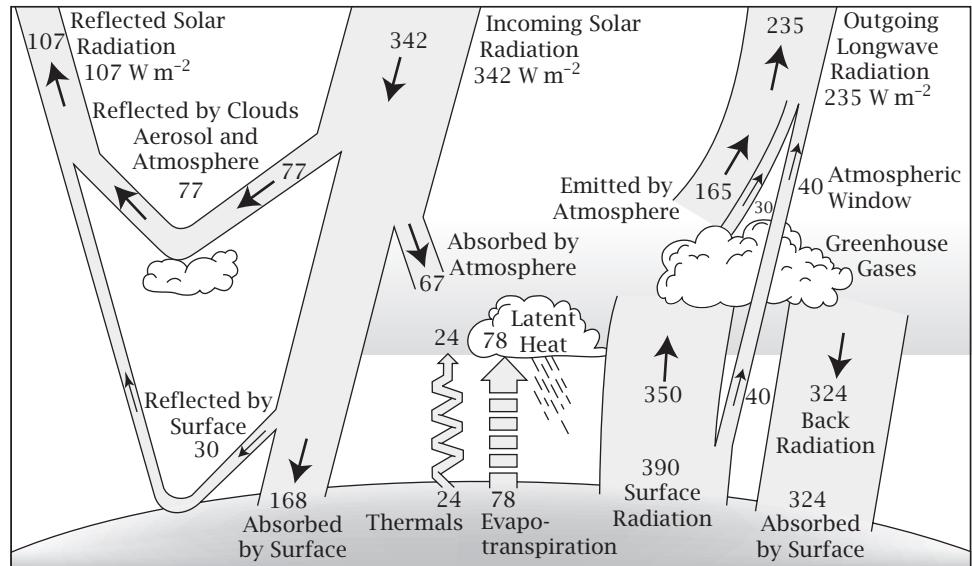


Figure 3.3 Earth's energy flows. (1) 342 W m^{-2} enters the Earth from sunlight (short wavelength). (2) 107 W m^{-2} of that energy is reflected away. (3) The Earth also radiates away 390 W m^{-2} of heat (longer wavelengths; see Problem 7). (4) 235 W m^{-2} of the radiation exits the Earth's atmosphere back into space, and (5) 324 W m^{-2} of it is absorbed by greenhouse gases and returned to the Earth. Source: redrawn from JT Kiehl and KE Trenberth, *Bull Am Meteor Soc* **78**, 197–208 (1997).

This conversion to longer wavelengths comes mainly from geophysical water cycles of evaporation and, to a lesser extent, from driving winds, waves, ocean currents, and photosynthesis in plants.

Figure 3.3 shows the Earth's energy balance. The main arrows on this diagram are constructed from a few facts and principles: (1) The incoming solar energy, as measured by satellites, is about 342 W m^{-2} . (2) The energy transfer rate *in*, 342 W m^{-2} , must equal the energy transfer rate *out*, $107 + 235 \text{ W m}^{-2}$, where 107 is an estimate of sunlight reflected from the Earth and 235 is an estimate of the net heat radiating away from the planet. (3) The Earth's core radiates away additional heat, about $P = 390 \text{ W m}^{-2}$ (based on the Earth's average temperature; see Problem 7 at the end of this chapter). Combining all these energy flows gives you one remaining large arrow on the diagram, indicating that a large amount (324 W m^{-2}) of the outgoing heat is reflected back to the Earth by *greenhouse gases*. What are greenhouse gases?

EXAMPLE 3.4 What is the greenhouse effect? The Earth's atmosphere acts like the glass roof on a *greenhouse*; see Figure 3.4. A greenhouse roof allows sunlight freely into the greenhouse, but it does not let the heat (infrared) freely out of the greenhouse. So, greenhouses are warm inside. Similarly the *greenhouse effect* warms up the earth. The Earth's 'glass roof' is its atmosphere of *greenhouse gases*—mostly water vapor (which accounts for about 40–70% of

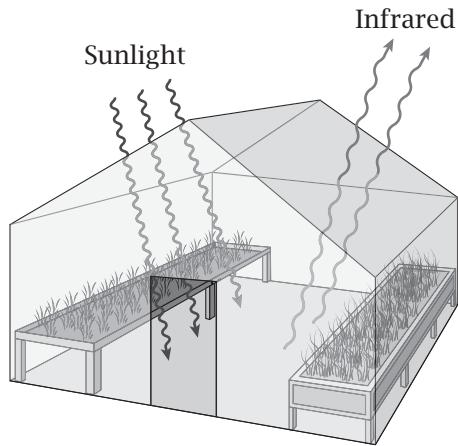


Figure 3.4 A greenhouse has a glass roof. Through glass, visible light passes freely, but infrared light (heat) inside is partially reflected. Sunlight enters a greenhouse and is converted to infrared radiation (heat) by evaporating water and photosynthesis by the plants. Because the infrared radiation cannot fully escape, it warms the inside of the greenhouse.

the greenhouse effect), carbon dioxide (10–25%), and smaller contributors such as methane, nitrous oxide, and ozone. The Sun's visible light reaches the Earth after passing freely through these atmospheric gases. But the heat that the Earth radiates outward is absorbed by vibrational motions of the greenhouse gases, and sent back down to Earth, as a greenhouse roof does, like a blanket warming the Earth. This greenhouse effect is responsible for the warming from $T = 255\text{ K}$ (the temperature that the Earth would have in the absence of greenhouse gases, due to the balance of energy fluxes) to $T = 288\text{ K}$. This greenhouse effect is natural and has been part of our planet's energy balance ever since the Earth first had an atmosphere. However, more recently, man-made chemicals, such as chlorofluorocarbons, and carbon compounds from fossil fuels, are adding to the greenhouse gases, trapping additional heat. It is believed that this may contribute to *global warming*.

Energy also plays a key role in biology and evolution. Your intake of food energy must equal the work you perform plus the heat you generate plus the energy you store as fat. At the cellular level, energy utilization drives molecular motors, transporters, and biochemical reactions. On a grander scale, energy balance governs the sizes and lifespans of animals. You might first guess that the total amount of energy consumed by an organism at rest, per unit time, called the *basal metabolic rate* (BMR), would scale in linear proportion to an organism's body mass M ($\text{BMR} \propto M$). You might reason that each cell has about equal size and burns energy at about equal rates. But this argument would overestimate the energy usage in large organisms. (The BMR for humans is about 1200 kcal per day; the BMR is the minimal level measured after fasting and does not account for exertion or keeping warm in cold weather.) As a better approximation, you might reason that an organism's energy utilization is governed by its surface area rather than its volume or mass, ($\text{BMR} \propto M^{2/3}$), because the surface area controls how much heat the organism can dissipate. Similar considerations are involved in designing computers.

But it is not clear that this argument is correct either. Figure 3.5 shows a different result, namely that $\text{BMR} \propto M^{3/4}$, over a wide range of sizes from single-celled mycoplasma to the largest mammals and whales [2]. Interestingly, many

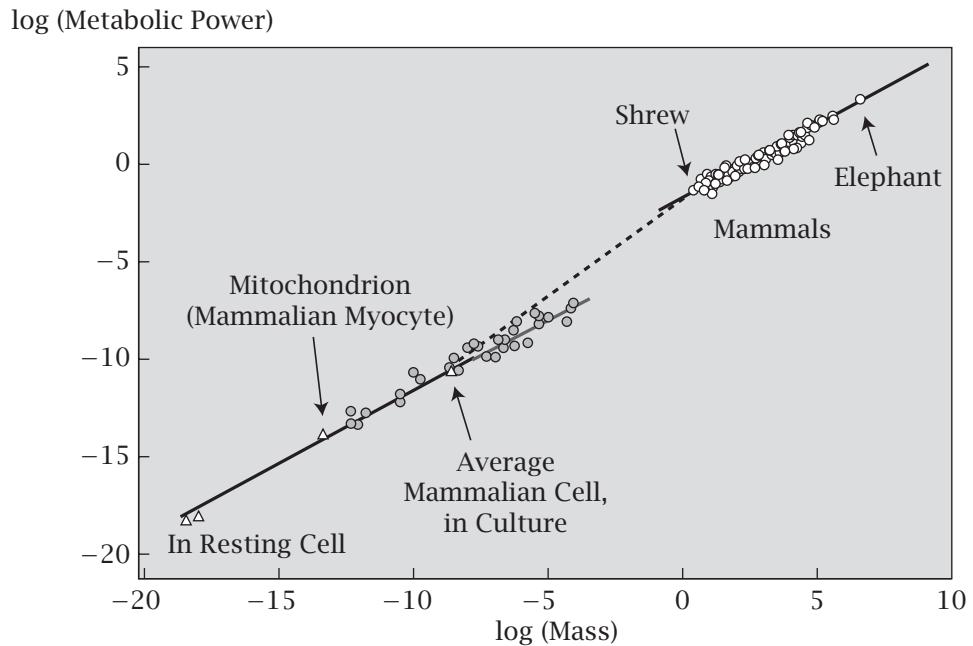


Figure 3.5 Basal metabolic rate (BMR) versus sizes of animals and cells. Over 27 orders of magnitude, this measure of energy utilization in organisms increases with $M^{3/4}$, where M is the body mass.

such *allometric scaling laws* are observed in biology. The powers in the scaling laws are multiples of approximately 1/4: heart rates in mammals $\propto M^{-1/4}$, growth rates $\propto M^{-1/4}$, DNA nucleotide substitution rates $\propto M^{-1/4}$, densities of mitochondria, chloroplasts, and ribosomes $\propto M^{-1/4}$, and lifespans $\propto M^{1/4}$. Are all these processes related to each other? Such 1/4-power scaling laws have been explained by the rates at which supply networks (such as networks of blood vessels) transport energy to cells. For optimal efficiency, blood vessels, like the streets and highways that supply cities, should have a hierarchy of sizes: large roads are needed for rapid global transport and small roads are needed for local transport.

EXAMPLE 3.5 Elephants metabolize more efficiently than mice. Dividing the BMR (energy per unit time used by the whole organism), which scales as $M^{3/4}$, by the body mass M shows that the energy used by an organism per unit mass is

$$\text{BMR}/M \propto M^{-1/4}, \quad (3.3)$$

so larger animals expend less energy per unit mass than smaller animals do.

The question of whether metabolic scaling exponents are 2/3 or 3/4 or some other value remains a matter of active research [3].

In short, the world's commerce, climate and environment, and living systems are dependent on the thermophysical principles of energy balance. Below, we develop the quantitative basis for these principles. We start with a little historical background.

Before 1800, a major mystery of physical science was why heat flows from hot objects to cold objects. Is heat like a fluid that equalizes its levels in two connected containers? This was the view until thermodynamics emerged. Heat was mistakenly thought to be *conserved*, like the volume of a flowing fluid. The correction to this misunderstanding was captured in the *First Law of Thermodynamics*.

Some Physical Quantities Obey Conservation Laws

The origin of ideas about molecules and their energies is deeply rooted in laws of conservation, which are central to physical science. The law of conservation of momentum was discovered in 1668 by J Wallis. Wallis observed that when objects collide, their velocities, positions, and accelerations can change, but the sum of the *mass × velocity* of all the objects does not. The quantity *mass × velocity*, called the momentum, is special because it is *conserved*.

Other functions of the position, time, velocity, and acceleration depend on the details of the collision process. For example, the quantity *velocity* or the quantity *mass × acceleration squared* will generally change when one particle collides with another. Such quantities are different for every situation and depend on the angles of the collisions and the shapes of the objects. However, conservation laws describe properties that are exceptionally simple because they are not dependent on particular details. The total momentum (*mass × velocity*) is the same before and after a collision, no matter how the collision occurs. Similar laws describe the conservation of mass, of angular momentum, and of energy. A property that is *conserved* is neither created nor destroyed as collisions take place. Because they are conserved, mass, momentum, and energy can only ‘flow’ from one place to another.

Conservation laws help to predict the behaviors of physical and chemical systems. For example, because the total momentum does not change in a collision, you can predict the final velocities of colliding objects on the basis of their masses and initial velocities. Another conserved quantity is the *energy*. Like momentum, the total energy cannot be created or destroyed. Properties that are conserved, such as matter, momentum, and energy, can flow from one place to another, since they cannot be destroyed. The importance of energy derives from two properties. Energy describes a capacity to perform work. And, because it is conserved, that capacity to perform work can move from one place to another.

TOTAL ENERGY. The total energy E of a mechanical system is defined by the **law of conservation of energy**:

$$K + V = \text{constant.} \quad (3.4)$$

where the kinetic energy K is the work that an object can perform by virtue of its *motion* and V is the work that an object can perform by virtue of its spatial position.

KINETIC ENERGY. The kinetic energy K of an object of mass m moving at velocity v is defined as

$$K = \frac{1}{2}mv^2. \quad (3.5)$$

A moving object slows down as it does work, losing kinetic energy.

POTENTIAL ENERGY. In contrast, the potential energy V is the work that an object can perform by virtue of its *position*, or the positions of its subcomponents. If an object is raised above the Earth, it gains the gravitational potential to do work when it comes back down. The potential to do work can be stored in a stretched or compressed spring, or in a battery through the separation of electrical charges. Like the kinetic energy, the potential energy of a system can change. The potential energy of a ball changes as it rolls downhill, as in Example 2.1. The potential energy of a mass attached to a spring changes as the mass bounces.

The conservation of energy is a fundamental observation about nature. Although the kinetic and potential energies each change throughout a process, their sum, the total energy E , is constant. The total energy is an **invariant of the motion**. That is, the sum of all the kinetic energies plus all the potential energies of a system is the same at any stage before, after, and during a process. Any change in the kinetic energy of an object as it does work, or as work is done upon it, is balanced by an equal and opposite change in the potential energy of the system. A system performs work by virtue of its ability to exert a force.

Forces and Work

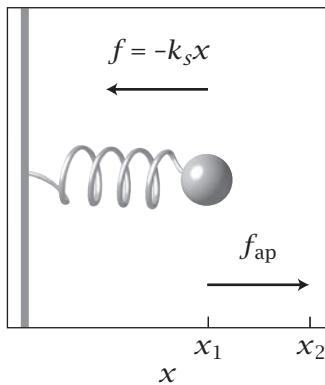


Figure 3.6 A spring pulls a ball to the left, with force $f = -k_s x$. To stretch the spring requires an applied force $f_{ap} = -f$ acting toward the right ($+x$ direction).

FORCE. Newton's second law defines the force in terms of the acceleration of an object:

$$f = ma = m \frac{d^2x}{dt^2}, \quad (3.6)$$

where the mass m of an object represents its resistance to acceleration, x is its position, the acceleration $a = d^2x/dt^2$, and the velocity $v = dx/dt$.

WORK. Consider a ball located at $x = 0$ that can move along the x axis (see Figure 3.6). One force acting on the ball is from a spring attached to a wall pulling the ball to the left. Call this the intrinsic force f . For a spring, you have

$$f = -k_s x, \quad (3.7)$$

called *Hooke's law*, named after 17th century British physicist R Hooke, where k_s is the spring constant. The minus sign shows that the spring acts in the $-x$ direction. Suppose you apply an approximately equal opposing force to stretch the spring $f_{ap} = -f$. (It must be slightly greater, and not exactly equal, to stretch the spring. However, the near equality ensures that there is approximately no net force or acceleration.) The work δw you perform *on* the system by moving the ball a distance dx to the right is

$$\delta w = f_{ap} dx = -f dx. \quad (3.8)$$

The total work w performed on the system in stretching the spring from x_1 to x_2 is the integral

$$w = \int_{x_1}^{x_2} f_{ap} dx = - \int_{x_1}^{x_2} f dx. \quad (3.9)$$

Examples 3.6 and 3.7 calculate the work of stretching a spring and the work of lifting a weight.

EXAMPLE 3.6 The work of stretching a spring. The work w to stretch a spring from 0 to x_0 is

$$w = \int_0^{x_0} k_s x \, dx = \frac{1}{2} k_s x_0^2.$$

EXAMPLE 3.7 The work of lifting a weight. Gravity acts downward with a force $f = mg$, where m is the mass of the weight and g is the gravitational acceleration constant (see Appendix H). Suppose that you lift a weight from the floor, $z = 0$, to a height $z = h$ (see Figure 3.7). The downward force defines the direction $x > 0$, but positive z is in the upward direction, so $dz = -dx$. Substituting $f = mg$ and $dx = -dz$ into Equation (3.9) gives the work of lifting the weight

$$w = - \int_0^h (mg)(-dz) = mgh.$$

CONSERVATIVE FORCES. The forces operating on the ball on the hill in Example 2.1 are *conservative forces*, because there is no frictional loss, no turbulence, and no other dissipation of energy. For a conservative force f , no net work is performed in moving an object through any cycle that returns it to its starting point. For a force acting in the x direction, the work performed in the cycle from A to B to A is zero,

$$\begin{aligned} w(A, B) + w(B, A) &= - \int_A^B f \, dx - \int_B^A f \, dx \\ &= - \int_A^B f \, dx + \int_A^B f \, dx = 0. \end{aligned} \quad (3.10)$$

In contrast, for nonconservative forces, the work around a cycle does not sum to zero, for example when the friction is different in one direction than the other. So far, we have described how work arises from the actions of forces. Work can be done by different types of forces: stretching springs, lifting weights, or moving electrical charges or magnets, for example. What is the role of heat? Let's look at the history of the thought that led to the First Law of Thermodynamics.

Heat Was Thought to Be a Fluid

The ability to *flow* is an aspect of conserved properties: mass, momentum, and energy can all flow. What about heat? When a hot object contacts a cold one, heat flows. Does it follow that heat is a conserved quantity? It does not. Heat is not a conserved quantity. All conserved properties can flow, but not all properties that flow are conserved.

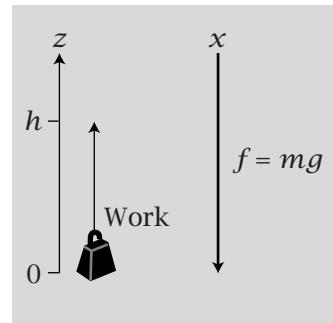


Figure 3.7 The intrinsic force f acts downward on a weight. The work to raise the weight to a height $z = h$ is positive.

Until the mid-1800s, heat was (incorrectly) thought to be a conserved form of matter, a fluid called *calorique*. The calorique model explained that materials expand upon heating because the added heat (calorique fluid) occupied space. The flow of heat from hot to cold objects was attributed to repulsions between the more crowded calorique particles in the hot material. Calorique particles escaped to cold materials, where they were less crowded. Because all known materials could absorb heat, it was thought that calorique particles were attracted to all materials, filling them in proportion to the volume of the material. The temperature of a substance was known to increase on heating, so it was thought that temperature directly measured the amount of calorique in a material.

The misconception that calorique obeyed a law of conservation had important implications for industry and commerce. The industrial revolution began with J Watt's steam engine around 1780. During the late 1700s and early 1800s, power generated from steam gained major economic importance. The industrial quest for efficient steam engines drove the development of thermodynamics. How is heat converted to work in steam engines? It was thought that the heat flow in a steam engine was like the water flow in a water wheel. In a water wheel, water goes in at the top, falls down to turn the wheel, and comes out the bottom (see Figure 3.8). The water is conserved: the flow of water from the bottom equals the water flow into the top. The amount of heat flowing out of the exhaust of a steam piston was believed to equal the amount of heat in the steam that enters the piston chamber.

The view of heat as a fluid began to change in the mid-1700s. The first step was the concept of *heat capacity*, developed by J Black around 1760: he heated mercury and water over the same flame and discovered that the temperature of the mercury was higher than the temperature of the water. He concluded that heat could not be a simple fluid, because the amount taken up depended on the material that contained it. Different materials have different capacities to take up heat. The heat capacity of a material was defined as the amount of heat required to raise its temperature by 1°C.

The next step was Black's discovery of *latent heat*, which showed that temperature and heat are different properties. Liquid water slightly below 100°C can absorb heat and boil to become steam slightly above 100°C. In this case, the uptake of heat causes a change of phase, not a change of temperature. This heat is latent because it is 'stored' in the vapor and can be recovered in a condenser that converts the steam back to liquid water. The melting of ice also involves a latent heat. Although the discovery of latent heat showed that heat and temperature are different properties, heat was still regarded as a conserved fluid until the 1800s. And the question remained: What is temperature? We develop a conceptual model of temperature in Chapter 12, after laying the groundwork in thermodynamics and statistical mechanics.

The development of quantitative thermodynamics ultimately caused the downfall of the calorique theory of heat. The early 1800s were years of intensive development of engines driven by steam (see Figure 3.9), by hot air, and later by internal combustion. In 1853, J Ericsson launched a magnificent engine, designed using the calorique theory, with pistons 14 feet in diameter driven by hot air. Ericsson hoped the calorique engine would achieve 600 horsepower and compete with the steam engine. A set of 'regenerator' coils were supposed to



Figure 3.8 Heat doesn't flow in and out of a steam engine like water flows in and out of a water wheel.

collect exhaust heat and re-use it, on the basis of the assumption that heat was a conserved and recyclable fluid. But the calorique engine and the calorique theory ultimately failed because heat is not a conserved fluid. One of the main uses of the Second Law of Thermodynamics in its early days was to debunk such ‘perpetual motion’ machines. The Second Law, now the centerpiece of statistical thermodynamics, is little needed for that purpose any longer.

Other problems with the theory of calorique emerged at that time. First, it was shown that *radiant heat* could be transmitted through a vacuum by electromagnetic radiation. If heat is a form of matter, how could it pass through a vacuum that is devoid of matter?

Second, careful measurements showed that work could be converted to heat quantitatively. In 1798, Count Rumford showed that the mechanical work involved in boring cannons was converted to heat. In 1799, H Davy showed that the mechanical work involved in rubbing ice cubes together was sufficient to melt them. In the 1850s, JP Joule produced heat from work in many ways: a falling weight rotating a paddle wheel and imparting heat to a liquid, electrical work heating a resistor, and others. These experiments were difficult to reconcile with the view that heat is a conserved fluid. How could heat be created by work?

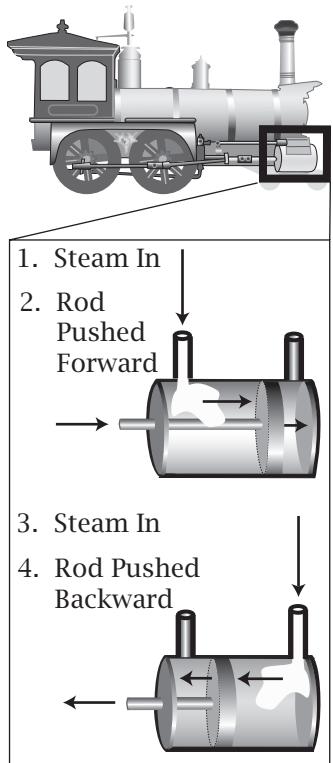


Figure 3.9 In a heat engine, hot gas enters a chamber, exerts pressure on a piston, and performs work. The gas cools on expansion and exits the chamber.

The First Law of Thermodynamics: Heat Plus Work Is Conserved

These experiments led to two paradigm-changing conclusions: (1) heat is *not* conserved, and (2) heat and work can be interconverted. This new understanding, first formulated by JR von Mayer in 1842, came to be known as the First Law of Thermodynamics. It is not the heat q that is conserved ($q_{in} \neq q_{out}$). Rather, it is the *sum* $q+w$ of the heat q *plus* the work w that has special properties. This sum is called the *internal energy change* $\Delta U = q+w$. ΔU of a system will rise or fall as a system does work or takes in or gives off heat, but any gain in ΔU of the system must balance a loss in energy somewhere else; ΔU is conserved between the system and its surroundings. Various forms of work were known before the 1800s, including mechanical, electrical, and magnetic. The advance embodied in the First Law was the recognition that heat can be added to this list, and that heat is a form of energy transfer, not a form of matter.

The First Law of Thermodynamics The internal energy ΔU changes when a system takes up or gives off heat q or work w :

$$\Delta U = q+w$$

This is a statement of the equivalence of heat and work. The internal energy is conserved; if ΔU increases in the system, the energy decreases in the surroundings.

Energy is a property of a *system*, while heat and work are properties of a process of energy transfer across a boundary. As a metaphor, the level of water in a lake is like the amount of internal energy in a system, and the different modes of water exchange between the lake and its surroundings are like heat and work [4]. Say that rainfall corresponds to the process of heat going *into* the system ($q > 0$), evaporation corresponds to the process of heat going *out of* the system ($q < 0$), streams flowing into the lake correspond to work done *on* the system ($w > 0$), and streams flowing out of the lake correspond to work done *by* the system ($w < 0$). The change in internal energy $\Delta U = q_{\text{in}} - q_{\text{out}} + w_{\text{in}} - w_{\text{out}}$ is the sum of the heat and work *into* the system minus the heat and work *out of* the system, just as the change in the water level of the lake is the sum of rainfall plus river flows in minus evaporation and river flows out. Once the water is in the lake, you cannot tell whether it came from rainfall or from streams. Similarly, you cannot tell whether the amount of internal energy in a system was acquired as heat or work. In this analogy, the internal energy corresponds to a property of the lake, while heat and work correspond to processes of the transfer across the boundaries of the lake.

Atoms and Molecules Have Energies

The Kinetic Theory of Gases

The modern conception of particulate atoms and molecules developed in parallel with the modern view of heat. What is the microscopic nature of heat? This question was addressed in the kinetic theory of gases, a radically different view of heat that superseded the model of calorique, owing to the efforts of Clausius, Maxwell, Boltzmann, and others during the late 1800s. The concept of calorique had been intimately tied to the concept of the *ether*, a hypothetical pervasive medium that could transmit vibrations. Atoms were thought to be at fixed positions in the ether. The kinetic theory introduced three novel concepts:

- (1) Matter is composed of molecules that are not located at fixed positions in space, but are free to move through a space that is otherwise empty.
- (2) Heat is the exchange of energy that takes place owing to the *motions and collisions of molecules*. In the kinetic theory, molecules collide and exchange energy like Newtonian billiard balls.
- (3) Electromagnetic radiation can influence the motions of molecules. This is the basis for radiant heat.

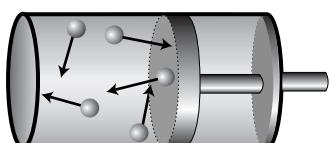


Figure 3.10 The kinetic view of heat: matter is composed of molecules that exchange kinetic energies in collisions.

How does the kinetic theory of gases explain the conversion of heat to work? How does a gas lose heat and push the piston in a steam engine? The kinetic theory of gases is a mechanical model on a microscopic scale (see Figure 3.10). According to the kinetic theory, when molecules collide, they exchange energy and momentum. At high temperature, gas particles move with high velocities. When a molecule collides with a piston, it imparts momentum, loses kinetic energy, moves the piston, and produces work. Alternatively, a high-energy particle colliding with its container can impart energy to the container wall, and the container wall can impart this kinetic energy to the surroundings as heat.

Because the collisions between the gas and the piston are random, some collisions will perform work, owing to motion in the direction of the piston's trajectory. Other collisions will produce only heat, molecular motions in all other directions that result in a flow of energy out of the container.

The kinetic theory is remarkably successful. It predicts the ideal gas law and other properties of gases, including diffusion rates, viscosities, thermal conductivities, and the velocities of sound in gases. It provides a model in which every molecule has its own energy, different molecules have different energies, and molecules can exchange their energies. **The kinetic theory predicts that the temperature T of a gas is proportional to the average kinetic energy of the gas particles.**

$$\frac{3}{2}kT = \frac{m\langle v^2 \rangle}{2}, \quad (3.11)$$

where m is the mass of one particle, $\langle v^2 \rangle$ is the square of the particle velocity averaged over all the particles, $m\langle v^2 \rangle/2$ is the average kinetic energy of the gas molecules, and k is Boltzmann's constant (see Appendix H).

A Better Model: Energy Is Quantized

Despite the tremendous advances that resulted from the kinetic theory of gases, it wasn't perfect either. The kinetic theory described gas particles as mechanical objects like billiard balls, with a continuum of possible energies. **But 20th century quantum theory**, describing the motions of particles at the atomic level, showed that the **energies of atoms and molecules are quantized**. Each particle has discrete amounts of energy associated with each of its allowed degrees of freedom, some of which are translations, rotations, vibrations, and electronic excitations. For example, quantum mechanics might dictate that the **molecules** in a given system have energies of 1, 5, 16, 21, and 26 units, and **no other energies are allowed**. The allowed energies for a given system are indicated in *energy-level diagrams*, as shown in Figure 3.11. Although these diagrams seem to contain very little information, this information is **sufficient to predict the thermodynamic properties**.

For simple systems of independent particles such as ideal gases, we can express the total internal energy of a thermodynamic system as the sum of the **particle energies**:

$$U = \sum_i N_i \varepsilon_i, \quad (3.12)$$

where ε_i is the energy of any particle at level i and N_i is the number of particles at energy level i . When the total internal energy of a system is increased by heating it, the energy levels do not change; the populations N_i change. Then **more particles occupy higher energy levels**.

What drives molecules or materials to exchange energy with each other? Why does heat flow? The answer is not found in an **equalization principle** or a **conservation law** like the First Law. Heat flow is a consequence of the **tendency toward maximum multiplicity**, which is the **Second Law of Thermodynamics**.

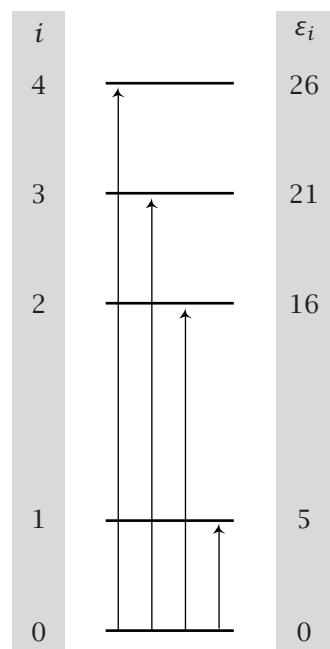


Figure 3.11 The quantum mechanical view: the quantized energies of particles are given by energy-level diagrams.

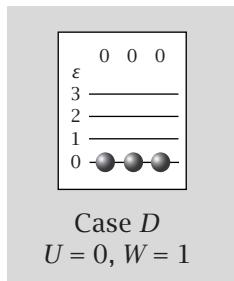
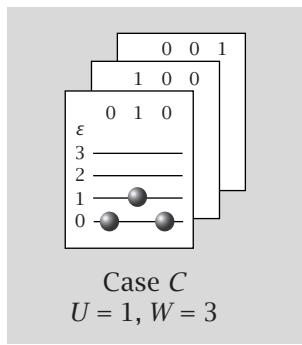
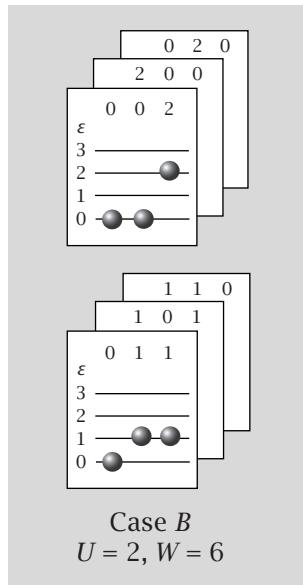
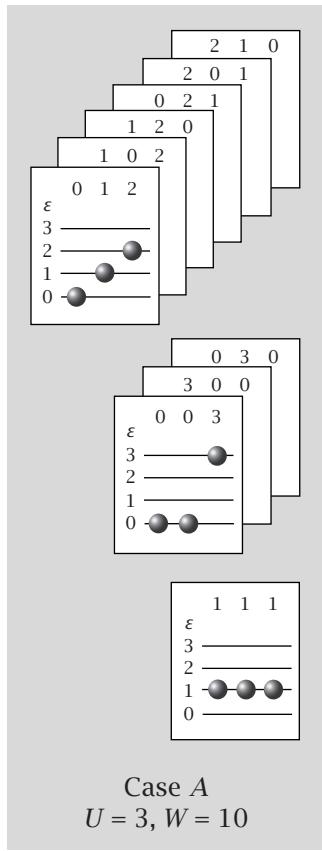


Figure 3.12 Each case represents a system with a different given energy U . Each card represents a different distribution of the three particles over the four energy levels. The numbers shown at the top of each card are the individual energies of the particles. W is the number of configurations (cards).

Why Does Heat Flow?

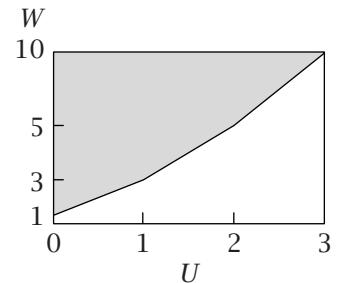
In Example 2.2, we found that gases expand because the multiplicity $W(V)$ increases with volume V . The dependence of W on V defines the force called **pressure**. In Example 2.3, we found that particles mix because the multiplicity $W(N)$ increases as the **particle segregation decreases**. This tendency defines the chemical potential. These are manifestations of the principle that systems tend toward their states of **maximum multiplicity**, also known as the **Second Law of Thermodynamics**.

The following examples illustrate in two steps that the flow of heat from hot to cold objects is also driven by a tendency to maximize multiplicity. First, Example 3.8 shows a model for how the multiplicity $W(U)$ of an object depends on its energy. Then Example 3.9 uses that result to show that heat flow maximizes multiplicity.

EXAMPLE 3.8 Why do materials absorb heat? Here is a miniaturized model of a material. The material has three distinguishable particles. Each particle can have energy $\epsilon = 0, 1, 2$, or 3 units. (For now, it doesn't matter that our energy rule is simple. In Chapter 11, we'll get energies from quantum mechanics.) There are different ways of getting different total energies U . What is the total number $W(U)$ of ways the system can partition its energy U among its component particles? Figure 3.12 shows four different states, $U = 0, 1, 2$, and 3 , and their different multiplicities.

Case D shows that there is only one way to achieve a total energy of $U = 0$: each particle must have zero energy. Case C shows the three ways of achieving

$U = 1$: two particles must have zero energy and one particle must have $\varepsilon = 1$. Case *B* shows the six different ways to get $U = 2$. And case *A* shows the 10 different ways of achieving a total energy $U = \sum_{i=0}^t N_i \varepsilon_i = 3$. Collecting these results together, Figure 3.13 shows that the number of configurations W is an increasing function of the total energy U , for the system shown in Figure 3.12. The higher the energy U the material has, the higher is its multiplicity of states.



So, because systems tend towards states having high multiplicities, materials tend to take up heat from their surroundings. Now why does heat flow from hot objects to cold ones? Example 3.9 addresses this question.

EXAMPLE 3.9 Why does energy exchange? Consider the two systems, *A* and *B*, shown in Figure 3.14. Each system has 10 particles. Each particle has two possible energies, $\varepsilon = 0$ or $\varepsilon = 1$. Suppose that system *A* starts with energy $U_A = 2$ and system *B* starts with energy $U_B = 4$. We show here that if the two systems were in thermal contact and could instead share their energies equally, $U_A = U_B = 3$, it would increase the multiplicity of the combined system, so heat will flow from *B* to *A* to reach this state. Because each system *A* and *B* has only two energy levels, the multiplicities $W(U)$ are given by binomial statistics. Count the number of ways that *A* can have two particles in its ‘excited state’ ($\varepsilon = 1$), and eight in its ‘ground state’ ($\varepsilon = 0$), for a total energy of $U_A = 2$. Then make the corresponding count for *B*.

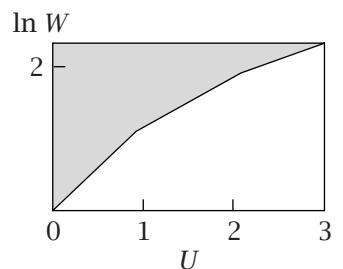


Figure 3.13 In Example 3.8, the multiplicity W increases with the total system energy U . Its logarithm $\ln W$ also increases with U .

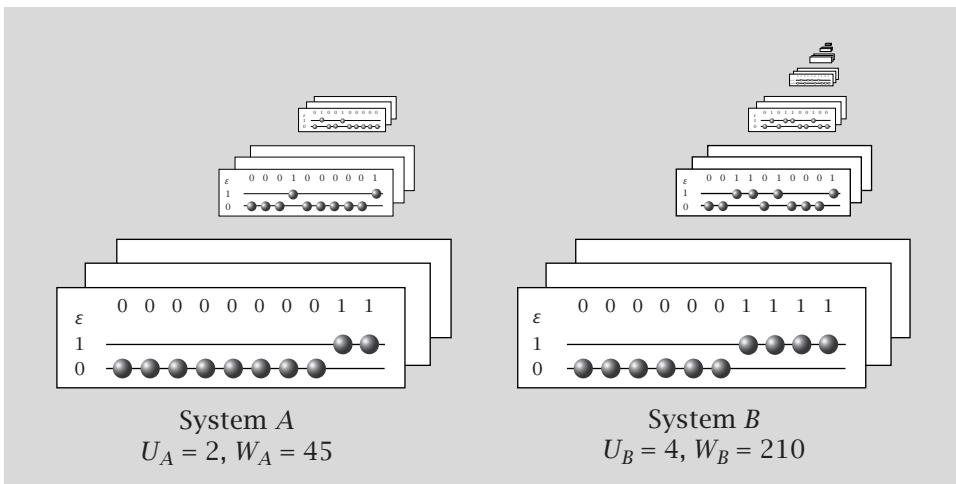


Figure 3.14 Energy-level diagrams for the two different systems in Example 3.9 with 10 particles each. System *A* has total energy $U_A = 2$, and *B* has $U_B = 4$. System *B* has the greater multiplicity of states.

The multiplicities of isolated systems A and B are

$$W_A = \frac{10!}{2!8!} = 45, \quad \text{and} \quad W_B = \frac{10!}{4!6!} = 210.$$

If A and B do not exchange energies, the total multiplicity is $W_{\text{total}} = W_A W_B = 9450$. Now suppose that you bring A and B into ‘thermal contact’ so that they can exchange energy. Now the system can change values of U_A and U_B subject to conservation of energy ($U_A + U_B$ will be unchanged). One possibility is $U_A = 3$ and $U_B = 3$. Then the total multiplicity W_{total} will be

$$W_{\text{total}} = \frac{10!}{3!7!} \frac{10!}{3!7!} = 14,400.$$

This shows that a principle of maximum multiplicity predicts that heat will flow from B to A to equalize energies in this case. Consider the alternative. Suppose A were to lower its energy to $U_A = 1$ while B wound up with $U_B = 5$. Then the multiplicity of states would be

$$W_{\text{total}} = \frac{10!}{1!9!} \frac{10!}{5!5!} = 2520.$$

A principle of maximal multiplicity predicts that this inequitable distribution is unlikely. That is, heat will not flow from the cold to the hot object.

EXAMPLE 3.10 However, energy doesn’t always flow downhill. Example 3.9 predicts that energies tend to equalize. But here’s a more interesting case, which shows that the tendency to maximize multiplicity does not always result in a draining of energy from higher to lower. Consider two systems having the same energies, but different particle numbers. Suppose system A has 10 particles and an energy $U_A = 2$ and system B is smaller, with only four particles, but also has energy $U_B = 2$. The multiplicity is

$$W = W_A W_B = \frac{10!}{2!8!} \frac{4!}{2!2!} = 45 \times 6 = 270.$$

Now suppose that A and B come into thermal contact and the larger system absorbs energy from the smaller one, so $U_A = 3$ and $U_B = 1$. This causes the multiplicity to increase:

$$W = W_A W_B = \frac{10!}{3!7!} \frac{4!}{1!3!} = 120 \times 4 = 480.$$

The tendency for heat to flow is not always a tendency to equalize energies. It is a tendency to maximize multiplicity. We will see later that the concept of *temperature* describes the driving force for energy exchange. The tendency toward maximum multiplicity is a tendency toward equal temperatures, not equal energies. Temperature is the quantity that describes the tendency toward maximum multiplicity when energies can exchange. This is the topic of Chapter 12.

The Second Law of Thermodynamics Is an Extremum Principle

In this and the preceding chapter we have used simple models to predict that gases expand to fill the volume available to them, molecules mix and diffuse to uniform concentrations, rubber retracts when pulled, and heat flows from hot bodies to colder ones. All these tendencies are predicted by a principle of maximum multiplicity: a system will change its degrees of freedom to reach the microscopic arrangement with the maximum possible multiplicity. This principle is the Second Law of Thermodynamics, and is much broader than the simple models that we have used to illustrate it.

Why do systems tend toward their states of maximum multiplicity? It is because every microstate is just as likely as every other. So, in the same way that a series of coin flips is most likely to lead to half heads and half tails, the probability of observing any particular macrostate in a physical system depends on how many microstates it has. In the following chapters, we explore the Second Law in more detail.

Summary

We have described the concept of energy and traced a little of the history of thermodynamics. The First Law of Thermodynamics says that heat is a form of energy exchange, and that the sum of heat plus work is a quantity ΔU that is conserved between a system and its surroundings. The First Law is a bookkeeping tool. It catalogs the balance of heat and work. It doesn't tell us why heat flows. The Second Law of Thermodynamics says that systems tend toward their states of maximum multiplicity. Heat flows to maximize multiplicity. A simple example shows that transferring internal energy from one system to another can change the multiplicity.

To make these principles more useful, we need some mathematical tools and then, beginning in Chapter 6, the definitions of thermodynamics.

Problems

1. The time dependence of a mass on a spring

- (a) For the harmonic motion of a mass on a spring, the kinetic energy is $K = (1/2)mv^2$, and the potential energy is $V = (1/2)k_s x^2$, where k_s is the spring constant. Using the conservation of energy, find the time-dependent spring displacement $x(t)$.

(b) Compute the force $f(t)$ on the mass.

2. Equalizing energies.

For the two 10-particle two-state systems of Example 3.9, suppose the total energy to be shared between the two objects is $U = U_A + U_B = 4$. What is the distribution of energies that gives the highest multiplicity?

3. Energy conversion. When you drink a cup of juice, you get about 100 Cal (1 food Cal = 1 kcal). You can work this off on an exercise bicycle in about 10 minutes. If you hook your exercise bicycle to a generator, what wattage of light bulb could you light up, assuming 100% efficiency? (1 watt = 1 Js^{-1} is power, i.e., energy per unit time.)

4. Kinetic energy of a car. How much kinetic energy does a 1700 kg car have, if it travels at 100 km h^{-1} .

5. Root-mean-square (RMS) velocity of a gas. Using $(1/2)kT = (1/2)m\langle v_x^2 \rangle$, for $T = 300 \text{ K}$, compute the RMS velocity, $\langle v_x^2 \rangle^{1/2}$, of O_2 gas.

6. Earth's energy balance. Figure 3.3 in the text shows energy inputs and outputs to the Earth, in W m^{-2} , which is a measure of energy per unit area per unit time ($1 \text{ W} = 1 \text{ V} \cdot 1 \text{ A}$).

- (a) Suppose your backyard has an area of 1000 m^2 . Your yard only receives sunlight about one-third of each day, and, because of overcast skies, let's say the total flux is reduced by another factor of onehalf. What average energy flux, in W m^{-2} , does your yard receive?

(b) Your cell phone uses about 2 W . Assuming 20% efficiency, how big a solar cell would you need to power it?

(c) If we had no *natural* greenhouse gases (CO_2 and water vapor), how would that shift the Earth's energy balance?

(d) How much energy, in Wh, is in a gallon of gasoline? Assume that 100% of the gasoline is octane (C_8H_{18}). The density of gasoline is about 1 g cm^{-3} , and each carbon-carbon covalent bond is worth about 60 kcal mol^{-1} . Assume 50% efficiency in burning.

7. Taking the Earth's temperature. A physical law, called the Stefan-Boltzmann (SB) law, relates the transfer rate of electromagnetic radiation of a body (called *black-body radiation*) to its temperature. The SB law is a relative of the *Wein law*, the principle that explains why an object that is hot enough will emit visible light, as when a burning ember in a fire glows red, or glows blue if is even hotter.

An object having a nonzero temperature will emit electromagnetic radiation. Conversely, an object that absorbs energy via radiation increases its temperature. The SB law says that the electromagnetic power P is related to the temperature T by

$$P = (5.67 \times 10^{-8} \text{ W m}^{-2} \text{ K}^{-4}) T^4 \quad (3.13)$$

- (a) If the incoming solar power hitting the Earth is $P = 342 \text{ W m}^{-2}$, what would you predict should be the Earth's temperature?
(b) Now, turn around this argument around. The Earth's actual average temperature is $T = 288 \text{ K}$ or $+15^\circ\text{C}$. Use the SB law to compute the power of radiated heat from the Earth's core.

8. Why do elephants live longer than mice? Let L represent the lifespan of an organism. It is known that all mammalian species live, on average, to have about 1.5×10^9 heartbeats (HB), independent of their size. Assume that an organism will have a total of $\text{HB} = L \times \text{HR}$ = constant heartbeats during its lifetime, where HR is its heart rate. What assumptions would you need to make to explain the observation that the lifespans of organisms grow with body mass M as $L \propto (\text{HB}/\text{HR}) \propto M^{1/4}$?

References

- [1] *Key World Energy Statistics*, International Energy Agency, Paris, 2006.
- [2] GB West and JH Brown, *J Exp Biol* **208**, 1575–1592 (2005).
- [3] CR White, TM Blackburn, and RS Seymour, *Evolution* **63**, 2658–2667 (2009).
- [4] HB Callen, *Thermodynamics and an Introduction to Thermostatistics*, 2nd edition. Wiley, New York, 1985.

Suggested Reading

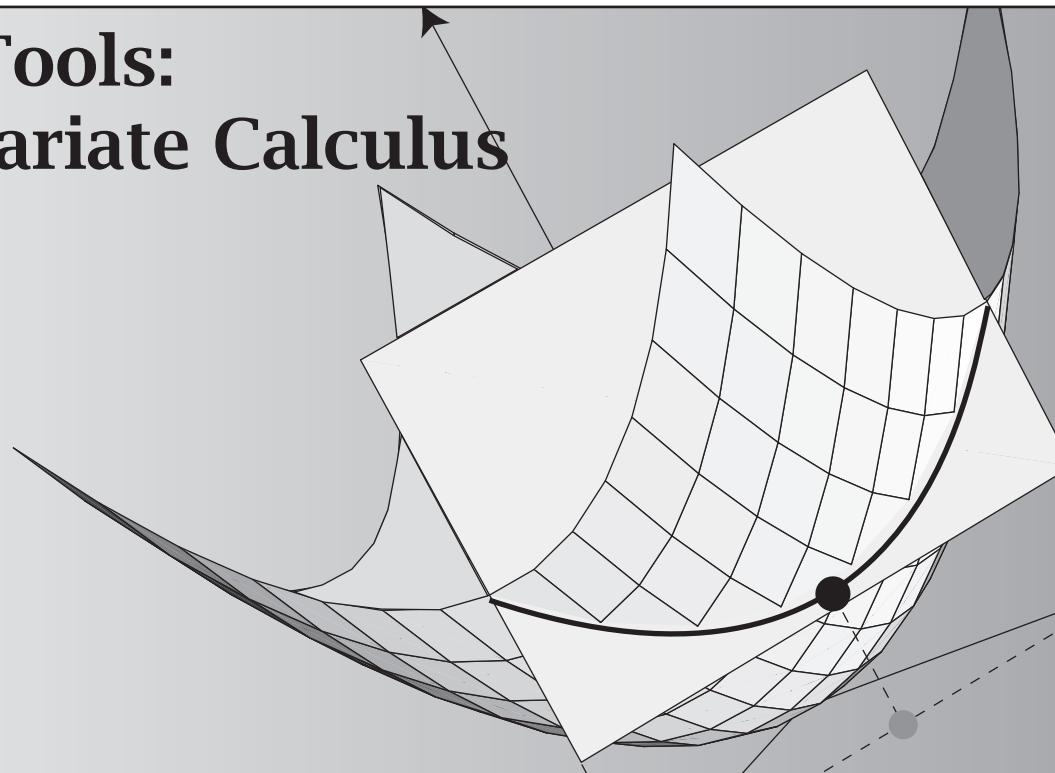
- RS Berry, SA Rice, and J Ross, *Physical Chemistry*, 2nd edition, Oxford University Press, New York, 2000. Physical chemistry covered in great breadth and depth.
- SG Brush, *The Kind of Motion We Call Heat*, Volumes 1 and 2, North-Holland, Amsterdam, 1976. Scholarly history of thought and concepts in thermodynamics.
- AP French, *Newtonian Mechanics*, WW Norton, New York, 1971. Elementary treatment of forces, mechanics, and conservation laws.
- D Halliday, R Resnick, and J Walker, *Fundamentals of Physics Extended*, 9th edition, J Wiley, New York, 2010. Broad survey of elementary physics, including forces and mechanics.

C Kittel and H Kroemer, *Thermal Physics*, 2nd edition, Wiley, New York, 1980. Excellent introduction to statistical thermodynamics for physicists. Introduces the lattice model for the ideal gas law.

F Reif, *Statistical Physics* (Berkeley Physics Course, Volume 5), McGraw-Hill, New York, 1967. Extensive discussion of the kinetic theory of gases.

This page is intentionally left blank.

4 Math Tools: Multivariate Calculus



Multivariate Calculus Applies to Functions of Multiple Variables

So far, we've described model systems with only one degree of freedom: the number of heads n_H , the volume V , the number of particles N , or the amount of internal energy U . In each case, we considered the multiplicity as a function of only one variable: $W = W(V)$, $W = W(N)$, or $W = W(U)$. However, problems in statistical thermodynamics often involve simultaneous changes in multiple degrees of freedom: internal combustion engines burn gasoline, change particle composition, create heat, and expand against a piston all at the same time; metabolites enter cells, and change cell volume and particle number at the same time; a change in the number of bound ligand particles can change the permeability of cell membranes and the shapes, activities, and stabilities of biomolecules.

Here we review the tools from multivariate calculus that you need to describe processes in which multiple degrees of freedom change together. You need these methods to solve two main problems: to find the extrema of multivariate functions and to integrate them.

A function of a single variable, $y = f(x)$, assigns one number to y for each value of x . The function $f(x)$ may be $y = 3x^2$ or $y = \log(x-14)$, for example. In contrast, a *multivariate* function $z = f(x_1, x_2, \dots, x_t)$ of t variables assigns one number to z for each set of values (x_1, x_2, \dots, x_t) . For instance, a function

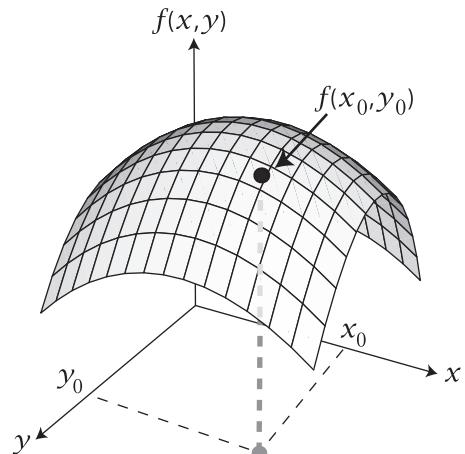


Figure 4.1 The surface represents the multivariate function $f(x, y)$. The point $f(x_0, y_0)$ is the value of f when $y = y_0$ and $x = x_0$.

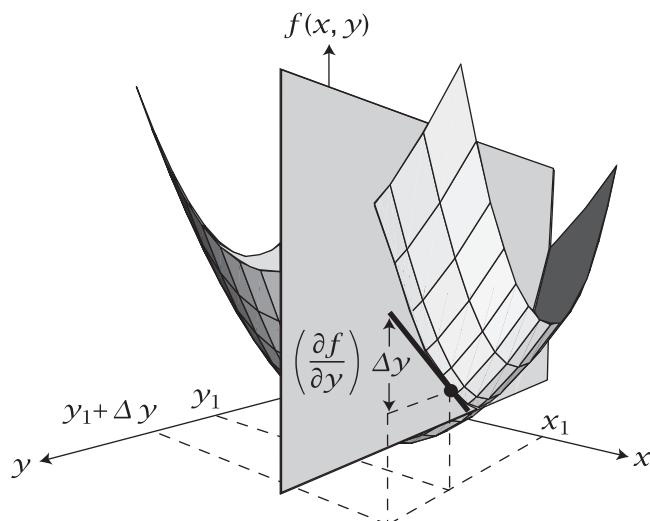


Figure 4.2 The partial derivative of f with respect to y at the point (x_1, y_1) is defined by the slope of the tangent to f at the point (x_1, y_1) , according to Equation (4.2). This is the slope of $f(x, y)$ in the plane $x = x_1$.

of two variables, $z = f(x, y)$, might be $z = 3x/y$ or $z = x + y^3$. A function of two or more variables is represented as a surface, not a curve (see Figure 4.1).

Consider a paraboloid given by the function

$$f(x, y) = (x - x_0)^2 + (y - y_0)^2. \quad (4.1)$$

Writing $f = f(x, y)$ indicates that we take f to be the dependent variable and x and y to be the independent variables. You can choose any x and any y independently, but then the result $f(x, y)$ is determined. Figure 4.2 shows how f depends on x and y , according to Equation (4.1). The plane sliced out of Figure 4.2 parallel to the f and y axes shows how f depends on y alone, with x held constant.

Partial Derivatives Are the Slopes of the Tangents to Multivariate Functions

The *partial derivative* of a multivariate function is the slope along one direction (i.e., taken with respect to only one variable) with all the other variables held

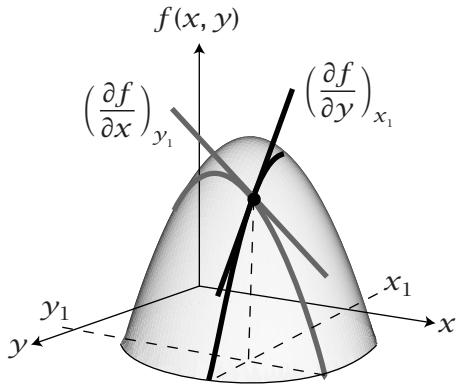


Figure 4.3 The function $f(x, y)$ has two partial derivatives at every point. The partial derivative with respect to y , $(\partial f / \partial y)_{x_1, y_1}$, is the slope of the tangent to the curve parallel to the (y, f) plane, evaluated at the point (x_1, y_1) . $(\partial f / \partial x)_{x_1, y_1}$ is the slope of the tangent parallel to the (x, f) plane, evaluated at (x_1, y_1) .

constant. For example, the partial derivatives of $f(x, y)$ are defined by

$$\left(\frac{\partial f}{\partial x}\right)_y = \lim_{\Delta x \rightarrow 0} \frac{f(x + \Delta x, y) - f(x, y)}{\Delta x}$$

and

$$\left(\frac{\partial f}{\partial y}\right)_x = \lim_{\Delta y \rightarrow 0} \frac{f(x, y + \Delta y) - f(x, y)}{\Delta y}, \quad (4.2)$$

where the symbol ∂ represents a partial derivative. $(\partial f / \partial y)_x$ is the partial derivative of f with respect to y while x is held constant (see Figures 4.2 and 4.3). To evaluate partial derivatives, use the rules for ordinary differentiation with respect to one variable, treating the other independent variables as constants. For example, if the function is $f(x, y) = 3xy^2 + x + 2y^2$, then $(\partial f / \partial x)_{y=1} = (3y^2 + 1)_{y=1} = 4$, and $(\partial f / \partial y)_{x=2} = (6xy + 4y)_{x=2} = 16y$.

Just like ordinary derivatives, partial derivatives are functions of the independent variables. They can be further differentiated to yield second- and higher-order derivatives. For example, $(\partial^2 f / \partial x^2)$ means that you first take the derivative $(\partial f / \partial x)$, then you take the derivative of that quantity with respect to x : $(\partial / \partial x)(\partial f / \partial x)$. For mixed derivatives, the ordering of variables in the denominator indicates what order you take the derivatives: $(\partial^2 f / \partial x \partial y) = (\partial / \partial x)(\partial f / \partial y)$ and $(\partial^2 f / \partial y \partial x) = (\partial / \partial y)(\partial f / \partial x)$. The following example shows how these rules are applied to the *ideal gas law*. We explore the physics of this law later. For now, our purpose is just to illustrate partial derivatives.

EXAMPLE 4.1 Partial derivatives and the ideal gas law. To illustrate the math, consider the ideal gas law, which describes the pressure p of a gas as a function of the independent variables volume V and temperature T :

$$p = p(V, T) = \frac{RT}{V}, \quad (4.3)$$

where R is a constant. The first partial derivatives are

$$\left(\frac{\partial p}{\partial V}\right)_T = -\frac{RT}{V^2} \quad \text{and} \quad \left(\frac{\partial p}{\partial T}\right)_V = \frac{R}{V}.$$

The second partial derivatives are

$$\begin{aligned}\left(\frac{\partial^2 p}{\partial V^2}\right)_T &= \frac{2RT}{V^3} \quad \text{and} \quad \left(\frac{\partial^2 p}{\partial T^2}\right)_V = 0, \\ \left(\frac{\partial^2 p}{\partial V \partial T}\right) &= \left(\frac{\partial(\partial p / \partial T)_V}{\partial V}\right)_T = -\frac{R}{V^2}, \quad \text{and} \\ \left(\frac{\partial^2 p}{\partial T \partial V}\right) &= \left(\frac{\partial(\partial p / \partial V)_T}{\partial T}\right)_V = -\frac{R}{V^2}.\end{aligned}$$

Small Changes and the Total Differential

If you know the value of $f(x)$ at some point $x = a$, you can use the Taylor series expansion (see Appendix A, Equation (A.1)) to compute $f(x)$ near that point:

$$\begin{aligned}\Delta f &= f(x) - f(a) \\ &= \left(\frac{df}{dx}\right)_{x=a} \Delta x + \frac{1}{2} \left(\frac{d^2 f}{dx^2}\right)_{x=a} \Delta x^2 + \frac{1}{6} \left(\frac{d^3 f}{dx^3}\right)_{x=a} \Delta x^3 + \dots,\end{aligned}\tag{4.4}$$

where the subscript $x = a$ here means the derivative is evaluated at $x = a$. For very small changes $\Delta x = (x - a) \rightarrow dx$, the terms involving Δx^2 , Δx^3 , etc., will become vanishingly small. The change in f will become small, and $\Delta f \rightarrow df$ will be given by the first term:

$$df = \left(\frac{df}{dx}\right)_{x=a} dx.\tag{4.5}$$

You can use derivatives to find the extrema of functions (maxima and minima), which predict equilibrium states (as described in Chapters 2 and 3). Maxima or minima in $f(x)$ occur at the points where $df/dx = 0$.

Now let's find the corresponding change for a multivariate function, say $f(x, y)$. In this case, too, the change Δf can be expressed as a Taylor series:

$$\begin{aligned}\Delta f &= f(x, y) - f(a, b) \\ &= \left(\frac{\partial f}{\partial x}\right)_{y=b} \Delta x + \left(\frac{\partial f}{\partial y}\right)_{x=a} \Delta y \\ &\quad + \frac{1}{2} \left[\left(\frac{\partial^2 f}{\partial x^2}\right) \Delta x^2 + \left(\frac{\partial^2 f}{\partial y^2}\right) \Delta y^2 + 2 \left(\frac{\partial^2 f}{\partial x \partial y}\right) \Delta x \Delta y \right] + \dots\end{aligned}$$

As $\Delta x = (x - a) \rightarrow dx$ and $\Delta y = (y - b) \rightarrow dy$, Δf becomes df and the higher-order terms in Δx^2 , Δy^2 , etc., become vanishingly small. Again, the Taylor series expansion reduces the total differential to a sum of the first-order terms:

$$df = \left(\frac{\partial f}{\partial x}\right)_y dx + \left(\frac{\partial f}{\partial y}\right)_x dy.\tag{4.6}$$

For a function of t variables such as $f(x_1, x_2, \dots, x_t)$, the total differential is defined as

$$df = \sum_{i=1}^t \left(\frac{\partial f}{\partial x_i} \right)_{x_j \neq i} dx_i. \quad (4.7)$$

The subscript $x_{j \neq i}$ indicates that all the other $t - 1$ variables x_j are kept constant when you take the derivative with respect to x_i . The total differential df is the sum of contributions resulting from small increments in the independent variables x_i .

The Extrema of Multivariate Functions Occur Where the Partial Derivatives Are Zero

To use extremum principles, which can identify states of equilibrium, you need to find points at which derivatives are zero. In mathematics, a *critical point* is where a first derivative equals zero. It could be a maximum, a minimum, or a saddle point. In statistical thermodynamics, ‘critical point’ has a different meaning, but in this chapter critical point is used only in its mathematical sense.

How do we find an extremum of a multivariate function? First, recall how to find an extremum when there is only one variable. Figure 4.4 shows a single-variable function, $f(x) = x^2 + b$, where b is a constant. To locate the extremum, find the point at which the first derivative equals zero:

$$\left(\frac{df}{dx} \right)_{x=x^*} = 0. \quad (4.8)$$

The critical point x^* occurs at the value $f = f(x^*)$ that satisfies Equation (4.8). If the second derivative d^2f/dx^2 is positive everywhere, the critical point is a minimum. If the second derivative is negative everywhere, the critical point is a maximum.

Now let’s find an extremum of a multivariate function $f(x, y)$. A critical point is where small changes in both x and y together cause the total differential df to equal zero:

$$df = \left(\frac{\partial f}{\partial x} \right)_y dx + \left(\frac{\partial f}{\partial y} \right)_x dy = 0. \quad (4.9)$$

Since x and y are *independent* variables, the small nonzero quantities dx and dy are unrelated to each other. This means that we cannot suppose that the dx and dy terms in Equation (4.9) balance perfectly to cause df to equal zero.

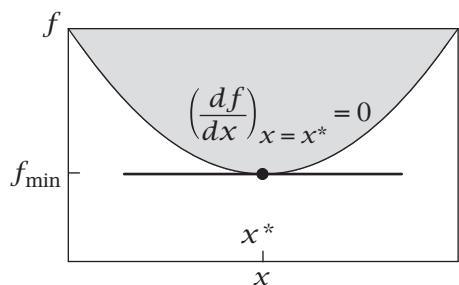


Figure 4.4 To find the extremum x^* of a function $f(x)$, find the point at which $df/dx = 0$.

The independence of the variables x and y requires that *both* terms $(\partial f / \partial x)dx$ and $(\partial f / \partial y)dy$ equal zero simultaneously to satisfy Equation (4.9):

$$\left(\frac{\partial f}{\partial x}\right)_y dx = 0 \quad \text{and} \quad \left(\frac{\partial f}{\partial y}\right)_x dy = 0. \quad (4.10)$$

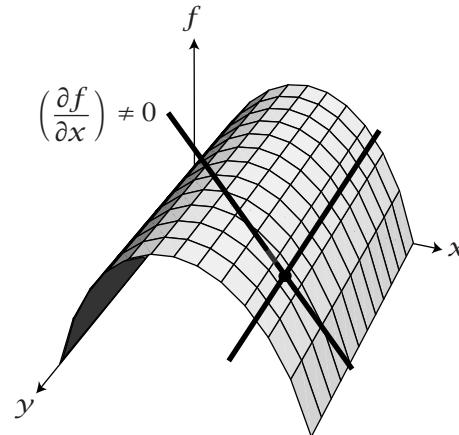
Furthermore, even though dx and dy are infinitesimal quantities, they are *not* equal to zero, so Equations (4.10) are satisfied only when the partial derivatives themselves equal zero:

$$\left(\frac{\partial f}{\partial x}\right)_y = 0 \quad \text{and} \quad \left(\frac{\partial f}{\partial y}\right)_x = 0. \quad (4.11)$$

To have $df = 0$, both Equations (4.11) must be satisfied. All the partial first derivatives must be zero for a point to be an extremum. Here is an example showing that it isn't sufficient for only one partial derivative to be zero.

EXAMPLE 4.2 Check all the partial derivatives. For the function f shown in Figure 4.5, f is independent of y so $(\partial f / \partial y)_x = 0$ everywhere. But $f(x, y)$ is only at a maximum where *both* $(\partial f / \partial x)_y$ and $(\partial f / \partial y)_x$ equal zero. In this case, $f(x, y)$ is maximal only along the line at the top of the function (see Figure 4.5(b)).

(a) Not a Maximum



(b) A Maximum

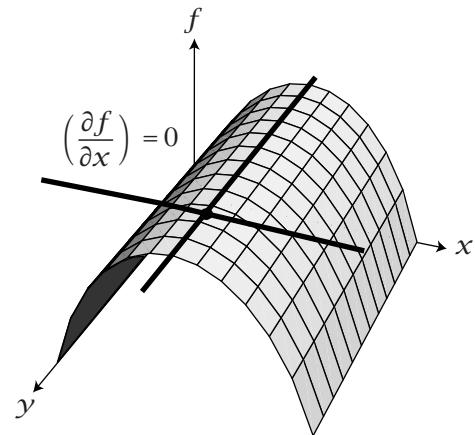


Figure 4.5 To identify the maximum of this function, both $(\partial f / \partial x)$ and $(\partial f / \partial y)$ must equal zero as they do in (b), but not in (a).

The function $f(x_1, x_2, \dots, x_t)$ of t independent variables will have extrema when $df(x_1, x_2, \dots, x_t) = 0$, which means satisfying t simultaneous equations:

$$\left(\frac{\partial f}{\partial x_i}\right)_{x_j \neq i} = 0 \quad \text{for all } i = 1, 2, \dots, t. \quad (4.12)$$

Suppose you want to reach the highest point on a hill. Traveling a path from east to west and finding the highest point on that path corresponds to setting one partial derivative to zero. However, the high point on the east-west path

might not be the top of the hill. From your east–west path, perhaps traveling north takes you still farther uphill. Follow that route to the peak. At the summit, both derivatives are zero. The caveat is that this process might lead to a saddle point. The final tests that we need to ensure that we are at a maximum or a minimum are described in Examples 4.3 and 4.4.

EXAMPLE 4.3 Finding the minimum of a paraboloid. To find the minimum of the paraboloid $f(x, y) = (x - x_0)^2 + (y - y_0)^2$ in Figure 4.6, evaluate each partial derivative:

$$\left(\frac{\partial f}{\partial x}\right)_y = 2(x - x_0) \quad \text{and} \quad \left(\frac{\partial f}{\partial y}\right)_x = 2(y - y_0). \quad (4.13)$$

The critical point (x^*, y^*) is the point at which

$$\left(\frac{\partial f}{\partial x}\right)_y = 0 \implies 2(x^* - x_0) = 0 \implies x^* = x_0$$

and

$$\left(\frac{\partial f}{\partial y}\right)_x = 0 \implies 2(y^* - y_0) = 0 \implies y^* = y_0.$$

The critical point is at $(x^*, y^*) = (x_0, y_0)$. Check the signs of the partial derivatives at the critical point, $(\partial^2 f / \partial x^2)_{x^*=x_0}$ and $(\partial^2 f / \partial y^2)_{y^*=y_0}$, which are both positive, so you might have a minimum. If those second partial derivatives were negative, you might have had a maximum. When some second partial derivatives are positive and some are negative, the function is at a saddle point and not at an extremum. For a function of two variables, look at the *Hessian* derivative expression,

$$M = \left(\frac{\partial^2 f(x, y)}{\partial x^2}\right)_{x^*, y^*} \left(\frac{\partial^2 f(x, y)}{\partial y^2}\right)_{x^*, y^*} - \left(\frac{\partial^2 f(x, y)}{\partial x \partial y}\right)_{x^*, y^*}^2,$$

evaluated at (x^*, y^*) . If $M > 0$ and $(\partial^2 f / \partial x^2) > 0$, you have a local minimum. If $M > 0$ and $(\partial^2 f / \partial x^2) < 0$, you have a local maximum. If $M < 0$, you have a saddle point. If $M = 0$, this test is indecisive—see Example 4.4.

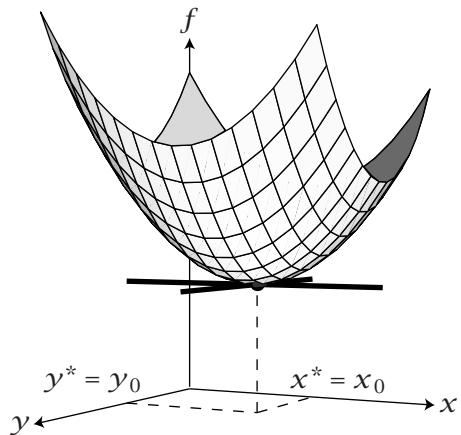
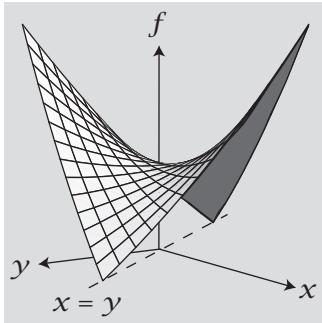
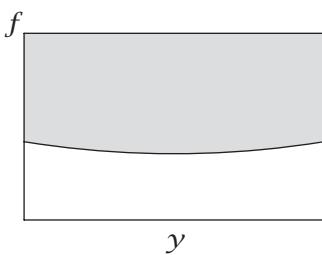


Figure 4.6 For this paraboloid $f(x, y) = (x - x_0)^2 + (y - y_0)^2$, the critical point (x^*, y^*) equals (x_0, y_0) . At (x^*, y^*) , both partial derivatives equal zero.

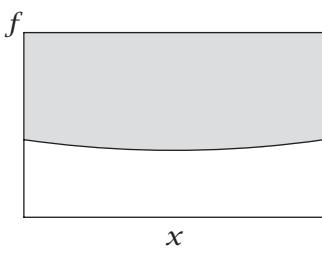
(a) $f(x, y) = x^2 + y^2 - 4xy$



(b) Slice Through $x = 0$



(c) Slice Through $y = 0$



(d) Slice Through $x = y$

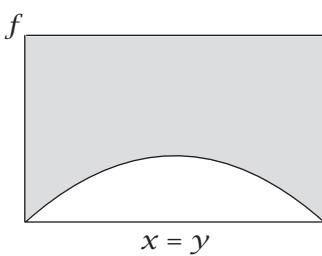


Figure 4.7 The origin is a saddle point of the function $f(x, y) = x^2 + y^2 - 4xy$, not a maximum or a minimum.

EXAMPLE 4.4 A simple function that fails the Hessian test. A function that satisfies the test of the second partial derivatives, with $(\partial^2 f / \partial x^2)$ and $(\partial^2 f / \partial y^2)$ both positive, but fails the Hessian test is $f(x, y) = x^2 + y^2 - 4xy$. The origin $(0, 0)$ is a critical point, and has positive second partials—both equal to 2—but the origin is not an extremum (because $M = -12$), as Figure 4.7 shows.

We have considered the extrema of functions of *independent* variables. Now consider the extrema of functions of variables that are not independent, but are subject to *constraints*. In these cases, you can no longer find the extrema by setting the derivatives $(\partial f / \partial x)$ and $(\partial f / \partial y)$ to zero, as you did in Equation (4.11).

How Can You Find the Extrema of Multivariate Functions That Are Subject to Constraints?

Constraints hold material systems at constant temperature or pressure, or at constant concentrations of chemical species, for example. Our aim in later chapters will be to find states of equilibrium by locating extrema of functions subject to such constraints. To find the extrema of a function $f(x, y)$ subject to constraints, you must find the set of values that satisfy *both* the extremum equation

$$df = 0 = \left(\frac{\partial f}{\partial x} \right)_y dx + \left(\frac{\partial f}{\partial y} \right)_x dy \quad (4.14)$$

and some constraint equation. A constraint is just a relationship between x and y . A constraint equation has the form $g(x, y) = \text{constant}$, for example,

$$\begin{aligned} g(x, y) &= x - y = 0 \\ \Rightarrow x &= y. \end{aligned} \quad (4.15)$$

Because x and y are related through the equation $g(x, y) = \text{constant}$, they are not independent variables. To satisfy both Equations (4.14) and (4.15) simultaneously, put the constraint equation (4.15) into differential form and combine it with the total differential equation (4.14). Since $g = \text{constant}$, the total differential dg is

$$dg = \left(\frac{\partial g}{\partial x} \right)_y dx + \left(\frac{\partial g}{\partial y} \right)_x dy = 0. \quad (4.16)$$

For the constraint equation (4.15), the partial derivatives are $(\partial g / \partial x)_y = 1$ and $(\partial g / \partial y)_x = -1$. Equation (4.16) then gives $dx - dy = 0$, so $dx = dy$. Replace the dy by dx in the extremum equation (4.14) to get

$$\begin{aligned} df &= 0 = \left(\frac{\partial f}{\partial x} \right)_y dx + \left(\frac{\partial f}{\partial y} \right)_x dx \\ \Rightarrow \left[\left(\frac{\partial f}{\partial x} \right)_y + \left(\frac{\partial f}{\partial y} \right)_x \right] dx &= 0. \end{aligned}$$

Since dx is not zero, we must have

$$\left(\frac{\partial f}{\partial x}\right)_y = - \left(\frac{\partial f}{\partial y}\right)_x. \quad (4.17)$$

Solving Equation (4.17) identifies the point that is both an extremum of f and also lies along the line $x = y$. This situation is different from that when x and y are independent. Then each of the partial derivatives equals zero (compare Equation (4.17) with Equation (4.12)).

EXAMPLE 4.5 Finding the minimum of a paraboloid subject to a constraint. Let's apply this reasoning to finding the minimum of the paraboloid $f(x, y) = x^2 + y^2$, subject to the constraint $g(x, y) = x + y = 6$ (see Figure 4.8). Look at this function geometrically: take a slice at $y = 6 - x$ through the paraboloid. Look for the lowest point on that slice (which is not necessarily the lowest point on the entire paraboloid).

To find the lowest point on $f(x, y)$ that satisfies $g(x, y)$, first evaluate the partial derivatives of g . In this case, $(\partial g / \partial x) = 1$ and $(\partial g / \partial y) = 1$. Equation (4.16) then gives $dx = -dy$. Substitute this into Equation (4.14) to get $(\partial f / \partial x)_{x^*, y^*} = (\partial f / \partial y)_{x^*, y^*}$. Taking these partial derivatives gives

$$2x^* = 2y^*. \quad (4.18)$$

The constraint equation requires $y^* = 6 - x^*$, so substitute y^* for x^* in Equation (4.18), with the result

$$2x^* = 2(6 - x^*) \implies x^* = y^* = 3.$$

This is the lowest point on the paraboloid that also lies in the plane $x + y = 6$. As shown in Figure 4.8, the constrained minimum of the function is $f(3, 3) = 18$, which is higher than the global minimum of the unconstrained function, $f(0, 0) = 0$.

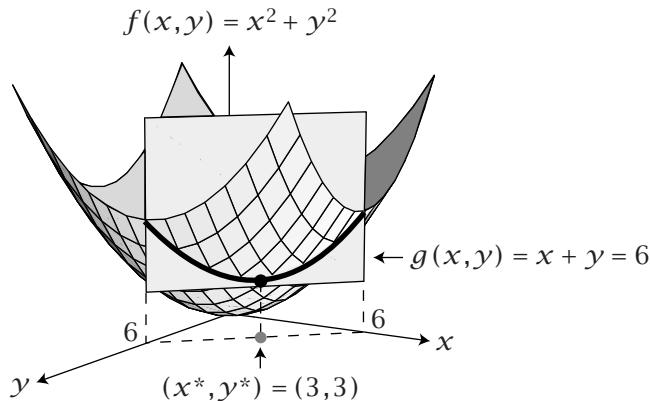


Figure 4.8 The global minimum of $f(x, y) = x^2 + y^2$ is $(x^*, y^*) = (0, 0)$. But the minimum of $f(x, y)$ on the plane $g(x, y) = x + y = 6$ is $(x^*, y^*) = (3, 3)$.

Example 4.5 shows one way to find the extrema of functions subject to constraints. A more general and powerful way is the method of Lagrange multipliers.

The Method of Lagrange Multipliers

Suppose you want to find the extremum of $f(x, y)$, subject to the constraint that $g(x, y) = \text{constant}$. Let's look at the problem graphically (Figure 4.9). In (a), $f(x, y)$ and $g(x, y)$ are intersecting surfaces. Because $g(x, y)$ relates to x and y , but does not relate to f , $g(x, y)$ is just a function in the (x, y) plane. However, in Figure 4.9, we have extended g up and down to look like a ribbon, so you can see how g intersects with f . Curve h is the intersection of f and g , the set of points on the surface $f(x, y)$ that are also on the surface $g(x, y)$. h_{\max} is the maximum value of f on the curve h . Figure 4.9(b) shows two tangent curves in the plane parallel to the (x, y) plane through h_{\max} : ℓ_g , the level curve of $g(x, y)$, and ℓ_f , the level curve of $f(x, y)$. The two level curves are tangent

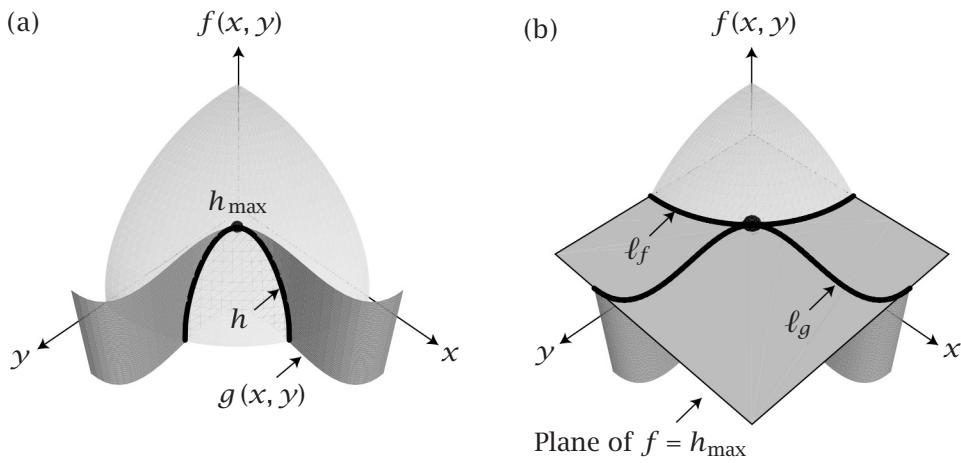


Figure 4.9 (a) The constraint $g(x, y) = 0$ is a curve in the (x, y) plane. However, to visualize it, we extend $g(x, y) = 0$ and show it here as a ribbon that intersects the function $f(x, y)$. (b) The h_{\max} plane slices the constraint ribbon g at the curve ℓ_g and the function f at the curve ℓ_f . (c) The extremum h_{\max} is where ℓ_f is tangent to ℓ_g . Equation (4.21) relates the partial derivatives of f and g to dy/dx at the point h_{\max} .

at h_{\max} , the extremum of f that satisfies $g(x, y)$. The point $h_{\max} = (x^*, y^*)$ must satisfy two conditions: the total differentials of both df and dg are zero,

$$\left(\frac{\partial f}{\partial x}\right)_y dx + \left(\frac{\partial f}{\partial y}\right)_x dy = 0 \quad (4.19)$$

and

$$\left(\frac{\partial g}{\partial x}\right)_y dx + \left(\frac{\partial g}{\partial y}\right)_x dy = 0. \quad (4.20)$$

We use the level curves to find the value of h_{\max} . Figure 4.9(c) shows that at h_{\max} , dy/dx is the slope of the tangent common to both ℓ_f and ℓ_g . The value of dy/dx at h_{\max} is found by combining Equations (4.19) and (4.20):

$$\frac{dy}{dx} = \frac{-\left(\frac{\partial f}{\partial x}\right)_y}{\left(\frac{\partial f}{\partial y}\right)_x} = \frac{-\left(\frac{\partial g}{\partial x}\right)_y}{\left(\frac{\partial g}{\partial y}\right)_x}. \quad (4.21)$$

Because Equation (4.21) requires the equality of two ratios, the derivatives of f and g need only be the same to within an arbitrary constant λ , called the Lagrange multiplier:

$$\left(\frac{\partial f}{\partial x}\right)_y = \lambda \left(\frac{\partial g}{\partial x}\right)_y \quad \text{and} \quad \left(\frac{\partial f}{\partial y}\right)_x = \lambda \left(\frac{\partial g}{\partial y}\right)_x. \quad (4.22)$$

The values $x = x^*$ and $y = y^*$ that satisfy Equations (4.22) are at the extremum, and satisfy the constraint.

EXAMPLE 4.6 Finding the minimum of a paraboloid with Lagrange multipliers. Again let's find the minimum of the paraboloid $f(x, y) = x^2 + y^2$ subject to $g(x, y) = x + y = 6$ (see Example 4.5). But now let's use the Lagrange multiplier method. We have $(\partial g/\partial x) = (\partial g/\partial y) = 1$. According to the Lagrange multiplier method, the solution will be given by

$$\left(\frac{\partial f}{\partial x}\right)_y = \lambda \left(\frac{\partial g}{\partial x}\right)_y \implies 2x^* = \lambda, \quad (4.23)$$

$$\left(\frac{\partial f}{\partial y}\right)_x = \lambda \left(\frac{\partial g}{\partial y}\right)_x \implies 2y^* = \lambda. \quad (4.24)$$

Combining Equations (4.23) and (4.24), and using $x + y = 6$, gives

$$2x^* = 2y^* \implies y^* = x^* = 3. \quad (4.25)$$

The Lagrange multiplier result in Equation (4.25) is identical to our earlier result in Equation (4.18).

The Lagrange multiplier method doesn't simplify this particular example. However, it does simplify problems involving many variables and more than one constraint.

EXAMPLE 4.7 Lagrange multipliers again: finding a maximum. Find the maximum of the function $f(x, y) = 50 - (x-3)^2 - (y-6)^2$ subject to the constraint $g(x, y) = y-x = 0$, as shown in Figure 4.10. The global maximum is at $(x^*, y^*) = (3, 6)$, where both partial derivatives equal zero:

$$\left(\frac{\partial f}{\partial x}\right)_y = -2(x^*-3) = 0 \implies x^* = 3,$$

$$\left(\frac{\partial f}{\partial y}\right)_x = -2(y^*-6) = 0 \implies y^* = 6.$$

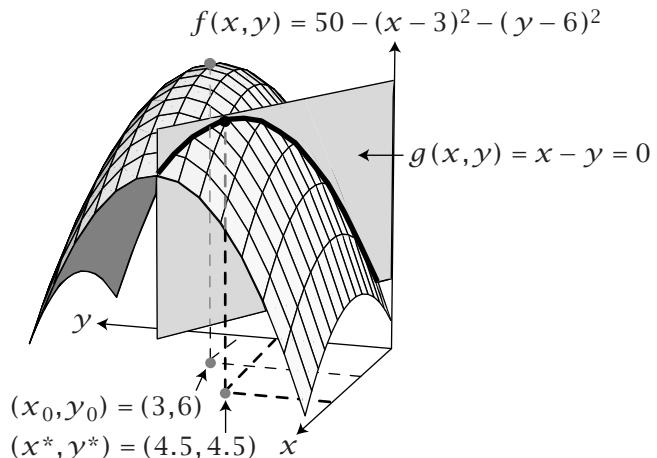
But, since $(\partial g/\partial x) = -1$ and $(\partial g/\partial y) = 1$, the maximum along the constraint plane is at $(x^*, y^*) = (4.5, 4.5)$, according to Equation (4.22):

$$\left(\frac{\partial f}{\partial x}\right)_y = -2(x^*-3) = -\lambda,$$

$$\left(\frac{\partial f}{\partial y}\right)_x = -2(y^*-6) = \lambda.$$

Eliminating λ gives $(x^*-3) = -(y^*-6)$. Because the constraint is $y^* = x^*$, the maximum is at $y^* = x^* = 4.5$.

Figure 4.10 The global maximum of the paraboloid $f(x, y) = 50 - (x-3)^2 - (y-6)^2$ is $f(x_0, y_0) = f(3, 6) = 50$. However, the maximum subject to the constraint $g(x, y) = x-y = 0$ is $f(x^*, y^*) = f(4.5, 4.5) = 45.5$.



EXAMPLE 4.8 Lagrange multipliers find the maximum area/perimeter ratio. Suppose you have a fence 40 feet long, and want to use it to enclose a rectangle of the largest possible area. The area is

$$f(x, y) = xy.$$

The perimeter is

$$g(x, y) = 2x + 2y = 40. \quad (4.26)$$

Use Lagrange multipliers to maximize f subject to g :

$$\left(\frac{\partial f}{\partial x}\right) = \gamma, \quad \left(\frac{\partial g}{\partial x}\right) = 2 \quad \Rightarrow \quad \gamma^* = 2\lambda, \quad (4.27)$$

$$\left(\frac{\partial f}{\partial y}\right) = x, \quad \left(\frac{\partial g}{\partial y}\right) = 2 \quad \Rightarrow \quad x^* = 2\lambda. \quad (4.28)$$

Substituting Equations (4.27) and (4.28) into Equation (4.26) and solving for λ gives $8\lambda = 40 \Rightarrow \lambda = 5$ and $x^* = y^* = 10$. Therefore the rectangle of maximal area is a square, in this case with area $x^*y^* = 100$ square feet.

In general, to find the extremum of $f(x_1, x_2, \dots, x_t)$ subject to a constraint $g(x_1, x_2, \dots, x_t) = c$, where c is a constant, you have

$$\begin{aligned} \left(\frac{\partial f}{\partial x_1}\right) - \lambda \left(\frac{\partial g}{\partial x_1}\right) &= 0, \\ \left(\frac{\partial f}{\partial x_2}\right) - \lambda \left(\frac{\partial g}{\partial x_2}\right) &= 0, \\ &\vdots \\ \left(\frac{\partial f}{\partial x_t}\right) - \lambda \left(\frac{\partial g}{\partial x_t}\right) &= 0. \end{aligned} \quad (4.29)$$

λ is eliminated by using the constraint equation

$$g(x_1, x_2, \dots, x_t) = \text{constant}. \quad (4.30)$$

For the extremum of $f(x_1, x_2, \dots, x_t)$ subject to more than one constraint, $g(x_1, x_2, \dots, x_t) = c_1$ and $h(x_1, x_2, \dots, x_t) = c_2$, etc., where the c_i are constants, the Lagrange multiplier method gives the solutions

$$\begin{aligned} \left(\frac{\partial f}{\partial x_1}\right) - \lambda \left(\frac{\partial g}{\partial x_1}\right) - \beta \left(\frac{\partial h}{\partial x_1}\right) - \dots &= 0, \\ \left(\frac{\partial f}{\partial x_2}\right) - \lambda \left(\frac{\partial g}{\partial x_2}\right) - \beta \left(\frac{\partial h}{\partial x_2}\right) - \dots &= 0, \\ &\vdots \\ \left(\frac{\partial f}{\partial x_t}\right) - \lambda \left(\frac{\partial g}{\partial x_t}\right) - \beta \left(\frac{\partial h}{\partial x_t}\right) - \dots &= 0. \end{aligned} \quad (4.31)$$

where λ, β, \dots are the Lagrange multipliers for each constraint. Each Lagrange multiplier is found from its appropriate constraint equation.

Here is another perspective on the Lagrange multiplier strategy embodied in Equations (4.31). We intend to find the extremum

$$df = \sum_{i=1}^t \left(\frac{\partial f}{\partial x_i}\right) dx_i = 0, \quad (4.32)$$

subject to the constraints

$$dg = \sum_{i=1}^t \left(\frac{\partial g}{\partial x_i}\right) dx_i = 0 \quad (4.33)$$

and

$$dh = \sum_{i=1}^t \left(\frac{\partial h}{\partial x_i} \right) dx_i = 0. \quad (4.34)$$

Incorporating Equations (4.33) and (4.34) into Equation (4.32) with Lagrange multipliers gives

$$d(f - \lambda g - \beta h) = \sum_{i=1}^t \left[\left(\frac{\partial f}{\partial x_i} \right) - \lambda \left(\frac{\partial g}{\partial x_i} \right) - \beta \left(\frac{\partial h}{\partial x_i} \right) \right] dx_i = 0. \quad (4.35)$$

Equations (4.31) show that each bracketed term in Equation (4.35) must equal zero, so the Lagrange multipliers have served to convert a problem involving constraints to a problem involving t independent quantities $[(\partial f / \partial x_i) - \lambda (\partial g / \partial x_i) - \beta (\partial h / \partial x_i)]$, each of which must equal zero.

Integrating Multivariate Functions Sometimes Depends on the Pathway of Integration, and Sometimes It Doesn't

Maxima or minima are not the only things you need from multivariate functions. Sometimes you need to integrate them. Suppose you have a topographic map that gives the altitude $f(x, y)$ of every point in the United States as a function of east-west and north-south coordinates, x and y . Finding the maxima and minima will locate the peaks of the mountains and the bottoms of the valleys. But maybe instead you want to know the difference in altitude between San Francisco and Chicago. For that quantity, you need to sum up all the ups and downs between those two cities. That is, you need to find $\Delta f = \int_A^B df$, where f is the altitude function, A represents the coordinates (x, y) of San Francisco, and B represents the coordinates of Chicago.

For integrating, there are two different types of multivariate functions: one type is called a *state* function and the other is called a *path-dependent* function. The altitude function (height above sea level) is a state function: the difference in altitude between Chicago and San Francisco does not depend on what route you take between the two cities. The altitude difference depends only on the coordinates (i.e., the two *states*) of the beginning point, San Francisco, and the end point, Chicago. All the ups and downs sum to exactly the same altitude difference whether you drive a northern route or a southern route. Contrast that with a different function, your *distance traveled*, as measured by your odometer. Distance traveled is a path-dependent function: you travel a longer distance along some routes from San Francisco to Chicago than along others [1]. For state functions, computing the integral is much simpler than for path-dependent functions. Both types of functions are important in thermodynamics.

What is the difference between these two types of functions? For any multivariate function, small changes can be expressed as $t(x, y)dx + s(x, y)dy$, where t and s are functions of x and y . For example, $y dx + x dy$ is a state function and $xy^2 dx + x dy$ is a path-dependent function. Why do you integrate

one differently than the other? To integrate any multivariate function, you sum over all the variations in both x and y :

$$\int_{y_A}^{y_B} \int_{x_A}^{x_B} [s(x, y) dx + t(x, y) dy]. \quad (4.36)$$

You could first integrate over x holding $y = y_A$, then integrate over y . Or you could first integrate over y holding $x = x_A$, then integrate over x . A *pathway* is defined as the particular sequence of values of the independent variables x and y that you use when you integrate over them. A few possible pathways for integration are shown as projections onto the (x, y) -plane in Figure 4.11. This matter of pathways never arises for functions of a single variable, $f(x)$, because a single coordinate is like having only a single road. For path-dependent functions like $xy^2 dx + x dy$ integrating over one pathway gives you a different answer for the integral than integrating over another pathway. Your distance traveled from San Francisco to Chicago is the sum of the lengths of all the specific roads you took, which is different for different routes. On the other hand, when $f(x, y)$ has the special simplicity of being a state function, integrating is easier. The definition of a state function is one in which you can express $s dx + t dy$ as the differential df of some function $f(x, y)$:

$$df = s dx + t dy = \left(\frac{\partial f}{\partial x} \right)_y dx + \left(\frac{\partial f}{\partial y} \right)_x dy. \quad (4.37)$$

That is, any expression that you can write as some function $f(x, y) = \dots$ has an exact differential. For a state function, you can substitute Equation (4.37) into Equation (4.36) to get

$$\int_{x_A}^{x_B} \int_{y_A}^{y_B} (s dx + t dy) = \int df = f(x_B, y_B) - f(x_A, y_A) = f_B - f_A. \quad (4.38)$$

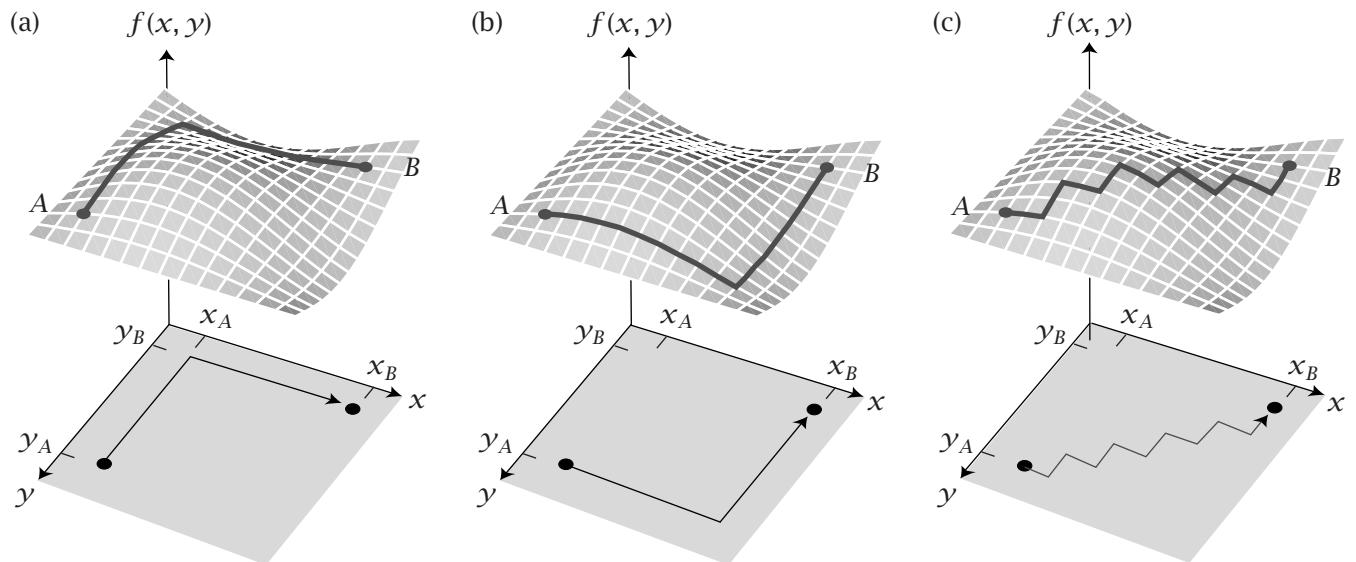


Figure 4.11 Different pathways for integrating $f(x, y)$ from (x_A, y_A) to (x_B, y_B) .

In short, evaluating this type of integral is very simple: you just compute the altitude f_B in Chicago and subtract the altitude f_A in San Francisco. Figure 4.11 shows that this difference in f does not depend on which pathway you take.

EXAMPLE 4.9 State functions. $x \, dy + y \, dx$ is a state function because it can be expressed as the differential df of the function $f = xy$,

$$df = \left(\frac{\partial f}{\partial x} \right)_y dx + \left(\frac{\partial f}{\partial y} \right)_x dy = y \, dx + x \, dy.$$

Similarly, $6xy \, dx + 3x^2 \, dy$ is a state function because it can be expressed as the differential of $f = 3x^2y$.

df is called an *exact* differential. On the other hand, quantities like $y^2 \, dx + x^2 \, dy$ or $x^2y^3 \, dx + 3x^2y^2 \, dy$ are called *inexact* differentials because they cannot be expressed as $d(\text{some function of } x \text{ and } y)$. For inexact differentials, Equation (4.38) does not apply and the value of the integral will depend on the integration pathway. How do you determine whether any particular differential $s(x, y) \, dx + t(x, y) \, dy$ is exact or inexact? Use the Euler test.

The **Euler reciprocal relationship** is

$$\left(\frac{\partial^2 f}{\partial x \partial y} \right) = \left(\frac{\partial^2 f}{\partial y \partial x} \right), \quad (4.39)$$

for any state function f (see Appendix D for a proof). To find out whether an expression $s(x, y) \, dx + t(x, y) \, dy$ is an exact differential, determine whether the following equality holds:

$$\left(\frac{\partial s(x, y)}{\partial y} \right) = \left(\frac{\partial t(x, y)}{\partial x} \right). \quad (4.40)$$

If Equation (4.40) holds, then $s = (\partial f / \partial x)$, $t = (\partial f / \partial y)$, and $s(x, y) \, dx + t(x, y) \, dy = df$ is an exact differential.

EXAMPLE 4.10 Applying the Euler test and integrating a state function. Is $6xy^3 \, dx + 9x^2y^2 \, dy$ an exact differential? Apply the test of Equation (4.40):

$$\left(\frac{\partial(6xy^3)}{\partial y} \right) = 18xy^2 \quad \text{and} \quad \left(\frac{\partial(9x^2y^2)}{\partial x} \right) = 18xy^2.$$

Because the two second partial derivatives are equal, $6xy^3 \, dx + 9x^2y^2 \, dy$ is an exact differential. The function $f(x, y)$ is

$$f(x, y) = 3x^2y^3. \quad (4.41)$$

That is, $d(3x^2y^3) = 6xy^3 dx + 9x^2y^2 dy$. You can confirm this by checking its derivatives $(\partial f/\partial x) = 6xy^3$ and $(\partial f/\partial y) = 9x^2y^2$. (To obtain the function f , either integrate $(\partial f/\partial x) = 6xy^3$ with respect to x or integrate $(\partial f/\partial y) = 9x^2y^2$ with respect to y ; for state functions, both give the same answer, which is f . Now, to integrate this state function from $(x_1, y_1) = (1, 1)$ to $(x_2, y_2) = (4, 5)$,

$$\int_{x=1}^4 \int_{y=1}^5 (6xy^3 dx + 9x^2y^2 dy), \quad (4.42)$$

we need not be concerned with the pathway of integration. The integral depends only on the difference between the initial and final states, (x_1, y_1) and (x_2, y_2) :

$$\begin{aligned} \Delta f &= f(x_2, y_2) - f(x_1, y_1) = 3x^2y^3 \Big|_{x_1, y_1}^{x_2, y_2} \\ &= (3 \cdot 4^2 \cdot 5^3) - (3 \cdot 1^2 \cdot 1^3) = 5997. \end{aligned} \quad (4.43)$$

EXAMPLE 4.11 Applying the Euler test again. Is $x^2 dx + xy dy$ an exact differential? Apply the test of Equation (4.40). You will find that it is not, because $(\partial x^2/\partial y) = 0$, which does not equal $(\partial(xy)/\partial x) = y$. The value of the integral of this differential depends on the choice of pathway.

Path functions, state functions, and the Euler relationships play a central role in thermodynamics, as we'll see in later chapters. We now give some useful rules for manipulating partial derivatives.

The Chain Rule Applies to Functions of Several Variables

The chain rule is useful for *composite functions*, functions of variables that are themselves functions of other variables. For example, the velocity v of a particle moving through a liquid will depend on the force F that is applied to it and on the viscosity η of the liquid, which further depends on the temperature T of the liquid. You can express this dependence as $v = v(F, \eta(T))$.

Suppose that the variables x , y , and z in a function $f(x, y, z)$ are each a function of another variable u , that is, $x = x(u)$, $y = y(u)$, and $z = z(u)$. Then $f(x, y, z)$ is a *composite* function $f = f(x(u), y(u), z(u))$. The total derivative df in terms of the variables (x, y, z) is

$$df = \left(\frac{\partial f}{\partial x}\right)_{y,z} dx + \left(\frac{\partial f}{\partial y}\right)_{x,z} dy + \left(\frac{\partial f}{\partial z}\right)_{x,y} dz. \quad (4.44)$$

The differentials of x , y , and z with respect to u are

$$dx = \frac{dx}{du} du, \quad dy = \frac{dy}{du} du, \quad dz = \frac{dz}{du} du. \quad (4.45)$$

Substitution of Equations (4.45) into Equation (4.44) produces the **chain rule for partial differentiation**:

$$df = \left[\left(\frac{\partial f}{\partial x} \right)_{y,z} \frac{dx}{du} + \left(\frac{\partial f}{\partial y} \right)_{x,z} \frac{dy}{du} + \left(\frac{\partial f}{\partial z} \right)_{x,y} \frac{dz}{du} \right] du, \quad (4.46)$$

or

$$\frac{df}{du} = \left(\frac{\partial f}{\partial x} \right)_{y,z} \frac{dx}{du} + \left(\frac{\partial f}{\partial y} \right)_{x,z} \frac{dy}{du} + \left(\frac{\partial f}{\partial z} \right)_{x,y} \frac{dz}{du}. \quad (4.47)$$

The symbol d/du indicates differentiation when a function depends only on a single variable, while $(\partial/\partial x)$ indicates partial differentiation of a multivariate function. In these expressions $x(u)$, $y(u)$, and $z(u)$ are functions of only one variable, while $f(x, y, z)$ is multivariate.

EXAMPLE 4.12 The chain rule. Let's return to the ball in the valley from Chapter 2, pages 29–30. The height of the ball $z = x^2$ is a function of lateral position x . However, the energy is a function of height $U = mgz$, so the energy is a composite function $U(z(x))$. The chain rule gives the change in energy with lateral position, dU/dx :

$$\frac{dU}{dx} = \left(\frac{\partial U}{\partial z} \right) \frac{dz}{dx} = (mg)(2x) = 2mgx.$$

EXAMPLE 4.13 Reshaping a cylinder while keeping its volume constant. You have a cylinder of radius r and height h . Its volume is $V = \pi r^2 h$. If you double the radius, how should you change the height to keep the volume constant? The constraint is expressed by

$$dV = 0 = \left(\frac{\partial V}{\partial r} \right)_h dr + \left(\frac{\partial V}{\partial h} \right)_r dh. \quad (4.48)$$

Now divide Equation (4.48) by dr :

$$0 = \left(\frac{\partial V}{\partial r} \right)_h + \left(\frac{\partial V}{\partial h} \right)_r \frac{dh}{dr} \implies \frac{dh}{dr} = -\frac{(\partial V/\partial r)_h}{(\partial V/\partial h)_r}. \quad (4.49)$$

Since $(\partial V/\partial r)_h = 2\pi rh$ and $(\partial V/\partial h)_r = \pi r^2$, Equation (4.49) gives

$$\frac{dh}{dr} = -\frac{2\pi rh}{\pi r^2} = -\frac{2h}{r}.$$

Rearrange to get the h 's on the left and the r 's on the right and integrate:

$$-\frac{1}{2} \int_{h_1}^{h_2} \frac{dh}{h} = \int_{r_1}^{r_2} \frac{dr}{r} \implies -\frac{1}{2} \ln\left(\frac{h_2}{h_1}\right) = \ln\left(\frac{r_2}{r_1}\right) \implies \frac{h_2}{h_1} = \left(\frac{r_1}{r_2}\right)^2.$$

Doubling the radius implies that the height must be reduced to one-quarter of its initial value to hold the same volume.

Summary

Some functions depend on more than a single variable. To find extrema of such functions, it is necessary to find where all the partial derivatives are zero. To find extrema of multivariate functions that are subject to constraints, the Lagrange multiplier method is useful. Integrating multivariate functions is different from integrating single-variable functions: multivariate functions require the concept of a pathway. *State functions* do not depend on the pathway of integration. The Euler reciprocal relation can be used to distinguish state functions from path-dependent functions. In the next few chapters, we will combine the First and Second Laws with multivariate calculus to derive the principles of thermodynamics.

Problems

1. Applying the Euler test. Which of the following are exact differentials?

- (a) $6x^5 dx + dy$
- (b) $x^2 y^2 dx + 3x^2 y^3 dy$
- (c) $(1/y) dx - (x/y^2) dy$
- (d) $y dx + 2x dy$
- (e) $\cos x dx - \sin y dy$
- (f) $(x^2 + y) dx + (x + y^2) dy$
- (g) $x dx + \sin y dy$

2. Differentiating multivariate functions. Compute the partial derivatives $(\partial f / \partial x)_y$ and $(\partial f / \partial y)_x$ for the following functions:

- (a) $f(x, y) = 3x^2 + y^5$
- (b) $f(x, y) = x^{10} y^{1/2}$
- (c) $f(x, y) = x + y^2 + 3$
- (d) $f(x, y) = 5x$

3. Minimizing a single-variable function subject to a constraint. Given the function $y = (x-2)^2$, find x^* , the value of x that minimizes y subject to the constraint $y = x$.

4. Maximizing a multivariate function without constraints. Find the maximum (x^*, y^*, z^*) of the function $f(x, y, z) = d - (x-a)^2 - (y-b)^2 - (z-c)^2$.

5. Extrema of multivariate functions with constraints.

- (a) Find the maximum of the function $f(x, y) = -(x-a)^2 - (y-b)^2$ subject to the constraint $y = kx$.
- (b) Find the minimum of the paraboloid $f(x, y) = (x-x_0)^2 + (y-y_0)^2$ subject to the constraint $y = 2x$.

6. Composite functions.

- (a) Given the functions $f(x, y(u)) = x^2 + 3y^2$ and $y(u) = 5u + 3$, express df in terms of changes dx and du .
- (b) What is $\left(\frac{\partial f}{\partial u}\right)_{x,u=1}$?

7. Converting to an exact differential. Given the expression $dx + (x/y) dy$, show that dividing by x results in an exact differential. What is the function $f(x, y)$ such that df is $dx + (x/y) dy$ divided by x ?

8. Propagation of error. Suppose that you can measure independent variables x and y and that you have a dependent variable $f(x, y)$. The average values are \bar{x} , \bar{y} , and \bar{f} . We define the *error* in x as the deviations $\varepsilon_x = x - \bar{x}$, in y as $\varepsilon_y = y - \bar{y}$, and in f as $\varepsilon_f = f - \bar{f}$.

- (a) Use a Taylor series expansion to express the error ε_f in f , as a function of the errors ε_x and ε_y in x and y .
- (b) Compute the mean-squared error $\langle \varepsilon_f^2 \rangle$ as a function of $\langle \varepsilon_x^2 \rangle$ and $\langle \varepsilon_y^2 \rangle$.

(c) Consider the ideal gas law with $f = p$, $x = T$, and $y = V$. If T and V have 10% random errors, how big will be the error in p ?

(d) How much increase in $\varepsilon(f)$ comes from adding n measurements that all have the same uncorrelated error $\varepsilon(x)$?

9. Small differences in large numbers can lead to nonsense. Using the results from Problem 8, show that the propagated error is larger than the difference itself for $f(x, y) = x - y$, with $x = 20 \pm 2$ and $y = 19 \pm 2$.

10. Finding extrema. Find the point (x^*, y^*, z^*) that is at the minimum of the function

$$f(x, y, z) = 2x^2 + 8y^2 + z^2$$

subject to the constraint equation

$$g(x, y, z) = 6x + 4y + 4z - 72 = 0.$$

11. Derivative of a multivariable composite function. For the function $f(x, y(v)) = x^2 y + y^3$, where $y = mv^2$, compute df/dv around the point where $m = 1$, $v = 2$, and $x = 3$.

12. Volume of a cylinder. For a cylinder of radius r and height h , the volume is $V = \pi r^2 h$, and the surface area is $A = 2\pi r^2 + 2\pi r h$.

- (a) Derive the height $h(r_0)$ that maximizes the volume of a cylinder with a given area $A = a_0$ and given radius r_0 .
- (b) Compute the change in volume, ΔV , from $(r_1, h_1) = (1, 1)$ to $(r_2, h_2) = (2, 2)$.
- (c) Compute the component volume changes ΔV_a and ΔV_b that sum to ΔV , where ΔV_a is the change from $(r_1, h_1) = (1, 1)$ to $(r_2, h_1) = (2, 1)$ and ΔV_b is the change from $(r_2, h_1) = (2, 1)$ to $(r_2, h_2) = (2, 2)$.
- (d) Should (b) equal (c)? Why or why not?

13. Equations of state. Which of the following could be the total derivative of an equation of state?

(a)

$$\frac{2nRT}{(V-nb)^2} dV + \frac{R(V-nb)}{nb^2} dT.$$

(b)

$$-\frac{nRT}{(V-nb)^2} dV + \frac{nR}{V-nb} dT.$$

14. Distance from the circumference of a circle. The circle shown in Figure 4.12 satisfies the equation $x^2 + y^2 = 4$. Find the point (x^*, y^*) on the circle that is closest to the point $(3, 2)$. (That is, minimize the distance $f(x, y) = \Delta r^2 = (x-3)^2 + (y-2)^2$.)

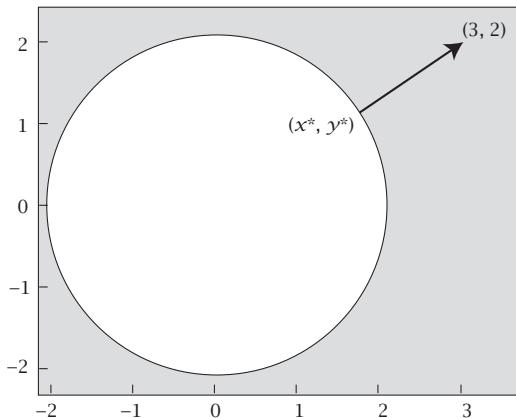


Figure 4.12

15. Find df and Δf .

- (a) If $f(x, y) = x^2 + 3y$, express df in terms of dx and dy .
 (b) For $f(x, y) = x^2 + 3y$, integrate from $(x, y) = (1, 1)$ to $(x, y) = (3, 3)$ to obtain Δf .

16. Derivative of a composite function. For $f(x, y) = x^2 y + y^3$, where $y = mr^2$, find df/dr .

17. Maximum volume of a solid with constraints. You have a rectangular solid of length x , width y , and height z . Find the values of x , y , and z that give the maximum volume, subject to a fixed surface area: $g(x, y, z) = 2xz + 2yz + 2xy - \text{constant} = 0$.

18. Short-answer questions.

- (a) Compute the partial derivatives: $(\partial f / \partial x)_y$ and $(\partial f / \partial y)_x$ for the following functions:

$$f(x, y) = \ln(2x) + 5y^3,$$

$$\begin{aligned} f(x, y) &= (x+a)^8 y^{1/2}, \\ f(x, y) &= e^{7y^2} + 9, \\ f(x, y) &= 13x + 6xy^3. \end{aligned}$$

- (b) Which of the following are exact differentials?

$$\begin{aligned} &5x^2 dx + 6y dy, \\ &5 \ln(y) dx + 5x^0 dy, \\ &(\sin x - y \cos x) dx + \sin(-x) dy, \\ &(e^{2x} + y) dx + (x + e^{2y}) dy, \\ &y^2 dy + x^2 dx. \end{aligned}$$

References

- [1] R Dickerson, *Molecular Thermodynamics*, WA Benjamin, New York, 1969.

Suggested Reading

The following are good calculus references:

- HB Callen, *Thermodynamics and an Introduction to Thermostatistics*, 2nd edition, Wiley, New York, 1985.
 JE Marsden and AJ Tromba, *Vector Calculus*, 5th edition, WH Freeman, New York, 2003.
 GB Thomas, MD Weir, and JR Hass, *Thomas' Calculus*, 12th edition, Addison-Wesley, Reading, MA, 2009.

This page is intentionally left blank.

5 Entropy & the Boltzmann Law

$$S = k \log W$$



What Is Entropy?

Carved on the tombstone of Ludwig Boltzmann in the Zentralfriedhof (central cemetery) in Vienna is the inscription

$$S = k \log W. \quad (5.1)$$

This equation is the historical foundation of statistical mechanics. It connects the microscopic and macroscopic worlds. It defines a macroscopic quantity called the *entropy* S , in terms of the multiplicity W of the microscopic degrees of freedom of a system. For thermodynamics, $k = 1.380662 \times 10^{-23} \text{ JK}^{-1}$ is a quantity called Boltzmann's constant, and Boltzmann's inscription refers to the natural logarithm, $\log_e = \ln$.

In Chapters 2 and 3, we used simple models to illustrate that the composition of coin flips, the expansion of gases, the tendency of particles to mix, rubber elasticity, and heat flow can be predicted by the principle that systems tend toward their states of maximum multiplicity W . However, states that maximize W will also maximize W^2 or $15W^3 + 5$ or $k \ln W$, where k is any positive constant. Any monotonic function of W will have a maximum where W has a maximum. In particular, states that maximize W also maximize the entropy, $S = k \ln W$. Why does $k \ln W$ deserve special attention as a prediction principle, and why should it have this particular mathematical form?

We first show that expressing the entropy instead in terms of a set of probabilities p_i ,

$$\frac{S}{k} = - \sum_{i=1}^t p_i \ln p_i, \quad (5.2)$$

is equivalent to the multiplicity expression $S = k \ln W$ in Equation (5.1). Roll a t -sided die N times. The multiplicity of outcomes is given by Equation (1.21):

$$W = \frac{N!}{n_1! n_2! \cdots n_t!},$$

where n_i is the number of times that side i appears face up. Use Stirling's approximation $x! \approx (x/e)^x$ (see Appendix B), and define the probabilities $p_i = n_i/N$, to convert Equation (1.21) to

$$\begin{aligned} W &= \frac{(N/e)^N}{(n_1/e)^{n_1} (n_2/e)^{n_2} \cdots (n_t/e)^{n_t}} \\ &= \frac{N^N}{n_1^{n_1} n_2^{n_2} \cdots n_t^{n_t}} = \frac{1}{p_1^{n_1} p_2^{n_2} \cdots p_t^{n_t}}. \end{aligned} \quad (5.3)$$

Take the logarithm of both sides and divide by N to get

$$\ln W = - \sum_{i=1}^t n_i \ln p_i \implies \frac{1}{N} \ln W = - \sum_{i=1}^t p_i \ln p_i = \frac{S_N}{Nk} = \frac{S}{k}, \quad (5.4)$$

where S_N is the total entropy for N trials, so the entropy per trial is $S = S_N/N$. For this dice problem and the counting problems in Chapters 2 and 3, the two expressions for the entropy, Equations (5.2) and (5.1), are equivalent. The flattest distributions have maximum multiplicity W in the absence of constraints. For example, in N coin flips, the multiplicity $W = N! / [(n_H!)(N-n_H)!]$ is maximized when $n_H/N \approx n_T/N \approx 0.5$, that is, when the probabilities of heads and tails are equal. We will sometimes prefer the counting strategy, and the use of $S/k = \ln W$, but $S/k = -\sum_i p_i \ln p_i$ will be more convenient in other cases.

Boltzmann's constant k puts entropy into units that interconvert with energy for thermodynamics and molecular science. k is the entropy per particle. Sometimes it is more convenient to express the entropy per mole of particles, $S = R \ln W$, where $R = \mathcal{N}k$ is called the *gas constant* and \mathcal{N} is Avogadro's number, the number of molecules per mole. In addition, entropy is a concept that is broader than molecular science; it applies to any type of probability distribution function. For other types of probability distributions, k is chosen to suit the purposes at hand. For some of the examples below, $k = 1$ is simplest.

Different probability distributions lead to different types of entropy. Sometimes molecules are distributed throughout different points in space. When those spatial distributions are described by probabilities p_i , the distribution describes the *translational* freedom and the entropy is called the *translational entropy* (see Examples 2.2 and 2.3). In other cases, objects or molecules populate different spatial orientations or angles; this leads to the *rotational* or *orientational* entropy. Example 5.1 applies Equation (5.2) to the orientations of *dipoles*. It illustrates a property of all types of entropies: flatter probability distributions have higher entropies than more peaked distributions.

EXAMPLE 5.1 Dipoles tend to orient randomly. Objects with distinguishable heads and tails such as magnets, chemically asymmetrical molecules, electrical dipoles with (+) charges at one end and (–) charges at the other, or even pencils with erasers at one end have rotational freedom as well as translational freedom. They can orient.

Spin a pencil on a table N times. Each time it stops, the pencil points in one of four possible directions: toward the quadrant facing north (n), east (e), south (s), or west (w). Count the number of times that the pencil points in each direction; label those numbers n_n , n_e , n_s , and n_w . Spinning a pencil and counting orientations is analogous to rolling a die with four sides labeled n , e , s , and w . Each roll of that die determines the orientation of one pencil or dipole. N die rolls correspond to the orientations of N dipoles. The number of configurations for systems with N trials, distributed with any set of outcomes $\{n_1, n_2, \dots, n_t\}$, where $N = \sum_{i=1}^t n_i$, is given by the multiplicity equation (1.21): $W(n_1, n_2, \dots, n_t) = N!/(n_1!n_2!\cdots n_t!)$. The number of different configurations of the system with a given composition n_n , n_e , n_s , and n_w is

$$W(N, n_n, n_e, n_s, n_w) = \frac{N!}{n_n!n_e!n_s!n_w!}.$$

The probabilities that a pencil points in each of the four directions are

$$p(n) = \frac{n_n}{N}, \quad p(e) = \frac{n_e}{N}, \quad p(s) = \frac{n_s}{N}, \quad p(w) = \frac{n_w}{N}.$$

Figure 5.1 shows some possible distributions of outcomes. Each distribution function satisfies the constraint that $p(n)+p(e)+p(s)+p(w)=1$. You can compute the entropy per spin of the pencil of any of these distributions by using Equation (5.2) (with $k=1$), $S = -\sum_{i=1}^t p_i \ln p_i$. The absolute entropy is never negative, that is, $S \geq 0$. The perfectly ordered state is defined as having $S=0$.

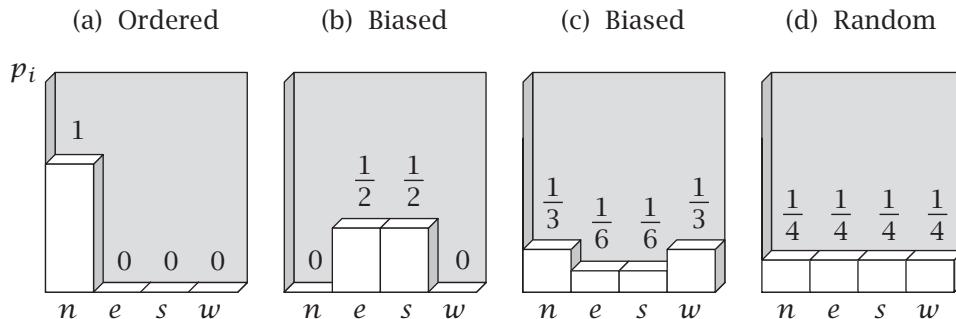


Figure 5.1 Spin a hundred pencils. Here are four (of a large number) of possible distributions of outcomes. (a) All pencils could point north (n). This is the most *ordered* distribution, $S = -1 \ln 1 = 0$. (b) Half the pencils could point east (e) and half could point south (s). This distribution has more entropy than (a), $S = -2[(1/2) \ln(1/2)] = 0.69$. (c) One-third of the pencils could point n , and one-third w , one-sixth e , and one-sixth s . This distribution has even more entropy, $S = -2[(1/3) \ln(1/3) + (1/6) \ln(1/6)] = 1.33$. (d) One-quarter of the pencils could point in each of the four possible directions. This is the distribution with highest entropy, $S = -4[(1/4) \ln(1/4)] = 1.39$.

Flat distributions have high entropy. Peaked distributions have low entropy. When all pencils point in the same direction, the system is perfectly ordered and has the lowest possible entropy, $S = 0$. Entropy does not depend on being able to order the categories along an x axis. For pencil orientations, there is no difference between the x -axis sequence *news* and *esnw*. To be in a state of low entropy, it does not matter which way the pencils point, just that they all point the same way. The flattest possible distribution has the highest possible entropy. In Figure 5.1 we have four states: the flattest distribution has $S = -4(1/4) \ln(1/4) = \ln 4 = 1.39$. Flatness in a distribution corresponds to *disorder* in a system. Also, the entropy of the flat distribution increases with the number t of states: $S = \ln t$.

The concept of entropy is broader than statistical thermodynamics. It is a property of any distribution function, as the next example shows.

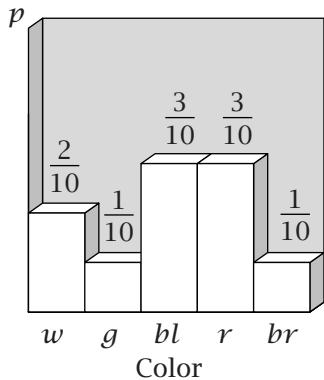


Figure 5.2 The entropy can be computed for *any* distribution function, even for colors of socks: white (*w*), green (*g*), black (*bl*), red (*r*), and brown (*br*).

EXAMPLE 5.2 Colors of socks. Suppose that, on a given day, you sample the distribution of the colors of the socks that 30 people are wearing. Figure 5.2 shows a possible outcome. The entropy of this distribution is

$$S = -0.2 \ln 0.2 - 0.6 \ln 0.3 - 0.2 \ln 0.1 = 1.50.$$

The entropy function just reports the relative flatness of a distribution function. The limiting cases are the most ordered, $S = 0$ (everybody wears the same color socks) and the most disordered, $S = \ln t = \ln 5 = 1.61$ (all five sock colors are equally likely).

Why should the entropy have the form of either Equation (5.1) or Equation (5.2)? Here is a simple justification. A deeper argument is given in Appendix E.

The Simple Justification for $S = k \ln W$

Consider a thermodynamic system having two subsystems, *A* and *B*, with multiplicities W_A and W_B , respectively. The multiplicity of the total system will be the product $W_{\text{total}} = W_A W_B$. Compatibility with thermodynamics requires that entropies be *extensive*, meaning that the system entropy is the sum of subsystem entropies: $S_{\text{total}} = S_A + S_B$. The logarithm function satisfies this requirement. If $S_A = k \ln W_A$ and $S_B = k \ln W_B$, then $S_{\text{total}} = k \ln W_{\text{total}} = k \ln W_A W_B = k \ln W_A + k \ln W_B = S_A + S_B$. This argument illustrates why S should be a logarithmic function of W .

Let's use the principle of maximization of $S/k = -\sum_i p_i \ln p_i$ to derive the exponential distribution law, called the Boltzmann distribution law, Equation (5.17), that is at the center of statistical thermodynamics. We'll see later that the Boltzmann distribution law describes the energy distributions of atoms and molecules.

Underdetermined Distributions

In the rest of this chapter, we illustrate the principles that we need by concocting a class of problems involving die rolls and coin flips instead of molecules.

How would you know if a die is biased? You could roll it N times and count the numbers of 1's, 2's, ..., 6's. If the probability distribution were perfectly flat, you should conclude the die is not biased. You could use the same test for the orientations of pencils or to determine whether atoms or molecules have biased spatial orientations or bond angle distributions. However, the options available to molecules are usually so numerous that you could not possibly measure each one. In statistical mechanics, you seldom have the luxury of knowing the full distribution, corresponding to all six numbers p_i for $i = 1, 2, 3, \dots, 6$ on die rolls.

Therefore, as a prelude to statistical mechanics, let's invent a dice problem that is underdetermined in the same way as the problems of molecular science. Suppose that you do not know the distribution of all six possible outcomes. Instead, you know only the total score (or, equivalently, the average score per roll) on the N rolls. In thousands of rolls, the average score per roll of an unbiased die will be $3.5 = (1+2+3+4+5+6)/6$. If you observe that the average score is 3.5, it is evidence (but not proof)¹ that the distribution is unbiased. In that case, your best guess consistent with the evidence is that the distribution is flat. All outcomes are equally likely.

However, if you observe the average score per roll is 2.0, then you must conclude that every outcome from 1 to 6 is *not* equally likely. You know only that low numbers are somehow favored. This one piece of data—the total score—is not sufficient to predict all six unknowns of the full distribution function. So we aim to do the next best thing. We aim to predict the least biased distribution function that is consistent with the known measured score. This distribution is predicted by the maximum entropy principle.

Maximum Entropy Predicts Flat Distributions When There Are No Constraints

You can apply the principle that systems tend to their states of maximum entropy to show that probabilistic systems will tend toward flat probability distributions when there are no constraints. Roll an unbiased t -sided die many times. The probabilities must sum to one. So, any changes, or differentials, must sum to zero. That is, if $p_1 + p_2 = 1$, then $dp_1 + dp_2 = 0$. More generally,

$$\sum_{i=1}^t p_i = 1 \implies \sum_{i=1}^t dp_i = 0. \quad (5.5)$$

We seek the distribution $(p_1, p_2, \dots, p_t) = (p_1^*, p_2^*, \dots, p_t^*)$ that causes the entropy function $S(p_1, p_2, \dots, p_t) = -k \sum_i p_i \ln p_i$ to be at its maximum possible value, subject to the normalization equation (5.5). For this problem, let $k = 1$. Let's solve it by the Lagrange multiplier method. Use multiplier α to impose the constraint given by Equation (5.5). Then solve either form of the Lagrange method, Equation (4.29),

$$\left(\frac{\partial S}{\partial p_i} \right)_{p_{j \neq i}} - \alpha = 0. \quad (5.6)$$

¹For example, that score could also arise from 50% 2's and 50% 5's.

or Equation (4.35),

$$\sum_{i=1}^t \left[\left(\frac{\partial S}{\partial p_i} \right)_{p_{j \neq i}} - \alpha \right] dp_i = 0. \quad (5.7)$$

Set the term inside the square brackets equal to zero for each i . Here's how you do this. First choose a particular i , say $i = 4$. Take the derivative of S with respect to that p , $\partial S / \partial p_4$, holding all the other p 's ($p_1, p_2, p_3, p_5, \dots$) constant. Taking this derivative gives $(\partial S / \partial p_4) = -1 - \ln p_4$ in this case, or $(\partial S / \partial p_i) = -1 - \ln p_i$ in general. So, the solution is

$$-1 - \ln p_i - \alpha = 0 \implies p_i^* = e^{(-1-\alpha)}. \quad (5.8)$$

To put this into a simpler form, divide Equation (5.8) by $\sum_i p_i^* = 1 = te^{(-1-\alpha)}$ to get

$$\frac{p_i^*}{\sum_i p_i^*} = \frac{e^{(-1-\alpha)}}{te^{(-1-\alpha)}} = \frac{1}{t}. \quad (5.9)$$

Maximizing the entropy predicts that when there is no bias, all outcomes are equally likely. However, what if there is bias? In the next section, we use the principle of maximum entropy again to show that the presence of constraints leads to an exponential, or *Boltzmann*, distribution.

Maximum Entropy Predicts Exponential Distributions When There Are Constraints

Roll a die having t sides, with faces numbered $i = 1, 2, 3, \dots, t$. You do not know the distribution of outcomes of each face, but you know the total score after N rolls. You want to predict the distribution function.

First, let's generalize our dice problem. Instead of having the numbers $i = 1, 2, \dots, 6$ painted on its six sides, the die has a more general set of numbers painted on its t sides. When side i appears face up, the score is ε_i . For now, you are completely free to choose those numbers ε_i any way you want. (Later, we'll get those numbers from physical models.) The total score after N rolls will be $E = \sum_{i=1}^t \varepsilon_i n_i$, where n_i is the number of times that you observe face i . Let $p_i = n_i/N$ represent the fraction of the N rolls on which you observe face i . The average score per roll $\langle \varepsilon \rangle$ is

$$\langle \varepsilon \rangle = \frac{E}{N} = \sum_{i=1}^t p_i \varepsilon_i. \quad (5.10)$$

What is the expected distribution of outcomes $(p_1^*, p_2^*, \dots, p_t^*)$ consistent with the observed average score $\langle \varepsilon \rangle$? We seek the distribution that maximizes the entropy, Equation (5.2), subject to two conditions: (1) that the probabilities sum to one and (2) that the average score agrees with the observed value $\langle \varepsilon \rangle$,

$$g(p_1, p_2, \dots, p_t) = \sum_{i=1}^t p_i = 1 \implies \sum_{i=1}^t dp_i = 0, \quad (5.11)$$

and

$$h(p_1, p_2, \dots, p_t) = \langle \varepsilon \rangle = \sum_{i=1}^t p_i \varepsilon_i \implies \sum_{i=1}^t \varepsilon_i dp_i = 0. \quad (5.12)$$

The solution is given by the method of Lagrange multipliers (see Chapter 4, pages 68–72):

$$\left(\frac{\partial S}{\partial p_i} \right) - \alpha \left(\frac{\partial g}{\partial p_i} \right) - \beta \left(\frac{\partial h}{\partial p_i} \right) = 0 \quad \text{for } i = 1, 2, \dots, t, \quad (5.13)$$

where α and β are the unknown multipliers. The partial derivatives are evaluated for each p_i :

$$\left(\frac{\partial S}{\partial p_i} \right) = -1 - \ln p_i, \quad \left(\frac{\partial g}{\partial p_i} \right) = 1, \quad \left(\frac{\partial h}{\partial p_i} \right) = \varepsilon_i. \quad (5.14)$$

Substitute Equations (5.14) into Equation (5.13) to get t equations of the form

$$-1 - \ln p_i^* - \alpha - \beta \varepsilon_i = 0, \quad (5.15)$$

where the p_i^* 's are the values of p_i that maximize the entropy. Solve Equations (5.15) for each p_i^* :

$$p_i^* = e^{-1-\alpha-\beta\varepsilon_i}. \quad (5.16)$$

To eliminate α in Equation (5.16), use Equation (5.11) to divide both sides by one. The result is an **exponential distribution law**:

$$p_i^* = \frac{p_i^*}{\sum_{i=1}^t p_i^*} = \frac{e^{(-1-\alpha)} e^{-\beta\varepsilon_i}}{\sum_{i=1}^t e^{(-1-\alpha)} e^{-\beta\varepsilon_i}} = \frac{e^{-\beta\varepsilon_i}}{\sum_{i=1}^t e^{-\beta\varepsilon_i}}. \quad (5.17)$$

In the examples below, we show how these equations can solve dice problems. They also solve problems in molecular science. In statistical mechanics, this is called the **Boltzmann distribution law** and the quantity in the denominator is called the **partition function q** :

$$q = \sum_{i=1}^t e^{-\beta\varepsilon_i}. \quad (5.18)$$

Using Equations (5.12) and (5.17) you can express the average score per roll $\langle \varepsilon \rangle$ (Equation (5.10)) in terms of the distribution:

$$\langle \varepsilon \rangle = \sum_{i=1}^t \varepsilon_i p_i^* = \frac{1}{q} \sum_{i=1}^t \varepsilon_i e^{-\beta\varepsilon_i}. \quad (5.19)$$

The next two examples show how Equation (5.19) predicts all t of the p_i^* 's from the one known quantity, the average score.

EXAMPLE 5.3 Finding bias in dice by using the exponential distribution law.

Here we illustrate how to predict the maximum entropy distribution when an average score is known. Suppose a die has $t = 6$ faces and the score on each face equals the face index number, $\varepsilon(i) = i$. You want to compute the probabilities of all six faces. You first need to find the Lagrange multiplier β . Let's change to a simpler variable, $x = e^{-\beta}$. Then Equation (5.18) gives $q = x + x^2 + x^3 + x^4 + x^5 + x^6$, and Equation (5.17) gives

$$p_i^* = \frac{x^i}{\sum_{i=1}^6 x^i} = \frac{x^i}{x + x^2 + x^3 + x^4 + x^5 + x^6}. \quad (5.20)$$

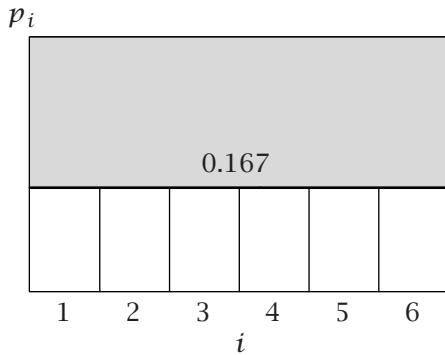
From the constraint equation (5.19), you have

$$\langle \varepsilon \rangle = \sum_{i=1}^6 i p_i^* = \frac{x + 2x^2 + 3x^3 + 4x^4 + 5x^5 + 6x^6}{x + x^2 + x^3 + x^4 + x^5 + x^6}. \quad (5.21)$$

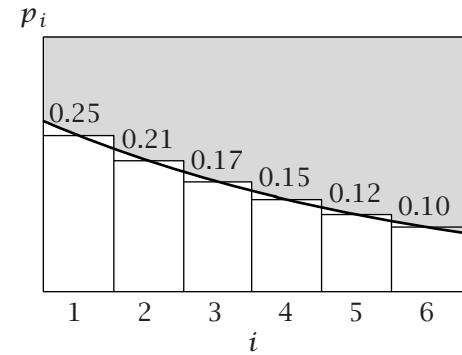
You have a polynomial, Equation (5.21), that you must solve for the one unknown x . You begin with knowledge of $\langle \varepsilon \rangle$. Compute the value x^* that solves Equation (5.21). Then substitute x^* into Equations (5.20) to give the distribution function $(p_1^*, p_2^*, \dots, p_t^*)$.

For example, if you observe the average score $\langle \varepsilon \rangle = 3.5$, then $x = 1$ satisfies Equation (5.21), predicting $p_i^* = 1/6$ for all i , indicating that the die is unbiased and has a flat distribution (see Figure 5.3(a)).

(a) $\langle \varepsilon \rangle = 3.5$



(b) $\langle \varepsilon \rangle = 3.0$



(c) $\langle \varepsilon \rangle = 4.0$

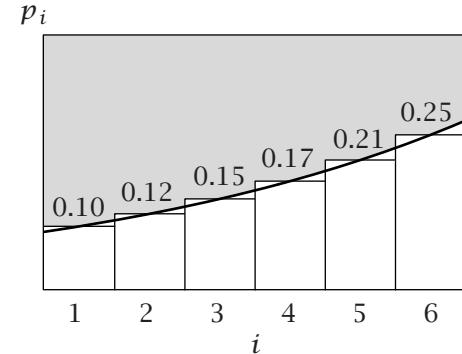


Figure 5.3 The probabilities of dice outcomes for known average scores. (a) If the average score per roll is $\langle \varepsilon \rangle = 3.5$, then $x = 1$ and all outcomes are equally probable, predicting that the die is low biased. (b) If the average score is low ($\langle \varepsilon \rangle = 3.0$, $x = 0.84$), maximum entropy predicts an exponentially diminishing distribution. (c) If the average score is high ($\langle \varepsilon \rangle = 4.0$, $x = 1.19$), maximum entropy implies an exponentially increasing distribution.

If, instead, you observe the average score $\langle \varepsilon \rangle = 3.0$, then $x = 0.84$ satisfies Equation (5.21), and you have $q = 0.84 + 0.84^2 + 0.84^3 + 0.84^4 + 0.84^5 + 0.84^6 = 3.41$. The probabilities are $p_1 = 0.84/3.41 = 0.25$, $p_2 = 0.84^2/3.41 = 0.21$, $p_3 = 0.84^3/3.41 = 0.17$, and so on, as shown in Figure 5.3(b).

If you observe $\langle \varepsilon \rangle < 3.5$, then the maximum entropy principle predicts an exponentially decreasing distribution (see Figure 5.3(b)), with more 1's than 2's, more 2's than 3's, etc. If you observe $\langle \varepsilon \rangle > 3.5$, then maximum entropy predicts an exponentially increasing distribution (see Figure 5.3(c)): more 6's than 5's, more 5's than 4's, etc. For any given value of $\langle \varepsilon \rangle$, an exponential or a flat distribution gives the most impartial distribution consistent with that score. The flat distribution is predicted if the average score is 3.5, or if you have no information at all about the score.

EXAMPLE 5.4 Biased coins? The exponential distribution again. Let's determine a coin's bias. A coin is just a die with two sides, $t = 2$. Score tails $\varepsilon_T = 1$ and heads $\varepsilon_H = 2$. The average score per toss $\langle \varepsilon \rangle$ for an unbiased coin would be 1.5.

Again, to simplify, write the unknown Lagrange multiplier β in the form $x = e^{-\beta}$. In this notation, the partition function, Equation (5.18), is $q = x + x^2$. According to Equation (5.17), the exponential distribution law for this two-state system is

$$p_T^* = \frac{x}{x+x^2} \quad \text{and} \quad p_H^* = \frac{x^2}{x+x^2}. \quad (5.22)$$

From the constraint equation (5.19), you have

$$\langle \varepsilon \rangle = 1p_T^* + 2p_H^* = \frac{x+2x^2}{x+x^2} = \frac{1+2x}{1+x}.$$

Rearranging gives

$$x = \frac{\langle \varepsilon \rangle - 1}{2 - \langle \varepsilon \rangle}.$$

If you observe the average score to be $\langle \varepsilon \rangle = 1.5$, then $x = 1$, and Equation (5.22) gives $p_T^* = p_H^* = 1/2$. The coin is fair. If instead you observe $\langle \varepsilon \rangle = 1.2$, then $x = 1/4$, and you have $p_H^* = 1/5$ and $p_T^* = 4/5$.

There are two situations that will predict a flat distribution function. First, the distribution will be flat if $\langle \varepsilon \rangle$ equals the value expected from a uniform distribution. For example, if you observe $\langle \varepsilon \rangle = 3.5$ in Example 5.3, maximum entropy predicts a flat distribution. Second, if there is no constraint at all, you expect a flat distribution. By the maximum entropy principle, having no information is the same as expecting a flat distribution.

What Is the Reason for Maximizing the Entropy?

Why should you maximize the multiplicity or the entropy in order to predict probabilities? You might call it a Principle of Fair Apportionment. Every *microstate*—each particular microscopic sequence of outcomes—is as likely to appear as any other. Because any given *macrostate* that you observe will be made up of a different number of microstates, the probability of observing that macrostate is proportional to its number of microstates. So, in modeling a system, you should ‘treat each outcome fairly’ in comparison with each other. If no value of $\langle \varepsilon \rangle$ is specified, then the maximum entropy approach makes the fair-apportionment prediction that all microscopic sequences are equally likely. However, if some value of $\langle \varepsilon \rangle$ is given, then the maximum entropy approach says that the system should be apportioned as fairly as possible while satisfying that constraint. Throughout a long history, the idea that every outcome is equivalent has gone by various names. In the 1700s, Bernoulli called it the Principle of Insufficient Reason; in the 1920s, Keynes called it the Principle of Indifference [1]. For a discussion in greater depth, see Appendix E.

Summary

The entropy $S(p_{11}, \dots, p_{ij}, \dots, p_{ab})$ is a function of a set of probabilities. You are most likely to observe the probabilities p_{ij}^* that cause S to be maximal and that otherwise satisfy any given average, such as $\langle \varepsilon \rangle$. This distribution is typically exponential. In Chapter 10, this exponential function will define the Boltzmann distribution law. With this law, you can predict thermodynamic and physical properties of atoms and molecules, and their averages and fluctuations. However, first we need the machinery of thermodynamics, the subject of the next three chapters.

Problems

1. Calculating the entropy of dipoles in a field. You have a solution of dipolar molecules with a positive charge at the head and a negative charge at the tail. When there is no electric field applied to the solution, the dipoles point north (n), east (e), west (w), or south (s) with equal probabilities. The probability distribution is shown in Figure 5.4(a). However when you apply a field to the solution, you now observe a different distribution, with more heads pointing north, as shown in Figure 5.4(b).

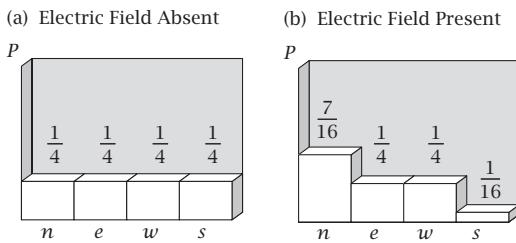


Figure 5.4

- (a) What is the polarity of the applied field? (In which direction does the field have its most positive pole?)
- (b) Calculate the entropy of the system in the absence of the field.
- (c) Calculate the entropy of the system in the applied field.
- (d) Does the system become more ordered or disordered when the field is applied?

2. Comparing the entropy of peaked and flat distributions. Compute the entropies for the spatial concentration shown in Figure 5.5.

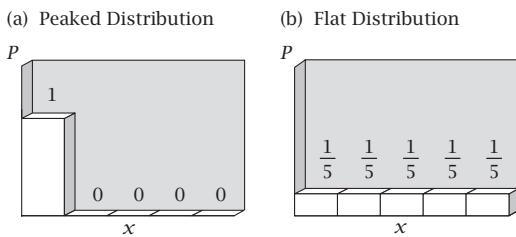


Figure 5.5

3. Calculating the entropy of mixing. Consider a lattice with N sites and n green particles. Consider another lattice, adjacent to the first, with M sites and m red particles. Assume that the green and red particles cannot switch lattices. This is state A .

- (a) What is the total number of configurations W_A of the system in state A ?
- (b) Now assume that all $N+M$ sites are available to all the green and red particles. The particles remain distinguishable by their color. This is state B . Now

what is the total number of configurations W_B of the system?

Now take $N = M$ and $n = m$ for the following two problems.

- (c) Using Stirling's approximation (see Appendix B), what is the ratio W_A/W_B ?
- (d) Which state, A or B , has the greatest entropy? Calculate the entropy difference given by

$$\Delta S = S_A - S_B = k \ln \left(\frac{W_A}{W_B} \right).$$

4. Comparing the entropy of two peaked distributions. Which of the two distributions shown in Figure 5.6 has the greater entropy?

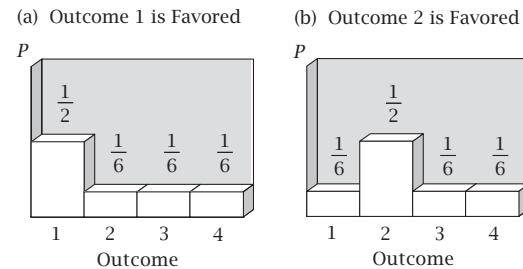


Figure 5.6

5. Proof of maximum entropy for two outcomes. Example 5.4 is simple enough that you can prove the answer is correct even without using the maximum entropy principle. Show this.

6. The maximum entropy distribution is Gaussian when the second moment is given. Prove that the probability distribution p_i that maximizes the entropy for die rolls, subject to a constant value of the second moment $\langle i^2 \rangle$, is a Gaussian function. Use $\varepsilon_i = i$.

7. Maximum entropy for a three-sided die. You have a three-sided die, with numbers 1, 2, and 3 on the sides. For a series of N rolls, you observe an average score, or constraint value, of α per roll.

- (a) Using the maximum entropy principle, write expressions that show how to compute the relative probabilities of occurrence of the three sides, n_1^*/N , n_2^*/N , and n_3^*/N , if α is given.
- (b) Compute n_1^*/N , n_2^*/N , and n_3^*/N , if $\alpha = 2$.
- (c) Compute n_1^*/N , n_2^*/N , and n_3^*/N , if $\alpha = 1.5$.
- (d) Compute n_1^*/N , n_2^*/N , and n_3^*/N , if $\alpha = 2.5$.

8. Maximum entropy in Las Vegas. You play a slot machine in Las Vegas. For every \$1 coin you insert, there are three outcomes: (1) you lose \$1; (2) you win \$1, so your profit is \$0; (3) you win \$5, so your profit is \$4. Suppose you believe that your average expected profit over many trials is \$0 (what an optimist!). Find the maximum entropy distribution for the probabilities p_1 , p_2 , and p_3 of observing outcomes (1), (2), and (3), respectively.

9. Flat distribution, high entropy. Consider the following possible distributions you might observe in $N = 4$ flips of a coin: $(p_H, p_T) = (0, 1), (1/4, 3/4), (1/2, 1/2), (3/4, 1/4), (1, 0)$. Compute W for each case, and show that the flattest distribution has the highest multiplicity.

Reference

- [1] ET Jaynes, Where do we stand on maximum entropy? *The Maximum Entropy Formalism*, RD Levine and M Tribus, eds, MIT Press, Cambridge, 1979, pp. 15-118.

Suggested Reading

ET Jaynes, *Phys Rev* **106**, 620-630 (1957). This is the classic article that introduces the maximum entropy approach to statistical thermodynamics. It is well written and includes a scholarly historical introduction.

JN Kapur and HK Kesavan, *Entropy Optimization Principles with Applications*, Academic Press, Boston, 1992. Excellent overview of maximum entropy applied to many fields of science.

RD Levine and M Tribus, eds, *The Maximum Entropy Formalism*, MIT Press, 1979. This is an advanced edited volume that discusses applications and other aspects of maximum entropy, including an extensive review by Jaynes (the reference cited above).

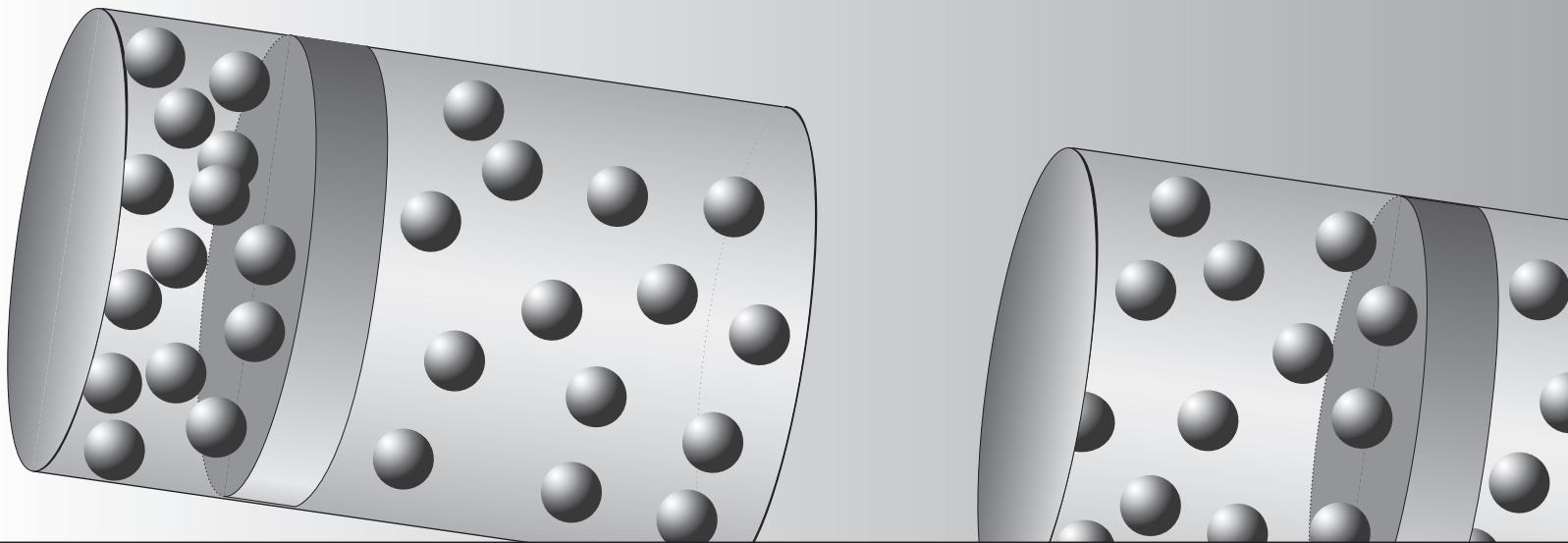
The following three articles show how the maximum entropy principle follows from the multiplication rule. A brief overview is given in J Skilling, *Nature* **309**, 748-749 (1984).

AK Livesay and J Skilling, *Acta Cryst A* **41**, 113-122 (1985).

JE Shore and RW Johnson, *IEEE Trans Inform Theory* **26**, 26-37 (1980).

Y Tikochinski, NZ Tishby and RD Levine, *Phys Rev Lett* **52**, 1357-1360 (1984).

6 Thermodynamic Driving Forces



Thermodynamics Is Two Laws and a Little Calculus

Thermodynamics is a set of tools for reasoning about energies and entropies. It enables you to predict the tendencies of atoms and molecules to react; to bind, absorb, or partition; to dissolve or change phase; and to change their shapes or chemical bonds. The three basic tools of thermodynamics are the First Law for the conservation of energy (see Chapter 3), the maximum entropy principle, also called the Second Law of Thermodynamics (Chapters 2 and 5), and multivariate calculus (Chapter 4). In this chapter, we explore the definitions of the forces that act in material systems: pressure, the tendency to exchange volume; temperature, the tendency to exchange energy; and chemical potential, the tendency to exchange matter or to change its form.

What Is a Thermodynamic System?

A thermodynamic system is a collection of matter in any form, delineated from its surroundings by real or imaginary boundaries. The system may be a biological cell, the contents of a test tube, the gas in a piston, a thin film, or a can of soup. How you define a system depends on the problem that you want to solve. In thermodynamics, defining the boundaries is important, because boundaries determine what goes in or out. Much of thermodynamics is bookkeeping, accounting for energy or matter exchange across the boundaries or

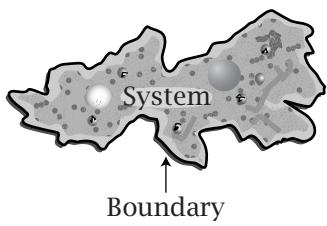


Figure 6.1 An amoeba is an open system.

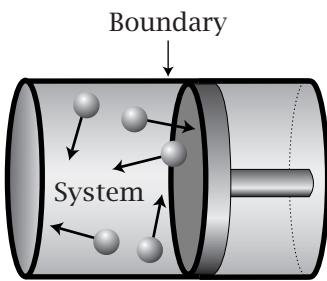


Figure 6.2 A piston can be a closed system.

for volume changes. For a biochemical experiment, the boundary might be the walls of a test tube or a cell membrane. For a steam engine, the interior of the piston might be the most convenient definition of the system. The boundary need not be fixed in space. It need not even surround a fixed number of molecules. The system might contain subsystems, and the subsystems might also be delineated by real or imaginary boundaries. Systems are defined by the nature of their boundaries:

OPEN SYSTEM. An open system can exchange energy, volume, and matter with its surroundings. The Earth and living organisms are open systems (see Figure 6.1).

CLOSED SYSTEM. Energy can cross the boundary of a closed system, but matter cannot. The boundaries can be stationary, as in a light bulb or the walls of a house that loses heat, or movable, as in a balloon or a piston (see Figure 6.2).

ISOLATED SYSTEM. Energy and matter cannot cross the boundaries of an isolated system. Also, the volume does not change: the boundaries do not move. The total internal energy of an isolated system is constant. In reality, there is never complete isolation. This kind of idealization is common in thermodynamics.

SEMPERMEABLE MEMBRANE. A semipermeable membrane is a boundary that restricts the flow of **some kinds of particle**, while allowing **others to cross**. Dialysis is performed with semipermeable membranes. Biological membranes are semipermeable.

ADIABATIC BOUNDARY. An adiabatic boundary prevents heat from flowing between a system and its surroundings. Thermos bottles have adiabatic boundaries, more or less. Adiabatic walls permit **you to measure work by** prohibiting heat flow.

PHASE. A phase is a homogeneous part of a system that is mechanically separable from the rest of the system. Homogeneous means that the pressure, temperature, and concentration are uniform, or continuous functions of position within a phase. In a simple system, a phase can be solid, liquid, gaseous, liquid crystalline, or other.

SIMPLE SYSTEM. A simple system is defined as homogeneous, with only a single phase. It must be uncharged, large enough so that **surface effects can be neglected**, and **not subject to changes due to electric, magnetic, or gravitational fields**.

These definitions, like much of thermodynamics, involve idealizations and approximations to reality. In this chapter and in Chapters 7 and 8, we will focus on simple systems.

Some Properties Are Extensive

Some properties of matter are *extensive*. An extensive property **P** of a system is the **sum of the properties $P_1 + P_2 + P_3 + \dots$** of its component subsystems.

Said differently, an extensive property increases with system size. For example, the volume of a glass of water is the sum of the volumes of two half glasses of water. Alternatively, properties can be *intensive*, or independent of the size of a system. Temperature is intensive: the temperature of a glass of water is not the sum of the temperatures of two half glasses. Pressures and concentrations are also intensive quantities. Pouring two half glasses of water together does not double the pressure; it doubles the volume. In thermodynamics, extensive properties have a special importance, **and each extensive property is associated with a companion intensive property**, as we'll see below. Extensive properties include the following:

SPATIAL EXTENT. A system has volume, area, or length determined by the positions of its boundaries. We will usually consider the volume V , but with films or surfaces you may be interested in the area. When you stretch rubber bands, you may be interested in the length.

NUMBER OF PARTICLES. A system is composed of some number of particles such as atoms, molecules, or other microscopic units. For a gas in a piston, a chemical reaction, or ice in water, a natural microscopic unit is the atom or molecule. The particles of a system can belong to different species. For example, the gases in a piston can include O_2 , N_2 , CO, etc. You can have N_i molecules of species i , that is, N_1 molecules of species 1, N_2 molecules of species 2, and so on. Rather than the number of molecules, we sometimes specify the numbers of moles, $n_i = N_i/\mathcal{N}$, where $\mathcal{N} = 6.022045 \times 10^{23}$ molecules per mole is Avogadro's number. We use bold type to denote the set of all species: $\mathbf{n} = n_1, n_2, \dots, n_M$ or $\mathbf{N} = N_1, N_2, \dots, N_M$, where M is the number of species.

INTERNAL ENERGY. The internal energy U of a system is extensive. For a system of independent non-interacting particles, the internal energy is the sum of the particle energies, according to Equation (3.12):

$$U = \sum_{i=1}^t N_i \varepsilon_i,$$

where t is the number of energy levels, N_i is the number of particles in energy level i , and ε_i is the energy at level i .

ENTROPY. **The entropy of a system is the sum of the entropies of its independent subsystems** (see Chapter 5).

The Fundamental Thermodynamic Equations, $S = S(U, V, N)$ or $U = U(S, V, N)$, Predict Equilibria

In Chapters 2 and 3, we used simple models to illustrate how the tendency toward maximum multiplicity W (and maximum entropy S) can account for some tendencies of physical systems. When the volume can change, the tendency toward maximum entropy $S(V)$ predicts the expansion of gases. When particle numbers can change, maximizing $S(N)$ predicts changes in composition. And when the energy can exchange, maximizing $S(U)$ predicts the tendency for heat to flow. However, many problems of interest involve multiple

degrees of freedom. With the tools of multivariate calculus, you can predict more complex processes. If the energy, volume, and particle number are all free to change, then the **fundamental thermodynamic equation for entropy** is multivariate:

$$S = S(U, V, \mathbf{N}).$$

$S = S(U, V, \mathbf{N})$ is just a statement that we are taking $U, V, N_1, N_2, \dots, N_M$ to be independent of each other, and that the entropy is dependent on them.

It is an inconvenient quirk of history (for us) that thermodynamics evolved instead with the arrangement of variables known as the **fundamental thermodynamic equation for energy**:

$$U = U(S, V, \mathbf{N}).$$

The fundamental definitions of pressure, chemical potential, and temperature are based on the form $U = U(S, V, \mathbf{N})$. Unfortunately, while the main definitions of thermodynamics originate from the energy equation, the microscopic driving forces are better understood in terms of the entropy equation $S = S(U, V, \mathbf{N})$. So we need to switch back and forth between $S = S(U, V, \mathbf{N})$ and $U = U(S, V, \mathbf{N})$. For example, in Chapter 3, we illustrated how $S(U)$ accounts for the tendency for heat to flow, but temperature is defined in terms of $U(S)$. This small inconvenience adds a few steps to our logic.

Either equation, $S = S(U, V, \mathbf{N})$ or $U = U(S, V, \mathbf{N})$, will completely specify the state of a simple system. Thermodynamics does not tell you the specific mathematical dependence of S on (U, V, \mathbf{N}) or U on (S, V, \mathbf{N}) . *Equations of state*, such as the ideal gas law, specify interrelations among these variables. Equations of state must come either from experiments or from microscopic models. Applications of the laws of thermodynamics to those equations give additional useful relationships.

The Fundamental Equations Define the Thermodynamic Driving Forces

The Definitions

Start with $S(U, V, \mathbf{N})$, the mathematical statement that the entropy is a function of independent variables. Taking the derivative shows how variation in S depends on variations in the independent variables:

$$dS = \left(\frac{\partial S}{\partial U} \right)_{V, \mathbf{N}} dU + \left(\frac{\partial S}{\partial V} \right)_{U, \mathbf{N}} dV + \sum_{j=1}^M \left(\frac{\partial S}{\partial N_j} \right)_{U, V, N_{i \neq j}} dN_j. \quad (6.1)$$

When you use the fundamental energy equation instead, you can describe small changes in energy $U(S, V, \mathbf{N})$ as

$$dU = \left(\frac{\partial U}{\partial S} \right)_{V, \mathbf{N}} dS + \left(\frac{\partial U}{\partial V} \right)_{S, \mathbf{N}} dV + \sum_{j=1}^M \left(\frac{\partial U}{\partial N_j} \right)_{S, V, N_{i \neq j}} dN_j. \quad (6.2)$$

So far, Equations (6.1) and (6.2) are just mathematical definitions based on which variables are dependent and which are independent. However, each of the

partial derivatives in the fundamental energy equation (6.2) also corresponds to a measurable physical property. For now, take the following to be mathematical definitions: the *temperature* T , the *pressure* p , and the *chemical potential* μ_j are *defined by* the partial derivatives given in Equation (6.2):

$$T = \left(\frac{\partial U}{\partial S} \right)_{V,N}, \quad p = - \left(\frac{\partial U}{\partial V} \right)_{S,N}, \quad \mu_j = \left(\frac{\partial U}{\partial N_j} \right)_{S,V,N_{i \neq j}}. \quad (6.3)$$

We will show, starting on page 99, that these mathematical definitions describe the familiar physical properties temperature, pressure, and chemical potential. According to Equation (6.3), the quantities T , p , and μ are partial derivatives of the internal energy. In total derivative expressions, each partial derivative term ($\partial U / \partial x$) is described as being *conjugate* to x . Temperature is conjugate to entropy, pressure is conjugate to volume, and chemical potential is conjugate to number of particles. Intensive quantities, such as T , p , and μ , are conjugate to extensive quantities U , S , V , and N .

Substituting the definitions given in Equation (6.3) into Equation (6.2) gives the **differential form of the fundamental energy equation**:

$$dU = T dS - p dV + \sum_{j=1}^M \mu_j dN_j. \quad (6.4)$$

Alternatively, you can rearrange Equation (6.4) to get a useful **differential form of the fundamental entropy equation**:

 $ds = \left(\frac{1}{T} \right) dU + \left(\frac{p}{T} \right) dV - \sum_{j=1}^M \left(\frac{\mu_j}{T} \right) dN_j. \quad (6.5)$

Comparison of Equations (6.5) and (6.1) gives three more definitions:

 $\frac{1}{T} = \left(\frac{\partial S}{\partial U} \right)_{V,N}, \quad \frac{p}{T} = \left(\frac{\partial S}{\partial V} \right)_{U,N}, \quad \frac{\mu_j}{T} = - \left(\frac{\partial S}{\partial N_j} \right)_{U,V,N_{i \neq j}}. \quad (6.6)$

Equations (6.4) and (6.5) are completely general statements that there are some functional dependences $S(U, V, N)$ and $U(S, V, N)$, and that these dependences define T , p , and μ . Equations (6.4) and (6.5) are *fundamental* because they completely specify all the changes that can occur in a simple thermodynamic system, and they are the bases for extremum principles that predict states of equilibria. An important difference between these fundamental equations and others that we will write later is that S and U are functions of only the extensive variables. Beginning on page 99, we will show how to use Equations (6.4) and (6.5) to identify states of equilibrium.

For learning thermodynamic reasoning, one of the simplest systems is a state of matter called the *ideal gas*.

The Ideal Gas Law Describes an Important Idealized Form of Matter

Studies of gases expanding against pistons in cylinders have played a central role in thermodynamics, and are important for understanding energy conversion in steam engines and internal combustion engines. Empirical observations

by R Boyle in 1662 and later studies by JL Gay-Lussac, J Charles, and J Dalton, were combined by E Clapeyron in 1834 into the **ideal gas law**, a relationship between pressure, volume, and temperature that is obeyed by all gases that are at **sufficiently low densities**,

$$pV = NkT \quad (6.7)$$

where p is the gas pressure, V is the volume of the container, N is the number of gas molecules, T is the temperature, and k is *Boltzmann's constant*. This law says that, at constant temperature, reducing pressure leads to gas expansion, or that, at constant pressure, heating the gas will increase its volume.

Sometimes **thermodynamic reasoning enters a problem only by providing a definition**. The law for ideal gas behavior, Equation (6.7), can be derived from just two ingredients: the definition of pressure given in Equation (6.6) and the simple lattice model of a gas from Example 2.2.

EXAMPLE 6.1 The ideal gas law derived from the lattice model. The lattice model in Example 2.2 shows that gases exert pressure in order to increase their multiplicity of states. For a lattice of M sites containing N particles, Equation (2.3) gives the multiplicity W and the entropy S as

$$\frac{S}{k} = \ln W(N, M) = \ln \left[\frac{M!}{N!(M-N)!} \right],$$

where $S = k \ln W$ comes from the Boltzmann expression. Now, thermodynamics gives you a way to compute the pressure if you know how the entropy depends on the volume, $S(V)$: $p/T = (\partial S / \partial V)_{U, N}$ (Equation (6.6)).

In our model, M is proportional to the volume V . That is, $V/M = v_0$, the volume per lattice site, which is a constant. You can get $S(V)$ from $S(M)$ using

$$\left(\frac{\partial S}{\partial V} \right)_N = \left(\frac{\partial S}{\partial M} \right)_N \left(\frac{dM}{dV} \right) = \left(\frac{\partial S}{\partial M} \right)_N \left(\frac{1}{v_0} \right). \quad (6.8)$$

To get $S(M)$, apply Stirling's approximation (Appendix B) to Equation (2.3):

$$\begin{aligned} \frac{S}{k} &\approx \ln \left[\frac{M^M}{N^N (M-N)^{M-N}} \right] \\ &= M \ln M - N \ln N - (M-N) \ln(M-N). \end{aligned} \quad (6.9)$$

Rewrite the first term on the right-hand side as a sum of two parts, $M \ln M = N \ln M + (M-N) \ln M$, and incorporate these parts into the second and third terms in Equation (6.9) to get

$$\frac{S}{k} = -N \ln \left(\frac{N}{M} \right) - (M-N) \ln \left(\frac{M-N}{M} \right). \quad (6.10)$$

Take the derivative of S with respect to M in Equation (6.9) to get

$$\begin{aligned} \left(\frac{\partial S}{\partial M} \right)_N &= k \left[1 + \ln M - \ln(M-N) - \left(\frac{M-N}{M} \right) \right] \\ &= -k \ln \left(1 - \frac{N}{M} \right). \end{aligned} \quad (6.11)$$

Equation (6.11) can be expressed by using the series expansion (J.4) of the logarithm from Appendix J: $\ln(1-x) \approx -x - x^2/2 + \dots$. If the density of molecules is small, $N/M \ll 1$, you need only the lowest-order terms. Substituting Equations (6.11) and (6.8) into the definition of pressure, and using $V/M = v_0$ gives

$$p = -kT \left(\frac{M}{V} \right) \ln \left(1 - \frac{N}{M} \right)$$

$$\approx \left(-\frac{MkT}{V} \right) \left(-\frac{N}{M} \right) \left[1 + \frac{1}{2} \left(\frac{N}{M} \right) + \dots \right] \approx \frac{NkT}{V}, \quad (6.12)$$

which is the ideal gas law.

This shows how to derive the ideal gas law using a model for how S depends on V and the thermodynamic definition of pressure. We will show in Chapter 24 that keeping the next higher-order term in the expansion gives a refinement toward the van der Waals gas law.

$1/T$, p/T , and μ/T Behave as Forces

Now we'll show that T , p , and μ , mathematically defined by Equation (6.3), coincide with the physically measurable quantities temperature, pressure, and chemical potential. We will do this in two steps. First we will show that T , p , and μ are *intensive* quantities, not dependent on system size. Then we will show that $1/T$, p/T , and μ/T describe forces. $1/T$ describes a tendency for heat flow, p/T represents a tendency for volume change, and μ/T represents a tendency for particle exchange.

T , p , and μ are intensive because $U(S, V, N)$ is a *first-order homogeneous function*. For some mathematical functions $f(x)$, if you increase the argument by a factor of λ to λx , this will cause the function to increase to $\lambda f(x)$. That is, $f(\lambda x) = \lambda f(x)$. These are called *first-order homogeneous functions*. The fundamental energy and entropy equations are examples. The fundamental energy equation has the form $U = U(x_1, x_2, x_3)$, where the x_i 's (S, V, N) are all extensive quantities. Also, U is extensive. So, if the system size is scaled up by a factor of λ , then each x_i will increase to λx_i and the energy will also increase from U to λU . The derivatives of such functions do not increase with size λ . Mathematically, $df(\lambda x)/d(\lambda x) = df(x)/dx$, because $df(\lambda x)/d(\lambda x) = [df(\lambda x)/dx][dx/d(\lambda x)] = (\lambda df/dx)(1/\lambda)$. Since T , p , and μ are defined as partial derivatives $(\partial U/\partial x_i)$ (see Equation (6.3)), they obey this relationship $(\partial(\lambda U)/\partial(\lambda x_i)) = (\partial U/\partial x_i)$, so T , p , and μ are intensive quantities that do not depend on the system size.

The Second Law Predicts Thermal, Mechanical, and Chemical Equilibria

Now, we introduce the Second Law of Thermodynamics. The Second Law is the principle that isolated systems tend toward their states of maximum entropy. We show how this predicts states of equilibrium. In the next four examples, we show that $1/T$ is a measure of a system's tendency for heat exchange, p/T is

a measure of a system's tendency for **volume change**, and μ/T is a measure of a system's **tendency for particle exchange**. We follow the same **strategy in each case**. We determine the relevant independent variables and apply the fundamental entropy equation, and the constraints. Then the maximum entropy principle ($dS_{\text{total}} = 0$) specifies the state of equilibrium.

EXAMPLE 6.2 Temperature describes the tendency for energy exchange. *Temperature* is the quantity that tells you when **heat exchange will occur**. Objects exchange heat to reach a state of maximum entropy (equilibrium), and in this state their temperatures are equal. Suppose two objects are brought into thermal contact. With *each other*, they can exchange energy but not volume or particles. Both objects are otherwise isolated from their surroundings. So, with the *surroundings*, the two objects cannot exchange energy, particles, or volume. We are interested in how the entropy of each object *A* and *B* depends on its energy: $S_A(U_A)$ and $S_B(U_B)$. Object *A* has energy U_A , entropy S_A , and temperature T_A , where $1/T_A = (\partial S_A / \partial U_A)$. Object *B* has energy U_B , entropy S_B , and $1/T_B = (\partial S_B / \partial U_B)$. Although entropy is not experimentally measurable, temperature is, so the temperature is a convenient surrogate quantity for determining the closeness to equilibrium. What temperatures will the two objects have at equilibrium? What state maximizes the entropy?

Entropy is an extensive property, so

$$S_{\text{total}} = S_A + S_B. \quad (6.13)$$

Equation (6.13) does not mean that the total entropy of the system remains fixed during the thermal exchange process. The total entropy can increase as a system approaches equilibrium. Equation (6.13) says only that the total entropy of the system is the sum of the entropies of the subsystems.

When the two subsystems are brought into contact, they can exchange energies U_A and U_B with each other (see Figure 6.3). U_A and U_B are the degrees of freedom and the entropy depends on them: $S_{\text{total}}(U_A, U_B)$. However, the system cannot exchange energy with the surroundings, so the total energy $U_A + U_B$ is constrained:

$$U_A + U_B = U_{\text{total}} = \text{constant}. \quad (6.14)$$

To find the state of thermal equilibrium, determine what variations in U_A and U_B will cause the total entropy of the system to be maximal: $dS = 0$.

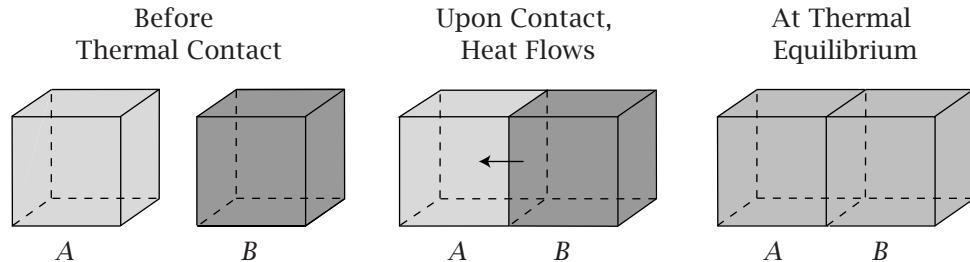


Figure 6.3 When objects *A* and *B* are brought into thermal contact, heat q (shown here as shading) can flow from one to the other. If there is no additional exchange with the surroundings, the process will conserve energy (see Equation (6.14)).

Because the entropy is extensive, the total entropy is the sum of entropies of the two objects: $S = S_A + S_B$. Write the fundamental entropy equation for the total differential dS_{total} in terms of the degrees of freedom, and set it equal to zero:

$$dS_{\text{total}} = dS_A + dS_B = \left(\frac{\partial S_A}{\partial U_A} \right)_{V,N} dU_A + \left(\frac{\partial S_B}{\partial U_B} \right)_{V,N} dU_B = 0. \quad (6.15)$$

The differential form of the constraint equation (6.14) is

$$dU_A + dU_B = dU_{\text{total}} = 0 \implies dU_A = -dU_B. \quad (6.16)$$

Substitute $-dU_A$ for dU_B in Equation (6.15) and rearrange:

$$\begin{aligned} dS_{\text{total}} &= \left[\left(\frac{\partial S_A}{\partial U_A} \right)_{V,N} - \left(\frac{\partial S_B}{\partial U_B} \right)_{V,N} \right] dU_A = 0 \\ \implies \left(\frac{\partial S_A}{\partial U_A} \right)_{V,N} &= \left(\frac{\partial S_B}{\partial U_B} \right)_{V,N}. \end{aligned} \quad (6.17)$$

Substituting the definition $1/T = (\partial S / \partial U)_{V,N}$ into Equation (6.17) gives the equilibrium condition

$$\frac{1}{T_A} = \frac{1}{T_B} \implies T_A = T_B. \quad (6.18)$$

When two objects at different temperatures are brought into thermal contact, they exchange heat. They exchange heat because it leads to maximization of the entropy of the combined system ($dS_{\text{total}} = 0$). This results in equalization of the temperatures: $T_A = T_B$. The quantity $1/T$ describes a tendency to transfer energy. When the temperatures are equal, there is no heat exchange.

Which way does the energy flow to reach equilibrium? Look at the signs on the differentials. Substituting $1/T = (\partial S / \partial U)_{V,N}$ into Equation (6.17) leads to

$$dS_{\text{total}} = \left(\frac{1}{T_A} - \frac{1}{T_B} \right) dU_A. \quad (6.19)$$

A change toward equilibrium increases the entropy: $dS_{\text{total}} \geq 0$. For dS_{total} to be positive, the signs on both factors dU_A and $[(1/T_A) - (1/T_B)]$ must be the same. If dU_A is negative, it means that object A loses energy to object B. So, that change will move the system toward equilibrium, $dS_{\text{total}} > 0$, if T_A is greater than T_B . Thus, the transfer of energy from object A to object B happens if A is the hotter object. The prediction that systems tend toward states of maximum entropy explains why heat flows from hot to cold objects.

What Is Temperature? A First Look

Example 6.2 illustrates the relationships among **energy, entropy, temperature, and heat**. A system that has energy has the capacity to do work. That capacity can flow from one place to another as heat. The entropy is a sort of **potential** or capability to move energy from one place to another, and $1/T$ represents the corresponding driving force. The higher the temperature of an object, the greater the tendency for energy **to escape from that object**.

Is the flow of heat due to a tendency to equalize energies? No. Heat flows to maximize entropy. Temperatures are equalized, not energies. The First Law describes energy bookkeeping of heat plus work. It does not tell us why, or when, or how much energy will be exchanged. For that we need the Second Law of Thermodynamics, which is the principle that systems tend to their states of maximum entropy.

Consider money as an analog to energy. Money represents a capacity to produce things, just as energy represents the capacity for work. Money can flow between people, like energy between objects. When person *A* pays money to person *B*, the sum of their combined dollars is unchanged, just as when the two bodies exchange energy in Example 6.2. This conservation of dollars corresponds to the First Law, the conservation of energy.

However, the drive for energy to flow from hot objects to cold objects is not a tendency to equalize the distribution of energy, just as the rich do not pay the poor to equalize their wealth. This is where entropy and temperature come in. Energy flows only if it leads to an increase in the entropy of the whole system.

What is the analog of entropy in the money example? Suppose person *A* has money and wants widgets and person *B* has widgets to sell. Person *A* will benefit in some way by *paying* money to receive a widget. On the other hand, person *B* will benefit by *receiving* money for a widget. Money will be exchanged because both parties benefit. Economists call this ‘maximizing a utility function.’ Money flows if it leads to an increase in the overall utility function for both parties, just as energy flows to maximize the entropy.

What is temperature in this analogy? Based on its definition, $1/T = (\partial S/\partial U)$, $1/T$ represents the incremental benefit to person *A* or *B* of getting a dollar or a widget. When these incremental benefits are equal, there is no further net redistribution of money and widgets. This analogy emphasizes that energy flows if it can increase another quantity, the entropy. When entropy can be increased no further by net changes in a degree of freedom, the system is in equilibrium. However, do not take this analogy too seriously. Money does not always flow to everyone’s maximum benefit.

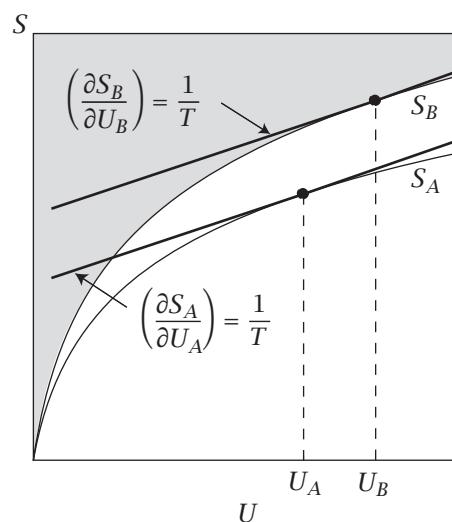


Figure 6.4 $1/T$ is the slope of S as a function of U for both systems *A* and *B*. Two systems at thermal equilibrium have the same temperature T , but not necessarily the same energy or entropy.

Figure 6.4 shows that at thermal equilibrium, equality of temperatures T_A and T_B is not the same as equality of energies or entropies. We interpret temperature in more detail in Chapter 12. Now let's see that *pressure* describes the thermodynamic driving force for changing the volume of a system.

Cont. M

TP

EXAMPLE 6.3 Pressure is a force for changing volume. How does volume change maximize entropy? Consider a cylinder partitioned into subsystems A and B by a movable piston as shown in Figure 6.5. The volumes of both subsystems change when the piston moves. In this problem, volume is the degree of freedom and there is no energy or matter exchange. You know from Example 6.2 that there will be no energy exchange if the temperatures on both sides are equal, $T_A = T_B$.

The total system is otherwise isolated, so it is not able to exchange energy, volume, or particles with the surroundings. Subsystem A has entropy $S_A(U_A, V_A)$ and subsystem B has entropy $S_B(U_B, V_B)$. Because the entropy is an extensive quantity, the total entropy is the sum of the entropies of the subsystems: $S = S_A + S_B$. Because the system is isolated, the volume of the total system is fixed: $V_A + V_B = V_{\text{total}} = \text{constant}$. Now allow subsystems A and B to exchange volume. For the moment, see what happens if they can also exchange energy. In differential form, the volume constraint is

$$dV_A = -dV_B. \quad (6.20)$$

The state of maximum entropy, $dS = 0$, defines the equilibrium. Write the total differential dS_{total} in terms of the degrees of freedom:

$$dS = \left(\frac{\partial S_A}{\partial V_A} \right) dV_A + \left(\frac{\partial S_B}{\partial V_B} \right) dV_B + \left(\frac{\partial S_A}{\partial U_A} \right) dU_A + \left(\frac{\partial S_B}{\partial U_B} \right) dU_B = 0, \quad (6.21)$$

where we have left off the implied subscripts to keep the notation simple, $(\partial S_A / \partial V_A) = (\partial S_A / \partial V_A)_{N_A, U_A}$, for example. Because no energy is exchanged with the surroundings, $dU_A = -dU_B$. Substitute dU_A for $-dU_B$ and $1/T$ for (dS/dU) :

$$dS = \left(\frac{\partial S_A}{\partial V_A} \right) dV_A + \left(\frac{\partial S_B}{\partial V_B} \right) dV_B + \left(\frac{1}{T_A} - \frac{1}{T_B} \right) dU_A = 0. \quad (6.22)$$

Since $T_A = T_B$, Equation (6.22) reduces to

$$dS = \left(\frac{\partial S_A}{\partial V_A} \right) dV_A + \left(\frac{\partial S_B}{\partial V_B} \right) dV_B = 0. \quad (6.23)$$

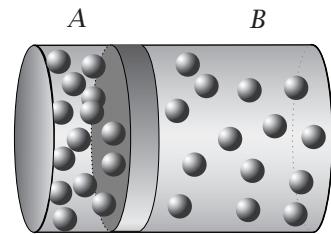
Substituting Equations (6.20) and (6.6) into Equation (6.23), the condition for equilibrium becomes

$$dS = \left(\frac{p_A}{T_A} - \frac{p_B}{T_B} \right) dV_A = 0. \quad (6.24)$$

Equation (6.24) is satisfied when

$$\frac{p_A}{T_A} = \frac{p_B}{T_B}. \quad (6.25)$$

Before Equilibration



After Equilibration

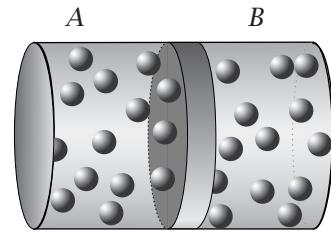


Figure 6.5 Before equilibration, the piston in Example 6.3 is fixed in place and the pressures on the two sides are unequal. After the piston is freed, the volumes will change (subject to the constraint $V_A + V_B = V_{\text{total}} = \text{constant}$) until the pressures equalize.

Since $T_A = T_B$, *mechanical equilibrium* occurs when the pressures of the two subsystems are equal, $p_A = p_B$. Just as a system with $T_B > T_A$ increases its entropy and moves to equilibrium by transferring heat from subsystem B to subsystem A , a system with $p_B > p_A$ increases its entropy and moves toward equilibrium by increasing volume V_B and decreasing V_A . Volumes change to equalize the pressures.

The thermodynamic reasoning in Example 6.3 is quite general, but it gives no insight into the microscopic nature of entropy. Let's return to the lattice model to see how equalizing the pressures maximizes the multiplicity of the arrangements of particles.

EXAMPLE 6.4 How does the equality of pressures maximize the multiplicity of states? Consider a gas contained on two sides of a movable piston as shown in Figure 6.5. Equilibrium is reached when the gas densities are equal on both sides. Why? This is the state of maximum multiplicity of the system. To see this, consider fixed numbers N_A of particles on the left and N_B on the right. The whole system has a fixed total volume of M lattice sites. The movable piston partitions the total volume into M_A sites on the left and M_B sites on the right, with the constraint that $M = M_A + M_B = \text{constant}$.

We want to find the volumes of the two sides, M_A^* and M_B^* , that maximize the multiplicity function, given by Equation (2.3):

$$W(N, M) = \frac{M!}{(M-N)! N!} \cdot 1.$$

First, notice that for low densities, $N/M \ll 1$, you can approximate the first factor in this expression as $M!(M-N)! \approx M^N$. (Try it: if $M = 1000$ and $N = 3$, you have $M!/(M-N)! = 1000!/997! = (1000)(999)(998) \approx 1000^3 = M^N$.) To get the total multiplicity from both sides of the piston, use this approximation and substitute $M_B = M - M_A$ to get

$$W = W_A W_B = \text{constant} \times M_A^{N_A} (M - M_A)^{N_B}. \quad (6.26)$$

We call $1/(N_A!N_B!)$ a constant here because we are only interested below in taking a derivative with respect to M , and this quantity doesn't contribute to it. Now find the value $M_A = M_A^*$ that maximizes $S/k = \ln W$. Taking M , N_A , and N_B as fixed, solve

$$\frac{d \ln W}{d M_A} = 0. \quad (6.27)$$

Taking the logarithm of Equation (6.26) gives $\ln W = N_A \ln M_A + N_B \ln(M - M_A)$. Now take the derivative, as indicated in Equation (6.27), to get

$$\frac{N_B}{M_B^*} = \frac{N_A}{M_A^*}. \quad (6.28)$$

The entropy is maximized and the system is at equilibrium when the density of particles N/M is the same on both sides. We already knew from

Before Equilibration

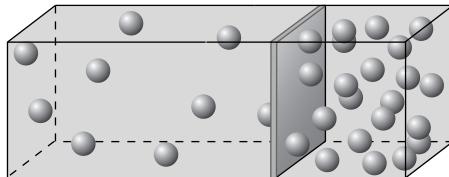
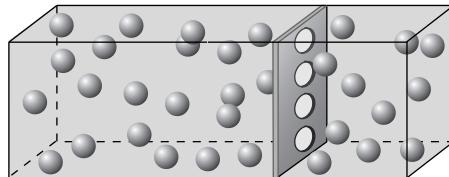


Figure 6.6 A permeation barrier prevents particle flow before equilibration. After the barrier is removed, the particles exchange until the chemical potential is the same on both sides (see Example 6.5).

After Equilibration



Equation (6.25) that the condition for equilibrium is $p_A/T_A = p_B/T_B$. Adding the ideal gas law $P/T = k(N/V)$ gives you the additional information that gas densities N/V are also equal at equilibrium.

Now we turn our attention from changing volumes to changing particle numbers.

EXAMPLE 6.5 The chemical potential is a tendency for particle exchange. The chemical potential (or, more accurately, μ/T) describes a tendency for matter to move from one place to another. Suppose that two compartments (A and B) are separated by a fixed barrier. Consider N identical particles, N_A of which are on side A and N_B of which are on side B (see Figure 6.6). The degrees of freedom are N_A and N_B . When particles are exchanged, they carry energy with them, so energy is also exchanged.

The subsystems have entropies $S_A(N_A)$ and $S_B(N_B)$. Again write the differential form of the fundamental entropy equation in terms of the degrees of freedom, introduce the constraints, and then find the state that has the maximum entropy. When the barrier is made permeable to particles, the particle numbers on both sides change, subject to the constraint $N_A + N_B = N = \text{constant}$. In differential form, the number constraint is

$$dN_A = -dN_B. \quad (6.29)$$

The condition for equilibrium is

$$dS_{\text{total}} = \left(\frac{\partial S_A}{\partial N_A} \right) dN_A + \left(\frac{\partial S_B}{\partial N_B} \right) dN_B = 0. \quad (6.30)$$

Substituting $-\mu/T$ for $(\partial S/\partial N)$ (Equation (6.6)) and using Equation (6.30) gives the equilibrium condition

$$dS_{\text{total}} = \left(\frac{\mu_B}{T_B} - \frac{\mu_A}{T_A} \right) dN_A = 0. \quad (6.31)$$

The condition for material balance equilibrium at temperature $T = T_A = T_B$ is

$$\mu_A = \mu_B. \quad (6.32)$$

A change toward equilibrium must increase the entropy:

$$dS_{\text{total}} = \left(\frac{\mu_B}{T} - \frac{\mu_A}{T} \right) dN_A > 0. \quad (6.33)$$

For dS_{total} to be positive, the signs of the factors $[(\mu_B/T) - (\mu_A/T)]$ and dN_A must be the same. Therefore, if the particle number on side A increases as the system approaches equilibrium, μ_B must be greater than μ_A . The chemical potential μ is sometimes called the *escaping tendency* because particles tend to escape from regions of high chemical potential to regions of low chemical potential. The chemical potential pertains not just to particles in different spatial locations, say on different sides of a membrane. It also pertains to molecules in different phases or in different chemical states, as we will see in Chapters 13 and 16. Chemical potentials are found by measuring concentrations.

Summary

The First Law of Thermodynamics is a bookkeeping tool that defines the internal energy change as a sum of heat plus work. The Second Law describes the tendencies of systems toward their states of maximum entropy. The fundamental equation $S(U, V, N)$ is $dS = (1/T) dU + (p/T) dV - \sum_{j=1}^M (\mu_j/T) dN_j$. The importance of this equation is twofold. It gives definitions of temperature, pressure, and chemical potential, and it is the basis for the maximum entropy extremum principle. Set $dS = 0$ subject to the appropriate constraints for the problem at hand to find the condition for equilibrium in terms of experimental quantities. In this chapter, we have used thermodynamics in two simple ways. First, in Example 6.1, we combined the thermodynamic definition of pressure with a molecular model to derive an equation of state, the ideal gas law. Second, in Examples 6.2, 6.3, and 6.5, we found that the condition ($dS = 0$) for different equilibria could be restated in terms of the experimentally accessible quantities T , p , and μ . $1/T$ is a tendency for exchanging energy in the form of heat, p/T is a tendency to change volume, and μ/T is a tendency for particles to exchange. For thermal, mechanical, and chemical equilibrium between subsystems A and B, $T_A = T_B$, $p_A = p_B$, and $\mu_A = \mu_B$. Combining the First and Second Laws gives ways to calculate the properties of engines, refrigerators, and heat pumps, where heat can convert to work or energy input can be used to warm up hot objects or cool down cold ones. In the following chapters, we look at the broader toolkit of thermodynamics.

Problems

1. One-dimensional lattice. You have a one-dimensional lattice that contains N_A particles of type A and N_B particles of type B. They completely fill the lattice, so the number of sites is $N_A + N_B$.

- (a) Express the entropy $S(N_A, N_B)$ as a function of N_A and N_B .
- (b) Give the relationship between the chemical potential μ_A and the quantity $(\partial S / \partial N_A)_{N_B}$.
- (c) Express $\mu_A(N_A, N_B)$ as a function of N_A and N_B .

2. The entropy of an ideal gas. Show that the entropy of an ideal gas is

$$S(V, N) = Nk \ln V. \quad (6.34)$$

3. Entropy changes don't depend on a process pathway. For an ideal gas, the entropy is $S = Nk \ln V$ (see above).

- (a) Express $\Delta S_V = S_2(V_2) - S_1(V_1)$, the entropy change upon changing the volume from V_1 to V_2 , at fixed particle number N .
 - (b) Express $\Delta S_N = S_2(N_2) - S_1(N_1)$, the entropy change upon changing the particle number from N_1 to N_2 , at fixed volume V .
 - (c) Write an expression for the entropy change, ΔS , for a two-step process $(V_1, N_1) \rightarrow (V_2, N_1) \rightarrow (V_2, N_2)$ in which the volume changes first at fixed particle number, then the particle number changes at fixed volume.
 - (d) Show that the entropy change ΔS above is exactly the same as for the two-step process in reverse order: changing the particle number first, then the volume.
- 4. Compute $\Delta S(V)$ for an ideal gas.** What is the entropy change if you double the volume from V to $2V$ in a quasi-static isothermal process at temperature T ?

Suggested Reading

CJ Adkins, *Equilibrium Thermodynamics*, 3rd edition, Cambridge University Press, Cambridge, 1983. Excellent introduction to thermodynamics and reversibility.

SG Brush, *The Kind of Motion We Call Heat*, Volumes 1 and 2, North-Holland, New York, 1976. An excellent history of thermodynamics and statistical mechanics.

HB Callen, *Thermodynamics and an Introduction to Thermostatistics*, 2nd edition, Wiley, New York, 1985. The classic text on the axiomatic approach to thermodynamics. The logic of thermodynamics is explained with great clarity.

G Carrington, *Basic Thermodynamics*, Oxford University Press, Oxford, 1994. Good general text with many engineering examples.

K Denbigh, *The Principles of Chemical Equilibrium*, 4th edition, Cambridge University Press, Cambridge, 1981. General text on thermodynamics.

IM Klotz and RM Rosenberg, *Chemical Thermodynamics: Basic Concepts and Methods*, 7th edition, Wiley, New York, 2008. Excellent simple and general text.

HS Leff and AF Rex, eds, *Maxwell's Demon: Entropy, Information, Computing*, Princeton University Press, Princeton, 1990. Essays on the relationship between entropy and information theory.

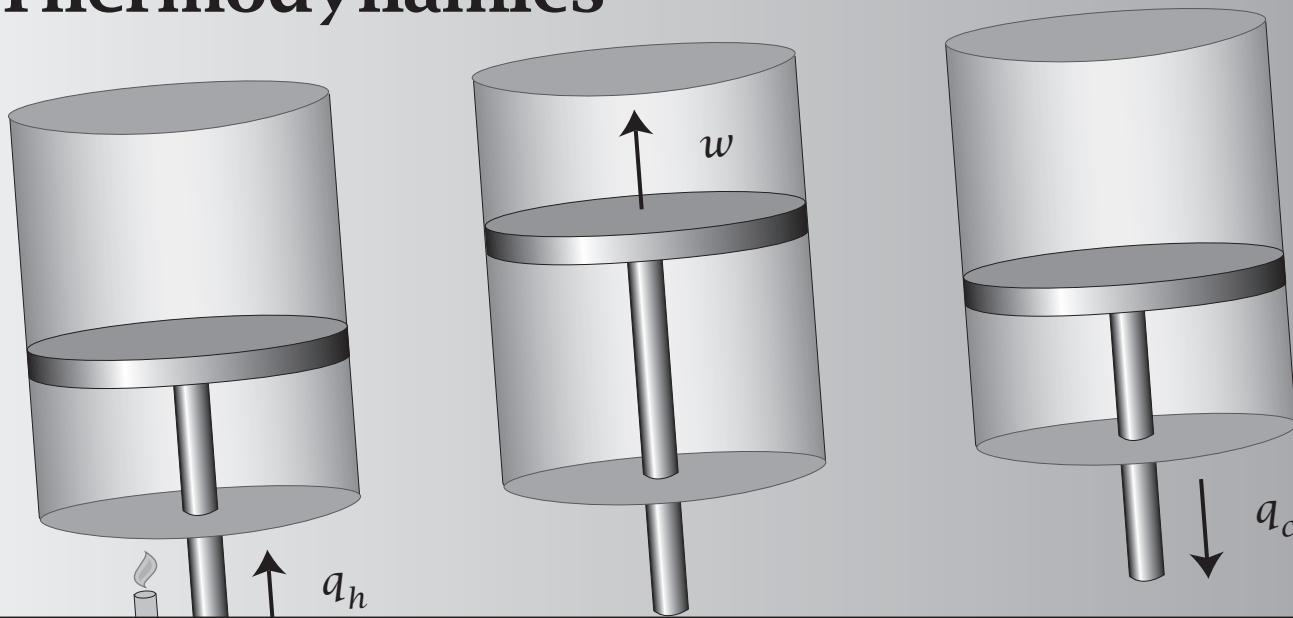
HS Leff and AF Rex, eds, *Maxwell's Demon 2: Entropy, Classical and Quantum Information, Computing*, CRC Press, Boca Raton, FL, 2002. Essays on the relationship between entropy and information theory.

EB Smith, *Basic Chemical Thermodynamics*, 5th edition, Imperial College Press, London, 2004. Excellent simple and general text.

MW Zemansky and RH Dittman, *Heat and Thermodynamics*, 7th edition, McGraw-Hill, New York, 1996. Excellent general text; extensive comparisons with experiments and detailed discussion of reversibility.

This page is intentionally left blank.

7 The Logic of Thermodynamics



Thermodynamics Helps You Reason About Heat, Work, Processes, and Cycles

So far, our discussion of thermodynamics has focussed on the **state functions** $S(U, V, N)$ and $U(S, V, N)$. Now, we broaden our toolkit to include *heat, work, processes, pathways, and cycles*. This is crucial for understanding cyclic energy conversions—in engines, motors, refrigerators, pumps (including your heart), rechargeable batteries, hurricanes, ATP-driven biochemical reactions, oxygen transport around your body, and geophysical cycles of carbon and water, for example.

Thermodynamic logic often seems complex. This apparent complexity arises because the fundamental quantities that predict equilibria **are not directly measurable**. Equilibria are governed by energy and entropy through the First and Second Laws, but unfortunately there are no ‘meters’ that measure energy or entropy. Instead, inferences about equilibria are indirect and drawn from observations of quantities that can be measured, such as temperature, pressure, work, heat capacities, concentrations, or electrical potentials. Thermodynamics is a **business of making clever inferences about unmeasurable quantities** from observable ones, by various means.

We begin with the question of how gases convert heat to work. Consider a gas in a cylinder that has a movable piston. The gas is at a given pressure, volume, and temperature (p_1, V_1, T_1). Now introduce an amount of heat q into the cylinder. We aim to compute, under various conditions, the work w resulting

state estimation

from expansion or compression of the gas, the changes ΔU and ΔS in the internal energy and entropy of the gas, respectively, and (p_2, V_2, T_2) , the final conditions of the gas. Let's illustrate with ideal gases, for which $pV = NkT$.

We Start with a Few Basic Premises

We begin with (1) the First Law, (2) the concept of performing processes that are slow, or *quasi-static*, (3) a way to measure work, (4) a way to measure heat, and (5) the concept of *reversible processes*. Then we show how to combine and reason with these concepts.

The First Law Interrelates Heat, Work, and Energy

The **First Law of Thermodynamics** relates a change in energy, dU , of a system to the heat transferred in, δq , and the work performed on it, δw :

$$dU = \delta q + \delta w. \quad (7.1)$$

The use of δ for a differential element indicates that heat and work are *path-dependent* quantities (more on this below), while the use of d indicates that their sum, the internal energy dU , is a *state function*. The signs of the quantities in the First Law are defined so that the internal energy increases when heat flows *into* a system, $\delta q > 0$, and when work is done *on* a system, $\delta w > 0$.

Performing Processes Slowly Avoids Speed-Dependent Complications

Thermodynamic reasoning is about equilibria, not rates. Imagine driving a car, riding a bike, or jogging. If you go too fast, your motion will be inefficient; some of your energy will be wasted battling wind resistance. Go slower, and the wind resistance is lower, so you can travel further per unit energy expended. By flying at high altitudes, 20,000–30,000 feet, commercial airlines can reduce their wind resistance and gain fuel efficiency. Fast processes are more wasteful of energy. They are also more complex and involve additional variables: time-dependent quantities; gradients of temperature, pressure, or concentration; and friction and dissipation. **The tools of thermodynamics provide more power to reason about slow processes than about fast ones.** Processes are *quasi-static* if they are performed slowly enough that their properties are independent of time and independent of the speed of the process. To be quasi-static, a process must be slower than the relaxation time of a system. That is, you should take data only after the property of interest stops changing.

Work Describes Mechanical Energy Transfer

Work describes energy transport through mechanical means. Work is measured as a force acting through a distance. For a gas in a piston expanding by a volume dV against an applied external pressure p_{ext} , the increment of work in a

quasistatic process is defined to be

$$\delta w = -p_{\text{ext}} dV. \quad (7.2)$$

With this sign convention, δw is positive when the volume inside a piston decreases ($dV < 0$). That is, work is positive when it is done *on the system* to compress the gas, and negative when it is done *by the system* to expand. If a process happens too rapidly, there will be additional velocity-dependent contributions to the work and Equation (7.2) will not apply. Equation (7.2) defines pV work, but there are corresponding expressions for other types of work in quasi-static processes: the work of molecular motors or stretching rubber bands, the pressure-area work involved in surfaces and liquid films, the electrical work of moving charges, or the magnetic work of orienting magnets in fields, for example.

Heat Describes Thermal Energy Transfer

Heat describes energy transfer by thermal means. Increase the temperature of an object by putting it in a heat bath. The object will take up heat. Different objects will take up different amounts of heat for a given increase in its temperature. The heat uptake per unit temperature change is called the *heat capacity*. You can measure an object's heat capacity in a *calorimeter* (see Figure 7.1). Heat is introduced into a thermally insulated chamber, often via an electrical current, and the temperature change is measured. Some calorimeters, called *bomb calorimeters*, have fixed volume. No work occurs, $\delta w = -p dV = 0$, upon heating quasi-statically, and the heat uptake increases the internal energy, $dU = \delta q$. At constant volume, the heat capacity C_V is

$$C_V = \left(\frac{\delta q}{dT} \right)_V = \left(\frac{\partial U}{\partial T} \right)_V. \quad (7.3)$$

Equation (7.3) holds only if you change the temperature quasi-statically; otherwise, the quantities you measure will also depend on how fast the heat

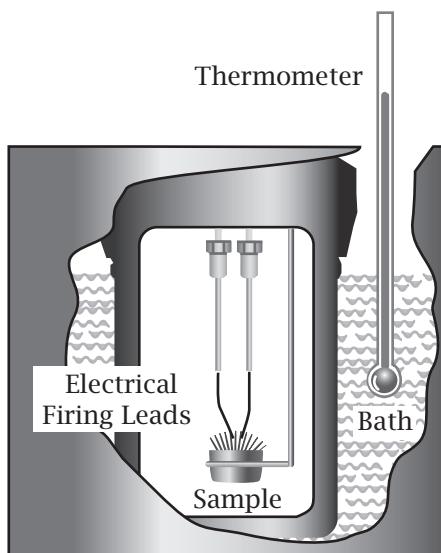


Figure 7.1 In a calorimeter, a measured electrical current heats a sample. A thermometer measures the corresponding change in temperature. The heat capacity is the ratio of heat input to temperature change.

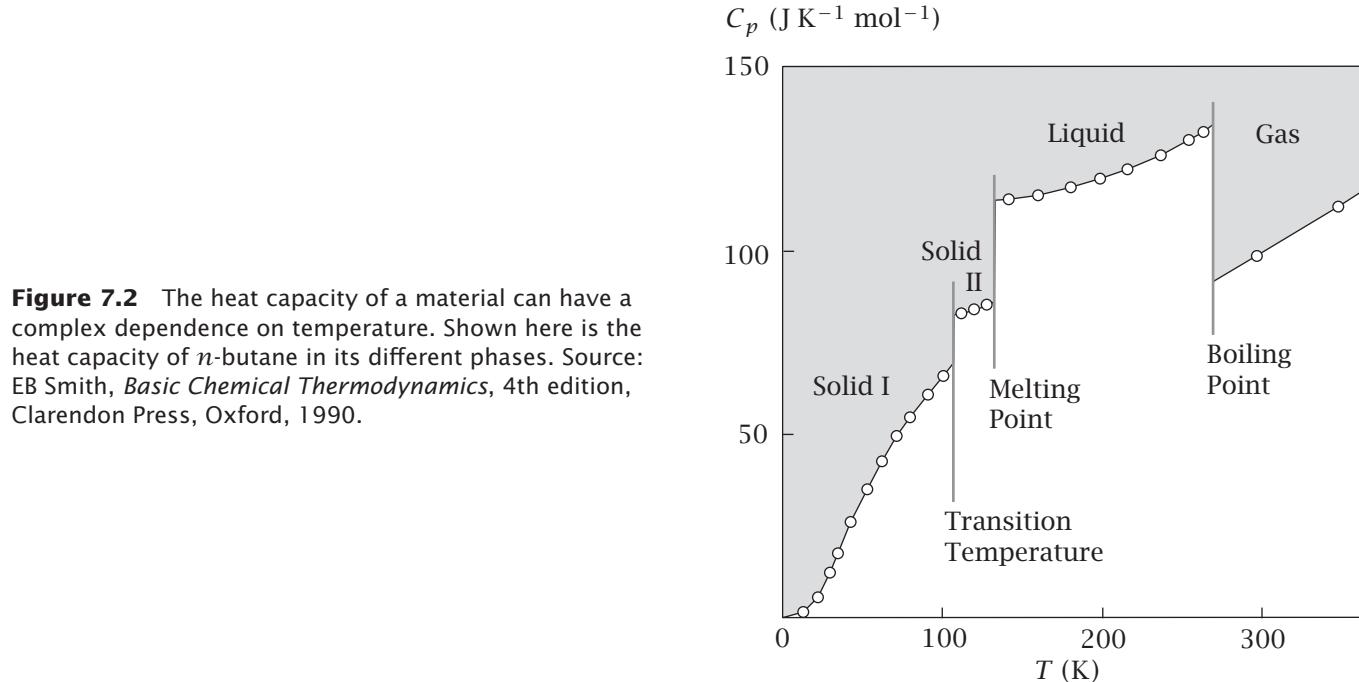


Figure 7.2 The heat capacity of a material can have a complex dependence on temperature. Shown here is the heat capacity of *n*-butane in its different phases. Source: EB Smith, *Basic Chemical Thermodynamics*, 4th edition, Clarendon Press, Oxford, 1990.

is conducted from one part of the calorimeter to another. A closely related quantity is C_p , the heat capacity of a system held at constant pressure, discussed in the next chapter.

The capacity of an object to take up heat can differ at different temperatures. The definition of heat capacity given by Equation (7.3) is only an operational recipe. It does not say how the heat capacity of any particular material will depend on its volume or temperature. As shown in Figure 7.2, the ability of a material to absorb heat can be high over some ranges of temperature and low over others, or it can be flat over the temperature range of interest.

In general, to find the energy change ΔU of an object as you raise its temperature from T_A to T_B , you need to know the functional dependence of C_V on T for the material. You can then integrate Equation (7.3):

$$\Delta U = \int_{T_A}^{T_B} C_V(T) dT = C_V(T_B - T_A), \quad (7.4)$$

where the last equality holds only if the heat capacity of a material is constant over the temperature range.

EXAMPLE 7.1 Heat capacities of gases. In general, the internal energy of a material can depend on the temperature and volume, $U = U(T, V)$, which means, in differential form,

$$dU = \left(\frac{\partial U}{\partial V}\right)_T dV + \left(\frac{\partial U}{\partial T}\right)_V dT. \quad (7.5)$$

What can be said about $U(T, V)$ for a gas? Experiments on ideal gases show that their internal energy depends on the temperature alone, not on the volume [$(\partial U / \partial V)_T = 0$]. This is because at low densities, where gases are ideal, the gas molecules rarely interact, so they aren't affected by density changes. So,

Equation (7.5) reduces to

$$dU = \left(\frac{\partial U}{\partial T}\right)_V dT = C_V dT \quad \text{for a gas at low density.} \quad (7.6)$$

For a monatomic ideal gas, a useful expression is derived in Chapter 11 that the heat capacity is $C_V = 3R/2$ per mole; it is a constant. This means that whenever the temperature of an ideal gas changes from T_1 to T_2 , the change in internal energy of the gas is

$$\Delta U = C_V(T_2 - T_1), \quad (7.7)$$

no matter whether the volume is constant or changing (and no matter whether the process is slow or fast). Equation (7.7) provides a useful general relationship for ideal gas processes.

Reversible Processes Are a Subset of Slow Processes

Central to the Industrial Revolution was the issue of how to convert heat to work efficiently. Steam engines convert heat to work by the expansion of a hot gas in a piston. James Watt's original steam engine converted only 7% of the energy (heat) input to work. The remaining 93% was exhausted to the surroundings. Understanding energy-conversion efficiency in the 1800s established the importance of thermodynamics. Is energy-conversion efficiency determined by the type of gas, its temperature, or the volume or shape of the piston?

In one respect, heat and work are completely interchangeable: ΔU is just the sum of heat plus work in the First Law. However, in another respect, heat and work are not equivalent. Consider three different situations: (1) converting one form of work to another form of work, (2) converting work to heat, or (3) converting heat to work. In principle, you could convert one type of work to another with up to 100% efficiency (if you eliminate friction). Typical electric motors convert electricity to motion with about 95% efficiency. Or, you can easily convert work to heat with 100% efficiency: just rub sandpaper on a door or spin a paddle wheel in water. However, the third situation is different: converting heat to work is essentially never done with efficiencies greater than 30–50%. This is important because a large fraction of energy conversion on the Earth results from converting heat to work in engines. Since work can fully convert to heat but heat cannot fully convert to work, work is somehow 'better' than heat—a sort of 'higher form' of energy transfer. Work is more reusable and more recoverable than heat. To explore these issues, we combine the Second Law with another thermodynamic idealization, called *reversibility*.

How can you transfer energy with maximum efficiency? First, consider interconverting work from one form to another, such as in the volume changes of a gas, the raising of a weight, the charging and discharging of an electrical capacity, or the stretching of a rubber band. To interconvert *work* with maximum efficiency, the process must be slow. However, to convert *heat* to work efficiently requires more. You need a process that is not only slow. It must also be *reversible*. A process is called *reversible* if returning the system to its initial condition *also returns the surroundings* to their initial condition. For example, you can push on a rusty piston to the right, then pull it back to the left

to its initial state, and you will have returned the *system* to its initial state. But you will not have returned the *surroundings* to their initial state. You will have heated the surroundings in both steps. So, even if done very slowly, pushing or pulling on a rusty piston is irreversible. In pushing the piston, some of your work dissipates as heat in the surroundings. Turning a paddle wheel in water is also irreversible. Turning a paddle wheel clockwise dissipates heat through viscous forces, but turning the wheel counterclockwise does not reabsorb the heat; it dissipates more heat. Or, poke a hole in a balloon to let the air escape. If the hole is small enough, this process can be slow. But it is also irreversible. The capacity to perform work that was stored in the pressurized air in the balloon has permanently escaped. So, reversible processes are not only slow, they also involve no friction or turbulence or leakage of material and no diffusion or *dissipation*, i.e., no conversion of energy in ways that cannot be recaptured.

In a reversible process—i.e., one that satisfies the conditions above—the heat exchange is defined as $q = q_{\text{rev}}$. An important idealization of thermodynamics is that you can compute the entropy, which is a state function, by measuring the heat exchange via a reversible process,

$$dS = \frac{\delta q_{\text{rev}}}{T}. \quad (7.8)$$

How do you get δq_{rev} ? One approach is to measure it calorimetrically.

EXAMPLE 7.2 Obtaining the entropy by measuring the heat capacity. You can compute the entropy of a material by measuring its heat capacity. Substitute $\delta q_{\text{rev}} = dU = C_V dT$ into Equation (7.8) to find

$$dS = \frac{C_V}{T} dT. \quad (7.9)$$

Integrating Equation (7.9) gives

$$\Delta S = \int_{T_A}^{T_B} \frac{C_V}{T} dT = C_V \ln\left(\frac{T_B}{T_A}\right), \quad (7.10)$$

where the latter equality holds only if the heat capacity is constant over the temperature range, for example for an ideal gas.

The premises above allow for a broad range of calculations in thermodynamics.

Four Different Processes Can Change an Ideal Gas

How can you change the state of a material? An ideal gas contained in a cylinder is described by an *equation of state*, such as $p = NkT/V$, a relationship between three variables: (p, V, T) . If you hold one of those quantities constant, say T , and control another, say p , then the third, V in this case, will vary accordingly. There are four ways to hold constant one variable; three options are to fix p or V or T . In such cases, heat will flow into or out of the cylinder. Or, in an

adiabatic process, you insulate the cylinder to prevent heat flow. Here's what you can learn by performing any of these processes quasi-statically.

EXAMPLE 7.3 Constant-volume process. Start with N molecules of a gas at (p_1, V_0, T_1) in a cylinder. Hold constant its volume V_0 . Now introduce an amount of heat q by heating the cylinder reversibly. Heating increases the gas pressure and temperature to (p_2, V_0, T_2) . When the volume is held constant (called an *isochoric* process), no work is performed, $w = 0$, by definition, because $dV = 0$; see Equation (7.2). In this case, the First Law gives $\Delta U = q$. Now use the ideal gas law and Equation (7.7) to get

$$q = \Delta U = C_V(T_2 - T_1) = \frac{C_V}{Nk} V_0(p_2 - p_1). \quad (7.11)$$

You are given the quantities q, C_V, N , and (p_1, V_0, T_1) . Equation (7.11) gives you a way to compute the others ($\Delta U, T_2$, and p_2) for an isochoric process. Equation (7.9) then gives the entropy change as $\Delta S = C_V \ln(T_2/T_1)$.

Equation (7.11) shows how q , which is a path-dependent quantity, becomes equal to a state function ΔU , which depends only on a property, T_1 and T_2 , of the two end states, if you hold the volume constant. This illustrates a general simplifying strategy of thermodynamics. Measuring path-dependent quantities when you fix certain conditions gives you a way to compute state quantities.

EXAMPLE 7.4 Constant-pressure process. Start with N molecules of a gas at (p_0, V_1, T_1) . Apply a constant external pressure $p_{\text{ext}} = p_0$ to a movable piston reversibly. Transfer in an amount of heat q to increase the gas volume and temperature to (p_0, V_2, T_2) . When the externally applied pressure is constant (called an *isobaric* process), the work in a quasi-static process of expansion from volume V_1 to V_2 is (see Equation (7.2))

$$w = - \int_{V_1}^{V_2} p_0 dV = -p_0(V_2 - V_1). \quad (7.12)$$

In a quasi-static gas expansion, each incremental step is so slow that the internal pressure of the gas equalizes with the external pressure, $p_{\text{int}} = NkT/V = p_0$, at every step. This relates a measurable external quantity, the applied pressure p_0 , to properties of the material inside the cylinder, such as its temperature T and particle number N . Use the ideal gas law and Equation (7.7) to get

$$\Delta U = C_V(T_2 - T_1) = \frac{C_V p_0}{Nk} (V_2 - V_1). \quad (7.13)$$

To relate ΔU to q , use

$$q = \Delta U - w = \left(\frac{C_V}{Nk} - 1 \right) p_0 (V_2 - V_1). \quad (7.14)$$

Given p_0, V_1, C_V, N , and q for an isobaric process, Equation (7.14) gives V_2 . Then, use Equation (7.12) to get w . Finally, use Equation (7.13) to get T_2 and ΔU . Equation (7.9) gives the entropy change as $\Delta S = C_V \ln(T_2/T_1)$.

EXAMPLE 7.5 Constant-temperature process. Start with N molecules of an ideal gas at (p_1, V_1, T_0) . Hold constant its temperature at T_0 . Reversibly introduce an amount of heat q (by inputting radiation or combustion inside a cylinder that is contained in a fixed-temperature heat bath, for example) to increase the gas pressure and volume to (p_2, V_2, T_0) . Do this quasi-statically, so $p_{\text{ext}} = p_{\text{int}} = NkT_0/V$. When the temperature is constant, the process is called *isothermal* and the work is given by integrating Equation (7.2):

$$\begin{aligned} w &= - \int_{V_1}^{V_2} p_{\text{ext}} dV = - \int_{V_1}^{V_2} p_{\text{int}} dV \\ &= - \int_{V_1}^{V_2} \frac{NkT_0}{V} dV = -NkT_0 \ln\left(\frac{V_2}{V_1}\right) = -NkT_0 \ln\left(\frac{p_1}{p_2}\right). \end{aligned} \quad (7.15)$$

In a constant-temperature process, the energy of an ideal gas does not change, $\Delta U = C_V \Delta T = 0$, so the heat is given by

$$q = -w = NkT_0 \ln\left(\frac{V_2}{V_1}\right). \quad (7.16)$$

These expressions for an isothermal process allow you to compute (p_2, V_2) , ΔU , and w , if you are given q and (p_1, V_1, T_0) . The entropy change is $\Delta S = q/T_0 = Nk \ln(V_2/V_1)$.

In the three processes above, heat enters the cylinder. Another approach is to change the state of the gas *adiabatically*, i.e., with no heat flow, $\delta q = 0$, by using a thermal insulating boundary, as in a thermos bottle. For any adiabatic process, fast or slow, simple or complex, the First Law gives $\Delta U = \delta w$.

EXAMPLE 7.6 Adiabatic process. Start with N molecules of an ideal gas at (p_1, V_1, T_1) in an insulated cylinder. No heat exchanges across it, $\delta q = 0$, so $dU = \delta w$. In an adiabatic process, reversibly change one variable—pressure, temperature, or volume—and the other two will change, to (p_2, V_2, T_2) . Only certain (p, V, T) changes are possible, as we'll see below. Substituting $dU = C_V dT$ and $\delta w = -p dV$ gives

$$C_V dT = dU = -p dV. \quad (7.17)$$

Substitute the ideal gas law $p = NkT/V$ into Equation (7.17):

$$C_V dT = -\frac{NkT}{V} dV. \quad (7.18)$$

Rearrange and integrate to relate the changes in temperature and volume:

$$\int_{T_1}^{T_2} \frac{C_V dT}{T} = - \int_{V_1}^{V_2} \frac{Nk}{V} dV. \quad (7.19)$$

For ideal gases, heat capacities are independent of temperature, so

$$C_V \ln\left(\frac{T_2}{T_1}\right) = -Nk \ln\left(\frac{V_2}{V_1}\right). \quad (7.20)$$

Thus, for a quasi-static adiabatic volume change in an ideal gas, the relation between temperature and volume is given by

$$\frac{T_2}{T_1} = \left(\frac{V_1}{V_2}\right)^{Nk/C_V}. \quad (7.21)$$

This shows how the temperature and volume changes are related in an adiabatic expansion or compression of an ideal gas.

Pathways Are Combinations of Processes

By themselves, the four processes above are limited. Fixing one variable does not allow you to reach any arbitrary final state (p_2, V_2, T_2) from an initial state (p_1, V_1, T_1) . A constant-volume process changes only (p, T) and not the volume, for example. However, you can make any arbitrary change to a gas by combining two or more processes into *pathways*. A pathway is a particular sequence in which you change the variables of a system. For example, you might first fix the pressure at p_1 and change from (p_1, V_1, T_1) to (p_1, V_2, T_x) , then fix the volume at V_2 and change from (p_1, V_2, T_x) to reach (p_2, V_2, T_2) . Or, you might first fix the volume at V_1 , then change the pressure. The pathways are different: the first passes through an intermediate temperature T_x and the second through T_y . Different pathways pass through different intermediate states. Recall from Chapter 4 (page 72) that for some functions $f(x, y)$, a change Δf , computed by integration, depends on the pathway of integration, while for others, called *state functions*, the integral Δf does not depend on the pathway of integration. Heat and work are path-dependent functions, while U and S are state functions.

A key principle follows from this logic. Consider a change of state from A to B . You want to know the difference in some state function, $\Delta U = U_B - U_A$ or $\Delta S = S_B - S_A$, that you cannot measure directly. How can you determine it? You can measure path-dependent quantities, such as heat or work, as long as you follow particular rules in performing those steps. Examples are given below. For computing state quantities, you can even invent pathways that have *fictitious* or non-physical intermediate states. For example, to compute the binding energy of a drug molecule to a protein, computational biologists sometimes start with an empty binding cavity on the protein, then ‘grow’ the atoms of the drug into the binding site in a continuous way, in the same way that you adjust the volume on a television set. While this process is physically impossible, the starting and ending states are real, and that is all that thermodynamic logic requires. You are free to invent such pathways. This strategy, illustrated below, often simplifies the computation of state functions of interest.

EXAMPLE 7.7 Work is a path-dependent function. You have one mole of an ideal gas in a cylinder. It undergoes a change from state (p_1, V_1, T_1) to (p_2, V_2, T_2) ; see Figure 7.3. To realize this change in state, consider a quasi-static pathway a involving two steps. First, the gas expands from volume $V_1 = 1$ liter to $V_2 = 5$ liters against a fixed applied pressure of $p_1 = 100$ atm. Then, lower the temperature, decreasing the pressure to $p_2 = 35$ atm,

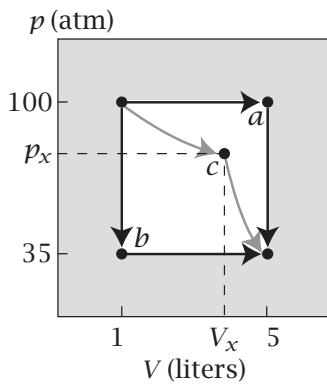


Figure 7.3 Different pathways for changing from (p_1, V_1) to (p_2, V_2) : (a) first change the volume, then the pressure through intermediate (p_1, V_2) , or (b) first change the pressure, then the volume through intermediate (p_2, V_1) , or (c) change both pressure and volume through intermediate (p_x, V_x) .

at constant volume V_2 . Along this pathway, a , the work of gas expansion is $w_a = -p_1(V_2 - V_1) + 0 = -100(5 - 1) = -400$ liter atm. This work comes entirely from the constant-pressure step, since no work is performed in the constant-volume step.

Along path b , the processes are performed in reverse order: change the volume first, then change the pressure (Figure 7.3). Now the work is $w_b = -p_2(V_2 - V_1) = -35(5 - 1) = -140$ liter atm. You see that the work is different along the two pathways, $w_a \neq w_b$, even though they have the same beginning and ending points, (p_1, V_1) and (p_2, V_2) . Or you could achieve the same state change using path c , adiabatic and isothermal steps. The amount of work performed depends the physical pathway a system takes.¹

EXAMPLE 7.8 Energy is a state function. ΔU is independent of the pathway. While the work is path-dependent, the *energy difference* is not. For the example above, the temperatures are $T_1 = p_1 V_1 / R = 1200\text{ K}$ and $T_2 = p_2 V_2 / R = 2100\text{ K}$, so the energy change is $\Delta U = C_V(T_2 - T_1) = (3/2)R(T_2 - T_1) = (1.5)(8.2 \times 10^{-2})(2100 - 1200) = 111$ liter atm. (We have used the result that the heat capacity per mole for a monatomic ideal gas equals $3R/2$, which we'll derive in Chapter 11.) This is the energy difference, no matter which pathway the system took from state 1 to 2, or even if the process was as fast as an explosion, or involved friction.

EXAMPLE 7.9 Quasi-static heat plus work gives the energy. The First Law says that q and w , two path-dependent quantities, sum to ΔU , a state function. Let's check this. Along route a , the quasi-static heat is the sum of the constant-pressure contribution, $q = C_V(T_2 - T_x) + p_1(V_2 - V_1)$, plus the constant-volume contribution, $q = C_V(T_x - T_1)$, for a total of $q_a = C_V(T_2 - T_1) + p_1(V_2 - V_1)$. Using $w_a = -p_1(V_2 - V_1)$ from Example 7.7 and summing gives $w_a + q_a = \Delta U = C_V(T_2 - T_1)$, as it should.

These examples illustrate the logic of thermodynamics. On the one hand, measuring a state quantity, such as ΔU or ΔS , tells you about a difference between the two states, no matter what pathway the system took, fast or slow. On the other hand, if you can't measure ΔU directly, you can obtain it by measuring the heat plus work along any convenient quasi-static pathway between the two end-states, and summing to get $\Delta U = q_a + w_a$. Or, you can use a reversible path to get $\Delta S = q_{\text{rev}}/T$.

A Cycle is a Set of Outbound and Returning Pathways

At the core of thermodynamics are *thermodynamic cycles*. A cycle is a set of processes that begin in one state, pass through other states, and then return to

¹Chapter 3 defines some forces as *conservative*. The force along each horizontal constant-pressure leg in Figure 7.3 is conservative, because the total work is $-p(V_2 - V_1) - p(V_1 - V_2) = 0$. But the forces around the whole cycle in Figure 7.3 are not conservative, because the net heat flow leads to nonzero work around the cycle.

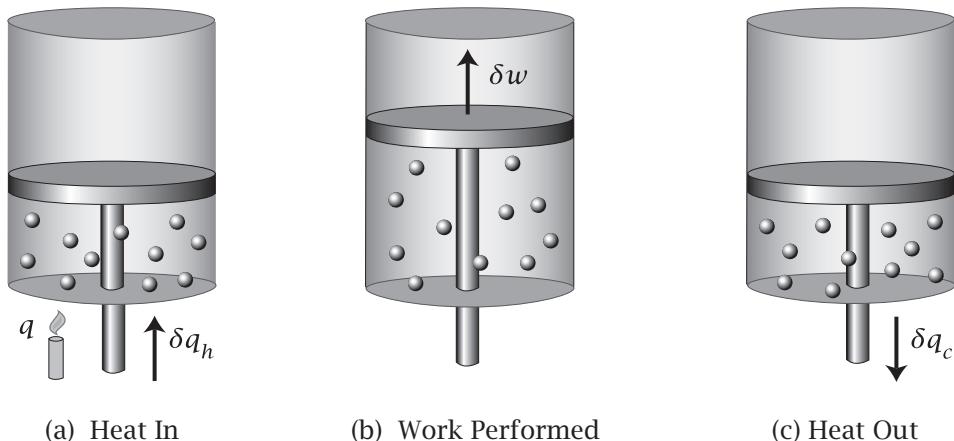


Figure 7.4 (a) In this idealized heat engine, a piston contains cooled gas. When heat q_h enters at high temperature T_h , the energies of the gas molecules increase. (b) Work is performed by the expansion of the heated gas, extracting energies from the molecules. (c) Heat q_c then flows out at a colder temperature T_c , while the volume in the piston decreases. Then the system is ready for the next cycle.

the initial state. In an engine, fuel enters; it is combusted to form a gas; the gas expands, performing work; the spent vapor exits; then new fuel enters to begin another cycle. In refrigerators, a working fluid flows through cooling coils to cool a compartment, then flows through other coils to dump heat outside the refrigerator, then repeats the cycle. Muscle proteins convert energy to motion in repetitive cycles. Cycles are described by *state diagrams*, which show how certain state variables change throughout the cycle. If a process is *cyclic*, its ending point is the same as its starting point. So, the sum of any state function around a cycle is zero: $\Delta U = (U_A - U_C) + (U_C - U_B) + (U_B - U_A) = 0$ for a cycle from state A to B to C and back to A again. The fact that state functions sum to zero around a cycle is useful for computing quantities you don't know from others that you do know.

Figure 7.4 shows a thermodynamic cycle for an idealized heat engine. Heat enters a cylinder, expanding the gas and performing work, then the system exhausts the heat and is ready for the next cycle. You can explore this conversion of heat to work using the *Carnot cycle*, invented by NLS Carnot in 1824, and described below.

EXAMPLE 7.10 Heat engines take in heat and perform work: the Carnot cycle. To model how engines convert heat to work efficiently, Carnot envisioned the cycle shown in Figure 7.5, called a *Carnot cycle*, which assumes four reversible steps. Two are adiabatic and two are isothermal. To keep it simple, we consider an ideal gas working fluid, even though the argument is more general. We want to compute the total work around the cycle, w_{tot} . Summing around the four steps of the cycle gives $\Delta U = q_{\text{tot}} + w_{\text{tot}} = 0$, so $w_{\text{tot}} = -q_{\text{tot}}$. You can compute q_{tot} from Figure 7.5. You have $q = 0$ for the two adiabatic steps, so

$$q_{\text{tot}} = q_h + q_c = NkT_h \ln\left(\frac{V_B}{V_A}\right) + NkT_c \ln\left(\frac{V_D}{V_C}\right). \quad (7.22)$$

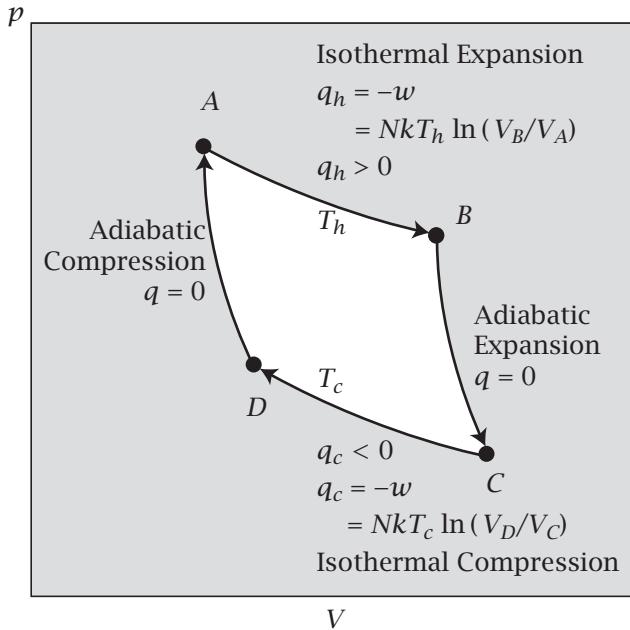


Figure 7.5 Carnot cycle. Any quasi-static change from (p_A, V_A, T_A) to (p_C, V_C, T_C) can be expressed in two steps (shown here for an ideal gas): an isothermal change and an adiabatic change from the hot temperature T_h to the cold temperature T_c . The cycle is completed by a return path.

You can simplify Equation (7.22) by using Equation (7.21), which relates temperatures to volumes for the adiabatic steps,

$$\frac{V_C}{V_B} = \left(\frac{T_h}{T_c}\right)^{C_V/Nk} = \frac{V_D}{V_A} \quad \Rightarrow \quad \frac{V_B}{V_A} = \frac{V_C}{V_D}. \quad (7.23)$$

Substituting Equation (7.23) into (7.22) gives

$$w_{\text{tot}} = -q_{\text{tot}} = -Nk(T_h - T_c) \ln\left(\frac{V_B}{V_A}\right). \quad (7.24)$$

A Carnot engine takes in heat and produces work. The amount of work it performs depends on the intake and exhaust temperatures T_h and T_c , respectively, and on V_B/V_A , which is called the *compression ratio*.

An example of a Carnot cycle is a hurricane.

EXAMPLE 7.11 The thermodynamics of a hurricane: a Carnot cycle. A hurricane takes up heat from warm ocean water (a source of high temperature T_h), converts the heat to the work of spiraling the winds, then expels the heat at high altitude in the cold atmosphere (a sink at low temperature T_c). There are three aspects to this problem: the angular momentum physics explaining why a hurricane spins in a particular direction; the thermodynamics of converting heat to work that drives the hurricane engine; and the aerodynamics of the winds flowing in at the bottom of the funnel and out the top.

Angular momentum physics. The reason that a hurricane always rotates in a particular direction (depending on which hemisphere it is in) is that a

(a)

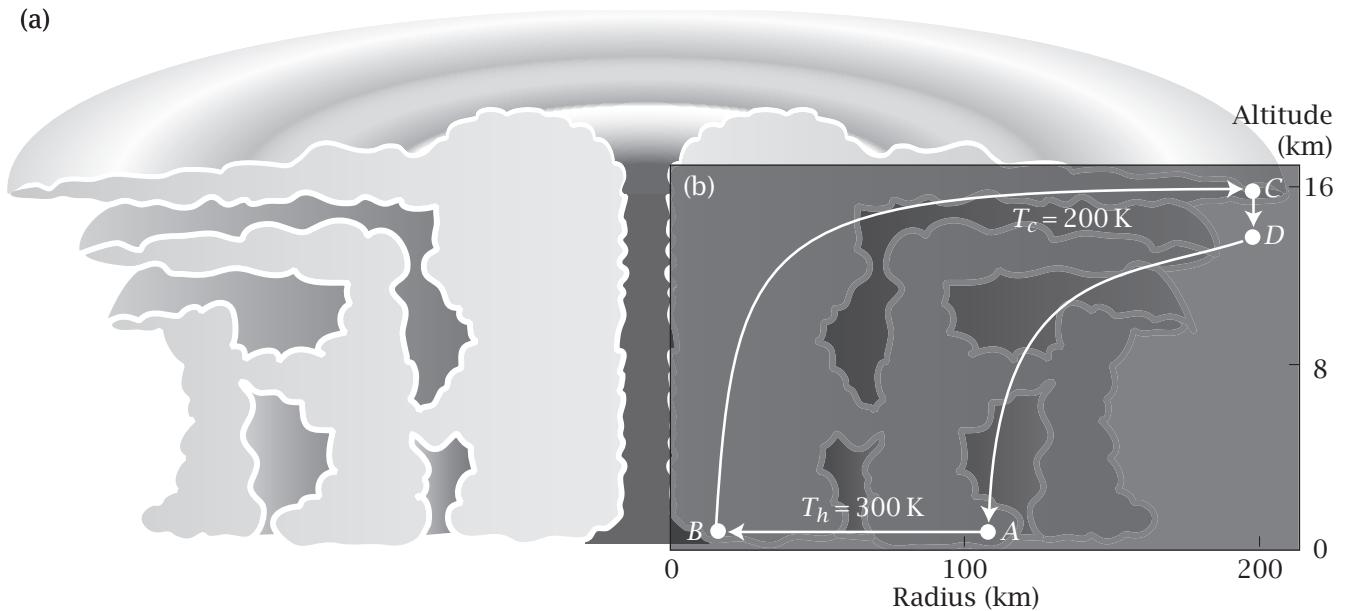


Figure 7.6 A hurricane is a Carnot ‘engine’: (a) its cross-section, (b) Carnot cycle, and (c) an aerial photo showing the size of a hurricane next to the US state of Florida. In (b), $A \rightarrow B$: Air moves along the warm ocean surface at constant temperature, taking up heat (heat source); $B \rightarrow C$: the hot air rises up the eye of the hurricane, rapidly and adiabatically, and gets flung away from the center; $C \rightarrow D$: the air dumps heat at high altitudes (cold exhaust); $D \rightarrow A$: cold air falls back to Earth, adiabatically. This engine drives the spinning of the air. Source: (b) redrawn from KA Emanuel, *Nature* **326**, 483–485 (1987). (c) Courtesy Kerry Emanuel.

hurricane covers a significant fraction of the Earth’s surface (Figure 7.6). Because the rotational speed of the Earth at the equator is greater than at more northern latitudes, there is a natural counterclockwise rotation of the winds in the northern hemisphere, clockwise in the southern hemisphere.

The thermodynamic driving force. Why does a hurricane drive the winds so fast? A hurricane is a powerful and systematic engine, fueled by the heat from warm ocean water. Here’s how it works.

Imagine a small fictitious volume element of air containing N molecules of air. Let’s call this volume element a ‘balloon.’ This balloon is our thermodynamic system: it can take up and give off water vapor and heat, change

its temperature, and expand or contract. We will follow this balloon through a full cycle as it moves around in the hurricane, comparing the hurricane in Figure 7.6 with the four idealized Carnot cycle steps in Figure 7.5.

Step $A \rightarrow B$: Small balloon takes up heat at the ocean surface. The volume element of air starts at the ocean's surface, with small volume V_A (high density). The ocean is a thermal reservoir of high temperature T_h . The balloon takes up heat q_h by absorbing warm water vapor isothermally from the ocean. The balloon expands to a larger volume $V_B = V_A \exp(q_h/NkT_h)$ (Equation (7.22)).

Step $B \rightarrow C$: Balloon expands adiabatically and rises. Because the balloon now has a lower density (larger volume), it rises up through the eye of the hurricane (hot air rises). This happens quickly, approximately adiabatically (Figures 7.5 and 7.6(b)). The balloon rises until it reaches an altitude at which the surrounding air has equivalently low density; then it stops rising. The temperature of the balloon drops from T_h to T_c .

Step $C \rightarrow D$: Large balloon releases heat at high altitude. The high altitude is a thermal reservoir at low temperature T_c . The balloon now dumps heat q_c into the surrounding cold air, approximately isothermally. The density of air in the balloon increases, poised the balloon to fall back to Earth. The balloon shrinks.

Step $D \rightarrow A$: Balloon falls back to Earth. The dense balloon now falls back to Earth (cold air falls), increasing its temperature as it approaches the ocean again, ready to start the cycle again. In this way, the thermal power source for the cyclic engine of a hurricane is the warm ocean.

Aerodynamics. The wind enters the hurricane at the bottom and flows out the top. As each fictitious balloon rises up in the eye of the hurricane in step $B \rightarrow C$ of Figure 7.6, the balloon leaves behind a low-pressure area, sucking in more surface balloons from surrounding areas above the ocean surface. The balloons are blown out at the top of the hurricane, forcing other high-altitude balloons radially away from the center. Because of conservation of angular momentum, slow rotation of an air mass of large radius at high altitudes converts to much faster rotation of the air masses of smaller radii at the surface of the ocean.

Figure 7.7 shows an *Otto cycle*, which illustrates how an internal combustion engine converts heat to work.

The Maximum Possible Work Is Achieved via Reversible Processes

Now, let's define reversibility more quantitatively. We consider the changes in a system and its surroundings in the process of a thermodynamic cycle. Figure 7.8 defines some terms: (1) f refers to some process $A \rightarrow B$ of a system, which we call forward; (2) f' refers to how the surroundings change as the

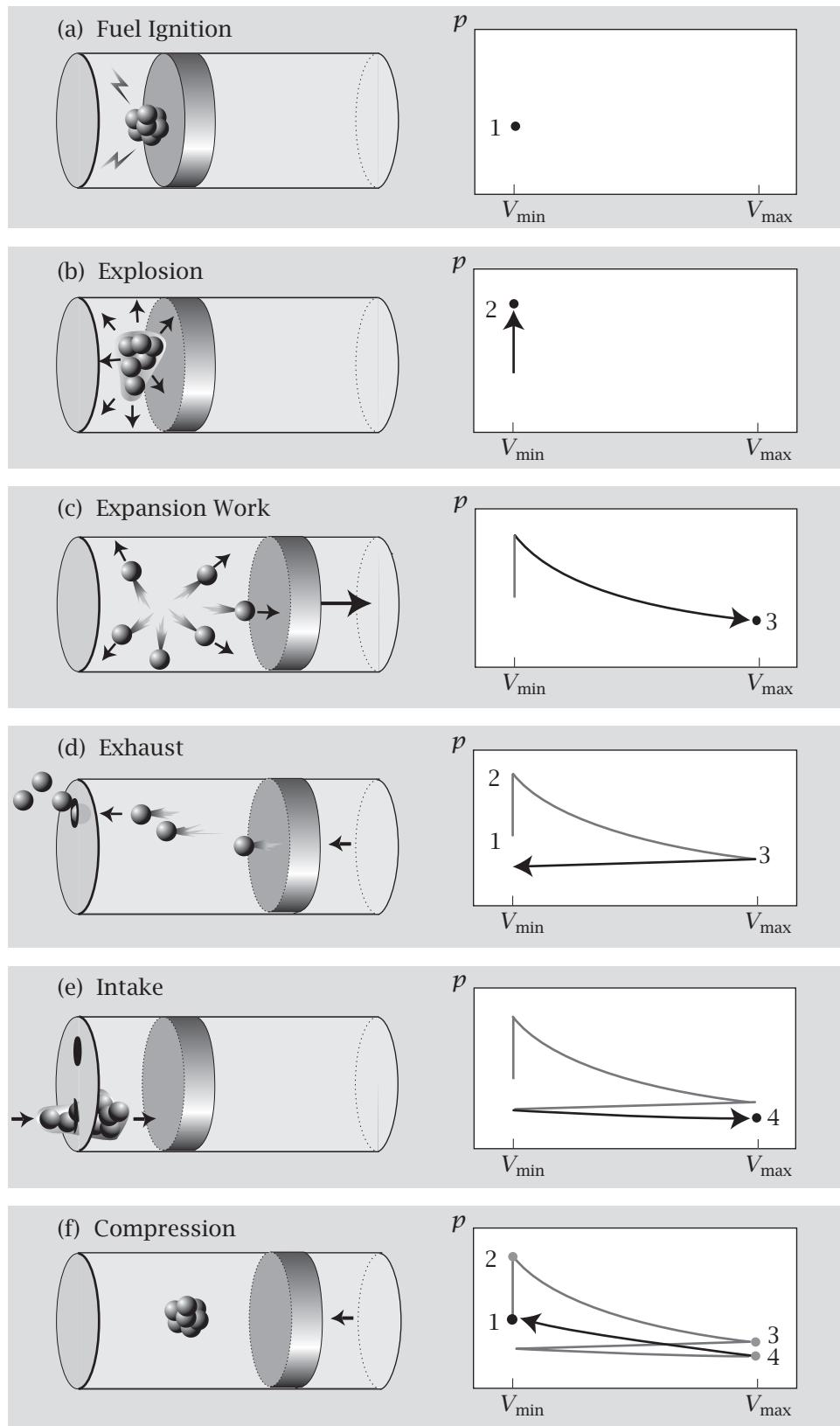


Figure 7.7 The Otto cycle (see Example 7.16). Combustion is complex. But a simple approximation is that the working fluid in the cylinder is air (not the combusting gases). The air is heated by the combustion of the fuel mixture.

(a) The fuel is ignited by a spark.

(b) Combustion of the fuel introduces heat into the cylinder. At this step, the combustion chamber has a small and constant volume. Burning the fuel converts it from a liquid to a gas and increases the pressure. $w_b = 0$, and $q_b = \Delta U_b = C_V(T_2 - T_1)$.

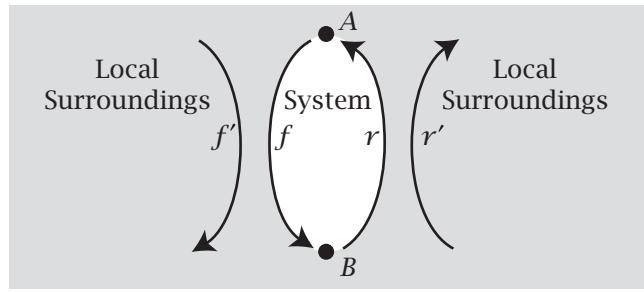
(c) The gas expands to perform work, increasing the volume and decreasing the pressure. The expansion is so rapid that it is approximately adiabatic. $q_c = 0$, and $w_c = \Delta U_c = C_V(T_3 - T_2)$.

(d) After the gas has expanded to the maximum volume V_2 allowed by the piston, the exhaust valve opens, and exhaust gases are forced out, reducing the volume and releasing heat.

(e) Fresh fuel is taken in. Little net work is performed on the cylinder to do this. For the combined steps (d) and (e), $w_{de} = 0$, and $q_{de} = \Delta U_{de} = C_V(T_4 - T_3)$.

(f) The fresh fuel is compressed by the action of the other cylinders. Like step (c), this step is so rapid that it is approximately adiabatic. $q_f = 0$, and $w_f = \Delta U_f = C_V(T_1 - T_4)$. Then the cycle repeats.

Figure 7.8 A cyclic process has a *forward* pathway $A \rightarrow B$ (called f) of the system and a corresponding forward process in the surroundings (f'). There is also a return pathway $B \rightarrow A$ (called r) and its corresponding process in the surroundings (r').



system undergoes the process f ; (3) r refers to some return process $B \rightarrow A$ of the system; and (4) r' refers to the corresponding process in the surroundings as the system returns to A . Because the entropy is a state function, it must sum to zero for the system around a cycle:

$$\Delta S_f + \Delta S_r = 0. \quad (7.25)$$

You get an additional relationship from the Second Law. If the system can exchange energy with its local surroundings and if the system plus local surroundings do not also exchange energy more globally (with the so-called *universe*; see Example 6.2), then the Second Law says that as the system approaches equilibrium, the total entropy of the system plus surroundings must increase:

$$\Delta S_{\text{total}} = \Delta S_f + \Delta S_{f'} \geq 0, \quad (7.26)$$

where the equality holds only at equilibrium. The Second Law also gives

$$\Delta S_r + \Delta S_{r'} \geq 0. \quad (7.27)$$

So far, we have three equations, (7.25)–(7.27), in the four unknown entropies. We need a fourth relationship. We invoke the definition of reversibility, $\Delta S = 0$, that a cyclic process must return both the system and its surroundings back to their original states:

$$\Delta S = \Delta S_f + \Delta S_r + \Delta S_{f'} + \Delta S_{r'} = 0. \quad (7.28)$$

This now gives all the relationships you need to describe efficient processes (such as Carnot cycles). First, note that reversibility can be expressed in different equivalent ways. For example, combining Equations (7.25), (7.26), and (7.28) says that the surroundings are not changed in a reversible cyclic process:

$$\Delta S_{f'} + \Delta S_{r'} = 0. \quad (7.29)$$

Here are examples of reversible and irreversible processes.

EXAMPLE 7.12 The Carnot cycle is reversible. The Carnot cycle has four steps. Each step is given to be reversible. If all the steps of a cycle are reversible, the whole cycle is reversible. To see this, note that the system reversibility equation (7.28) follows immediately when there are equalities in the relationships (7.26) and (7.27). Here, we illustrate for the Carnot cycle (Figure 7.5) that $\Delta S = \Delta S_{AB} + \Delta S_{BC} + \Delta S_{CD} + \Delta S_{DA} = 0$.

Two steps are adiabatic, so they involve no change in entropy, $\Delta S_{\text{adiabatic}} = q/T = 0$. Two steps are isothermal. One isothermal step (AB in Figure 7.5) occurs

at the hot temperature T_h . Compute the entropy for that step by combining Equations (7.8) and (7.16) to get

$$\Delta S_{AB} = \frac{q_h}{T_h} = Nk \ln\left(\frac{V_B}{V_A}\right). \quad (7.30)$$

The other isothermal step (*CD* in Figure 7.5) has entropy

$$\Delta S_{CD} = \frac{q_c}{T_c} = Nk \ln\left(\frac{V_A}{V_B}\right), \quad (7.31)$$

since $V_B/V_A = V_C/V_D$ (see Equation (7.23)). Summing these four terms around the cycle gives $\Delta S = 0$, showing that the Carnot cycle is reversible.

In contrast, the free expansion of a gas, like the leaking of a balloon, is irreversible.

EXAMPLE 7.13 Proving that the free expansion of a gas is irreversible. A gas expanding against a pressure can perform reversible work. But in *free expansion*, also called *Joule expansion*, a gas expands against zero applied pressure, performing no work (substitute $p = 0$ into $\delta w = -p dV$). Here is the experiment first performed in 1852 by JP Joule, a British beer brewer and amateur physicist. He performed free expansion adiabatically, $q = 0$. Because $w = 0$ in this case, you also have $\Delta U = C_V \Delta T = 0$, so this free expansion occurs isothermally, $\Delta T = 0$. To compute the entropy change in free expansion at constant temperature, use a reversible process, Equation (7.30), to get $\Delta S_{\text{system}} = Nk \ln(V_2/V_1)$, which is greater than zero because $V_2 > V_1$. Because the process is adiabatic, the environment provided no heat, so $\Delta S_{\text{surroundings}} = q_{\text{rev}}/T = 0$. Therefore, free expansion of a gas leads to a total increase in entropy

$$\Delta S_{\text{total}} = S_{\text{system}} + \Delta S_{\text{surroundings}} > 0 \quad (7.32)$$

and therefore is irreversible.

EXAMPLE 7.14 Examples of irreversible processes.

Diffusion and mixing. A drop of dye expanding in water is irreversible. Example 2.3 shows that $\Delta S_{\text{system}} > 0$ when molecules diffuse or mix, even when there is no change in the surroundings, so $\Delta S_{\text{total}} > 0$.

Heat flow from a hot object to a cold object. Example 3.9 shows that $\Delta S_{\text{system}} > 0$ when heat flows from a hot to a cold object, even when there is no change in the surroundings of the two objects, so $\Delta S_{\text{total}} > 0$.

Friction. A gas expanding in a piston can raise a weight through some distance. But if there is friction, it will only lift the weight through a lesser distance, performing less work, heating the surroundings, and leading to $\Delta S_{\text{total}} > 0$.

Other irreversible processes. Other processes, such as stretching clay, or dinosaurs decomposing over the eons into hydrocarbons, can be slow, but are not reversible. A biological cell grows irreversibly. Boiling an egg tangles its proteins together irreversibly. All spontaneous processes of fluid flows, heat flows, and material flows are irreversible.

EXAMPLE 7.15 Why do engines waste heat? Why can't you design an engine that converts 100% of the heat input to work output? This question was resolved by S Carnot in the 1820s, when he determined the maximum efficiency of a heat engine by the following logic. Figure 7.4 shows the three steps of the cycle: taking in heat q_h at fixed high temperature, T_h ; performing work w adiabatically; and expelling heat q_c at fixed low temperature, T_c . The energy, a state function, sums to zero around the cycle, $\Delta U = 0$, so the First Law gives

$$w = q_h - q_c. \quad (7.33)$$

Define the *efficiency* η of the system as the work output divided by the heat input:

$$\eta = \frac{w}{q_h} = 1 - \frac{q_c}{q_h}. \quad (7.34)$$

The Second Law gives the entropy change toward equilibrium of the whole system, ΔS_{total} , as the sum of the three components: ΔS_h from the hot gas, ΔS_c from the cooled gas, and ΔS_w from the work source. Because the work is performed adiabatically, $\Delta S_w = q_w/T = 0$. In a reversible cycle,

$$\Delta S_{\text{total}} = \Delta S_h + \Delta S_c = 0. \quad (7.35)$$

For reversible heat transfers, the entropy changes of the gas are

$$\Delta S_h = \frac{q_h}{T_h} \quad \text{and} \quad \Delta S_c = \frac{-q_c}{T_c}. \quad (7.36)$$

Substitute Equations (7.36) into (7.35) and rearrange to get

$$\frac{q_c}{q_h} = \frac{T_c}{T_h}. \quad (7.37)$$

The maximum possible efficiency is

$$\eta = 1 - \frac{T_c}{T_h}. \quad (7.38)$$

Equation (7.38) implies that heat is most efficiently converted to work by engines that bring in heat at the highest possible temperature and exhaust it at the lowest possible temperature. The calculation of the efficiency of a reversible engine answered two historically important questions. First, it said that the efficiency of a reversible engine depends only on the temperatures of the heat intake and exhaust, and not on the type of gas or the volume of the piston. Second, the only way to reach 100% efficiency is to exhaust the waste heat at $T_c = 0$ K. The inability to achieve 100% efficiency in engines is *not* a consequence of friction or turbulence. We are assuming slow non-dissipative processes here. Since $T_c/T_h > 0$, a heat engine has an inherent inefficiency. Now, let's compute the efficiency of a car engine.

EXAMPLE 7.16 The efficiency of an internal combustion engine. Like a steam engine, an internal combustion engine is a heat engine, converting heat to work in a cyclic process. You can compute the maximum efficiency of an

internal combustion engine from the intake and exhaust temperatures and the heat capacity of the fuel vapor, assuming certain idealizations: (1) although the cylinder contains a complex mixture of fuel and air, the working fluid is assumed to be the air, which is caused to expand by the burning fuel, (2) combustion is viewed as a process of heating the air from an external source, (3) the working fluid is an ideal gas, and (4) work is performed during adiabatic reversible steps. The steps of the cycle are shown in Figure 7.7.

The efficiency η of the idealized internal combustion engine is the net work performed *by* the system ($-w$) for the given heat influx:

$$\eta = \frac{-(w_f + w_c)}{q_b} = \frac{-C_V(T_1 - T_4) - C_V(T_3 - T_2)}{C_V(T_2 - T_1)} = 1 - \frac{T_3 - T_4}{T_2 - T_1}. \quad (7.39)$$

The subscripts b , c , and f in Equation (7.39) refer to the steps in the Otto cycle in Figure 7.7. The subscripts 1–4 refer to the points in Figure 7.7 at which the temperatures can be calculated. Assuming ideal adiabatic steps (c) and (f), we can compute the temperatures by using Equation (7.21),

$$\frac{T_3}{T_2} = \left(\frac{V_1}{V_2}\right)^{Nk/C_V} \quad \text{and} \quad \frac{T_4}{T_1} = \left(\frac{V_1}{V_2}\right)^{Nk/C_V}, \quad (7.40)$$

so $T_3 = T_2(V_1/V_2)^{Nk/C_V}$ and $T_4 = T_1(V_1/V_2)^{Nk/C_V}$. Therefore the difference you need for Equation (7.39) is given by $T_3 - T_4 = (T_2 - T_1)(V_1/V_2)^{Nk/C_V}$. Substitution into Equation (7.39) gives

$$\eta = 1 - \left(\frac{V_1}{V_2}\right)^{Nk/C_V}. \quad (7.41)$$

The efficiency of an Otto-cycle engine depends on the compression ratio and on the heat capacity of the vaporized fuel. Typical compression ratios for gasoline engines range from 4 to 10.

This example shows two main features of thermodynamic cycles. First, because the internal energy is a state function, it sums to zero around the cycle: you can check that $\Delta U_b + \Delta U_c + \Delta U_{de} + \Delta U_f = C_V[(T_2 - T_1) + (T_3 - T_2) + (T_4 - T_3) + (T_1 - T_4)] = 0$. Second, work and heat are not state functions and do not sum to zero around the cycle. The engine performs work on each cycle. This is evident either from computing the sum $w_b + w_c + w_{de} + w_f$ or from the graph of the pressure versus volume in Figure 7.7. Because work is the integral of $-p dV$, the total work is the area inside the cycle.

For a typical automobile engine, the Carnot (maximum) efficiency is about 50–60%, but the typical efficiency is closer to 25%. A considerable fraction of the heat input is lost to the environment as heat and is not converted to work. The problem is that perfect efficiency would require extracting all the motion and all the internal energy from the gas molecules, $T_c = 0\text{ K}$.

The Second Law helps you understand inefficiencies in converting heat to work. Our statement of the Second Law is that isolated systems tend toward their states of maximum entropy. For example, heat does not spontaneously flow from cold to hot objects, because that would decrease the entropy. An alternative early statement was that heat will not flow from a colder body to a hotter one without the action of some external agent. For example, refrigerators

require electrical energy to move heat from cold places (inside the refrigerator) to warmer places (the rest of the kitchen). Spontaneous processes in isolated systems can lead to an increase in the entropy. But, a system can also decrease its entropy when heat is allowed to flow across its boundary.

Why Is There an *Absolute* Temperature Scale?

How should we define a practical measure of temperature? Three scales are popular: Kelvin, Celsius, and Fahrenheit. The Kelvin scale is the most fundamental because it defines *absolute* temperature T . The idea of an absolute temperature scale derives from Equation (7.38), which relates the efficiency of converting heat to work to a ratio of temperatures, T_c/T_h . An absolute zero of the temperature scale, $T_c = 0$, is the point at which heat is converted to work with 100% efficiency, $\eta = 1$. This defines the zero of the Kelvin scale.

Other scales of temperature T' are ‘shifted’ so that their zero points coincide, for example, with the freezing point of water (Celsius scale) or with a point of coexistence of salt, water, and ice (Fahrenheit),

$$T' = aT - b,$$

where a and b are constants. For example, for the Celsius scale, $a = 1$ and $b = 273.15$. On these scales, the efficiency of a Carnot engine approaches some value other than one as the cold temperature approaches the zero of that scale, $T'_c \rightarrow 0$.

Summary

This chapter explored the principles of pathways and cycles in thermodynamics. By measuring work and heat in slow processes, you can compute state functions, such as energy and entropy. State functions sum to zero around thermodynamic cycles. But path functions, like heat and work, do not sum to zero around cycles. This is why engines can convert heat to work. The Second Law, combined with the idea of reversible cycles, helps to explain why it is so difficult to convert heat to work with high efficiency, a problem that is central to the world’s energy needs. In the next chapter, we will introduce enthalpies and free energies, and move from heat engines and historical engineering problems to laboratory experiments in chemistry and biology.

Problems

1. The work of compression. One mole of a van der Waals gas is compressed quasi-statically and isothermally from volume V_1 to V_2 . For a van der Waals gas, the pressure is

$$p = \frac{RT}{V-b} - \frac{a}{V^2},$$

where a and b are material constants, V is the volume, and RT is the gas constant \times temperature.

- (a) Write the expression for the work done.
- (b) Is more or less work required in the low-density limit than for an ideal gas? What about the high-density limit? Why?

2. Deriving the ideal gas law in two dimensions. Molecules at low density on a surface, such as surfactants at an interface of air and water, often obey a two-dimensional equivalent of the ideal gas law. The two-dimensional equivalent of p is π , where π is a lateral two-dimensional pressure. A is area. Using

$$\pi = T \left(\frac{\partial S}{\partial A} \right)_N \quad (7.42)$$

and assuming no energetic interactions, derive the two-dimensional equivalent of the ideal gas law by using a lattice model in the low-density limit.

3. The work of expansion in freezing an ice cube. At 1 atm, freeze an amount of liquid water that is $2\text{ cm} \times 2\text{ cm} \times 2\text{ cm}$ in volume. The density (mass per unit volume) of liquid water at 0°C is 1.000 g cm^{-3} and the density of ice at 0°C is 0.915 g cm^{-3} .

- (a) What is the work of expansion upon freezing?
- (b) Is work done *on* the system or *by* the system?

4. The work of expanding a gas. Compute the total work performed when expanding an ideal gas, at constant temperature, from volume V to $2V$.

5. Pulling out the partition between two chambers of a gas. A partition separates equal volumes containing equal numbers N of ideal gas molecules of the same species at the same temperature. Using a simple lattice model for ideal gases, evaluate the relevant multiplicity terms to show that the entropy of the composite system does not change when the partition is removed. (Hint: use Stirling's approximation from Appendix B.)

6. The work in a thermodynamic cycle. A thermodynamic cycle is a series of steps that ultimately returns to its beginning point. Compute the total work performed around the thermodynamic cycle of quasi-static processes in Figure 7.9.

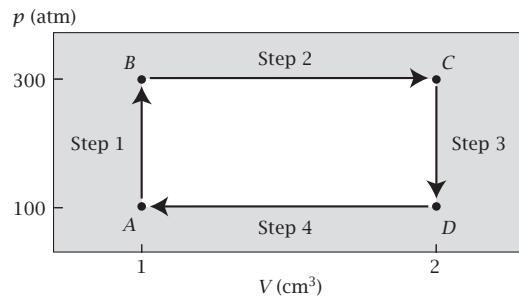


Figure 7.9

7. Engine efficiencies. Consider a Carnot engine that runs at $T_h = 380\text{ K}$.

- (a) Compute the efficiency if $T_c = 0^\circ\text{C} = 273\text{ K}$.
- (b) Compute the efficiency if $T_c = 50^\circ\text{C} = 323\text{ K}$.

8. Hadley Cycles—what powers the trade winds? The Earth's trade winds arise from the differences in buoyancy between hot air and cold air. Air is heated at the Earth's surface near the equator (see (1) in Figure 7.10), lowering its density; the air rises (2), pushing upper air away toward the Northern latitudes, where the air cools (3); then the air drops back down to Earth (4), pushing the cold surface air back toward the equator.

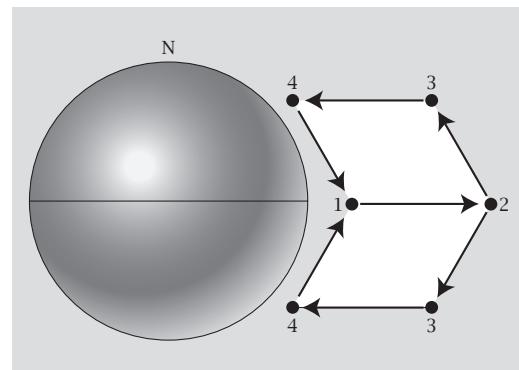


Figure 7.10

Consider an imaginary balloon containing 1 m^3 of an ideal gas.

- (a) At $p = 1\text{ atm}$ and $T = 300\text{ K}$, what is the number of moles n of air contained in the balloon?
- (b) If that balloon of n moles of air remains at $p = 1\text{ atm}$ but is now heated to $T = 330\text{ K}$, its volume increases. What is the new density $\rho = n/V$?
- (c) Assuming no heat transfer, how high will the balloon in part (b) rise? Use Figure 10.2 to make a rough guess (Useful conversion: $1\text{ atm} \approx 1\text{ bar}$).

9. Stirling engine. A Stirling engine has the pV cycle shown in Figure 7.11 and uses an ideal gas working fluid. The steps are quasi-static.

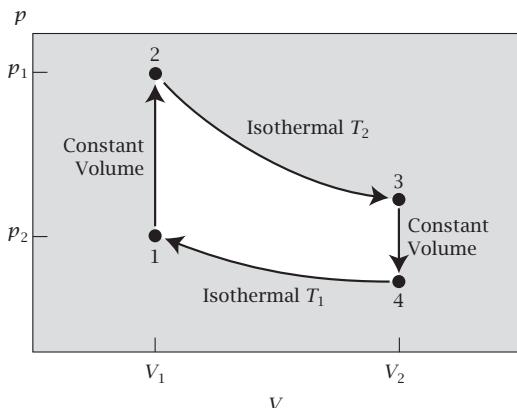


Figure 7.11

- How much work is done in the constant-volume segments, w_{12} and w_{34} ?
- What is the energy change around the cycle, $\Delta U_{\text{tot}} = \Delta U_{12} + \Delta U_{23} + \Delta U_{34} + \Delta U_{41}$?
- What is the total work performed around the cycle, $w_{\text{tot}} = w_{12} + w_{23} + w_{34} + w_{41}$?
- If $T_2 = 2000\text{ K}$, $T_1 = 300\text{ K}$, $V_2 = 1\text{ L}$, $V_1 = 0.01\text{ L}$, and $n = 0.1\text{ mol}$ of ideal gas, compute w_{tot} .

- 10. Ideal efficiency of a car engine.** Suppose the compression ratio in your car engine is $V_2/V_1 = 8$. For a diatomic gas, $C_V = (5/2)Nk$ and for ethane, $C_V \approx 5Nk$.
- What is the efficiency of your engine if the working fluid is a diatomic gas?
 - Which is more efficient: a diatomic gas or ethane?
 - Would your engine be more or less efficient with a higher compression ratio?

Suggested Reading

CJ Adkins, *Equilibrium Thermodynamics*, 3rd edition, Cambridge University Press, Cambridge, 1983. Excellent introduction to thermodynamics and reversibility.

SG Brush, *The Kind of Motion We Call Heat*, Volumes 1 and 2, North-Holland, New York, 1976. An excellent history of thermodynamics and statistical mechanics.

HB Callen, *Thermodynamics and an Introduction to Thermostatistics*, 2nd edition, Wiley, New York, 1985. The classic text on the axiomatic approach to thermodynamics. The logic of thermodynamics is explained with great clarity.

G Carrington, *Basic Thermodynamics*, Oxford University Press, Oxford, 1994. Good general text with many engineering examples.

K Denbigh, *The Principles of Chemical Equilibrium*, 4th edition, Cambridge University Press, Cambridge, 1981. General text on thermodynamics.

IM Klotz and RM Rosenberg, *Chemical Thermodynamics: Basic Concepts and Methods*, 7th edition, Wiley, New York, 2008. Excellent simple and general text.

HS Leff and AF Rex, eds, *Maxwell's Demon: Entropy, Information, Computing*, Princeton University Press, Princeton, 1990. Essays on the relationship between entropy and information theory.

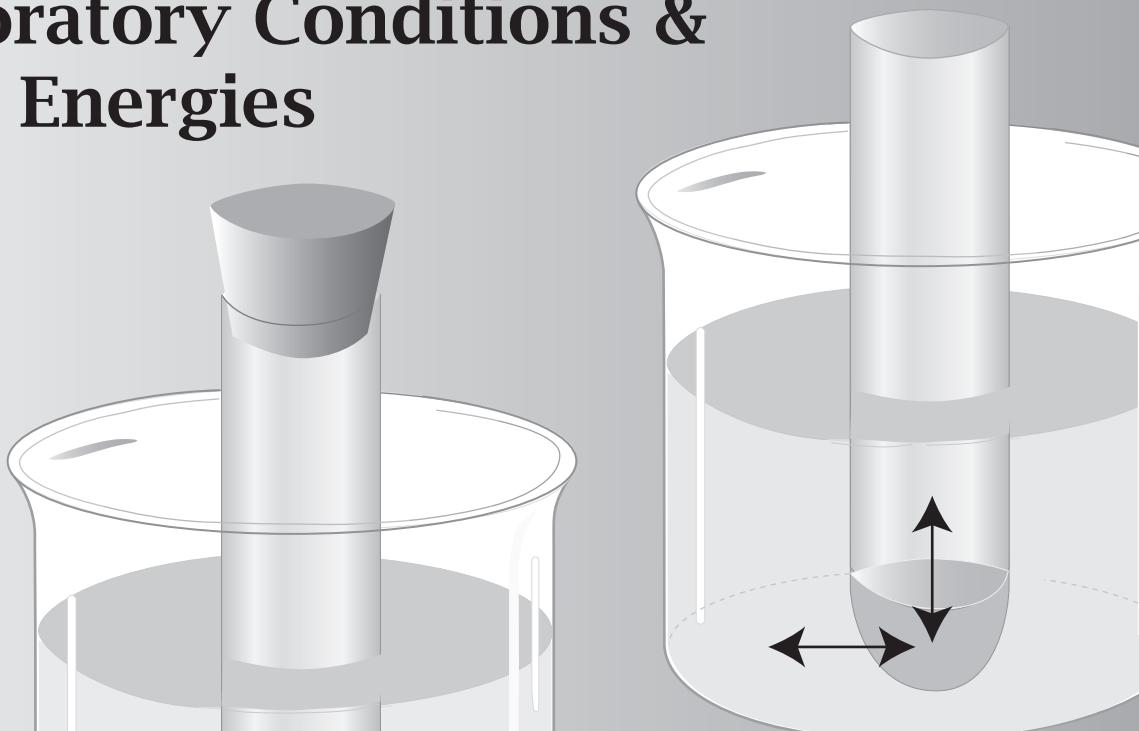
HS Leff and AF Rex, eds, *Maxwell's Demon 2: Entropy, Classical and Quantum Information, Computing*, CRC Press, Boca Raton, FL, 2002. Essays on the relationship between entropy and information theory.

M Samiullah, *Am J Phys* 75, 608–609 (2007). Succinct statement of the distinction between quasi-static and reversible processes.

EB Smith, *Basic Chemical Thermodynamics*, 5th edition, Imperial College Press, London, 2004. Excellent simple and general text.

MW Zemansky and RH Dittman, *Heat and Thermodynamics*, 7th edition, McGraw-Hill, New York, 1996. Excellent general text; extensive comparisons with experiments and detailed discussion of reversibility.

8 Laboratory Conditions & Free Energies



Switch from Maximum Entropy to Minimum Free Energy

In Chapters 6 and 7, we reasoned with the principle that systems tend toward states of maximum entropy. We considered systems with known energy exchange across their boundaries. That logic helped explain gas expansion, particle mixing, and the interconversion of heat and work in engines. If we continued no further, however, we would miss much of the power of thermodynamics for physics, chemistry, and biology. For processes in test tubes in laboratory heat baths, or processes open to the air, or processes in biological systems, it is not the *work* or *heat flow* that you control at the boundaries. It **is the temperature** and **the pressure**. This apparently slight change in conditions actually requires new thermodynamic quantities—the **free energy** and the **enthalpy**—and reformulated extremum principles. Systems held at constant temperature do not tend toward their states of **maximum entropy**. They tend toward their states of **minimum free energy**.

Why Introduce New Independent Variables?

In Chapter 6, we were interested in problems for which U , V , and N were the independent variables. Independent variables represent quantities that you can measure or control at the boundary of the system. You choose independent

variables based on the type of boundaries enclosing the system. So far, we have only considered systems having boundaries that controlled heat flow, work, or energy changes. In Example 6.2, you are given two conditions: between the two objects, the energy can freely exchange, $U_{\text{total}} = U_A + U_B = \text{constant}$, and between the system and its surroundings, no energy exchanges. The Carnot and Otto engine cycles (Examples 7.10, 7.11, and 7.16) involve boundaries across which we know the heat and work flows. In all these cases, U or q or w were known across the boundaries.

However, when an *intensive* variable, such as T , p , or μ , is controlled or measured at the boundary, it means that the conjugate variables, energy U , volume V , or particle number N of the system, are not the quantities that are known or controlled. In such cases, U , V , or N can exchange freely back and forth across the boundary with the *bath*, i.e., the external reservoir that is large enough that it can hold T , p , or μ fixed, no matter what happens in the system. Such exchanges are called *fluctuations*. When T is constant, heat can exchange between the system and the surroundings (the heat bath), so the energy of the system fluctuates. Constant p implies an action like a piston stroke through which the system can exchange volume with the surroundings. In that case, the volume of the system fluctuates. Constant μ implies that a *particle bath* is in contact with the system. Particles leave or enter the system to and from the particle bath. In this case, the number of particles in the system can fluctuate.

Consider a process in a *system* that we will call the *test tube*, immersed in a *heat bath*. The system need not be a *real test tube in a water jacket*. The system could be molecules in a liquid or in the air; it just needs to be distinguished from its surroundings in some way. *Heat bath* refers to any surroundings of a system that hold the temperature of the system constant. A heat bath controls the temperature T , not the energy U , at the boundary around the subsystem. If the combined test tube plus heat bath are isolated from the greater surroundings, equilibrium will be the state of maximum entropy for the total system. However, we are not interested in the state of the total system. We are interested in what happens in the test tube itself. We need a new extremum principle that applies to the test tube, where the independent variables are (T, V, N) .

If the extremum of a function such as $S(U)$ predicts equilibrium, the variable U is called the *natural variable* of S . T is not a natural variable of S . Now we show that (T, V, N) are natural variables of a function F , the *Helmholtz free energy*. An extremum in $F(T, V, N)$ predicts equilibria in systems that are constrained to constant temperature at their boundaries.

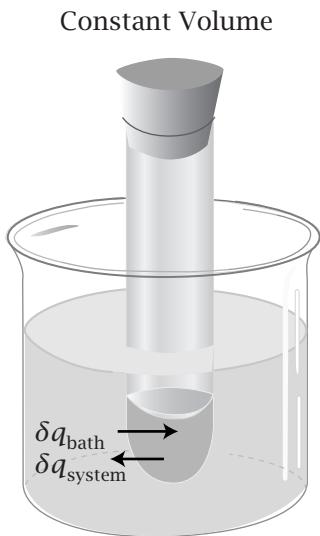


Figure 8.1 A heat bath is a reservoir that holds the system (the test tube in this case) at constant temperature by allowing heat flow in or out, as required. The properties that do not change inside the system are temperature T , volume V , and particle number N , denoted (T, V, N) . The condition for equilibrium inside the test tube is that the Helmholtz free energy $F(T, V, N)$ is at a minimum.

Free Energy Defines Another Extremum Principle

The Helmholtz Free Energy

Consider a process inside a test tube, sealed so that it has constant volume V and no interchange of its N particles with the surroundings (see Figure 8.1). A heat bath holds the test tube at constant temperature T . The process inside the test tube might be complex. It might vary in rate from a quasi-static process to an explosion. It might or might not involve chemical or phase changes. It might give off or absorb heat. Processes within the test tube will influence the heat

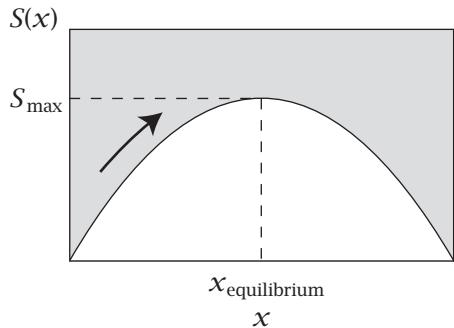


Figure 8.2 A system moving toward its equilibrium value of a degree of freedom $x = x_{\text{equilibrium}}$ increases its entropy to $S(x) = S_{\text{max}}$.

bath only through heat exchange, because its volume does not change and no work is done.

Now we reason with the First and Second Laws of Thermodynamics to find an expression that describes the condition for equilibrium in terms of the changes in the test tube alone.

If the *combined system* (the subsystem plus the heat bath) is isolated from its surroundings, equilibrium will be the state of maximum entropy $S(U, V, N)$ of the combined system. (Our strategy is to assume that the whole combined system is isolated from the broader surroundings. In the end, we will obtain properties of only the test tube system, so we will find that it does not matter how the bath interfaces with the greater surroundings.) Any change toward equilibrium must increase the entropy of the combined system, $dS_{\text{combined system}} \geq 0$ (see Figure 8.2).

Because the entropy is extensive, the entropy change of the total system is the sum of the entropy changes of its parts:

$$dS_{\text{combined system}} = dS_{\text{system}} + dS_{\text{bath}} \geq 0, \quad (8.1)$$

where the subscript ‘system’ indicates the test tube contents. Since the combined system is isolated,

$$dU_{\text{bath}} + dU_{\text{system}} = 0. \quad (8.2)$$

Our aim is to relate dS_{bath} to some property of the test tube system. Use the fundamental equation, $dS_{\text{bath}} = (1/T) dU + (p/T) dV - (\mu/T) dN = (1/T) dU_{\text{bath}}$ for a bath in which V and N are constant. Combine this with Equation (8.2) to get

$$dS_{\text{bath}} = -\frac{dU_{\text{system}}}{T}. \quad (8.3)$$

Substitute Equation (8.3) into Equation (8.1) to get

$$dS_{\text{system}} - \frac{dU_{\text{system}}}{T} \geq 0 \implies dU_{\text{system}} - T dS_{\text{system}} \leq 0. \quad (8.4)$$

You now have an expression describing the approach to equilibrium in terms of the test tube subsystem alone. Define a quantity F , the **Helmholtz free energy**:

$$F = U - TS. \quad (8.5)$$

Its differential is

$$\begin{aligned} dF &= dU - T dS - S dT \\ &= dU - T dS \quad \text{at constant temperature.} \end{aligned} \quad (8.6)$$

Comparison of Equation (8.6) with Equation (8.4) shows that when a system in which (T, V, N) are constant is at equilibrium, the quantity F is at a minimum.

The definition $F = U - TS$ shows that F is determined by a balance between internal energy and entropy, and that the position of the balance is determined by the temperature. To minimize its Helmholtz free energy, the system in the test tube will tend toward both high entropy and low energy, depending on the temperature. At high temperatures, the entropy dominates. At low temperatures, the energy dominates. Let's return to simple lattice models to illustrate the Helmholtz free energy.

EXAMPLE 8.1 How to use free energies: a model of 'dimerization.' Suppose that $N = 2$ gas particles are contained in a test tube having a volume of V lattice sites arranged in a row, at temperature T . Under what conditions will the two particles associate into a dimer? Because systems at constant (T, V, N) seek their states of minimum free energy, we compute the free energy of the two states—the dimer state and the dissociated monomer state—and we compare them. Whichever state has the lower free energy is the stable (equilibrium) state.

Dimer: Suppose that the dimer (shown in Figure 8.3) is held together by a 'bond energy' $U = -\varepsilon$ (where $\varepsilon > 0$). That is, when the two monomers are sitting on adjacent sites, there is an attractive energy between them. For now, we will not be concerned about whether this bond is covalent or due to weaker interactions. At this level of simplicity, the model is equally applicable to many different types of bond.

On a linear lattice of V sites, there are $W_{\text{dimer}} = V - 1$ possible placements of a dimer. (The first monomer could be in site 1, or 2, or 3, ..., or $V - 1$, but not in site V because then the second monomer would be off the end of the lattice. To keep it simple, let's assume a large lattice, $V - 1 \approx V$). Since $S = k \ln W$, the Helmholtz free energy for the dimer is

$$F_{\text{dimer}} = U_{\text{dimer}} - TS_{\text{dimer}} = -\varepsilon - kT \ln V. \quad (8.7)$$

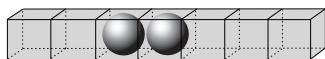


Figure 8.3 In the dimer state, the two particles of a one-dimensional gas are adjacent.

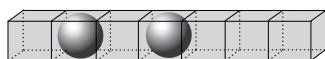


Figure 8.4 In the dissociated state, the two particles are not adjacent.

Monomers: The two dissociated monomers (shown in Figure 8.4) have no energy of interaction. The number of ways in which two particles can be placed on V sites, *not adjacent* to each other (because adjacent placements are the dimer states) is

$$W_{\text{monomer}} = W_{\text{total}} - W_{\text{dimer}} = \frac{V!}{(2!(V-2)!)} - (V-1) = \left(\frac{V}{2} - 1\right)(V-1) \approx \frac{V^2}{2},$$

and the Helmholtz free energy of the monomer-pair state is

$$\begin{aligned} F_{\text{monomer}} &= U_{\text{monomer}} - TS_{\text{monomer}} = -TS_{\text{monomer}} \\ &= -kT \ln \left(\frac{V^2}{2}\right). \end{aligned}$$

Figure 8.5 shows the free energies for monomers and dimer as a function of temperature. To determine which state is more stable, find the state with the lowest free energy. In this case, you choose between two options (associated or dissociated), rather than finding where a derivative is zero. Figure 8.5 shows that the dimer is the more stable state at low temperatures, while dissociation

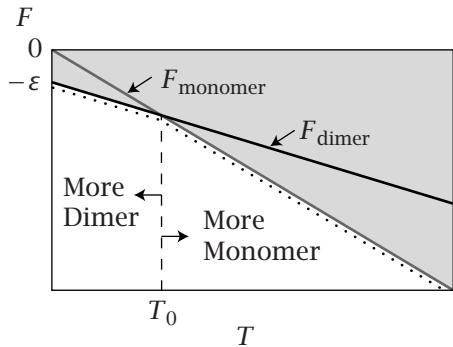


Figure 8.5 Free energy F versus temperature T for dimer association. The dotted line shows that more molecules are dimers at low temperature ($F_{\text{dimer}} < F_{\text{monomer}}$), and more are monomers at high temperature ($F_{\text{monomer}} < F_{\text{dimer}}$). For this simple model to correspond to the physical meanings of monomer and dimer, it must have more than four sites, $V > 4$.

into monomers is favored at high temperatures. The monomers and dimer are equally stable (have the same free energy) at the temperature $T = T_0$:

$$\begin{aligned} -\varepsilon - kT_0 \ln V &= -kT_0 \ln \left(\frac{V^2}{2} \right) \\ \Rightarrow \quad \varepsilon &= kT_0 \ln \left(\frac{V}{2} \right) \\ \Rightarrow \quad T_0 &= \frac{\varepsilon}{k \ln(V/2)}. \end{aligned} \quad (8.8)$$

The stronger the particle attraction (the more positive ε is), the higher is the dissociation temperature, and the more thermal energy is required to break the bond. We will develop methods in Chapter 10 to compute the relative amounts of monomer and dimer as functions of temperature.

On the one hand, we have said that equilibrium is the state of minimum free energy when T is constant. But in Figure 8.5 we have plotted the free energy versus temperature, $F(T)$. Figure 8.5 describes a series of experiments. The temperature of each experiment is fixed.

Although this model is very simple, it represents the essentials of dimer dissociation and many other processes involving bond breakage, like boiling, melting, and the unfolding of polymers. The association-dissociation equilibrium is a balance between the energetic tendency of the particles to stick together, which dominates at low temperatures, and the entropic tendency of the particles to disperse, which dominates at high temperatures. If we had maximized the entropy instead of minimizing the free energy, we would have concluded that dimers should never form. We would have missed the importance of interaction energies in systems at lower temperatures.

Here is another example of free energies.

EXAMPLE 8.2 Free energies again: a toy model of polymer collapse. When a polymer molecule is put into certain solvents (called *poor solvents*), it collapses into a *compact structure* owing to attractions similar to those that cause monomers to dimerize. For example, *proteins fold to compact structures in water*. Consider a two-dimensional model polymer having *four monomers in a closed test tube solution in a water bath* (Figure 8.6). The monomer units of the polymer chain are attracted to each other by a *sticking energy* $-\varepsilon$, where

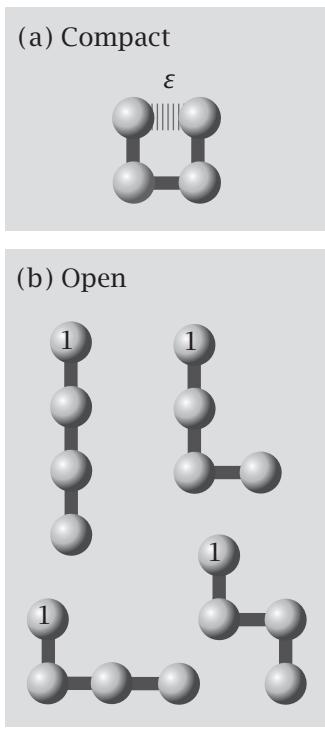
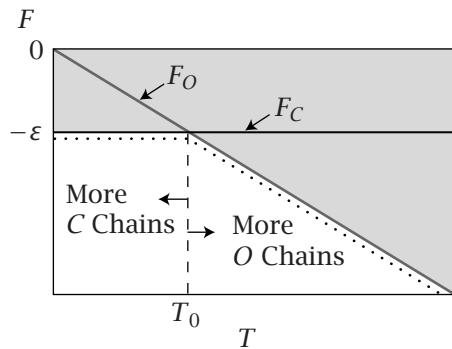


Figure 8.6 (a) The single compact (collapsed) conformation, and (b) the four open conformations of a toy model polymer. “1” indicates the first monomer. Other models distinguish mirror symmetric and rotated chains, leading to different numbers of conformations; the principles are the same.

Figure 8.7 Free energy F versus temperature T for the collapse of the four-unit toy model polymer shown in Figure 8.6. Open (O) chains are more stable at high temperature, while the compact (C) chain is more stable at low temperature. T_0 is the collapse temperature.



$\varepsilon > 0$. Suppose that experiments could distinguish between two conformational states: compact (Figure 8.6(a)) and open (Figure 8.6(b)). The single compact conformation ($W_C = 1$) is stabilized by the energy $U_C = -\varepsilon$ that attracts the ends to each other. The free energy of the compact state is $F_C = U_C - TS_C = -\varepsilon - kT \ln 1 = -\varepsilon$. The open state is the collection of the other four possible conformations, so $W_O = 4$. Open conformations have no pairs of monomers stuck together, $U_O = 0$, so $F_O = U_O - TS_O = -kT \ln 4$.

At low temperatures, the molecule is compact owing to the favorable sticking energy. At high temperatures, the molecule is unfolded owing to the favorable entropy. Figure 8.7 shows the free energies of the lattice polymer versus temperature. There is a collapse temperature, $T_0 = \varepsilon / (k \ln 4)$, at which there are equal populations of open and compact chains. Strong sticking energy (large $\varepsilon > 0$) implies a high collapse temperature.

Now consider the collapse process, the transformation from open to compact states. The free energy difference for this change is $\Delta F_{\text{collapse}} = F_C(T) - F_O(T) = -\varepsilon + kT \ln 4$. The energy change for this process is $\Delta U_{\text{collapse}} = -\varepsilon$. The entropy change is $\Delta S_{\text{collapse}} = -k \ln 4$.

Because the volume of the test tube is constant, a quasi-static collapse process in this model involves no work, so the First Law gives $\Delta U = q = -\varepsilon < 0$.

According to this model, the collapse process causes heat to be given off from the test tube to the bath. Processes that give off heat are called exothermic. Processes that take up heat are called endothermic. This model collapse process is exothermic at every temperature.

The Fundamental Equation for the Helmholtz Free Energy

Just as the functional form $S(U, V, N)$ implies a fundamental entropy equation for dS , the form $F(T, V, N)$ implies a fundamental equation for dF :

$$dF = d(U - TS) = dU - T dS - S dT. \quad (8.9)$$

Substitute the fundamental energy equation (6.4) for dU into Equation (8.9):

$$\begin{aligned} dF &= \left(T dS - p dV + \sum_{j=1}^M \mu_j dN_j \right) - T dS - S dT \\ &= -S dT - p dV + \sum_{j=1}^M \mu_j dN_j. \end{aligned} \quad (8.10)$$

We will use Equation (8.10) later to develop Maxwell's relations (see Chapter 9). Because dF is also defined by its partial derivative expression,

$$dF = \left(\frac{\partial F}{\partial T}\right)_{V,N} dT + \left(\frac{\partial F}{\partial V}\right)_{T,N} dV + \sum_{j=1}^M \left(\frac{\partial F}{\partial N_j}\right)_{V,T,N_{i\neq j}} dN_j, \quad (8.11)$$

you get additional thermodynamic relations by comparing Equation (8.11) with Equation (8.10):

$$S = - \left(\frac{\partial F}{\partial T}\right)_{V,N}, \quad p = - \left(\frac{\partial F}{\partial V}\right)_{T,N}, \quad \mu_j = \left(\frac{\partial F}{\partial N_j}\right)_{V,T,N_{i\neq j}}. \quad (8.12)$$

We derived $F(T, V, N)$ from $S(U, V, N)$ by physical arguments. You can also switch from one set of independent variables to another by purely mathematical arguments, called Legendre transforms (see Appendix F).

The Enthalpy

We now have three useful fundamental functions for systems with fixed V and N . Each is associated with an extremum principle: $S(U, V, N)$ is a maximum when U is controlled at the boundaries; $U(S, V, N)$ is a minimum when S is controlled at the boundaries (which we have not shown here); and $F(T, V, N)$ is a minimum when T is controlled at the boundaries. There are two other fundamental functions that are particularly important: the enthalpy $H(S, p, N)$ and the Gibbs free energy $G(T, p, N)$.

The enthalpy H is a function of the natural variables S , p , and N . Enthalpy is seldom used as an extremum principle, because it is not usually convenient to control the entropy. However, the enthalpy is important because it can be obtained from calorimetry experiments, and it gives an experimental route to the Gibbs free energy, which is of central importance in chemistry and biology. To find the enthalpy, you could reason in the same way as we did for the Helmholtz free energy, but instead let's use a simple math argument. Start with the internal energy $U(S, V, N)$. We seek to replace a dV term in the energy function with a dp term to get the enthalpy function dH . Add a pV term to the energy so that when you differentiate it, the dV term will disappear and a dp term will appear:

$$H = H(S, p, N) = U + pV. \quad (8.13)$$

Now differentiate:

$$dH = dU + p dV + V dp. \quad (8.14)$$

Substitute Equation (6.4), $dU = T dS - p dV + \sum_{j=1}^M \mu_j dN_j$, into Equation (8.14):

$$\begin{aligned} dH &= T dS - p dV + \sum_{j=1}^M \mu_j dN_j + p dV + V dp \\ \Rightarrow dH &= T dS + V dp + \sum_{j=1}^M \mu_j dN_j. \end{aligned} \quad (8.15)$$

If we had tried defining H with $H = U - pV$, the only other alternative for finding a function of dS , dp , and dN , we would have failed because our alternative to Equation (8.15) would have contained a fourth variable dV . Therefore the

only choice that will yield a function of only (S, p, N) is $U + pV$. The enthalpy is a minimum at equilibrium when S, p , and N are the independent variables.

The Gibbs Free Energy

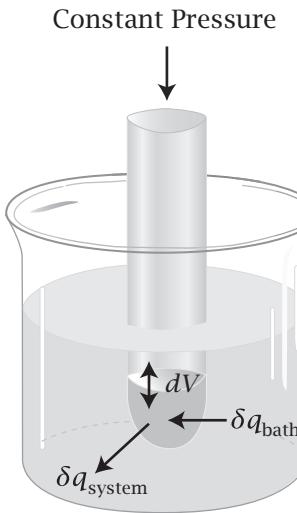


Figure 8.8 As in Figure 8.1, a heat bath holds the temperature constant. In addition, the pressure is now held constant and the volume fluctuates. In this case, the quantities (T, p, N) of the system are constant. Equilibrium results when the Gibbs free energy $G(T, p, N)$ is at a minimum.

The *Gibbs free energy* is one of the most important fundamental functions. Constant temperature and pressure are the easiest constraints to impose in the laboratory, because the atmosphere provides them. T, p , and N are the natural variables for the Gibbs free energy $G = G(T, p, N)$, which has a minimum at equilibrium. To find the fundamental equation, start with the enthalpy, $H = H(S, p, N)$. You want to replace the dS term with a dT term in the equation $dH = T dS + V dp + \sum_{j=1}^M \mu_j dN_j$. Define a function G :

$$G = H - TS. \quad (8.16)$$

The total differential dG is

$$dG = dH - T dS - S dT. \quad (8.17)$$

Substitute Equation (8.15) for dH into Equation (8.17) to get

$$dG = -S dT + V dp + \sum_{j=1}^M \mu_j dN_j. \quad (8.18)$$

(You could also have started with $F(T, V, N)$ instead, and replaced V with p ; $G = F + pV$.) The logic of the Gibbs free energy is similar to the logic of the Helmholtz free energy. If a process occurs in a test tube held at constant pressure and temperature (see Figure 8.8), it will be at equilibrium when the Gibbs free energy is at a minimum. Chemical reactions, phase changes, and biological or physical processes can take place in the test tube. The test tube exchanges energy with the surroundings by volume changes and heat transfer. Equilibrium is the state at which the entropy of the combined system plus surroundings is at a maximum. However, for the test tube system itself, which is at constant (T, p, N) , equilibrium occurs when the Gibbs free energy is at a minimum. dG can be expressed as

$$dG = \left(\frac{\partial G}{\partial T} \right)_{p,N} dT + \left(\frac{\partial G}{\partial p} \right)_{T,N} dp + \sum_{j=1}^M \left(\frac{\partial G}{\partial N_j} \right)_{p,T,N_{i \neq j}} dN_j. \quad (8.19)$$

Comparison of Equation (8.19) with Equation (8.18) shows that

$$S = - \left(\frac{\partial G}{\partial T} \right)_{p,N}, \quad V = \left(\frac{\partial G}{\partial p} \right)_{T,N}, \quad \mu_j = \left(\frac{\partial G}{\partial N_j} \right)_{p,T,N_{i \neq j}}. \quad (8.20)$$

For equilibrium phase changes at constant temperature, pressure, and particle number, the Gibbs free energy does not change, as shown in the next example.

EXAMPLE 8.3 Melting ice and freezing water. Consider a reversible phase change such as the melting of ice or the freezing of water in a test tube held at fixed constant pressure p_0 . The phase change occurs at a temperature T_0 . When the system is ice just below T_0 , the free energy is at a minimum (T, p, N constant, $dG = 0$). Let us call this free energy G_{solid} . Now add heat. This can cause a change in phase with a negligible change in temperature. Ice melts

to liquid water at a temperature just above 0°C. The free energy of this new equilibrium is G_{liquid} . What is the free energy of melting, $\Delta G = G_{\text{liquid}} - G_{\text{solid}}$? Because the temperature, pressure, and mole numbers are unchanged ($dT = dp = dN_1 = dN_2, \dots, dN_M = 0$) inside the test tube, the free energy does not change: $\Delta G = 0$. The relative amounts of ice and water can change inside the test tube, but we are given a system in which no water molecules escape from the test tube, $dN = 0$. The enthalpy and entropy of melting balance is

$$\Delta H = T_0 \Delta S \quad \text{when } \Delta G = 0. \quad (8.21)$$

Melting, boiling, and other phase changes involve an increase in entropy, $T_0 \Delta S$, that compensates for the increase in enthalpy ΔH resulting from the breakage or weakening of interactions.

A Simple Way to Find Relationships

We now have several partial derivative relationships, such as Equations (8.20), (8.12), (6.3), and (6.6). How can you determine them when you need them? Suppose that you want to know $(\partial G / \partial T)_{p,N}$. First, look at the quantities in the denominator and in the subscripts. In this case, these are (T, p, N) . Then use Table 8.1 to find the fundamental function that applies to those constraints. The table shows that the fundamental function of variables (T, p, N) is $G(T, p, N)$. Finally, identify the term you want in the equation for the differential, dG in this case. In this case, the result you are looking for is $(\partial G / \partial T)_{p,N} = -S$. Below, we sort out this alphabet soup of thermodynamic functions of dependent and independent variables.

The Limits on Constructing Thermodynamic Functions

What are the limits on constructing functions of T, S, p, V, N, U, F, H , and G ? You can divide thermodynamic functions into four categories:

FUNDAMENTAL AND USEFUL. Table 8.1 lists the main fundamental thermodynamic functions and their natural variables. The states of equilibrium are identified by extrema in these functions.

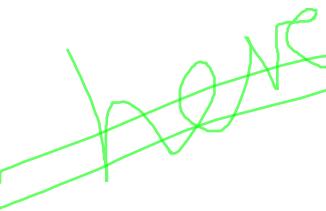
USEFUL BUT NOT FUNDAMENTAL. $U(T, V, N)$, $S(T, V, N)$, $H(T, p, N)$, and $S(T, p, N)$ are not functions of natural variables. These functions do not have corresponding extremum principles, but they are useful because they are components of $F(T, V, N)$ and $G(T, p, N)$.

COMPLETE BUT NOT USEFUL. Rearrangements of the dependent and independent variables from a fundamental thermodynamic function are possible, but not often useful. For example, $T(F, V, N)$ is a rearrangement of the fundamental Helmholtz free energy function, $F(T, V, N)$. This function is not very useful because you usually cannot constrain F at the system boundary.

INCOMPLETE. Additional functions could be constructed, such as $U(p, V, N)$ or $S(U, \mu, N)$ but because these involve conjugate pairs p and V , or μ and N , and are missing other variables, they do not uniquely specify the state of a system. Such functions cannot be obtained by Legendre transforms of the fundamental equations.

Table 8.1 Fundamental equations and their natural variables.

Function	Extremum at Equilibrium	Fundamental Equation	Definition
$U(S, V, N)$	Minimum	$dU = T dS - p dV + \sum_j \mu_j dN_j$	
$S(U, V, N)$	Maximum	$dS = \left(\frac{1}{T}\right) dU + \left(\frac{p}{T}\right) dV - \sum_j \left(\frac{\mu_j}{T}\right) dN_j$	
$H(S, p, N)$	Minimum	$dH = T dS + V dp + \sum_j \mu_j dN_j$	$H = U + pV$
$F(T, V, N)$	Minimum	$dF = -S dT - p dV + \sum_j \mu_j dN_j$	$F = U - TS$
$G(T, p, N)$	Minimum	$dG = -S dT + V dp + \sum_j \mu_j dN_j$	$G = H - TS = F + pV$



How are functions such as $U(T, V, N)$, $S(T, V, N)$, $H(T, p, N)$, or $G(T, p, N)$ useful? They are not fundamental functions. Although T , V , and N are natural variables for F , those are not the natural variables for the components U and S in

$$F(T, V, N) = U(T, V, N) - TS(T, V, N). \quad (8.22)$$

That is, minimizing $U(T, V, N)$ or maximizing $S(T, V, N)$ individually does not predict the state of equilibrium. Only their sum, in the form of F , has an extremum that identifies equilibrium if the constraints are (T, V, N) . Similarly, the Gibbs free energy is composed of two functions of non-natural variables, $H(T, p, N)$ and $S(T, p, N)$:

$$G(T, p, N) = H(T, p, N) - TS(T, p, N). \quad (8.23)$$

The enthalpy H plays the same role in the Gibbs free energy G that the energy U plays in the Helmholtz free energy F .

These non-fundamental component functions are important because they can be measured in calorimeters, and can be combined to give the fundamental functions, such as $F(T, V, N)$ and $G(T, p, N)$. You measure the temperature dependence of a material's heat capacity $C_V(T)$ in a constant-volume calorimeter. Then, using Equations (7.4) and (7.10), you get the component quantities $\Delta U = \int C_V(T) dT$ and $\Delta S = \int (C_V/T) dT$. Finally, you combine them to get the free energy $F(T) = U(T) - TS(T)$ using Equation (8.22).

You may prefer to work with $G(T)$ instead of $F(T)$. When you know p at the boundary, rather than V , you should measure the material's heat capacity in a constant-pressure calorimeter instead of in a constant-volume calorimeter; see Figures 7.1 and 7.2. Use the definition of enthalpy (Equation (8.13)) and the First Law to get

$$dH = d(U + pV) = dU + p dV + V dp = \delta q + \delta w + p dV + V dp.$$

For quasi-static processes ($\delta w = -p dV$) at constant pressure ($dp = 0$), this gives

$$dH = \delta q. \quad (8.24)$$

The heat capacity C_p measured at constant pressure is

$$C_p = \left(\frac{\delta q}{dT} \right)_p = \left(\frac{\partial H}{\partial T} \right)_p = T \left(\frac{\partial S}{\partial T} \right)_p. \quad (8.25)$$

Figure 7.2 shows an example of a heat capacity of a material. Rearranging, you see that H and C_p have the same relation as U and C_V :

$$\Delta H(T, p) = \int_{T_A}^{T_B} C_p(T) dT. \quad (8.26)$$

Correspondingly, to get the entropy if you know C_p instead of C_V , integrate $dS = (C_p/T) dT$ to get

$$\Delta S(T, p) = \int_{T_A}^{T_B} \frac{C_p}{T} dT. \quad (8.27)$$

From constant-pressure heat capacity experiments, you can get $G(T, p) = H(T, p) - TS(T, p)$.

The Third Law of Thermodynamics

Suppose you want to know the absolute entropy of a material at a temperature T . You can integrate the heat capacity, $S(T) = \int_0^T (C_V/T') dT' + S(0)$, where $S(0)$ is the entropy at absolute zero temperature, $T = 0$ K. How do you know the value of $S(0)$? The Third Law of Thermodynamics defines the absolute entropy to be zero, $S(0) = 0$ for a pure perfectly crystalline substance at a temperature of 0 K.

Table 8.2 shows the heat capacities of some materials. Within a given phase, heat capacities are often approximately independent of temperature.

Heat capacities of materials are also useful for problems of thermal equilibrium. Let's revisit Example 6.2, in which two objects are brought into thermal contact.

EXAMPLE 8.4 The equilibrium temperature of objects in thermal contact. Suppose objects A and B have different heat capacities, C_A and C_B (measured, let's say, at constant pressure), both independent of temperature. Initially, object A is colder with temperature T_A , and object B is hotter with temperature T_B . A and B are brought into thermal contact with each other, but they are isolated from the surroundings. At equilibrium, Example 6.2 shows that both objects will have the same temperature T . What is the final temperature? Because the objects are isolated from the surroundings, there is no net change in the enthalpy of the total system:

$$\Delta H = \Delta H_A + \Delta H_B = 0. \quad (8.28)$$

Table 8.2 Heat capacities at constant pressure of various substances near 25°C.

Gases	Molar Heat Capacities (cal deg ⁻¹ mol ⁻¹)	Liquids	Molar Heat Capacities (cal deg ⁻¹ mol ⁻¹)	Solids	Specific Heat Capacities (cal deg ⁻¹ g ⁻¹)
He	3.0	Hg	6.7	Au	0.0308
O ₂	7.0	H ₂ O	18.0	Fe	0.106
N ₂	7.0	Ethanol	27.0	C (diamond)	0.124
H ₂ O	7.9	Benzene	32.4	Glass (Pyrex)	0.2
CH ₄	8.6	<i>n</i> -Heptane	53.7	Brick	~0.2
CO ₂	9.0			Al	0.215
				Glucose	0.30
				Urea	0.50
				H ₂ O (0°C)	0.50
				Wood	~0.5

Source: I Tinoco, Jr, K Sauer, and JC Wang, *Physical Chemistry: Principles and Applications in Biological Sciences*, Prentice-Hall, New York, 1978.

The change in enthalpy of *A*, ΔH_A , is

$$\Delta H_A = \int_{T_A}^T C_A dT = C_A(T - T_A), \quad (8.29)$$

because C_A is independent of temperature. Similarly, ΔH_B is

$$\Delta H_B = C_B(T - T_B). \quad (8.30)$$

A given enthalpy change can be caused by either a smaller temperature change in a material with a larger heat capacity or a larger temperature change in a material with a smaller heat capacity. Substitute Equations (8.29) and (8.30) into (8.28) to get

$$C_A(T - T_A) + C_B(T - T_B) = 0. \quad (8.31)$$

Rearranging Equation (8.31) gives the final temperature *T*:

$$T = \frac{C_A T_A + C_B T_B}{C_A + C_B}.$$

If the two objects have the same heat capacity, $C_A = C_B$, the final temperature equals the average of the initial temperatures $T = (T_A + T_B)/2$. Note that heat capacities are usually given on a per mass or per volume basis, so such quantities need to be multiplied by the masses or volumes of the objects to give the total heat capacity of the object.

Combine Heat Capacities with Thermodynamic Cycles to Get Hard-to-Measure Quantities

The following example shows how to combine a thermodynamic cycle with heat capacities to compute properties for which measurements are difficult or impossible.

EXAMPLE 8.5 Measuring enthalpies under standard conditions and computing them for other conditions. Suppose you want to know the enthalpy of boiling water, $\Delta H_{\text{boil}(0^\circ\text{C})}$, at the freezing point of water 0°C and $p = 1 \text{ atm}$. Since you cannot boil water at water's freezing point, why would you want to know that quantity? We will see later that $\Delta H_{\text{boil}(0^\circ\text{C})}$ can tell you about the vapor pressure of water over a cold lake. You can get $\Delta H_{\text{boil}(0^\circ\text{C})}$ from the heat capacities of water and steam and the enthalpy of vaporization of water, $\Delta H_{\text{boil}(100^\circ\text{C})}$, under more 'standard' boiling conditions ($T = 100^\circ\text{C}$, $p = 1 \text{ atm}$) by using a simple thermodynamic cycle. The standard state enthalpy has been measured to be $\Delta H_{\text{boil}(100^\circ\text{C})} = 540 \text{ cal g}^{-1}$. The heat capacity of steam is $C_p = 0.448 \text{ cal K}^{-1} \text{ g}^{-1}$ and the heat capacity of liquid water is $C_p = 1.00 \text{ cal K}^{-1} \text{ g}^{-1}$. To obtain $\Delta H_{\text{boil}(0^\circ\text{C})}$, construct the thermodynamic cycle shown in Figure 8.9.

With the directions of the arrows shown, summing to zero around a cycle means that $\Delta H_{\text{boil}(0^\circ\text{C})} = \Delta H_{\text{boil}(100^\circ\text{C})} - \Delta H_{\text{liquid}} + \Delta H_{\text{steam}}$. Because there is no phase change for the steam or liquid, and because the heat capacities are reasonably independent of temperature, you have

$$\begin{aligned}\Delta H_{\text{liquid}} &= \int_{100}^0 C_{p,\text{liquid}} dT = C_{p,\text{liquid}} \Delta T \\ &= \left(1.00 \frac{\text{cal}}{\text{K g}} \right) (-100 \text{ K}) = -100 \text{ cal g}^{-1}\end{aligned}\quad (8.32)$$

and

$$\begin{aligned}\Delta H_{\text{steam}} &= \int_{100}^0 C_{p,\text{steam}} dT = C_{p,\text{steam}} \Delta T \\ &= \left(0.448 \frac{\text{cal}}{\text{K g}} \right) (-100 \text{ K}) = -44.8 \text{ cal g}^{-1}.\end{aligned}$$

Thus,

$$\Delta H_{0^\circ\text{C}} = (540 + 100 - 44.8) \text{ cal g}^{-1} = 595.2 \text{ cal g}^{-1}. \quad (8.33)$$

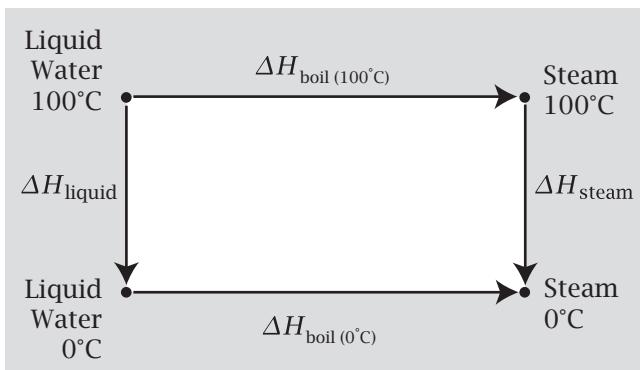


Figure 8.9 A thermodynamic cycle allows you to calculate the enthalpy for boiling water at the *freezing temperature* of water if you have measured the enthalpy at the *boiling temperature* (see Equation (8.33)).

Thermochemistry: The Enthalpy of a Molecule Is the Sum of the Enthalpies of Its Covalent Bonds

Mix up chemicals in a test tube. Will they react? You can predict this by comparing the Gibbs free energies of the products with the reactants. Whichever is lower predicts the stable state. (We consider here only the equilibrium. There is also the matter of kinetics. Sometimes, even when a reaction is favored, it is too slow to observe.) For covalently bonded molecules, the Gibbs free energies are mainly enthalpic. So, based on the following principles, you can predict stabilities of covalent compounds using tables of bond enthalpies. First, the enthalpy of a molecule is the sum of the enthalpies of its bonds. Therefore, you can predict quite accurately, for example, how much work an engine can produce from burning some type of hydrocarbon molecule by adding up the enthalpies of the carbon–carbon bonds and others. Second, you can obtain the covalent enthalpies of molecules by combusting them in calorimeters. Third, suppose the reaction you are interested in is $A+B \rightarrow C$, but it is difficult to perform. *Hess's law* says that because the enthalpy is a state function, you can use a different reaction to obtain those same enthalpies, meaning you can often

Table 8.3 Average bond enthalpies.

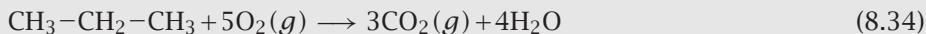
Bond	Bond Enthalpy (kJ mol ⁻¹)	Bond	Bond Enthalpy (kJ mol ⁻¹)
H–H	436.4	C–S	255
H–N	393	C=S	477
H–O	460	N–N	393
H–S	368	N=N	418
H–P	326	N≡N	941.4
H–F	568.2	N–O	176
H–Cl	430.9	N–P	209
H–Br	366.1	O–O	142
H–I	298.3	O=O	498.8
C–H	414	O–P	502
C–C	347	O=S	469
C=C	619	P–P	197
C≡C	812	P=P	490
C–N	276	S–S	268
C=N	615	S=S	351
C≡N	891	F–F	150.6
C–O	351	Cl–Cl	242.7
C=O	724	Br–Br	192.5
C–P	264	I–I	151.0
C=O (in CO ₂)	799		

Source: R Chang, *Physical Chemistry for the Biosciences*, University Science Books, Mill Valley, CA, 2005, p. 72.

find simple ways to compute quantities that are hard to measure. These principles are powerful for predicting what covalent chemical reactions will occur spontaneously from only a small body of experimental data. Table 8.3 gives some covalent bond enthalpies.

These principles are powerful. They led to the discovery of *aromaticity* in benzene. The earliest structure of benzene showed three single C–C bonds alternating with three double C=C bonds. But benzene turns out to be 150 kJ mol⁻¹ more stable than you would expect by just adding up those bond enthalpies. It was concluded that there must be some additional source of stabilization of molecules like benzene. The idea of aromatic bonds arose from such discrepancies in the additivities of bond enthalpies.

EXAMPLE 8.6 Converting bond enthalpies to heat in your propane stove.
The heat from your camping stove comes from the combustion of propane:



To compute the enthalpy change in this reaction, first count the bonds of each type and multiply by their bond enthalpies using Table 8.3. On the right side are the products: 8 HO bonds ($\Delta H = 8 \times 460 = 3680 \text{ kJ mol}^{-1}$) and 6 C=O bonds (4794 kJ mol⁻¹), totaling $\Delta H_{\text{prod}} = 8474 \text{ kJ mol}^{-1}$. On the left side are the reactants: 8 CH bonds, 2 C–C bonds, and 5 O=O bonds, totaling $\Delta H_{\text{reac}} = 6500 \text{ kJ mol}^{-1}$. These bond enthalpies are all positive because they correspond to the breaking of bonds. Since the reaction *makes* product bonds and *breaks* reactant bonds, you can compute the enthalpy change for the combustion reaction indicated above as $\Delta H_{\text{total}} = -\Delta H_{\text{prod}} + \Delta H_{\text{reac}} = -1974 \text{ kJ mol}^{-1}$. Because this quantity is negative, it indicates that the products are more stable than the reactants. This example shows how to compute thermodynamic properties of reactions by using additivities of the component covalent bond contributions. In contrast, noncovalent interactions, which are much weaker, are sometimes not additive.

Summary

In this chapter, we have shown that different fundamental functions and extremum principles are required to define the state of equilibrium, depending on what quantities are known or controlled at the boundaries. Often you can control temperature T , rather than energy U , so the condition for equilibrium is that the free energy is at a minimum. Such systems have tendencies toward both low energies (or enthalpies) and high entropies, depending on T . Heat capacities, which are measurable quantities, can be used to obtain energies, enthalpies, and entropies, which in turn help to predict free energies. Thermochemical cycles are valuable tools for computing unknown quantities from known quantities. The next chapter shows a few additional relationships for computing unknown quantities.

Problems

1. Finding a fundamental equation. While the Gibbs free energy G is the fundamental function of the natural variables (T, p, N) , growing biological cells often regulate not the numbers of molecules N , but the chemical potentials μ_j . That is, they control *concentrations*. What is the fundamental function Z of natural variables (T, p, μ) ?

2. Why does increasing temperature increase disorder? Systems become disordered as the temperature is increased. For example, liquids and solids become gases, solutions mix, adsorbates desorb. Why?

3. The difference between the energy and enthalpy changes in expanding an ideal gas. How much heat is required to cause the quasi-static isothermal expansion of one mole of an ideal gas at $T = 500\text{ K}$ from $P_A = 0.42\text{ atm}$, $V_A = 100\text{ liters}$ to $P_B = 0.15\text{ atm}$?

- (a) What is V_B ?
- (b) What is ΔU for this process?
- (c) What is ΔH for this process?

4. The work and the heat of boiling water. For the reversible boiling of five moles of liquid water to steam at 100°C and 1 atm pressure, calculate q . Is w positive or negative?

5. The entropy and free energy of gas expansion. Two moles of an ideal gas undergo an irreversible isothermal expansion from $V_A = 100\text{ liters}$ to $V_B = 300\text{ liters}$ at $T = 300\text{ K}$.

- (a) What is the entropy change for this process?
- (b) What is the Gibbs free energy change?

6. The free energy and entropy of membrane melting. Pure membranes of dipalmitoyl lecithin phospholipids are models of biological membranes. They melt at $T_m = 41^\circ\text{C}$. Reversible melting experiments indicate that $\Delta H_m = 9\text{ kcal mol}^{-1}$. Calculate

- (a) the entropy of melting ΔS_m and
- (b) the free energy of melting ΔG_m .
- (c) Does the membrane become more or less ordered upon melting?
- (d) There are 32 rotatable CH_2-CH_2 bonds in each molecule. What is the increase in multiplicity on melting one mole of bonds?

7. State and path-dependent functions. Which quantities sum to zero around a thermodynamic cycle?

- (a) q , heat
- (b) w , work
- (c) $-p dV$
- (d) ΔU
- (e) ΔG

8. Computing enthalpy and entropy with a temperature-dependent heat capacity. The heat capacity for liquid n -butane depends on temperature:

$$C_p(T) = a + bT,$$

where $a = 100\text{ JK}^{-1}\text{ mol}^{-1}$ and $b = 0.1067\text{ JK}^{-2}\text{ mol}^{-1}$, from its freezing temperature $T_f \approx 140\text{ K}$ to $T_b \approx 270\text{ K}$, its boiling temperature.

- (a) Compute ΔH for heating liquid butane from $T_A = 170\text{ K}$ to $T_B = 270\text{ K}$.
- (b) Compute ΔS for the same process.

9. Cycle for determining the enthalpy of vaporization of water at 473K. Suppose you want to know how much heat it would take to boil water at 473 K , rather than 373 K . At $T = 373\text{ K}$, $\Delta H_{\text{boiling}}(100^\circ\text{C}) = 40.7\text{ kJ mol}^{-1}$ is the enthalpy of vaporization. Assuming that the heat capacities of the liquid ($C_{p,\text{liquid}} = 75\text{ JK}^{-1}\text{ mol}^{-1}$) and the vapor ($C_{p,\text{vapor}} = 3.5\text{ JK}^{-1}\text{ mol}^{-1}$) are constant over this temperature range, calculate the enthalpy of vaporization at 473 K using the thermodynamic cycle shown in Figure 8.10.

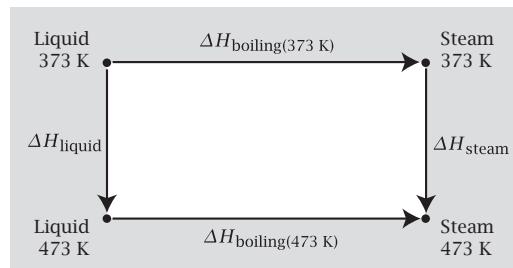


Figure 8.10 A thermodynamic cycle for calculating the enthalpy of boiling water at a temperature higher than the boiling point.

10. Heating a house. If your living room, having a volume of $6\text{ m} \times 6\text{ m} \times 3\text{ m} \approx 100\text{ m}^3$, were perfectly insulated, how much energy would be needed to raise the temperature inside the room from $T_{\text{initial}} = 0^\circ\text{C}$ to $T_{\text{final}} = 25^\circ\text{C}$? Note that $C_V = C_p - nR$ for an ideal gas.

11. Objects in thermal contact. Suppose two objects A and B , with heat capacities C_A and C_B and initial temperatures T_A and T_B , are brought into thermal contact. If $C_A \gg C_B$, is the equilibrium temperature T closer to T_A or to T_B ?

12. ΔS for an adiabatic expansion of a gas. In an adiabatic quasi-static expansion of an ideal gas, how do you reconcile the following two facts: (1) the increase in volume should lead to an increase in entropy, but (2) in an adiabatic process, $\delta q = 0$, so there should be no change in entropy (since $dS = \delta q/T = 0$)?

13. A thermodynamic cycle for mutations in protein folding. Suppose you can measure the stability of a wild-type protein, $\Delta G_1 = G_{\text{folded}} - G_{\text{unfolded}}$, the free

energy difference between folded and unfolded states. A mutant of that protein has a single amino acid replacement. Design a thermodynamic cycle that will help you find the free energy difference $\Delta G_2 = G_{\text{unfolded, mutant}} - G_{\text{unfolded, wildtype}}$, the effect of the mutation on the unfolded state.

14. Free energy of an ideal gas.

- For an ideal gas, calculate $F(V)$, the free energy versus volume, at constant temperature.
- Compute $G(V)$.

15. Heat capacity of an ideal gas.

The energy of an ideal gas does not depend on volume,

$$\left(\frac{\partial U}{\partial V}\right)_T = 0.$$

Use this fact to prove that the heat capacities $C_p = (\partial H / \partial T)_p$ and $C_V = (\partial U / \partial T)_V$ for an ideal gas are both independent of volume.

16. Computing entropies.

The heat capacity of an ideal diatomic gas is $C_V = (5/2)Nk$ for N molecules. Assume $T = 300\text{ K}$ where needed.

- One mole of O_2 gas fills a room of 500 m^3 . What is the entropy change ΔS for squeezing the gas into 1 cm^3 in the corner of the room?
- An adult takes in about 3000 kcal per day from food (1 food Cal = 1 kcal). What is ΔS for this process?
- One mole of O_2 gas is in a room of 500 m^3 . What is the entropy change ΔS for heating the room from $T = 270\text{ K}$ to 330 K ?
- The free energy of a conformational motion of a loop in a protein is $\Delta G = 2\text{ kcal mol}^{-1}$. The enthalpy change is $\Delta H = 0.5\text{ kcal mol}^{-1}$. Compute ΔS .

17. What powers a hurricane?

Example 7.11 of Chapter 7 describes the thermodynamics of a hurricane, called a Hadley cycle: air takes up heat and water vapor from a warm ocean, then rises, then releases the heat high in the atmosphere. The reason a hurricane is so strong is that its Hadley cycle is augmented by an additional ingredient, the vaporization of warm water. You can saturate 1 kg of air with 20 g of water vapor. (Consider air as consisting of N_2 gas.)

- If the density of saturated air is 1.25 kg m^{-3} , how many grams of water vapor will occupy 1 m^3 of air?
- If the enthalpy of vaporization of water is $2.3 \times 10^6\text{ J (kg of water)}^{-1}$, how much heat q is given off by the water when its vapor condenses into rain, per m^3 of air?
- If the heat q that is given off by water's condensation all goes into raising the temperature of the surrounding air (at $p = 1\text{ atm}$), what is the temperature increase ΔT ?

18. Lattice model of ligand binding to a protein. You have a ligand L and a protein P . When the ligand binds, the energy change is favorable by $5\varepsilon_0$ ($\varepsilon_0 < 0$) because it makes 5 contacts with the protein. As shown in Figure 8.11, the protein is modeled using two-dimensional lattice particles. The protein has a rigid portion (squares containing a P) and a flexible loop (circles containing a P at one end). The ligand (rectangle containing an L) is rigid and occupies two lattice sites.

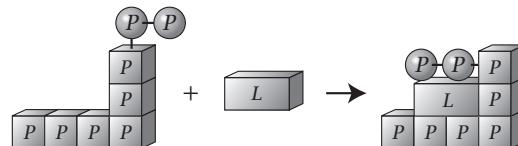


Figure 8.11

- Counting only lateral contacts (and not diagonal contacts), find the number of conformers W of the flexible loop.
- What are the energetic and entropic contributions to the free energy change due to ligand binding? Write an expression for $\Delta F_{\text{binding}}$.
- For $T = 300\text{ K}$, what is $\Delta F_{\text{binding}} = F_{\text{bound}} - F_{\text{unbound}}$ if $\varepsilon_0 = -1\text{ kcal mol}^{-1}$?
- What is the dissociation temperature T_{dissoc} ?
- If the flexible loop were made rigid instead (and in the proper configuration for ligand binding), what would be the new $\Delta F_{\text{binding}}$?
- Does the ligand bind more tightly or more weakly with the rigid loop?

Suggested Reading

CJ Adkins, *Equilibrium Thermodynamics*, 3rd edition Cambridge University Press, Cambridge, 1983. Simple and concise.

HB Callen, *Thermodynamics and an Introduction to Thermostatistics*, 2nd edition, Wiley, New York, 1985. This book is the classic development of the axiomatic approach to thermodynamics, in which energy and entropy, rather than temperature, are the primary concepts.

G Carrington, *Basic Thermodynamics*, Oxford University Press, Oxford, 1994. Describes the basics very clearly.

K Denbigh, *The Principles of Chemical Equilibrium*, 4th edition, Cambridge University Press, Cambridge, 1981. An excellent standard and complete text on thermodynamics.

EA Guggenheim, *Thermodynamics: An Advanced Treatment for Chemists and Physicists*, 7th edition, North-Holland, Amsterdam, 1985. An advanced treatment.

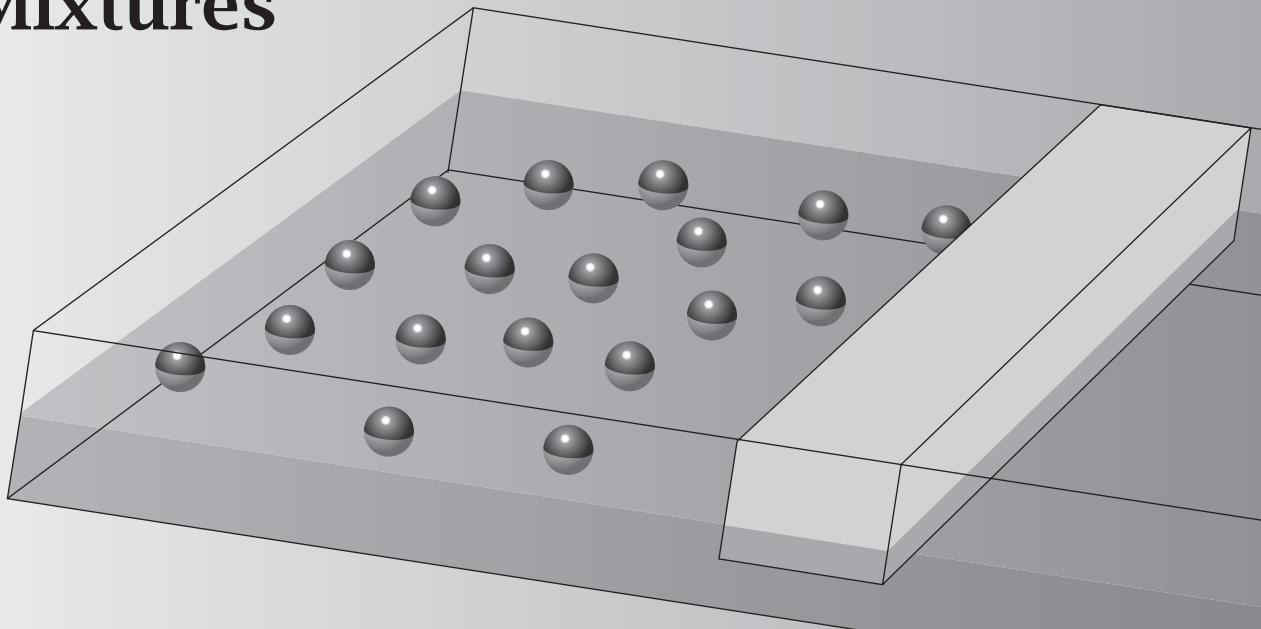
IM Klotz and RM Rosenberg, *Chemical Thermodynamics: Basic Concepts and Methods*, 7th edition, Wiley, New York, 2008. Well written, popular discussion of thermodynamics for chemists.

GN Lewis and M Randall, revised by KS Pitzer and L Brewer, *Thermodynamics*, McGraw-Hill, New York, 1961. A traditional classic text.

EB Smith, *Basic Chemical Thermodynamics*, 5th edition, Imperial College Press, London, 2004. Simple and concise presentation of thermodynamics, with emphasis on experiments.

MW Zemansky and RH Dittman, *Heat and Thermodynamics*, 7th edition, McGraw-Hill, New York, 1996. A classic text for chemists and chemical engineers, with extensive discussion of experimental data.

9 Maxwell's Relations & Mixtures



Thermodynamics Applies to Various Types of Forces

In this chapter, we treat forces other than pressures in pistons. Think, for example, about the force exerted by a rubber band, the forces that act on surfaces and membranes, the forces that move electrical charges, or the forces that rotate molecular magnets. We also introduce Maxwell's relations, which give insights about those forces. These relations give you hard-to-measure quantities of interest from properties that are easy to measure. For example, with Maxwell's relations, you can find the entropic and energetic components of forces. This type of information played a key historical role in understanding the molecular basis for the stretching of rubber, and how it is different from the stretching of metal in springs. This was critical for inventing synthetic elastic materials during World War II, when natural rubber was scarce.

For any force in general, what extremum principle will predict the state of equilibrium? For example, consider surface forces. For experiments in test tubes and beakers, the surface of the material is a small part of the system. Surface forces don't contribute much to the thermodynamic properties of typical bulk systems. But for experiments on soap films in a dish pan, on cell membranes, or on interfaces, the surface properties are of interest. In such cases, you have additional degrees of freedom. Even if you fix the values of (T, p, N) , a droplet can change its surface area A . How can we understand the forces that cause a droplet of water to become spherical? A water droplet minimizes its

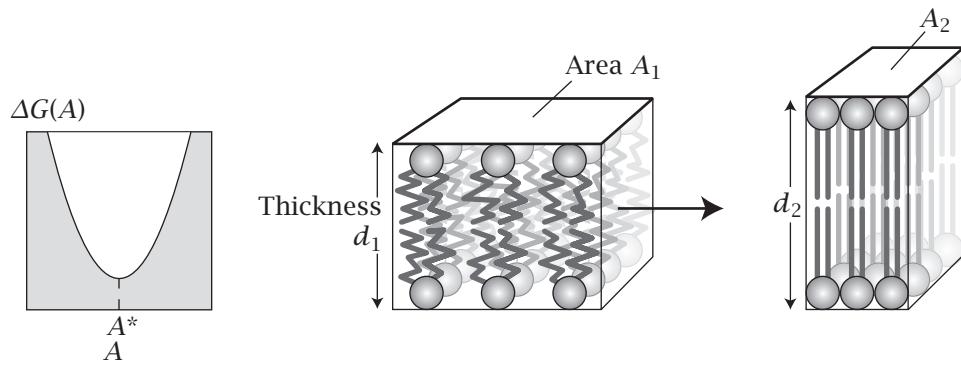


Figure 9.1 A lipid bilayer can vary its area A at fixed volume by changing its thickness d (volume = $Ad = A_1d_1 = A_2d_2$). The state of equilibrium occurs for the value of A that minimizes the free energy $G(T, p, N, A)$.

surface area, with fixed volume, driven by a tendency to minimize a generalized version of the free energy, $G(T, p, N, A)$. Similarly, the following example illustrates how lipid bilayers also minimize this function $G(T, p, N, A)$.

EXAMPLE 9.1 Lipid bilayer membranes can vary their surface areas. Cell membranes are composed of two opposing monolayers of lipid molecules; see Figure 9.1. Lipid bilayers and monolayers have an ability to change their surface area per lipid molecule, A , even when temperature, pressure, and the numbers of molecules are fixed. A lipid bilayer adjusts its area to have an equilibrium area per molecule, $A = A^*$, that minimizes the free energy function $G(A)$. The area A^* is a result of a compromise between opposing forces. On the one hand, if A^* were too large, the hydrocarbon chains of the lipids would form a large area of contact with neighboring water, which is unfavorable. On the other hand, if A^* were too small, the electrostatically charged head groups of the lipid molecules would be squeezed too close together, which is also unfavorable. Knowledge of $G(A)$ can help you model these forces.

Why is $G(T, p, N, A)$ the right function to minimize for water droplets or lipid bilayers?

How to Design a Fundamental Equation

Here is the recipe for determining a fundamental equation and its corresponding extremum principle for any type of force. Each extensive degree of freedom in the fundamental energy equation has a conjugate force. Pressure is conjugate to volume, temperature is conjugate to entropy, and chemical potential is conjugate to particle number. Other pairs include [force, length] = [f, L] for solid or elastic materials, [surface tension, area] = [γ, A] for liquids and interfaces, [electrical potential, charge] = [ψ, Q] for charged particles, and [magnetic field, magnetic moment] = [B, I] for magnetic systems. [\mathcal{F}, X] represents any conjugate pair [generalized force, generalized extensive variable].

To design an extremum function, introduce your *extensive* variables X into the fundamental energy equation to get $U = U(S, V, N, X)$. It is always *extensive* variables that you add, and you add them only to the *fundamental energy*

equation, because this is the key equation for defining the conjugate force quantities such as T , p , and μ . dU will now be a sum of work terms such as

$$dU = T dS - p dV + \sum_k \mu_k dN_k + f dL + \gamma dA + \psi dQ + B dI + \sum_j \mathcal{F}_j dX_j. \quad (9.1)$$

The generalized force \mathcal{F} for each new term is

$$\mathcal{F}_j = \left(\frac{\partial U}{\partial X_j} \right)_{S,V,N,X_{i \neq j}}. \quad (9.2)$$

With this augmented function dU , you can follow the Legendre transform recipes of Appendix F to find the fundamental function and extremum principle for the appropriate independent variables. Here is an example.

EXAMPLE 9.2 Why is $G(T, p, N, A)$ the fundamental function when surface area is an independent variable? Let's continue with Example 9.1. Suppose your system can change its surface area, independently of its volume. What extremum function identifies the state of equilibrium for the variables (T, p, N, A) , where A is the surface area of the system? First, in Equation (9.1), add the term γdA to dU to account for surface forces:

$$dU = T dS - p dV + \sum_j \mu_j dN_j + \gamma dA, \quad (9.3)$$

where γ is the *surface tension*,

$$\gamma = \left(\frac{\partial U}{\partial A} \right)_{S,V,N}. \quad (9.4)$$

Equation (9.4) is not particularly useful, because it requires an experiment in which you fix the variables (S, V, N, A) . It would be easier experimentally to fix the variables (T, p, N, A) . So, let's transform to those variables. Because (T, p, N) are the natural variables for the function G , use $G = U + pV - TS$, take the differential to get $dG = dU + p dV + V dp - T dS - S dT$, and substitute Equation (9.3) into this expression to get the differential form

$$dG = -S dT + V dp + \sum_j \mu_j dN_j + \gamma dA, \quad (9.5)$$

where

$$\gamma = \left(\frac{\partial G}{\partial A} \right)_{T,p,N}. \quad (9.6)$$

This says that $G(T, p, N, A)$ is the extremum function when (T, p, N, A) are the degrees of freedom of your system. G will be minimal at equilibrium. This transformation defines the surface tension in terms of a simple experiment in which you can fix T and p , rather than S and V .

The methods above give you a general prescription for finding the extremum function to predict equilibrium for various types of forces. A powerful way to make use of such expressions is through Maxwell's relations.

Maxwell's Relations Interrelate Partial Derivatives

Maxwell's relations are relationships between partial derivatives. An example is

$$\left(\frac{\partial S}{\partial V}\right)_T = \left(\frac{\partial p}{\partial T}\right)_V. \quad (9.7)$$

How is this useful? Recall from Example 6.1 that the quantity $(\partial S / \partial V)_T$ previously helped us model the driving force for gases to expand. But, you can't measure that quantity. On the other hand, the quantity on the right side of Equation (9.7) is easy to measure. For a gas, you observe the pressure as a function of temperature, holding the volume constant. Maxwell's relations give you ways to get unmeasurable quantities that are useful for modeling from other quantities that you can readily measure. There are many such relationships (see Table 9.1). They follow from Euler's reciprocal relation, Equation (4.39).

To derive Equation (9.7), start with the function $F(T, V)$ for fixed N. Because F is a state function, the Euler equation (4.39) gives

$$\left(\frac{\partial^2 F}{\partial T \partial V}\right) = \left(\frac{\partial^2 F}{\partial V \partial T}\right).$$

Rewrite this as

$$\left(\frac{\partial}{\partial T} \left(\frac{\partial F}{\partial V}\right)_T\right)_V = \left(\frac{\partial}{\partial V} \left(\frac{\partial F}{\partial T}\right)_V\right)_T. \quad (9.8)$$

The inner derivatives are given by Equation (6.3):

$$\left(\frac{\partial F}{\partial V}\right)_T = -p \quad \text{and} \quad \left(\frac{\partial F}{\partial T}\right)_V = -S.$$

Substituting these quantities into Equation (9.8) gives Maxwell's relation, Equation (9.7).

A Recipe for Finding Maxwell's Relations

Suppose you want $(\partial S / \partial p)_{T,N}$. This could be helpful for understanding how the entropies of materials change as you squeeze them.

To find Maxwell's relation for any quantity, first identify what independent variables are implied. In this case, the independent variables are (T, p, N) because these are the quantities that either are given as constraints in the subscript or are in the denominator of the partial derivative. Second, find the *natural function* of these variables (see Table 8.1). For (T, p, N) , the natural function is $G(T, p, N)$. Third, express the total differential of the natural function:

$$dG = -S dT + V dp + \sum_{j=1}^M \mu_j dN_j. \quad (9.9)$$

Fourth, based on Euler's reciprocal relation, set equal the two *cross-derivatives* you want. In this case, set the derivative of S (from the first term on the right side of Equation (9.9)) with respect to p (from the second term) equal

Table 9.1 Maxwell's relations.

Fundamental Function	Variables to Relate	Maxwell's Relation
$U(S, V, N)$	S, V	$(\partial T / \partial V)_{S,N} = -(\partial p / \partial S)_{V,N}$
$dU = T dS - p dV + \mu dN$	S, N	$(\partial T / \partial N)_{S,V} = (\partial \mu / \partial S)_{V,N}$
	V, N	$-(\partial p / \partial N)_{S,V} = (\partial \mu / \partial V)_{S,N}$
$F(T, V, N)$	T, V	$(\partial S / \partial V)_{T,N} = (\partial p / \partial T)_{V,N}$
$dF = -S dT - p dV + \mu dN$	T, N	$-(\partial S / \partial N)_{T,V} = (\partial \mu / \partial T)_{V,N}$
	V, N	$-(\partial p / \partial N)_{T,V} = (\partial \mu / \partial V)_{T,N}$
$H(S, p, N)$	S, p	$(\partial T / \partial p)_{S,N} = (\partial V / \partial S)_{p,N}$
$dH = T dS + V dp + \mu dN$	S, N	$(\partial T / \partial N)_{S,p} = (\partial \mu / \partial S)_{p,N}$
	p, N	$(\partial V / \partial N)_{S,p} = (\partial \mu / \partial p)_{S,N}$
(S, V, μ)	S, V	$(\partial T / \partial V)_{S,\mu} = -(\partial p / \partial S)_{V,\mu}$
	S, μ	$(\partial T / \partial \mu)_{S,V} = -(\partial N / \partial S)_{V,\mu}$
	V, μ	$(\partial p / \partial \mu)_{S,V} = (\partial N / \partial V)_{S,\mu}$
$G(T, p, N)$	T, p	$-(\partial S / \partial p)_{T,N} = (\partial V / \partial T)_{p,N}$
$dG = -S dT + V dp + \mu dN$	T, N	$-(\partial S / \partial N)_{T,p} = (\partial \mu / \partial T)_{p,N}$
	p, N	$(\partial V / \partial N)_{T,p} = (\partial \mu / \partial p)_{T,N}$
(T, V, μ)	T, V	$(\partial S / \partial V)_{T,\mu} = (\partial p / \partial T)_{V,\mu}$
	T, μ	$(\partial S / \partial \mu)_{T,V} = (\partial N / \partial T)_{V,\mu}$
	V, μ	$(\partial p / \partial \mu)_{T,V} = (\partial N / \partial V)_{T,\mu}$
(S, p, μ)	S, p	$(\partial T / \partial p)_{S,\mu} = (\partial V / \partial S)_{p,\mu}$
	S, μ	$(\partial T / \partial \mu)_{S,p} = -(\partial N / \partial S)_{p,\mu}$
	p, μ	$(\partial V / \partial \mu)_{S,p} = -(\partial N / \partial p)_{S,\mu}$
(T, p, μ)	S, p	$-(\partial S / \partial p)_{T,\mu} = (\partial V / \partial T)_{p,\mu}$
	T, μ	$(\partial S / \partial \mu)_{T,p} = (\partial N / \partial T)_{p,\mu}$
	p, μ	$(\partial V / \partial \mu)_{T,p} = -(\partial N / \partial p)_{T,\mu}$

Source: HB Callen, *Thermodynamics and an Introduction to Thermostatistics*, 2nd edition. Wiley, New York, 1985.

to the derivative of V (from the second term) with respect to T (from the first term). $-S$ in the first term on the right side of Equation (9.9) is $-S = (\partial G / \partial T)$. Take its derivative with respect to p , to get

$$-\left(\frac{\partial S}{\partial p}\right) = \left(\frac{\partial^2 G}{\partial p \partial T}\right). \quad (9.10)$$

The second term contains $V = (\partial G / \partial p)$. Take its derivative with respect to T , to get

$$\left(\frac{\partial V}{\partial T}\right) = \left(\frac{\partial^2 G}{\partial T \partial p}\right). \quad (9.11)$$

According to Euler's relation, the two second derivatives must be equal, so

$$\left(\frac{\partial S}{\partial p}\right)_{T,N} = -\left(\frac{\partial V}{\partial T}\right)_{p,N}. \quad (9.12)$$

This is the Maxwell relation you want. It gives you a quantity you cannot measure, $(\partial S / \partial p)$, from a quantity that is easy to measure, $(\partial V / \partial T)$. Suppose you have a block of metal or a liquid in a beaker. You measure its volume at constant pressure and plot how the volume changes with temperature. The slope $-(\partial V / \partial T)$ from that simple experiment will give $(\partial S / \partial p)_{T,N}$ through this Maxwell relationship. Or suppose you have an ideal gas held at constant pressure. The gas volume increases with temperature, so $(\partial V / \partial T)_{p,N} = Nk/p$ is a positive quantity; therefore, Equation (9.12) says that the entropy decreases with increasing pressure (integrating $(\partial S / \partial p)$ gives $S = -Nk \ln p$). Quantities such as $(\partial V / \partial T)$ or $(\partial V / \partial p)$ are readily measured, as noted below.

Susceptibilities Are Measurable and Useful

Some thermal quantities, called *susceptibilities*, are readily measured and useful. They include the heat capacity, the thermal expansion coefficient, the isothermal compressibility, and the thermal pressure coefficient. When combined with Maxwell's relations, these quantities can give insights into more fundamental quantities. The thermal expansion coefficient α ,

$$\alpha = \frac{1}{V} \left(\frac{\partial V}{\partial T}\right)_p, \quad (9.13)$$

is the fractional change in the volume of a system with temperature at constant pressure.

Because of Maxwell's relation $(\partial S / \partial p)_T = -(\partial V / \partial T)_p$, measuring how a material's volume depends on temperature $V(T)$ gives you information about $S(p)$. In particular, if $\alpha > 0$, it means that increasing the pressure orders the system, decreasing the entropy. Most materials have $\alpha > 0$. Thermal expansion coefficients are usually positive because increasing temperature causes a loosening up of the intermolecular bonds in the material. One example is an ideal gas, for which Equation (9.13) gives $\alpha = (p/NkT)(Nk/p) = 1/T$, which is positive. Increasing the pressure decreases the entropy of an ideal gas. Another example, shown in Figure 9.2, is the volume-versus-temperature relationship

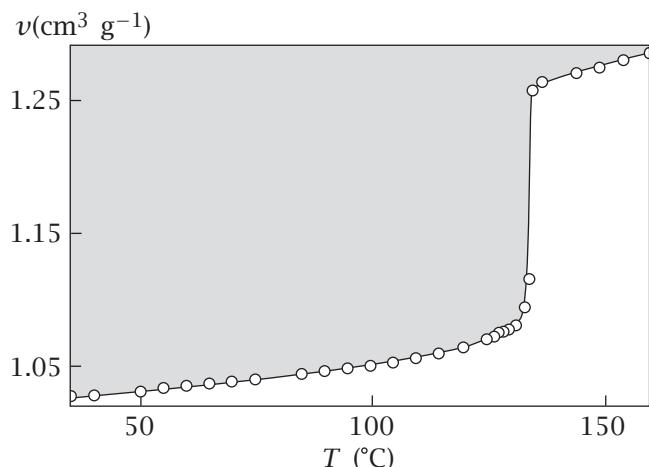


Figure 9.2 The specific volume v (volume per unit mass) of polyethylene versus temperature T . According to Equation (9.13), the thermal expansion coefficient α is proportional to the slope dv/dT . At low temperature, polyethylene is a hard crystalline plastic material. On melting at around 130°C , the specific volume v increases sharply and α is large. Source: JE Mark, A Eisenberg, WW Graessley, L Mandelkern, and JL Koenig, *Physical Properties of Polymers*, 2nd edition, American Chemical Society, Washington, DC, 1993. Data are from FA Quinn Jr and L Mandelkern, *J Am Chem Soc* **83**, 2857 (1961).

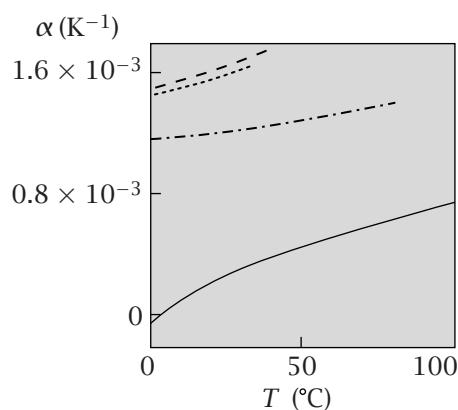


Figure 9.3 Thermal expansion coefficients α measured at atmospheric pressure for water (—), benzene (— · —), *n*-pentane (····), and diethyl ether (— – —). Source: A Hvidt, *Acta Chemica Scand A* **32**, 675–680 (1978). Data for water are from GS Kell, *J Chem Eng Data* **20**, 97 (1975). Data for benzene, *n*-pentane, and diethyl ether are from Landolt-Börnstein, *Physikalisch-Chemisch Tabellen II*, 5. Auflage, 1232 (1923).

for polyethylene. Thermal expansion coefficients are derived from the slopes of such plots by using Equation (9.13). The slope of the curve below about 130°C gives the thermal expansion coefficient for crystalline polyethylene, which is a hard plastic material. The volume expands sharply at the melting temperature. Above about 140°C , the slope gives the thermal expansion coefficient of the plastic liquid. Figure 9.3 shows thermal expansion coefficients for liquids, indicating that most liquids expand with temperature at most temperatures.

However, Figure 9.3 also shows that cold liquid water is anomalous. Just above its freezing temperature, water has $\alpha < 0$. So, Maxwell's relation above tells you that applying pressure to cold liquid water *disorders it*, increasing its entropy. The explanation is that cold water is highly cage-like, and applying pressure distorts and disorders the cages.

Example 9.3 shows how to determine $S(p)$ by integrating and using a Maxwell relationship.

EXAMPLE 9.3 Entropy changes with pressure. Suppose you want to know how the entropy $S(p)$ of a system depends on the pressure applied to it, at constant temperature:

$$dS = \left(\frac{\partial S}{\partial p} \right)_{T,N} dp. \quad (9.14)$$

Combine Maxwell's relation $(\partial S / \partial p)_{T,N} = -(\partial V / \partial T)_{p,N}$ with the definition of the thermal expansion coefficient given by Equation (9.13), $\alpha = V^{-1}(\partial V / \partial T)_{p,N}$, to get

$$dS = -\left(\frac{\partial V}{\partial T}\right)_{p,N} dp = -\alpha V dp. \quad (9.15)$$

Because the thermal expansion coefficients α of materials can be measured as functions of pressure, you can compute the entropy change from a series of thermal expansion experiments at different pressures:

$$\Delta S = - \int_{p_1}^{p_2} \alpha(p)V(p) dp. \quad (9.16)$$

Another readily measured property of materials is the isothermal compressibility κ ,

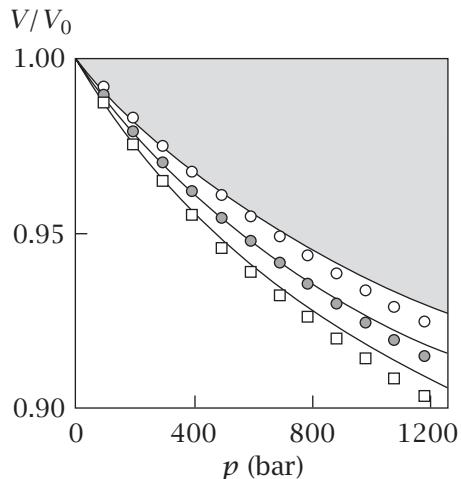
$$\kappa = -\frac{1}{V} \left(\frac{\partial V}{\partial p}\right)_T, \quad (9.17)$$

which is the fractional change in volume of a system as the pressure changes at constant temperature.

For an ideal gas, $\kappa = (p/NkT)(NkT/p^2) = 1/p$. Figure 9.4 shows the volume of liquid hexadecane as a function of the applied pressure. Isothermal compressibilities are derived from the slopes of such plots. Because the slope of this curve is negative, Equation (9.17) shows that the isothermal compressibility is positive: increasing the applied pressure decreases the volume of hexadecane. Liquids and solids are sometimes modeled as being *incompressible*, $\kappa \approx 0$, where the volume is assumed to be approximately independent of the applied pressure. They are not perfectly incompressible, as Figure 9.4 shows. Nevertheless, gases are much more compressible.

Volume measurements give clues to the forces within materials. Figures 9.3 and 9.5 suggest that the bonding in liquid water is more 'rigid' than that in organic liquids. Figure 9.3 shows that organic liquids undergo larger volume expansions on heating than water does, while Figure 9.5 shows that water is

Figure 9.4 The relative volume V/V_0 of hexadecane decreases with applied pressure p (V_0 is the volume of the material extrapolated to zero pressure). The slopes of the curves give the isothermal compressibilities in the form of Equation (9.17). Compressibility increases with temperature (\circ 60°C, \bullet 90°C, \square 120°C). Source: RA Orwoll and PJ Flory, *J Am Chem Soc* **89**, 6814–6822 (1967).



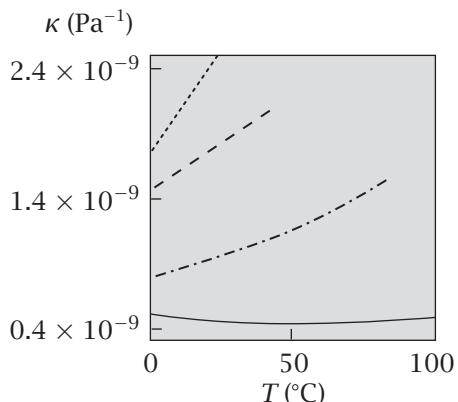


Figure 9.5 Compressibilities κ measured at atmospheric pressure for water (—), benzene (—·—), *n*-pentane (···), and diethyl ether (— —). Source: A Hvidt, *Acta Chemica Scand A* **32**, 675–680 (1978). Data for water are from GS Kell, *J Chem Eng Data* **20**, 97 (1975). Data for benzene are from *CRC Handbook of Chemistry and Physics*, 55th edition, CRC Press, Boca Raton, FL, 1974–1975, F-16. Data for pentane and diethyl ether are from Landolt-Börnstein, *Physikalisch-Chemisch Tabellen I*, 5. Auflage, 97–100 (1923).

less compressible than organic liquids. Also, the compressibility κ of water changes quite unusually as a function of temperature: the compressibilities of the organic liquids all increase between 0°C and 100°C, but water becomes less compressible between 0°C and 46°C. We examine the unusual properties of water in more detail in Chapters 30 and 31.

Another of Maxwell's relations relates internal energy to volume to describe the attractive and repulsive forces within materials.

EXAMPLE 9.4 Energy change with volume. The function $(\partial U / \partial V)_T$ tells you about the cohesive forces in materials. How can you obtain it from experiments? Start with the functions $U(T, V)$ and $S(T, V)$. The differentials are

$$dU = \left(\frac{\partial U}{\partial V}\right)_T dV + \left(\frac{\partial U}{\partial T}\right)_V dT,$$

and

$$dS = \left(\frac{\partial S}{\partial V}\right)_T dV + \left(\frac{\partial S}{\partial T}\right)_V dT.$$

Substitute these into the fundamental energy equation $dU = T dS - p dV$ to get

$$\left(\frac{\partial U}{\partial V}\right)_T dV + \left(\frac{\partial U}{\partial T}\right)_V dT = T \left[\left(\frac{\partial S}{\partial V}\right)_T dV + \left(\frac{\partial S}{\partial T}\right)_V dT \right] - p dV.$$

When T is constant, this reduces to

$$\left(\frac{\partial U}{\partial V}\right)_T = T \left(\frac{\partial S}{\partial V}\right)_T - p.$$

Using Maxwell's relation $(\partial S / \partial V)_T = (\partial p / \partial T)_V$ (see Table 9.1) gives

$$\left(\frac{\partial U}{\partial V}\right)_T = T \left(\frac{\partial p}{\partial T}\right)_V - p.$$

The measurable quantity $(\partial p / \partial T)_V$ is called the *thermal pressure coefficient*. For an ideal gas, $(\partial p / \partial T)_V = Nk/V$, so $(\partial U / \partial V)_T = 0$. For typical liquids, $T(\partial p / \partial T)_V - p$ is negative at high densities and positive at low densities. When you squeeze a material that is already compact, you are pushing against

repulsive forces (the energy goes up as the volume decreases). When you stretch that material, you are pulling against attractive forces (the energy goes up as the volume increases).

Now, let's see how Maxwell's relations are useful when types of work other than pV changes are involved.

EXAMPLE 9.5 The thermodynamics of a rubber band. Is the retraction of a rubber band driven by a change in enthalpy or in entropy? The answer to this question will help us to construct a model for the microscopic behavior of polymeric materials in Chapter 33. Suppose you apply a quasi-static stretching force that increases the length L of a rubber band. The force of retraction f exerted by the rubber band is equal and opposite to the applied stretching force. Follow the recipe of Equation (9.1) for finding the fundamental thermodynamic equation when elastic forces are involved. You have $U = U(S, V, L)$ (N is fixed), so

$$dU = T dS - p dV + f dL. \quad (9.18)$$

We are interested in experiments at constant T and p , so we want the Gibbs free energy $dG = d(H - TS) = d(U + pV - TS)$. Substitute Equation (9.18) into this expression to get

$$dG = -S dT + V dp + f dL. \quad (9.19)$$

It follows from Equation (9.19) and the definition $G = H - TS$ that the force f can be defined in terms of enthalpic and entropic components:

$$f = \left(\frac{\partial G}{\partial L} \right)_{T,p} = \left(\frac{\partial H}{\partial L} \right)_{T,p} - T \left(\frac{\partial S}{\partial L} \right)_{T,p}. \quad (9.20)$$

To get a Maxwell relationship for $(\partial S / \partial L)_{T,p}$, take the cross-derivatives in Equation (9.19):

$$\left(\frac{\partial S}{\partial L} \right)_{T,p} = - \left(\frac{\partial f}{\partial T} \right)_{p,L}. \quad (9.21)$$

Equation (9.21) implies that you can find the entropic component of the force, $(\partial S / \partial L)_T$, from a simple experiment. Hold the rubber band at a fixed stretched length L (and at constant pressure: atmospheric, for example) and measure how the retractive force depends on the temperature (see Figure 9.6). The slope of that line, $(\partial f / \partial T)_L$, will give $-(\partial S / \partial L)_T$. The positive slope of $f(T)$ in Figure 9.6 indicates that $S(L)$ will have a negative slope and therefore stretching rubber involves some type of molecular ordering upon stretching.

How do you get the enthalpy component, $(\partial H / \partial L)_T$? Substituting Equation (9.21) into (9.20) gives

$$\left(\frac{\partial H}{\partial L} \right)_{T,p} = f - T \left(\frac{\partial f}{\partial T} \right)_{p,L}, \quad (9.22)$$

which you can determine from the same experiment. Figure 9.6 shows that the retraction of rubber is stronger at higher temperatures. This

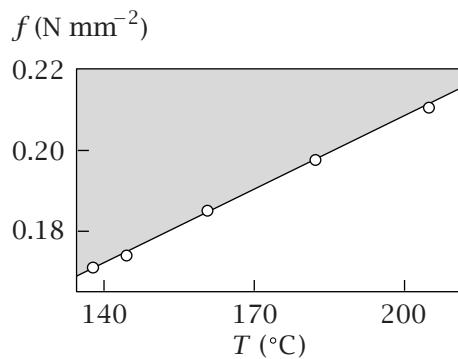


Figure 9.6 The retractive force f of a rubbery polymer, amorphous polyethylene, held at constant length, as a function of temperature T . The slope is $(\partial f / \partial T)_L$. Source: JE Mark, A Eisenberg, WW Graessley, L Mandelkern, and JL Koenig, *Physical Properties of Polymers*, 2nd edition, American Chemical Society, Washington, DC (1993). The data are from JE Mark, *Macromol Rev* **11**, 135 (1976).

observation, first made in 1806 by J Gough, a British chemist, distinguishes rubber from metal. The entropy of rubber *decreases* on stretching, while the entropy of metal *increases* on stretching. Stretching metal loosens the bonds, increasing the volume per atom and the entropy. Stretching rubber decreases the multiplicity of polymer conformations (see Example 2.4). By using the data shown in Figure 9.6 and the expressions above, you see that rubber elasticity is mostly entropic.

Example 9.6 explores the lateral forces among surfactant molecules on a two-dimensional liquid surface, using the same methods as in Example 9.5.

EXAMPLE 9.6 Surface forces in a Langmuir trough. Following Example 9.5, let's apply Maxwell's relations to explore a molecular model of surface forces. A *Langmuir trough* is a fancy dishpan with a bar floating in it; see Figure 9.7. On one side of the bar is water. On the other side is water plus a surfactant, such as soap or phospholipids. If you push sideways on the bar with a fixed lateral pressure π , the left-side surface can change its area A , in much the same way as pushing on a piston changes the volume of a gas in a three-dimensional cylinder. In this trough experiment, you can measure $\pi(T, A, N)$, the lateral pressure that you must apply as a function of temperature T , the number of surfactant molecules N , and the surface area A .

Let's make a simple model that you can test experimentally. Suppose that N surfactant molecules occupy A lattice sites on the two-dimensional surface on the left side of the trough. The molecules are dilute, so $N \ll A$. Assume the surfactant molecules on the two-dimensional surface each occupy one of A lattice sites, so the translational multiplicity of states is $W \approx A^N$. So, this model predicts that the entropy increases with area:

$$S(A) = k \ln W = Nk \ln A. \quad (9.23)$$

How can you test this model? If T and p are constant, equilibrium is predicted by a minimum of the following free energy function,

$$dG = -S dT + V dp - \pi dA. \quad (9.24)$$

(We use a minus sign here, $-\pi$, to indicate that the free energy increases as you compress the molecules; this is the same convention we used for the pV

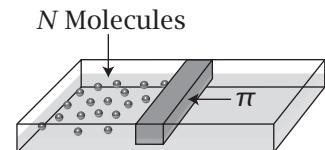


Figure 9.7 Langmuir trough. Two compartments are separated by a floating bar. Both sides contain water. The left side also contains surfactant molecules. A pressure π is applied to the bar (a force per unit length of the bar), pushing toward the left. Measuring the force that is needed to hold a given number of surfactant molecules in a given area gives the *pressure-area isotherm*, which provides an equation of state for two-dimensional systems.

term.) You can use Equation (9.24) and the definition $G = H - TS$ to partition the lateral pressure π into its enthalpic and entropic components:

$$\pi = -\left(\frac{\partial G}{\partial A}\right)_{T,p} = -\left(\frac{\partial H}{\partial A}\right)_{T,p} + T\left(\frac{\partial S}{\partial A}\right)_{T,p}. \quad (9.25)$$

To get a Maxwell relation for $(\partial S/\partial A)_{T,p}$, take the cross-derivatives in Equation (9.24):

$$\left(\frac{\partial S}{\partial A}\right)_{T,p} = \left(\frac{\partial \pi}{\partial T}\right)_{p,A}. \quad (9.26)$$

Now taking the derivative in Equation (9.26) of $S(A)$ from Equation (9.23) predicts that

$$\left(\frac{\partial \pi}{\partial T}\right)_{p,A} = \frac{Nk}{A} \quad (9.27)$$

$$\Rightarrow \pi = \frac{NkT}{A}, \quad (9.28)$$

which is the two-dimensional version of the ideal gas law. This pressure is purely entropic. By measuring $\pi(A, T)$, you can determine whether this model satisfactorily explains your data. If the surfactant molecules are not sufficiently dilute, you may observe enthalpic components, indicating that the surfactant molecules interact with each other. Maxwell's relations give you relationships for using experimental observables to test microscopic models.

The Tortuous Logic of Thermodynamics: a Graphical Perspective

We noted in Chapters 7 and 8 that fundamental thermodynamic quantities are seldom measurable and that measurable quantities are seldom fundamental. Figure 9.8 shows a graphical approach illustrating the chain of thermodynamic logic starting from the fundamental but unmeasurable function $S(V)$ and working down to the thermodynamic expansion coefficient α , a measurable quantity.

Multicomponent Systems Have Partial Molar Properties

We now develop some useful thermodynamic relationships for systems having more than one chemical component, such as liquid mixtures or metal alloys. First, let's look at a simple property, the volume V of a system. Consider a system having n moles of just one type, a one-component system. The *molar volume* (the volume per mole) is $v = V/n$. Other molar properties, such as the molar free energy, $g = G/n$, are also found by dividing by n .

Partial Molar Volumes

But what if your system has more than one chemical species, with mole numbers $\mathbf{n} = n_1, n_2, \dots, n_M$? Now each species of molecule can have a different molar volume. Equally important, the numbers of moles of each species can be specified independently, because different components can be present in

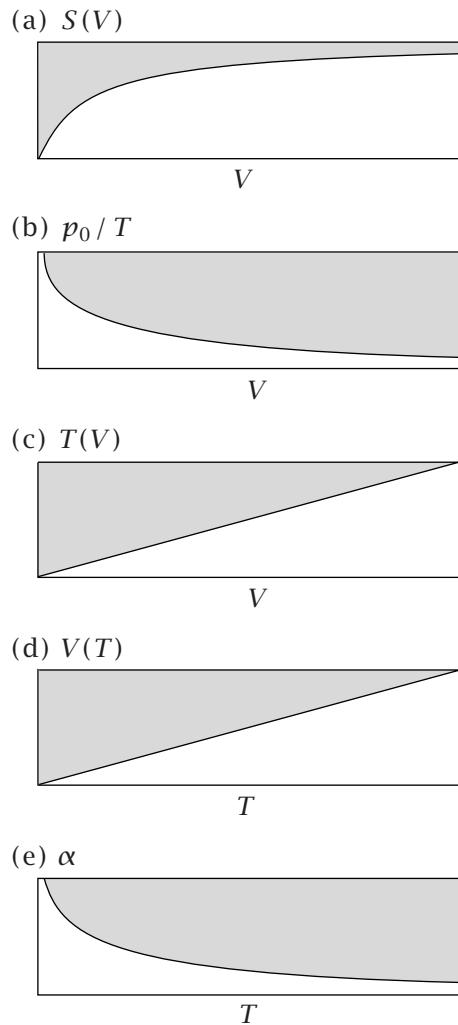


Figure 9.8 The top figure (a) shows the entropy, $S(V) = Nk \ln V$ for an ideal gas (see step (b)). To get (b), take the slope of curve (a), $p/T = (\partial S/\partial V)_{U,N} = Nk/V$, and fix the pressure at a constant value, $p = p_0$. (b) is a plot of this function, $p_0/T = Nk/V$. To get from (b) to (c), invert the y axis so you have $T(V) = p_0 V/Nk$. To get from (c) to (d), interchange the dependent and independent variables by interchanging x and y axes in (c): $V(T) = NkT/p_0$. Finally, the isothermal compressibility shown in (e) is the slope of (d), divided by V , $\alpha = V^{-1}(\partial V/\partial T)_p = (p_0/NkT)(Nk/p_0) = 1/T$.

different amounts. So a single quantity, the molar volume, is not sufficient to describe how the total volume depends on the components in multicomponent systems. Rather, you need the *partial molar volumes* v_j ,

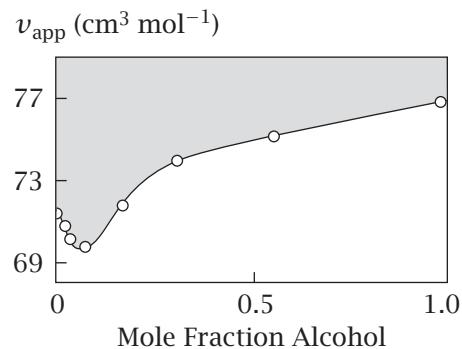
$$v_j = \left(\frac{\partial V}{\partial n_j} \right)_{T,p,n_{i \neq j}} . \quad (9.29)$$

The partial molar volume v_j describes how much the volume V of the system increases when you add dn_j moles of molecules of type j , with all other components held constant:

$$dV = \sum_{j=1}^M \left(\frac{\partial V}{\partial n_j} \right)_{T,p,n_{i \neq j}} dn_j = \sum_{j=1}^M v_j dn_j. \quad (9.30)$$

A related quantity is the *partial specific volume*, which is the derivative with respect to the *weights* w_j of species j , $(\partial V/\partial w_j)_{T,p,w_{i \neq j}}$, rather than with respect to mole numbers.

Figure 9.9 Adding alcohol to water at first decreases the apparent partial molar volume v_{app} . Adding more alcohol then increases this volume.
Source: SG Bruun and A Hvidt, *Ber Bunsenges Phys Chem* **81**, 930–933 (1977).



In ideal cases, partial molar volumes are independent of the composition of the mixture, and are equal to the molar volumes of the individual pure components. But this is not true in general. For example, if you mix a volume V_a of alcohol with a volume V_w of water, the total volume of the mixture may not equal the sum, $V_a + V_w$. For small concentrations of alcohol, each alcohol molecule pulls water molecules to itself, so the solution volume is smaller than the sum of the volumes of the pure water and pure alcohol components.

This is illustrated in Figure 9.9, where there are two unknown quantities: the partial molar volume of the water and the partial molar volume of the alcohol. However, if you fix a quantity, say v_w , equal to the molar volume of pure water, then you can represent the volume of the alcohol in terms of a single variable v_{app} , the *apparent* partial molar volume of the alcohol:

$$v_{\text{app}} = \frac{V - n_w v_w}{n_a},$$

where n_a is the number of moles of alcohol and n_w is the number of moles of water. At low concentrations in water, the apparent partial molar volumes of alcohols are smaller than for pure alcohols, indicating the shrinkage of the total volume when these two liquids are combined.

Ions such as magnesium sulfate have negative partial molar volumes in water. Stirring the ions into water gives less total volume than the sum of the starting ions and pure water before mixing. Water molecules are strongly attracted to the ions, through electrostatic interactions, so adding such salts to water can shrink the volume of the liquid. This is called *electrostriction*.

Figure 9.10 shows how the partial molar volume can depend on composition, using a barrel of bowling balls and adding sand. To begin, the barrel contains bowling balls and no sand. Now sand goes into the barrel. At first, the sand adds no volume to the system because it fills only the cavities between the bowling balls. As long as there is not enough sand to go beyond the tops of the bowling balls, the partial molar volume of the sand equals zero because it does not change the volume of the system, which is defined by the floor and walls of the barrel and the height to which the barrel is filled. Once the sand fills all the cavities and extends above the bowling balls it begins to increase the volume of the system. Now the *partial molar* volume equals the *molar* volume of the sand.

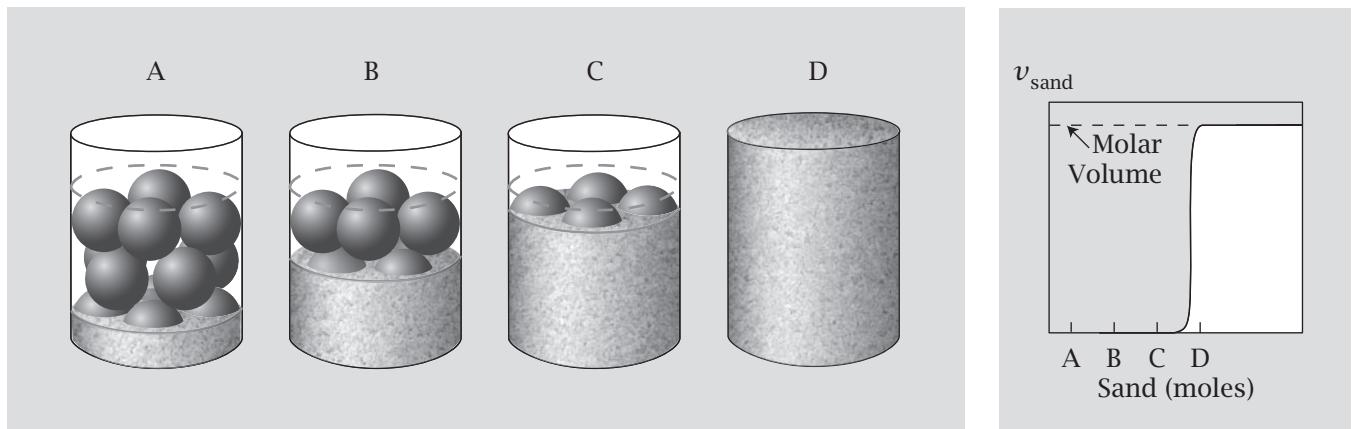


Figure 9.10 Adding sand to a barrel of bowling balls illustrates the idea of partial molar volume. At first, when the sand is at low ‘concentration,’ adding sand just fills in the holes between the bowling balls without increasing the barrel volume that is needed to contain the bowling balls and sand. However, when all the space between the bowling balls is filled, adding sand does add volume. At that point D, the partial molar volume v_{sand} equals the molar volume ν .

Chemical Potentials Are Partial Molar Free Energies

Perhaps the most useful partial molar quantity is the chemical potential (see Chapters 13 and 16), which is the partial free energy, either per mole (if you prefer to work with numbers of moles n_j) or per molecule (if you prefer numbers of molecules N_j). Here we will use numbers of molecules N_j . The chemical potential is given in Table 8.1 in terms of each of the quantities U , F , H , and G :

$$dU = T dS - p dV + \sum_{j=1}^M \mu_j dN_j,$$

$$dF = -p dV - S dT + \sum_{j=1}^M \mu_j dN_j,$$

$$dH = T dS + V dp + \sum_{j=1}^M \mu_j dN_j,$$

$$dG = -S dT + V dp + \sum_{j=1}^M \mu_j dN_j.$$

It follows that the chemical potential has several equivalent definitions:

$$\begin{aligned} \mu_j &= \left(\frac{\partial U}{\partial N_j} \right)_{V,S,N_{i \neq j}} = \left(\frac{\partial G}{\partial N_j} \right)_{T,p,N_{i \neq j}} \\ &= \left(\frac{\partial F}{\partial N_j} \right)_{T,V,N_{i \neq j}} = \left(\frac{\partial H}{\partial N_j} \right)_{S,p,N_{i \neq j}}. \end{aligned} \quad (9.31)$$

For example, you could get μ_j by measuring either how U depends on N_j with V and S held constant, or how G depends on N_j with T and p constant. However, *partial molar* quantities are defined specifically to be quantities measured at constant T and p . So only the derivative of the Gibbs free energy $(\partial G / \partial n_j)_{T,p,N_{i\neq j}}$ is called a *partial molar* quantity.

You can divide the partial molar free energy into its partial molar enthalpy and entropy components by using the definition $G = H - TS$:

$$\begin{aligned}\mu_j &= \left(\frac{\partial G}{\partial N_j} \right)_{T,p,N_{i\neq j}} \\ &= \left(\frac{\partial H}{\partial N_j} \right)_{T,p,N_{i\neq j}} - T \left(\frac{\partial S}{\partial N_j} \right)_{T,p,N_{i\neq j}} = h_j - Ts_j,\end{aligned}\quad (9.32)$$

where h_j and s_j are the partial molar enthalpy and entropy, respectively.

When Equation (9.32) is used in expressions such as

$$dG = -S dT + V dp + \sum_{j=1}^M (h_j - Ts_j) dN_j, \quad (9.33)$$

you see that entropy quantities appear in both the first and third terms. The first term, $-S dT$, describes how the free energy of a system changes if you change the temperature, without changing the pressure or the chemical constitution of the system. In contrast, the quantity $(h_j - Ts_j) dN_j$ describes how the free energy changes if dN_j moles of species j are transferred into the system, with both temperature and pressure held constant. Partial molar enthalpies and entropies pertain to changes in chemical composition, in the absence of thermal changes. Changes in one partial molar quantity can affect other partial molar quantities, when the quantities are ‘linked.’ Partial molar quantities are useful for mixtures or solutions or alloys when you want to focus on what one particular component of the system is doing.

Below, we use partial molar quantities to show that pV effects are typically small in liquids and solids. For condensed phases, you can use Gibbs or Helmholtz free energies interchangeably.

EXAMPLE 9.7 For condensed phases, the Helmholtz and Gibbs free energies are approximately equal. For many condensed-phase properties, Gibbs and Helmholtz free energies are interchangeable. Said differently, it doesn’t matter whether you fix the pressure or the volume when measuring those properties. To see this, let’s compute the free energy of inserting one water molecule into liquid water. Start with two of the equivalent definitions of the chemical potential, Equation (9.31):

$$\mu = \left(\frac{\partial G}{\partial N} \right)_{T,p} = \left(\frac{\partial F}{\partial N} \right)_{T,V}. \quad (9.34)$$

This shows that, per molecule, ΔG at constant p is exactly equal to ΔF at constant V , by definition. Now, let's make two comparisons. First, compare ΔF with ΔG , both at fixed pressure. To do this, use $F = G - pV$ (see Table 8.1) and take the derivative of both sides with respect to N , holding T and p constant, to get

$$\left(\frac{\partial F}{\partial N}\right)_{T,p} = \left(\frac{\partial G}{\partial N}\right)_{T,p} - p \left(\frac{\partial V}{\partial N}\right)_{T,p}, \quad (9.35)$$

Second, let's see that Equation (9.35) also gives the same result as comparing ΔF at constant volume with ΔF at constant pressure. Substitute Equation (9.34) into Equation (9.35) to get

$$\begin{aligned} \left(\frac{\partial F}{\partial N}\right)_{T,p} - \left(\frac{\partial F}{\partial N}\right)_{T,V} &= -p \left(\frac{\partial V}{\partial N}\right)_{T,p} \\ &= -(1 \text{ atm}) \left(18 \frac{\text{cm}^3}{\text{mol}}\right) \left(2.442 \times 10^{-2} \frac{\text{cal}}{\text{cm}^3 \text{ atm}}\right) \\ &= -0.44 \text{ cal mol}^{-1}. \end{aligned} \quad (9.36)$$

We used the molar volume of water $(\partial V / \partial N)_{T,p} = 18 \text{ cm}^3 \text{ mol}^{-1}$ and $p = 1 \text{ atm}$. The constant on the right is just the ratio of gas constants R in two different units, to provide a units conversion. The difference we computed of $-0.44 \text{ cal mol}^{-1}$, is very small. (If, instead of a condensed phase, you had computed this quantity for an ideal gas, you would have found $-p(\partial V / \partial N)_{T,p} = -RT = -600 \text{ cal mol}^{-1}$). So, inserting water into liquid water at constant pressure, the two types of free energy are nearly identical, $\Delta G - \Delta F \approx 0$, whereas for inserting water into the gas phase gives a much larger value, $\Delta G - \Delta F = RT$.

Summary

Thermodynamics is quite general and can treat different types of forces and work. Maxwell's relations are equalities among partial derivatives that derive from Euler's reciprocal relation for state functions. Maxwell's relations provide a powerful way to predict useful but unmeasurable quantities from measurable ones. In the following chapters, we will develop microscopic statistical mechanical models of atoms and molecules.

Problems

1. How do thermodynamic properties depend on surface area? The surface tension of water is observed to decrease linearly with temperature (in experiments at constant p and a): $\gamma(T) = b - cT$, where T = temperature (in $^{\circ}\text{C}$), $b = 75.6 \text{ erg cm}^{-2}$ (the surface tension at 0°C) and $c = 0.1670 \text{ erg cm}^{-2} \text{ deg}^{-1}$.

- If γ is defined by $dU = T dS - p dV + \gamma da$, where da is the area change of a pure material, give γ in terms of a derivative of the Gibbs free energy at constant T and p .
- Using a Maxwell relation, determine the quantitative value of $(\partial S / \partial a)_{p,T}$ from the relationships above.
- Estimate the entropy change ΔS from the results above if the area of the water/air interface increases by 4 \AA^2 (about the size of a water molecule).

2. Water differs from simple liquids. Figures 9.3 and 9.5 show that the thermal expansion coefficient $\alpha = (1/V)(\partial V / \partial T)_p$ and isothermal compressibility $\kappa = -(1/V)(\partial V / \partial p)_T$ are both much smaller for water, which is hydrogen-bonded, than for simpler liquids like benzene, which are not. Give a physical explanation for what this implies about molecular packing and entropies in water versus simple liquids.

3. The heat capacity of an ideal gas. For an ideal gas, $(\partial U / \partial V)_T = 0$. Show that this implies the heat capacity C_V of an ideal gas is independent of volume.

4. Using Maxwell's relations. Show that $(\partial H / \partial p)_T = V - T(\partial V / \partial T)_p$.

5. Pressure dependence of the heat capacity.

- Show that, in general, for quasi-static processes,

$$\left(\frac{\partial C_p}{\partial p} \right)_T = -T \left(\frac{\partial^2 V}{\partial T^2} \right)_p .$$

- Based on (a), show that $(\partial C_p / \partial p)_T = 0$ for an ideal gas.

6. Relating Gibbs free energy to chemical potential. Prove that $G = \sum_{i=1}^N \mu_i N_i$.

7. Piezoelectricity. Apply a mechanical force f along the x axis of a piezoelectric crystal, such as quartz, which has dimension ℓ in that direction. The crystal will develop a polarization p_0 , a separation of charge along the x axis, positive on one face and negative on the other. Applying an electric field E along the x axis causes a mechanical deformation. Such devices are used in microphones, speakers, and pressure transducers. For such systems, the energy equation is $dU = T dS + f d\ell + E dp_0$. Find a Maxwell relation to determine $(\partial p_0 / \partial f)_{T,E}$.

8. Relating C_V and C_p . Show that $C_p = C_V + Nk$ for an ideal gas.

9. Rubber bands are entropic springs. Experiments show that the retractive force f of polymeric elastomers as a function of temperature T and expansion L is approximately given by $f(T, L) = aT(L - L_0)$ where a and L_0 are constants.

- Use Maxwell's relations to determine the entropy and enthalpy, $S(L)$ and $H(L)$, at constant T and p .
- If you adiabatically stretch a rubber band by a small amount, its temperature increases, but its volume does not change. Derive an expression for its temperature T as a function of L , L_0 , a , and its heat capacity $C = (\partial U / \partial T)$.

10. Metal elasticity is due to energy, not entropy. Experiments show that the retractive force f of a metal rod as a function of temperature T and extension L relative to undeformed length L_0 is given by $f(T, L) = Ea\Delta L / L_0$, where $\Delta L = L[1 - \alpha(T - T_0)] - L_0 = L - L\alpha(T - T_0) - L_0$. a is the cross-sectional area of the rod, E (which has the role of a spring constant) is called *Young's modulus*, and $\alpha \approx 10^{-5}$ is the linear expansion coefficient. Compute $H(L)$ and $S(L)$. Is the main dependence on L due to enthalpy H or entropy S ?

11. Pressures in surfactant monolayers. A Langmuir trough is shown in Figure 9.7. N surfactant molecules float on water on the left side of a bar. When you push the bar to the left with pressure π , the N surfactant molecules occupy an equilibrium area A , where

$$\pi = - \left(\frac{\partial G}{\partial A} \right)_{T,p,N} .$$

Consider a two-dimensional van der Waals model, in which the entropy is given by

$$S = k \ln(A - A_0),$$

where A_0 is the minimum area the molecules can occupy when they crowd together. The enthalpy is given by

$$H = - \frac{b}{a} .$$

- You can express the pressure in terms of enthalpic (π_H) and entropic (π_S) components:

$$\pi = \pi_H + \pi_S .$$

Write a Maxwell relation for

$$\pi_S = T \left(\frac{\partial S}{\partial A} \right)_{T,p,N} .$$

- Express π_H in terms of measurable quantities T and A .

- (c) Write the equation of state of the model, giving the quantity $\pi(T, A, N)$.
 (d) Write an expression for the free energy $\Delta G(T, A, N)$ at constant p .

12. Torque on DNA. One end of a DNA molecule is fixed to a surface. Using single-molecule methods, a torque (twisting force) is applied to the other end, twisting the molecule as shown in Figure 9.11.

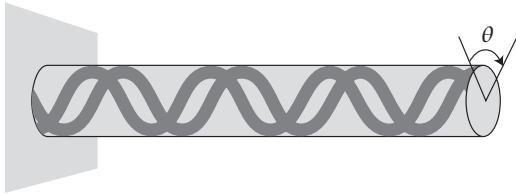


Figure 9.11

When the molecule is twisted through an angle θ , the DNA gives a restoring torque τ that depends on temperature T :

$$\tau = (k_0 + k_1 T) \theta,$$

where k_0 and k_1 are measured constants.

- (a) Write an expression for the free energy $G(\theta)$ at fixed T and p .
 (b) Use a Maxwell relation to express $(\partial S / \partial \theta)_{T,p}$ in terms of experimental observables, where S is entropy and p is pressure.
 (c) Derive an expression for the entropy $S(\theta)$ at constant T and p .
 (d) Derive an expression for the enthalpy $H(\theta)$ at constant T and p .
 (e) Under what conditions, i.e., for what values of k_0 , k_1 , and T , will this torsional retractive force be entropy-driven?

13. The Massieu function $J(\beta, V, N)$. Whereas the entropy is an extremum function $S(U, V, N)$, it is sometimes useful to use a related extremum function $J(\beta, V, N)$, where $\beta = 1/T$ and T is temperature. For this function, β and U are conjugate variables.

- (a) Derive an expression for the differential quantity dJ in terms of variations $d\beta$, dV , and dN . (Hint: as an initial guess for J , try either $J = S + \beta U$ or $J = S - \beta U$).
 (b) Using additional ‘reduced’ variables $\pi = p/T$ and $m = \mu/T$, write Maxwell relations for $(\partial \pi / \partial N)_{\beta, V}$, $(\partial U / \partial V)_{\beta, N}$, and $(\partial U / \partial N)_{\beta, V}$.
 (c) What is the relation between J and the Helmholtz free energy $F(T, V, N)$?

14. Thermodynamics of a single molecule. One end of a molecule is fixed to a surface. Another end is pressed down by the tip of an atomic force microscope, as shown in Figure 9.12. Using the tip of an atomic force microscope as a probe, we measure the force of molecule retraction.

The force depends on the temperature at which the experiment is performed:

$$f = (aT + b)x,$$

where a and b are constants.

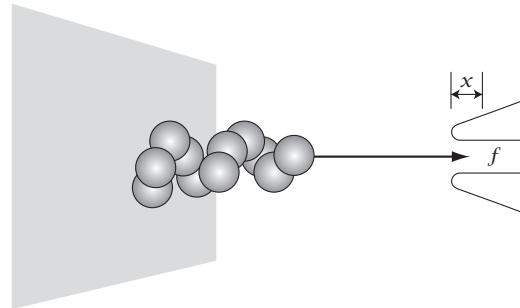


Figure 9.12

- (a) Write a fundamental equation describing the energy of the molecule. Write an expression for dG at fixed T and P . Derive an expression showing how G depends on x .
 (b) Use a Maxwell relation to express $(\partial S / \partial x)_{T,P}$.
 (c) Derive an expression for $S(x)$, at fixed T and P .
 (d) Derive an expression for enthalpy $H(x)$ at fixed T and P .
 (e) What values of a and b will cause the retraction force to be driven by a change in entropy?

15. Gibbs-Helmholtz Equation. From the definition of $G = H - TS$, derive the Gibbs-Helmholtz equation

$$\frac{\partial}{\partial T} \left(\frac{G}{T} \right) = -\frac{H}{T^2},$$

which relates the temperature variation of the Gibbs free energy G to the enthalpy H .

16. Surface tension components. Figure 9.13 shows the surface tension of water as a function of temperature.

- (a) From Figure 9.13, determine the numerical value of $(\partial \gamma / \partial T)_{p,N,A}$, where T = temperature, p = pressure, and A = surface area.
 (b) Find the Maxwell relation for the partial derivative equal to $(\partial \gamma / \partial T)_{p,N,A}$.
 (c) Write an expression for the enthalpic and entropic components of the surface tension γ .
 (d) Combining the results from above, compute the numerical values of the enthalpic and entropic parts of γ at $T = 300$ K, and comment on which component dominates the surface tension.

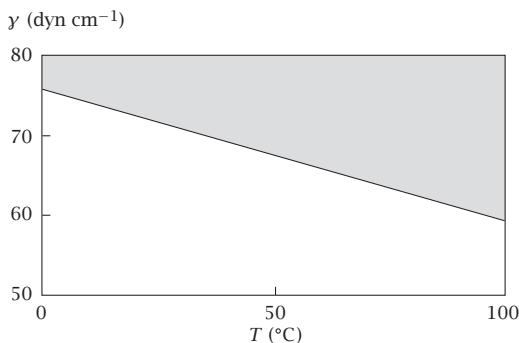


Figure 9.13 Surface tension of water versus temperature.
Source: redrawn from data in *CRC Handbook of Chemistry and Physics*, 47th edition, RC Weast, ed., The Chemical Rubber Company, Cleveland, OH, 1962.

17. Exploring the quantity F/T . Sometimes a quantity of interest is the Helmholtz free energy $F(T, V, N)$ divided by T . (For example, this quantity is proportional to the logarithms of equilibrium constants or solubilities.)

- (a) Derive a relationship showing that

$$\frac{\partial(F/T)}{\partial T} \propto U.$$

Find the constant of proportionality.

- (b) Suppose $F(T)$ depends on temperature in the following way: $F(T) = 2aT^2 + bT$ (so $F/T = 2aT + b$). Find $S(T)$ and $U(T)$.

(c) For this and the next parts of this problem, consider a transition between a helix and a coil state of a polymer. F/T is related to the ratio of populations as $F/T = -k \ln(K_{\text{eq}})$, where $K_{\text{eq}} = N_{\text{coil}}/N_{\text{helix}}$. Using $F/T = 2aT + b$ from above, at what temperature are the helix and coil in equilibrium (e.g., half of the molecules are helical and half are unfolded).

- (d) Assuming $a > 0$, does increasing the temperature favor the helix or the coil?

Suggested Reading

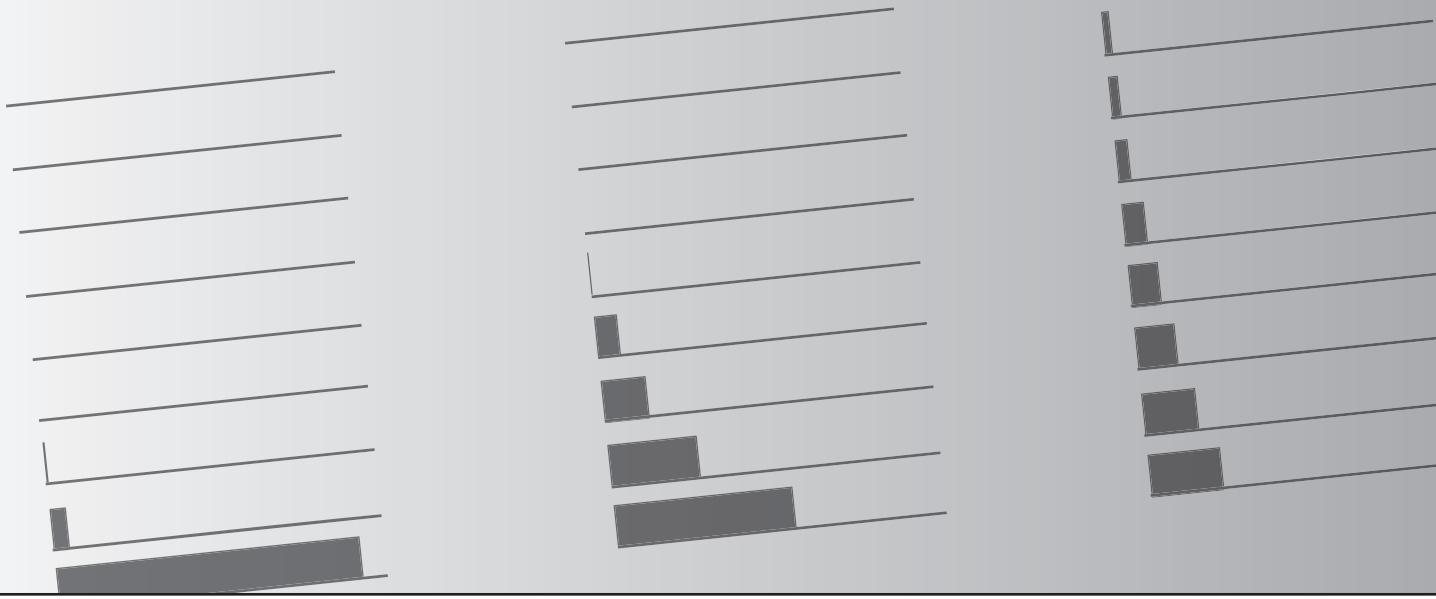
RS Berry, SA Rice, and J Ross, *Physical Chemistry*, 2nd edition, Oxford University Press, New York, 2000. A complete and advanced text on physical chemistry, including thermodynamics.

HB Callen, *Thermodynamics and an Introduction to Thermostatistics*, 2nd edition, Wiley, New York, 1985. Very clearly written discussion of Maxwell's relations.

G Carrington, *Basic Thermodynamics*, Oxford University Press, Oxford, 1994. Good discussion of the principles of thermodynamics, including many applications to models and equations of state.

RT DeHoff, *Thermodynamics in Materials Science*, 2nd edition, CRC Press, Boca Raton, FL, 2006. A useful practical guide for finding thermodynamic relations by simple recipes.

10 The Boltzmann Distribution Law



Statistical Mechanics Gives Probability Distributions for Atoms and Molecules

Now we begin statistical mechanics, the modeling and prediction of the properties of materials from the structures of the atoms and molecules of which they are composed. The core of **statistical mechanics** is modeling the **probability distributions** of the energies of **atoms and molecules**. The various averages over those distributions are what experiments measure. For example, to compute the **properties of gases**, you need the distributions of their energies and velocities (Chapter 11). You can predict **chemical reaction equilibria** if you know the distributions of the energies of the reactants and products (Chapter 13). And you can predict the average number of ligands bound to a DNA molecule if you know the distribution of energies of all the **ligation states** (Chapter 28).

The central result of this chapter is the Boltzmann distribution law, which gives probability distributions from the underlying energy levels. We derive this result by bringing together the two main threads from earlier chapters. First, the principles of thermodynamics in Chapters 6–9 describe how to predict the state of equilibrium. Second, Chapter 5 relates a macroscopic property of the equilibrium (the entropy) to a microscopic property, the probability distribution.

Here's the kind of problem we want to solve. Example 8.2 describes a two-dimensional model of a four-bead polymer chain that has four open conformations and one compact conformation. In that example, we computed the free

energies of the open and compact states. Now we want to compute the probability distribution, **the fraction of molecules that are in each conformation**.

To begin, we need to define the system and its **energy levels**. The system is one four-bead chain that has two energy levels (see Figure 10.1). Each energy level represents the number of bead-bead contacts that the chain can make. Let's use the convention that zero is the lowest energy, the lowest rung on the ladder. Conformations with the maximum number of bead-bead contacts have zero energy. Breaking a contact increases the energy by an amount $\varepsilon = \varepsilon_0 > 0$. We use ε to indicate the energy in general and ε_0 to represent some particular constant value. We seek the distribution of probabilities p_1, p_2, \dots, p_5 that the four-bead chain is in its various conformations.

The state of lowest energy is called the **ground state**. States of higher energy are called **excited states**. Each of the five configurations is called a **microstate**, to distinguish it from a *state* or *macrostate*, which is a collection of microstates. Each microstate is a snapshot of the system. When you measure system properties, you measure **averages over multiple microstates**. In Figure 10.1, you can see two macrostates of a one-chain system: the open macrostate, composed of four microstates, and the compact macrostate, composed of one microstate. For a lattice gas, a particular arrangement of N particles on M lattice sites is one microstate. A corresponding macrostate could include, for example, all such arrangements of particles that have the same density N/M .

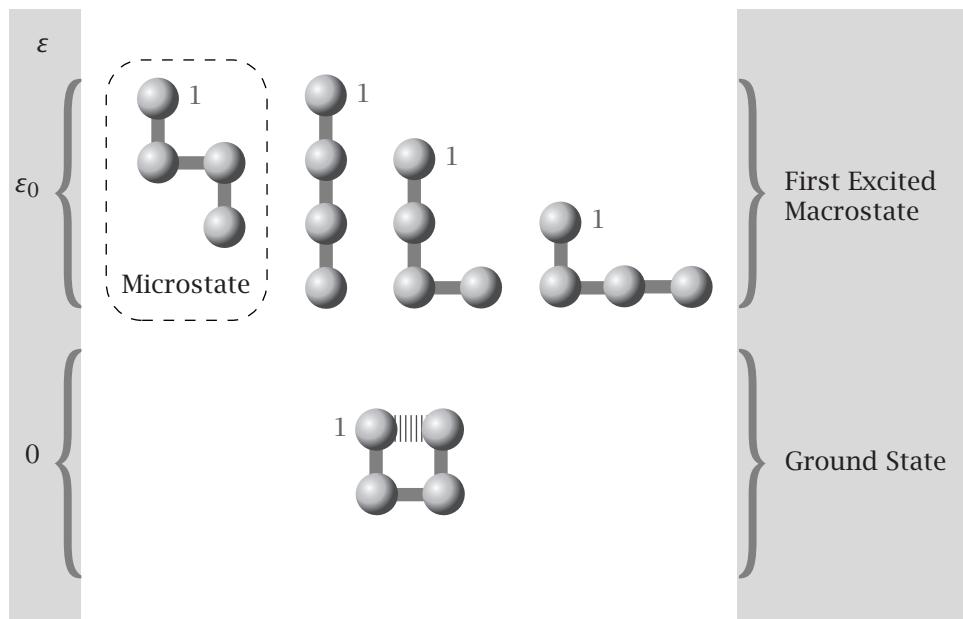


Figure 10.1 The five conformations of the four-bead chain in Example 8.2 are grouped on an energy ladder. Conformations with one bead-bead contact, $n_C = 1$, are taken to have energy $\varepsilon = 0$. The other four conformations have no bead-bead contacts, so $\varepsilon = \varepsilon_0$, where ε_0 is a constant. The number 1 next to the first bead indicates that the chain has a head and a tail. (Sometimes symmetry matters. Here, we counted only left turns. The numbers are different if you also count right turns separately. Here, for illustrating general principles, it doesn't matter.)

We will compute the probability distribution for this four-bead chain after we derive the Boltzmann law. But first, let's look at a more general version of the problem that we are trying to solve. Consider a complex system having N particles. Each particle has some **internal energy** due to its intrinsic properties. For example, different internal energies result from different **rotational** or **vibrational states** of molecules, or from the different conformations like the **four-bead polymer molecule**. In addition, there may be **interaction energies** between pairs or triples or larger collections of particles. The sum total of all the **internal and interaction energies**, taken for one particular arrangement j of the whole system, is E_j . The lowest energy over all the possible arrangements of the system is E_1 . The next lowest energy arrangement will have energy E_2 , etc. We will use two different notations for the energy: E_j for any system in general, no matter how complex, and ϵ_j for the specific simple case of independent particles (such as ideal gas atoms).

Now we derive the probability distribution for any system with known energy level spacings. We do this by **combining the definition of entropy**, $S/k = -\sum p_j \ln p_j$, **with the definition of equilibrium**. We focus here on a system with variables (T, V, N) . Our approach follows the dice problems of Chapter 5, but now, instead of knowing an average score, you know the **average energy** of the system.

The **Boltzmann Distribution Law Describes the Equilibria Among Atoms and Molecules**

Consider a system having N particles. (To make the math simple in this chapter, consider N particles of a single type, rather than a set of $\mathbf{N} = (N_1, N_2, \dots, N_M)$ particles of multiple types.) Suppose the system has **t different energy levels**, E_j , $j = 1, 2, 3, \dots, t$, defined by the physics of the problem that you want to solve. A given energy level will be composed of **many different microstates**. Energies may be equally spaced, as in the polymer problem above, or they may be spaced differently. They may come from quantum mechanics, as we'll see in Chapter 11. Given the energies E_j , we aim to compute the probabilities p_j that the system is in **each level j** . (Computing these populations of the energy levels is metaphorically like computing the fraction of dice rolls that have a particular score j .) Suppose (T, V, N) are held constant. Then the condition for equilibrium is $dF = dU - T dS = 0$. Apply the Lagrange multiplier method as in Chapter 5. We need dS and dU .

To get dS , use Equation (5.2) for the entropy as a function of the probabilities p_j :

$$\frac{S}{k} = - \sum_{j=1}^t p_j \ln p_j.$$

Differentiating with respect to p_j (holding the $p_{i \neq j}$'s constant) gives

$$dS = -k \sum_{j=1}^t (1 + \ln p_j) dp_j. \quad (10.1)$$

To get dU , we postulate that the internal energy U , which is the macroscopic quantity from thermodynamics, is the average over all the microscopic states (see Equation (1.35)):

$$U = \langle E \rangle = \sum_{j=1}^t p_j E_j. \quad (10.2)$$

Now take the derivative of Equation (10.2):

$$dU = d\langle E \rangle = \sum_{j=1}^t (E_j dp_j + p_j dE_j). \quad (10.3)$$

Like the macroscopic energy U , the energy levels $E_j = E_j(V, N)$ depend on V and N . But, unlike U , the energy levels E_j do not depend on S or T . We take as a fundamental principle of quantum mechanics that only the populations $p_j(T)$, and not the energies E_j , depend on temperature. However, the *average energy* $\langle E \rangle = \sum p_j(T) E_j$ does depend on temperature. $dE_j = (\partial E_j / \partial V) dV + (\partial E_j / \partial N) dN = 0$ because both V and N are held constant here, and Equation (10.3) becomes

$$d\langle E \rangle = \sum_{j=1}^t E_j dp_j. \quad (10.4)$$

The First Law of Thermodynamics gives $dU = \delta q + \delta w$, which reduces to $d\langle E \rangle = dU = \delta q$ when V and N are constant. Because Equation (10.4) applies when V is constant, it follows that the term $\sum_j E_j dp_j$ is the heat and $\sum_j p_j dE_j$ is the work.

We want the probability distribution that satisfies the equilibrium condition $dF = d\langle E \rangle - T dS = 0$ subject to the constraint that the probabilities sum to one, $\sum_{j=1}^t p_j = 1$. The constraint can be expressed in terms of a Lagrange multiplier α :

$$\alpha \sum_{j=1}^t dp_j = 0. \quad (10.5)$$

Substitute Equations (10.1) and (10.3)–(10.5) into $dF = dU - T dS = 0$ to get

$$dF = \sum_{j=1}^t \left[E_j + kT(1 + \ln p_j^*) + \alpha \right] dp_j^* = 0. \quad (10.6)$$

According to the Lagrange multiplier equation (4.31), the term in the brackets in Equation (10.6) must equal zero for each value of j , so you have t equations of the form

$$\ln p_j^* = -\frac{E_j}{kT} - \frac{\alpha}{kT} - 1. \quad (10.7)$$

Exponentiate Equation (10.7) to find

$$p_j^* = e^{-E_j/kT} e^{(-\alpha/kT)-1}. \quad (10.8)$$

To eliminate α from Equation (10.8), write the constraint equation

$$\sum_{j=1}^t p_j^* = 1$$

as

$$1 = \sum_{j=1}^t e^{-E_j/kT} e^{(-\alpha/kT)-1}.$$

Divide Equation (10.8) by this form of the constraint equation, to get the **Boltzmann distribution law**

$$p_j^* = \frac{e^{-E_j/kT}}{\sum_{j=1}^t e^{-E_j/kT}} = \frac{e^{-E_j/kT}}{Q}, \quad (10.9)$$

where Q is the **partition function**,

$$Q = \sum_{j=1}^t e^{-E_j/kT}. \quad (10.10)$$

The relative populations of particles in energy levels i and j at equilibrium are given by

$$\frac{p_i^*}{p_j^*} = e^{-(E_i - E_j)/kT}. \quad (10.11)$$

Equation (10.9) gives an exponential distribution law, just as Equation (5.17) does, but here the energy levels E_j replace the scores on individual die rolls.

It is interesting to compare Equation (10.7), which we obtained by minimizing the free energy, with Equation (5.15), which we obtained by the conceptually identical procedure of maximizing the entropy subject to a constraint of the form of Equation (10.4). This comparison shows that the Lagrange multiplier that enforces the constraint of average energy is $\beta = 1/(kT)$.

The Boltzmann distribution says that more particles will have low energies and fewer particles will have high energies. Why? Particles don't have an intrinsic preference for lower energy levels. Fundamentally, all energy levels are equivalent. Rather, *there are more arrangements of the system that way*. It is extremely unlikely that one particle would have such a high energy that it would leave all the others no energy. There are far more arrangements in which most particles have energies that are relatively low, but nonzero. If each particle takes only a small fraction of the total energy, it leaves a great many more ways for the other particles to distribute the remaining energy.

Applications of the Boltzmann Law

Example 10.1 illustrates an application of the Boltzmann distribution law. We compute how the atmospheric pressure depends on the altitude above the Earth's surface.

EXAMPLE 10.1 Barometric pressure of the atmosphere. The energy ε of a gas molecule in the Earth's gravitational field is a function of altitude z :

$$\varepsilon(z) = mgz, \quad (10.12)$$

where g is the gravitational acceleration constant and m is the molecular mass. In this case, the energy is a continuous function (of z), not a discrete ladder, but Boltzmann's law still applies. We assume that the atmosphere is in equilibrium and is at constant temperature (valid only approximately; see below). The populations, or numbers, of molecules $N(z)$ at altitude z relative to the number $N(0)$ at sea level are given by the Boltzmann law, Equation (10.11):

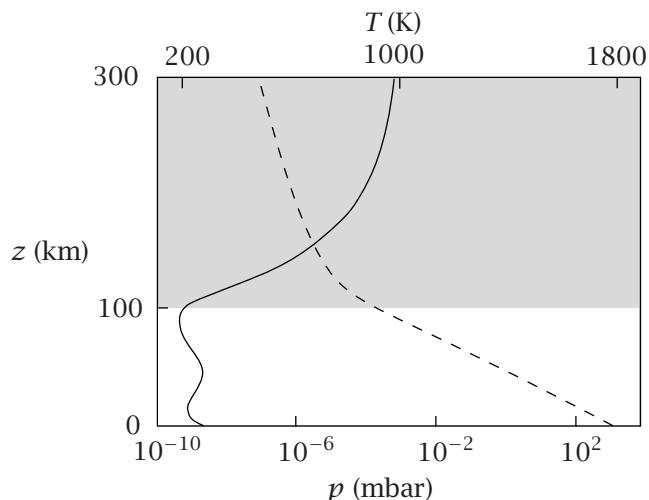
$$\frac{N(z)}{N(0)} = e^{-[\varepsilon(z)-\varepsilon(0)]/kT} = e^{-mgz/kT}. \quad (10.13)$$

If the gas is ideal and the temperature is constant, the pressure $p(z)$ is proportional to the number of molecules per unit volume, so the pressure should decrease exponentially with altitude:

$$\frac{p(z)}{p(0)} = \frac{N(z)kT/V}{N(0)kT/V} = \frac{N(z)}{N(0)} = e^{-mgz/kT}. \quad (10.14)$$

Figure 10.2 shows experimental evidence that the temperature is reasonably constant for the Earth's atmosphere up to about 100 km above the Earth's surface, and that the pressure decreases exponentially with altitude, as predicted. Above about 100 km, the equilibrium assumption no longer holds, because the atmosphere becomes too thin for normal wind turbulence to mix the gases, and the temperature is no longer independent of altitude.

Figure 10.2 The temperature T (—) of the atmosphere is approximately constant up to about 100 km in altitude z . The pressure p (---) decreases exponentially with altitude, following the Boltzmann law. Source: ML Salby, *Fundamentals of Atmospheric Physics*, Academic Press, San Diego, 1996. Data are from: *US Standard Atmosphere*, NOAA, US Air Force, US Government Printing Office, NOAA-S/T 76-1562, Washington, DC, 1976.



Example 10.2 gives another application of the Boltzmann distribution law: the distribution of the velocities of gas molecules. This is the basis for the kinetic theory of gases, a classical model of great historical importance. In our previous lattice modeling, our states have been discrete. We have evaluated discrete sums over Boltzmann factors. Now, we consider a continuous distribution of energies, so we replace sums by integrals.

EXAMPLE 10.2 The Maxwell–Boltzmann distribution of particle velocities. According to the kinetic theory, gases and liquids can be regarded as miniature billiard balls, i.e., Newtonian particles having mass m , velocity v , and kinetic energy ε , where

$$\varepsilon(v) = \frac{1}{2}mv^2.$$

According to the Boltzmann law, the probability $p(v_x)$ that a particle in a container at constant volume and temperature will have velocity v_x in the x direction is

$$p(v_x) = \frac{e^{-\varepsilon(v_x)/kT}}{\int_{-\infty}^{\infty} e^{-\varepsilon(v_x)/kT} dv_x} = \frac{e^{-mv_x^2/2kT}}{\int_{-\infty}^{\infty} e^{-mv_x^2/2kT} dv_x}$$

$$= \left(\frac{m}{2\pi kT} \right)^{1/2} e^{-mv_x^2/2kT}, \quad (10.15)$$

because $\int_{-\infty}^{\infty} e^{-ax^2} dx = \sqrt{\pi/a}$ (see Appendix K, Equation (K.1)) and $a = m/2kT$. This is called *the Maxwell–Boltzmann distribution*. Figure 10.3 shows the excellent agreement between this predicted distribution (multiplied by v^2 to convert to speeds) and experimental observations for the velocities of potassium atoms. The mean velocity is zero, because for any particular velocity v_x in the $+x$ direction, there is a corresponding velocity in the $-x$ direction. The *mean-square velocity* $\langle v^2 \rangle$ defines the width of the distribution:

$$\langle v_x^2 \rangle = \int_{-\infty}^{\infty} v_x^2 p(v_x) dv_x = \left(\frac{m}{2\pi kT} \right)^{1/2} \int_{-\infty}^{\infty} v_x^2 e^{-mv_x^2/2kT} dv_x.$$

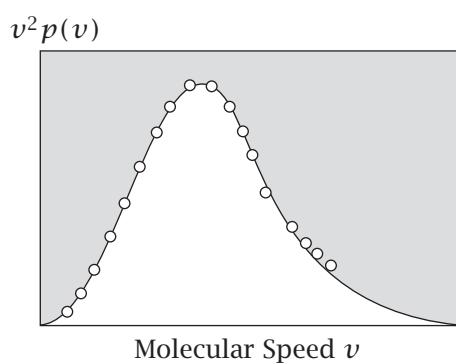


Figure 10.3 Experiments showing $v^2 p(v) = \text{constant} \times v^2 e^{-mv^2/2kT}$ versus average speed v , confirming the Maxwell–Boltzmann distribution of the speeds of potassium atoms in the gas phase. (The factor v^2 converts from velocity, which is a vector with a sign, to speed, which is not.) Source: DA McQuarrie and JD Simon, *Physical Chemistry a Molecular Approach*, University Science Books, Sausalito, 1997. Data are from RC Miller and P Kusch, *Phys Rev* **99**, 1314–1320 (1953).

You can evaluate such integrals using Equations (K.1) and (K.3) of Appendix K:

$$\langle x^2 \rangle = \frac{\int_{-\infty}^{\infty} x^2 e^{(-ax^2)} dx}{\int_{-\infty}^{\infty} e^{(-ax^2)} dx} = \frac{1/2a\sqrt{\pi/a}}{\sqrt{\pi/a}} = \frac{1}{2a}. \quad (10.16)$$

Because $a = m/(2kT)$, you have

$$\langle v_x^2 \rangle = \frac{kT}{m} \implies \frac{1}{2}m\langle v_x^2 \rangle = \frac{1}{2}kT.$$

Now suppose that instead of measuring the x component of velocity, you measure the three-dimensional velocity v , where $v^2 = v_x^2 + v_y^2 + v_z^2$. In an ideal gas, the components are independent of each other, so $\langle v^2 \rangle = \langle v_x^2 \rangle + \langle v_y^2 \rangle + \langle v_z^2 \rangle$, leading to:

$$\frac{1}{2}m\langle v^2 \rangle = \frac{3}{2}kT. \quad (10.17)$$

Because the velocity components are independent, the probabilities multiply, giving

$$\begin{aligned} p(v) &= p(v_x)p(v_y)p(v_z) = \left(\frac{m}{2\pi kT}\right)^{3/2} e^{-m(v_x^2+v_y^2+v_z^2)/2kT} \\ &= \left(\frac{m}{2\pi kT}\right)^{3/2} e^{-mv^2/2kT}. \end{aligned} \quad (10.18)$$

Equations (10.17) and (10.18) are the central results of the kinetic theory of gases, describing the motions of atoms in terms of the kinetic energies of Newtonian particles. They provide a fundamental relationship between temperature and the velocities of the gas molecules.

What Does a Partition Function Tell You?

The partition function is the connection between macroscopic thermodynamic properties and microscopic models. It is a sum of *Boltzmann factors* $e^{-E_j/kT}$ that specify how *particles are partitioned* throughout the accessible states. Equation (10.10) gives

$$Q = \sum_{j=1}^t e^{-E_j/kT} = e^{-E_1/kT} + e^{-E_2/kT} + e^{-E_3/kT} + \dots + e^{-E_t/kT}. \quad (10.19)$$

It is also common to express Q in an alternative form. Experiments rarely give information about absolute energies; they give information about energy differences. So, it is often convenient to define the ground-state energy as zero, $E_1 = 0$, and write the partition function instead in terms of energy differences:

$$Q = 1 + e^{-(E_2-E_1)/kT} + e^{-(E_3-E_1)/kT} + \dots + e^{-(E_t-E_1)/kT}. \quad (10.20)$$

Both forms give exactly the same results for probabilities, which are the quantities we are interested in.

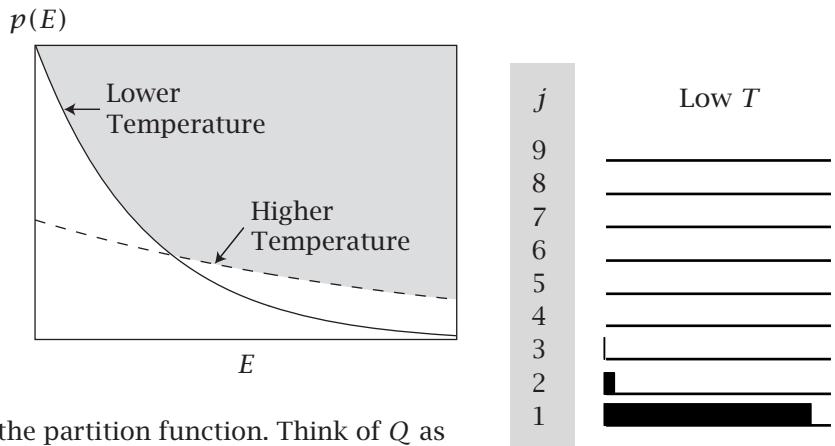


Figure 10.4 According to the Boltzmann distribution, states of lower energy are more populated than states of higher energy. As temperature increases, higher energy states become more populated.

Here's an intuitive way to think about the partition function. Think of Q as the number of states that are *effectively* accessible to the system. To see this, look at the limiting values (see Figures 10.4 and 10.5). When the energies are small, or the temperature is high, all the states become equally populated:

$$\left. \begin{array}{l} E_j \rightarrow 0 \\ \text{or} \\ T \rightarrow \infty \end{array} \right\} \Rightarrow \frac{E_j}{kT} \rightarrow 0 \Rightarrow p_j^* \rightarrow \frac{1}{t} \Rightarrow Q \rightarrow t. \quad (10.21)$$

In this case, all t states become accessible.

At the other extreme, as the energy intervals become large or as the temperature approaches zero, the particles occupy only the ground state:

$$\left. \begin{array}{l} E_{j \neq 1} \rightarrow \infty \\ \text{or} \\ T \rightarrow 0 \end{array} \right\} \Rightarrow \frac{E_{j \neq 1}}{kT} \rightarrow \infty \Rightarrow \begin{cases} p_1^* \rightarrow 1 \\ \text{and} \\ p_{j \neq 1}^* \rightarrow 0 \end{cases} \Rightarrow Q \rightarrow 1. \quad (10.22)$$

In this case, only the ground state becomes accessible.

The magnitude of E_j/kT determines whether or not the state j is 'effectively accessible.' So kT is an important reference unit of energy. States that have energies higher than kT are relatively inaccessible and unpopulated at temperature T , while states having energies lower than kT are well populated. Increasing the temperature makes the higher energy levels effectively more accessible.

The number of *effectively accessible states* is not the same as the number of accessible states, which is always t . The number t is fixed by the underlying physics of the system. In contrast, the effective accessibility Q also depends on the temperature.

The Density of States

Sometimes you may want to compute different probabilities than the ones we computed above. Sometimes, there are intrinsically a different number of ways that a system can occupy one energy level than another. For example, for the four-bead polymer, each configuration shown in Figure 10.1 is one microstate. But for experiments, a more meaningful grouping may be into two macrostates: the *open* state (four microstates) or the *compact* state (one microstate). We have previously defined the quantity W as the multiplicity of

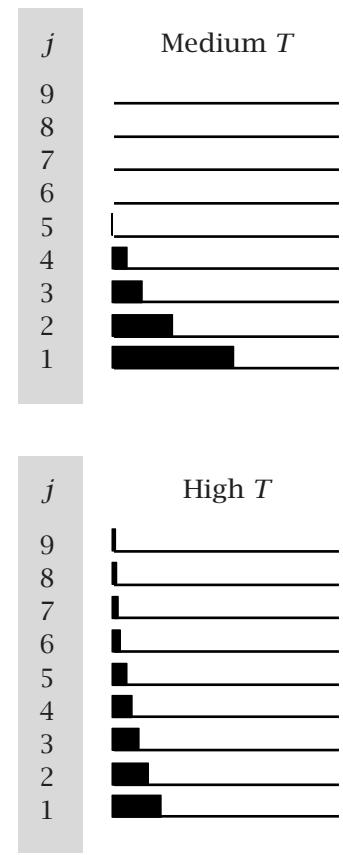


Figure 10.5 The Boltzmann distribution law. At low temperatures T , only the lowest energy states j are populated. Increasing T increases populations of higher energy states. The distribution is exponential.

states. We now generalize this. We now define $W(E)$ to be the *density of states*; that is, $W(E)$ is the total number of ways a system can occur in energy level E . When $W(E) > 1$, an energy level is called *degenerate*. (Think of an unusual die that has nine faces: three faces are 1's, two faces are 2's, and one face each has numbers 3–6. Its density of states would be $W(1) = 3$; $W(2) = 2$; $W(3) = W(4) = W(5) = W(6) = 1$.)

For the four-bead chain, there is one compact conformation and four open conformations (see Figure 10.1), so the density of states is $W(0) = 1$ and $W(\varepsilon_0) = 4$. Since W is the number of microstates per level, the partition function can be expressed as a sum over the two levels (open and compact, in this case), rather than over the five microstates:

$$Q = 1e^{-0/kT} + 4e^{-\varepsilon_0/kT},$$

where the first term describes the one microstate in level 1 ($\varepsilon = 0$) and the second describes the four microstates in level 2 ($\varepsilon = \varepsilon_0$).

When you have a density of states, you can express the partition function as a sum over energy levels $\ell = 1, 2, \dots, \ell_{\max}$ as

$$Q = \sum_{\ell=1}^{\ell_{\max}} W(E_\ell) e^{-E_\ell/kT}. \quad (10.23)$$

The factor of $W(E_\ell)$ in this expression counts the number of microstates in the given macrostate. You can compute the probability that the system is in macrostate energy level ℓ as

$$p_\ell = Q^{-1} W(E_\ell) e^{-E_\ell/kT}. \quad (10.24)$$

You are free to choose whether to focus on microstates or energy levels, or some other alternative, depending on what is most convenient for the problem that you are trying to solve. Examples 10.3 and 10.4 illustrate the density of states with the lattice polymer collapse model.

EXAMPLE 10.3 The collapse distribution for the four-bead polymer chain. How does the population of collapsed conformations change as the temperature is changed? We focus here on the two macrostates, open and compact, rather than on the individual microstates, so the partition function for this two-state system is

$$Q = 1 + 4e^{-\varepsilon_0/kT}.$$

where $\varepsilon_0 > 0$ is the energy increase if the bond is broken. At low temperatures, $T \rightarrow 0$, $Q \rightarrow 1$, so only the compact state is populated. But as the temperature increases, $T \rightarrow \infty$, $Q \rightarrow 5$, indicating that all conformations become equally populated. Cooling collapses the chain. The fractions of the molecules p_O that

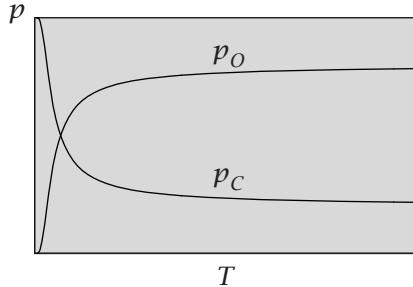


Figure 10.6 In the four-bead-chain polymer collapse model of Example 10.3, the collapsed population (probability p_C) diminishes with temperature T while the population of open conformations (probability p_O) grows.

are open and p_C that are compact are

$$p_C = \frac{1}{Q} \quad \text{and} \quad p_O = \frac{4e^{-\varepsilon_0/kT}}{Q} = \frac{4e^{-\varepsilon_0/kT}}{1+4e^{-\varepsilon_0/kT}}.$$

Figure 10.6 shows how the populations change with temperature. This figure describes a series of different experiments, each one at a fixed temperature. This kind of curve is sometimes called an *unfolding* or *denaturation* profile for a polymer or protein [1]. If you define $Q_C = 1$ as the partition function for the compact state alone and $Q_O = 4 \exp(-\varepsilon_0/kT)$ for the open state alone, then you can compute free energies using $F = -kT \ln Q$. At the cross-over point, where $p_C = p_O = 1/2$, the free energy of collapse is zero: $\Delta F = F_C - F_O = -kT \ln(Q_C/Q_O) = 0$.

Another example is the collapse of a six-bead chain.

EXAMPLE 10.4 Collapse of the six-bead polymer chain. For the six-bead chain shown in Figure 10.7, there are three equally spaced energy levels because a chain can have 0, 1, or 2 bead-bead contacts, corresponding to energies $\varepsilon = 2\varepsilon_0, 1\varepsilon_0$, and 0. The density of states is $W(0) = 4$, $W(\varepsilon_0) = 11$, and $W(2\varepsilon_0) = 21$. The partition function is

$$Q = 4 + 11e^{-\varepsilon_0/kT} + 21e^{-2\varepsilon_0/kT},$$

and the populations $p(\ell)$ of energy levels $\ell = 0, 1, 2$ are

$$p(0) = \frac{4}{Q}, \quad p(1) = \frac{11e^{-\varepsilon_0/kT}}{Q}, \quad p(2) = \frac{21e^{-2\varepsilon_0/kT}}{Q}.$$

Figure 10.8 shows the populations versus temperature for the collapse of the six-bead polymer. This six-bead model differs from the four-bead model in having an *intermediate* state that is populated near the middle of the transition from C at low temperatures to O at high temperatures.

Partition Functions for Independent and Distinguishable Particles

The Boltzmann distribution law applies to systems of any degree of complexity. The probabilities p_j can represent states $j = 1, 2, 3, \dots, t$ of atoms in ideal gases. Or the probabilities can represent states with energies that depend on the complex intermolecular system configurations. Calculations are simplest whenever

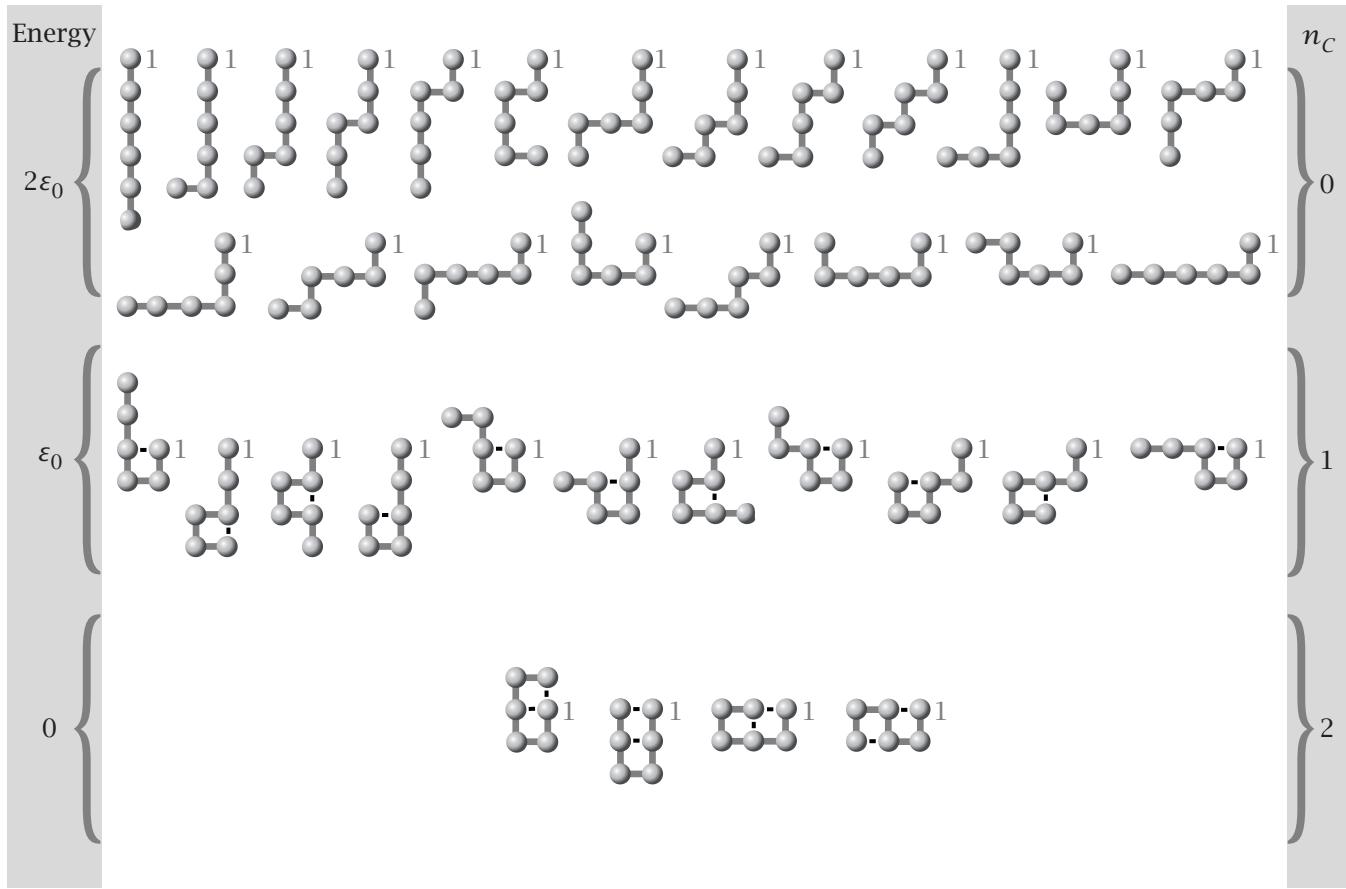
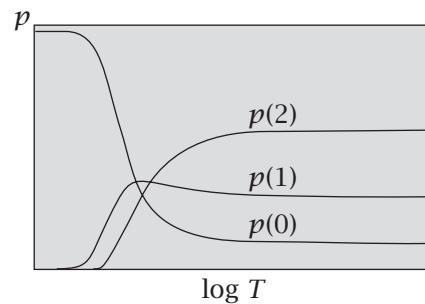


Figure 10.7 For the six-bead polymer chain, there are $W(0) = 4$ microstates having $n_C = 2$ contacts (contacts are shown as short dashes), $W(\varepsilon_0) = 11$ microstates having 1 contact, and $W(2\varepsilon_0) = 21$ microstates having 0 contacts.

Figure 10.8 Six-bead-chain polymer collapse of Example 10.4. At low temperatures T , the ground state is most stable ($p(0)$ is largest). At high temperatures, the open state is most stable ($p(2)$ is largest). Near the middle of the thermal transition, the intermediate state is most populated ($p(1)$ is largest).



a system is composed of independent subsystems. For example, each molecule in an ideal gas is independent of every other molecule. Or suppose you have N four-bead polymer molecules that do not interact with each other. In such cases, the system energy is the sum of the energies of each of the particles, and the system partition function is the product of particle partition functions.

In this section, we show this, first for the case of *distinguishable*, then for *indistinguishable*, particles. What is the difference? The atoms in a crystal are spatially distinguishable because each one has its own private location in the

crystal over the timescale of a typical experiment. Its location serves as a marker that distinguishes one particle from another. In contrast, according to quantum mechanics, the particles in a gas are indistinguishable from each other over the typical experimental timescale. Particles can interchange locations, so you cannot tell which particle is which.

First, consider distinguishable particles in a system with energy levels E_j . Suppose the system has two independent subsystems (e.g., two particles), distinguishable by labels A and B , with energy levels ε_i^A and ε_m^B , respectively, $i = 1, 2, \dots, a$ and $m = 1, 2, \dots, b$. The system energy is

$$E_j = \varepsilon_i^A + \varepsilon_m^B.$$

Because the subsystems are independent, you can write partition functions q_A for subsystem A and q_B for subsystem B according to Equation (10.10):

$$q_A = \sum_{i=1}^a e^{-\varepsilon_i^A/kT} \quad \text{and} \quad q_B = \sum_{m=1}^b e^{-\varepsilon_m^B/kT}. \quad (10.25)$$

The partition function Q for the entire system is the sum of Boltzmann factors $e^{-E_j/kT}$ over all $j = ab$ energy levels:

$$Q = \sum_{j=1}^t e^{-E_j/kT} = \sum_{i=1}^a \sum_{m=1}^b e^{-(\varepsilon_i^A + \varepsilon_m^B)/kT} = \sum_{i=1}^a \sum_{m=1}^b e^{-\varepsilon_i^A/kT} e^{-\varepsilon_m^B/kT}. \quad (10.26)$$

Because the subsystems are independent and distinguishable by their labels, the sum over the i levels of A has nothing to do with the sum over the m levels of B . The partition function Q in Equation (10.26) can be factored into subsystem partition functions q_A and q_B :

$$Q = \sum_{i=1}^a e^{-\varepsilon_i^A/kT} \sum_{m=1}^b e^{-\varepsilon_m^B/kT} = q_A q_B. \quad (10.27)$$

More generally, for a system having N independent and distinguishable particles, each with partition function q , the partition function Q for the whole system will be

$$Q = q^N. \quad (10.28)$$

Partition Functions for Independent and Indistinguishable Particles

Gas molecules are *indistinguishable*. They have *no labels* A or B that distinguish them from each other. For a system of two indistinguishable particles, the total energy is

$$E_j = \varepsilon_i + \varepsilon_m,$$

where $i = 1, 2, \dots, t_1$ and $m = 1, 2, \dots, t_2$. The system partition function is

$$Q = \sum_{j=1}^t e^{-E_j/kT} = \sum_{i=1}^{t_1} \sum_{m=1}^{t_2} e^{-(\varepsilon_i + \varepsilon_m)/kT}. \quad (10.29)$$

You cannot now factor the system partition function into particle partition functions as we did before. Here's the problem. If one particle occupied energy

level 27 and other particle occupied energy level 56, you could not distinguish that from the reverse. Because of this indistinguishability, you would have overcounted by a factor of $2!$.¹

For this system, you have $Q = q^2/2!$ to a good approximation. For N indistinguishable particles, this argument generalizes to give a system partition function Q ,

$$Q = \frac{q^N}{N!}. \quad (10.30)$$

We will use Equation (10.30) for gases.

If you know the partition function for a system or model, you can compute all the macroscopic thermodynamic properties.

Thermodynamic Properties Can Be Predicted from Partition Functions

Computing the Internal Energy from the Partition Function

Consider a system having fixed (T, V, N) . To get the internal energy for a system with energies E_j , substitute Equation (10.9) for p_j^* into Equation (10.2):

$$\begin{aligned} U &= \sum_{j=1}^t p_j^* E_j \\ &= Q^{-1} \sum_{j=1}^t E_j e^{-\beta E_j}, \end{aligned} \quad (10.31)$$

where $\beta = 1/kT$ is a useful quantity for simplifying the next few steps. Notice that the sum on the right-hand side of Equation (10.31) can be expressed as a derivative of the partition function in Equation (10.10):

$$\left(\frac{dQ}{d\beta} \right) = \frac{d}{d\beta} \sum_{j=1}^t e^{-\beta E_j} = - \sum_{j=1}^t E_j e^{-\beta E_j}. \quad (10.32)$$

Substituting Equation (10.32) into (10.31) simplifies it:

$$U = -\frac{1}{Q} \left(\frac{dQ}{d\beta} \right) = - \left(\frac{d \ln Q}{d\beta} \right). \quad (10.33)$$

¹That's close, but not exactly right. Suppose *both* particles were in energy level 27: then you would need no indistinguishability factor correction, because there's only one way to have that happen. To compute Equation (10.29) correctly in general for indistinguishable particles is challenging. But fortunately a most important case is simple. Suppose you have a huge number of energy levels, say 100,000, and only two particles. The chance that those particles would have coincident energies is exceedingly small, so the $2!$ correction would be a very good approximation. In reality, this is often valid: the number of accessible states is often much larger than the number of particles. You will see in Chapter 11 that translational partition functions are of the order of 10^{30} , while usually the number of particles is much smaller, 10^{20} or less. Therefore you are often justified in neglecting this small correction.

Since $\beta = 1/kT$, you have

$$\left(\frac{d\beta}{dT}\right) = -\frac{1}{kT^2}. \quad (10.34)$$

So you can multiply the left side of Equation (10.33) by $-1/kT^2$ and the right side by $d\beta/dT$ to get

$$\frac{U}{kT^2} = \left(\frac{d \ln Q}{dT}\right). \quad (10.35)$$

A useful alternative expression is $U/kT = d \ln Q / d \ln T = (T/Q) dQ/dT$.

Computing the Average Particle Energy

If particles are independent and distinguishable ($Q = q^N$), Equation (10.35) gives the average energy $\langle \varepsilon \rangle$ per particle:

$$\begin{aligned} \langle \varepsilon \rangle &= \frac{U}{N} = \frac{kT^2}{N} \left(\frac{\partial \ln q^N}{\partial T} \right)_{V,N} \\ &= kT^2 \left(\frac{\partial \ln q}{\partial T} \right) = - \left(\frac{\partial \ln q}{\partial \beta} \right). \end{aligned} \quad (10.36)$$

Computing the Entropy

The entropy of a system is defined by Equation (5.2):

$$\frac{S}{k} = - \sum_{j=1}^t p_j \ln p_j.$$

Substituting the Boltzmann distribution $p_j^* = Q^{-1} e^{-E_j/kT}$ from Equation (10.9) into Equation (5.2) gives

$$\frac{S}{k} = - \sum_{j=1}^t \left(\frac{1}{Q} e^{-E_j/kT} \right) \left[\ln \left(\frac{1}{Q} \right) - \frac{E_j}{kT} \right]. \quad (10.37)$$

Substituting Equation (10.10) for Q and Equation (10.31) for U into Equation (10.37) gives

$$S = k \ln Q + \frac{U}{T} = k \ln Q + kT \left(\frac{\partial \ln Q}{\partial T} \right). \quad (10.38)$$

For systems of N independent distinguishable particles, for which $Q = q^N$,

$$S = kN \ln q + \frac{U}{T}. \quad (10.39)$$

Because S increases linearly with N , the system entropy is the sum of the entropies of the independent particles.

Table 10.1 Thermodynamic quantities derived from the partition function for constant (T, V, N).

Internal energy U	$U = kT^2 \left(\frac{\partial \ln Q}{\partial T} \right)_{V,N}$	(10.40)
---------------------	---	---------

Entropy S	$S = k \ln Q + \frac{U}{T}$	(10.41)
-------------	-----------------------------	---------

Helmholtz free energy F	$F = U - TS = -kT \ln Q$	(10.42)
---------------------------	--------------------------	---------

Chemical potential μ	$\mu = \left(\frac{\partial F}{\partial N} \right)_{T,V} = -kT \left(\frac{\partial \ln Q}{\partial N} \right)_{T,V}$	(10.43)
--------------------------	---	---------

Pressure p	$p = - \left(\frac{\partial F}{\partial V} \right)_{T,N} = kT \left(\frac{\partial \ln Q}{\partial V} \right)_{T,N}$	(10.44)
--------------	--	---------

Computing the Free Energy and Chemical Potential

From U and S in Equations (10.35) and (10.38), thermodynamic relationships can produce the rest—the Helmholtz free energy, chemical potential, and pressure, for example. Table 10.1 lists the main relationships.

Now we illustrate these relationships by computing the thermodynamic properties of one of the simplest statistical mechanical models, the two-state model.

EXAMPLE 10.5 The Schottky two-state model. Consider a system that has N distinguishable particles with two energy levels for each particle: a ground state with energy zero and an excited state with energy $\varepsilon = \varepsilon_0 > 0$. This model is useful for describing many different problems: our polymer or dimer lattice models in Chapter 8, the behaviors of atoms or molecules that are excited by electromagnetic radiation, conduction of electrons in semiconductors, or the behavior of spins in magnetic fields (see Example 10.6).

Here we'll keep the model general and won't specify ε_0 in terms of any particular microscopic structure or property. We want to find the average particle energy $\langle \varepsilon \rangle$, the heat capacity C_V , the entropy, and the free energy per particle from the partition function. The partition function for a two-level system is the sum of two Boltzmann factors, one for each level:

$$q = 1 + e^{-\beta \varepsilon_0}. \quad (10.45)$$

The partition function approaches 1 at low temperatures and 2 at high temperatures. The relative populations of the two states are given by the Boltzmann distribution, Equation (10.9):

$$p_1^* = \frac{1}{q} \quad \text{and} \quad p_2^* = \frac{e^{-\beta \varepsilon_0}}{q}. \quad (10.46)$$

The average energy is given by Equation (10.36):

$$\langle \varepsilon \rangle = \sum p_j^* \varepsilon_j = 0p_1^*(0) + \varepsilon_0 p_2^* = \frac{\varepsilon_0 e^{-\varepsilon_0/kT}}{1 + e^{-\varepsilon_0/kT}} \quad (10.47)$$

(or you can get this by taking the derivative $\langle \varepsilon \rangle = -q^{-1}(\partial q/\partial \beta)$). Figure 10.9(a) shows the energy of the two-state system as a function of temperature. At low temperatures, most molecules are in the ground state, so the system has low energy. As the temperature increases, the energy of the system increases and approaches the value $\varepsilon_0/2$ per particle because energy levels 0 and ε_0 become equally populated.

To compute the heat capacity, use the definition $C_V = (\partial U/\partial T)$ from thermodynamics. Using Equation (10.36) to convert the total energy to the average energy per particle, $U = N\langle \varepsilon \rangle$, you have

$$C_V = N \left(\frac{\partial \langle \varepsilon \rangle}{\partial T} \right)_{V,N} = N \left(\frac{\partial \langle \varepsilon \rangle}{\partial \beta} \right) \left(\frac{d\beta}{dT} \right) = -\frac{N}{kT^2} \left(\frac{\partial \langle \varepsilon \rangle}{\partial \beta} \right), \quad (10.48)$$

where the right-hand expressions convert from T to β to make the next step of the differentiation simpler. Take a derivative of the form $d(u/v) = (v u' - u v')/v^2$, where $u = \varepsilon_0 e^{-\beta \varepsilon_0}$ and $v = 1 + e^{-\beta \varepsilon_0}$, to get

$$\begin{aligned} \left(\frac{\partial \langle \varepsilon \rangle}{\partial \beta} \right) &= \frac{(1 + e^{-\beta \varepsilon_0})(-\varepsilon_0^2 e^{-\beta \varepsilon_0}) - \varepsilon_0 e^{-\beta \varepsilon_0}(-\varepsilon_0 e^{-\beta \varepsilon_0})}{(1 + e^{-\beta \varepsilon_0})^2} \\ &= \frac{-\varepsilon_0^2 e^{-\beta \varepsilon_0}}{(1 + e^{-\beta \varepsilon_0})^2}. \end{aligned} \quad (10.49)$$

Substitute Equation (10.49) into the right side of Equation (10.48) to find the heat capacity C_V in terms of the energy level spacing ε_0 :

$$C_V = \frac{N\varepsilon_0^2}{kT^2} \frac{e^{-\beta \varepsilon_0}}{(1 + e^{-\beta \varepsilon_0})^2}. \quad (10.50)$$

The heat capacity is plotted in Figure 10.9(b), and is discussed in more detail in Chapter 12. Heat capacity peaks are characteristic of bond-breaking and melting processes. At low temperatures, the thermal energy kT from the bath is too small to excite the system to its higher energy level. At intermediate temperatures, the system can absorb heat from the bath, and particles are excited into the higher-energy state. At the highest temperatures, the system takes up no further energy from the bath because it has already taken up the maximum energy it can contain.

To get the entropy, substitute Equation (10.45) for q and $Q = q^N$ into Equations (10.41) and (10.40):

$$\frac{S}{N} = \frac{\varepsilon_0 e^{-\beta \varepsilon_0}}{T(1 + e^{-\beta \varepsilon_0})} + k \ln(1 + e^{-\beta \varepsilon_0}). \quad (10.51)$$

To get the free energy, substitute $Q = q^N$ into $F = -kT \ln Q$ (Equation (10.42)):

$$\frac{F}{NkT} = -\ln q = -\ln(1 + e^{-\beta \varepsilon_0}). \quad (10.52)$$

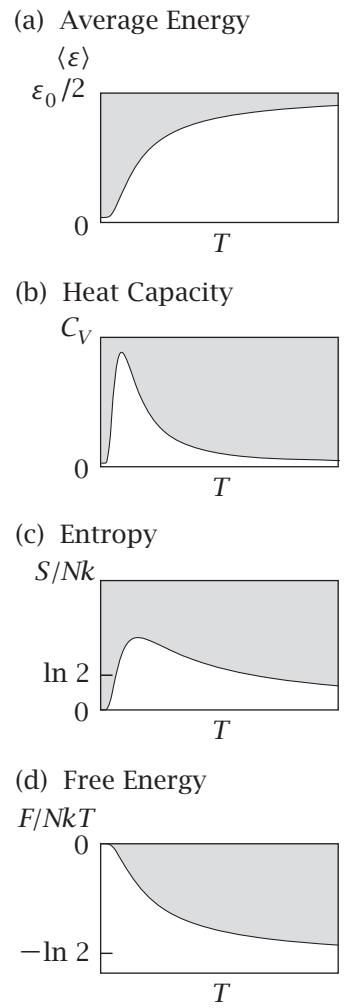


Figure 10.9 (a) The average energy per particle $\langle \varepsilon \rangle$ for the Schottky model (Equation (10.47)) saturates at $\varepsilon_0/2$ as temperature T approaches infinity. (b) The two-state model has a peak in the heat capacity C_V as a function of temperature (Equation (10.50)). (c) The entropy per particle divided by k has a peak and then approaches $\ln 2$ for large temperatures T (Equation (10.51)). (d) The free energy per particle divided by kT approaches $-\ln 2$ for large temperatures (Equation (10.52)).

As $\varepsilon_0 \rightarrow \infty$, the excited state becomes inaccessible, so $S \rightarrow 0$ and $F \rightarrow 0$. On the other hand, as $\varepsilon_0 \rightarrow 0$, both states become accessible, so $S \rightarrow Nk \ln 2$ and $F \rightarrow -NkT \ln 2$.

Now let's consider a specific example of a two-level system: a magnetic material in a magnetic field.

EXAMPLE 10.6 Curie's law of paramagnetism, a two-level system. Some materials are magnetic. Inside their atoms are unpaired spins (miniature magnets) that can be driven to align by an external magnetic force. The magnetic dipole moment describes the degree to which a spin aligns with a given magnetic field $B \geq 0$. Some materials, like refrigerator magnets, are ferromagnetic. They are magnetic even when there is no applied magnetic field. One spin points in a favored direction because its neighboring spins also point in that direction, helping to keep it oriented. Others, called paramagnets, are only aligned when a field is applied. In a paramagnet, the magnetic dipole-dipole interaction between neighboring atoms is small compared with the interaction with an applied field. Paramagnetic materials include hydrogen gas atoms and the heme iron atoms in hemoglobin. If you heat a paramagnet, it loses magnetization. The Schottky model illustrates this. Consider a simple paramagnetic material, in which every one of the N independent atoms has a magnetic dipole moment of magnitude $\mu_0 > 0$. (Don't confuse μ here with the chemical potential: the same symbol is used for both quantities, but they are unrelated.)

Here's a two-state model. Each spin can be either aligned or antialigned with the magnetic field (see Figure 10.10). The aligned spins have a lower energy than the antialigned spins. A spin that is parallel to the field is in the ground state, with energy $\varepsilon_1 = -\mu_0 B$. A spin that is antiparallel to the field is in the excited state, with higher energy: $\varepsilon_2 = +\mu_0 B$. Using the convention that the ground state defines the zero of energy, the energy difference between these two states is $+2\mu_0 B$, which defines the excited state energy, so the partition function is

$$q = 1 + e^{-2\mu_0 B/kT}. \quad (10.53)$$

We want to calculate the average magnetic moment (the material's magnetization) as a function of temperature. At equilibrium, the probability that an atom's magnetic moment is parallel to B is p_1^* and the probability that it is antiparallel is p_2^* :

$$p_1^* = \frac{1}{q} \quad \text{and} \quad p_2^* = \left(\frac{1}{q} \right) e^{-2\mu_0 B/kT}. \quad (10.54)$$

Because the magnetic moment of the aligned state is $+\mu_0$ and that of the antialigned state is $-\mu_0$, the average magnetic moment is

$$\begin{aligned} \langle \mu \rangle &= \sum_{j=1}^2 \mu_j p_j^* \\ &= \mu_0 p_1^* + (-\mu_0) p_2^* = \frac{\mu_0}{q} (1 - e^{-2\mu_0 B/kT}) = \mu_0 \frac{1 - e^{-2\mu_0 B/kT}}{1 + e^{-2\mu_0 B/kT}}. \end{aligned} \quad (10.55)$$

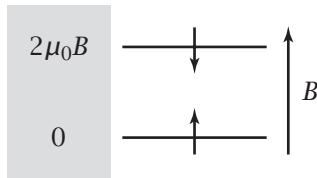


Figure 10.10 The two-state model of spins. The B arrow shows the direction of an externally applied magnetic field. The ground state occurs when an atomic spin is aligned with the direction of B . The excited state occurs when the atomic spin is anti-aligned with B .

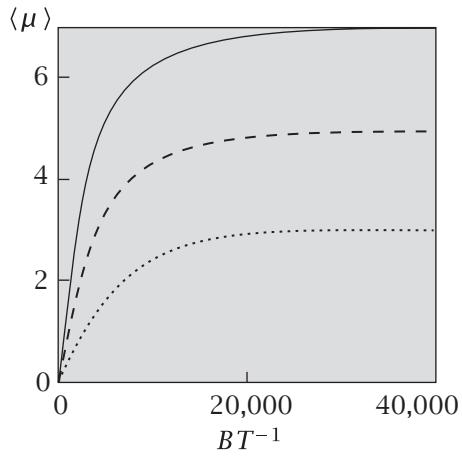


Figure 10.11 The average magnetic moment $\langle \mu \rangle$ saturates with increasing magnetic field B/kT . At low fields, $\langle \mu \rangle$ is linear in B/kT . Data for gadolinium sulphate octahydrate (—), ferric ammonium alum (---), chromium potassium alum (····). Source: JD Fast, *Entropy; The Significance of the Concept of Entropy and Its Applications in Science and Technology*, McGraw-Hill, New York, 1962. Data are from CJ Gorter, WJ deHaas and J van den Handel, *Proc Kon Ned Akad Wetensch* **36**, 158 (1933) and WE Henry, *Phys Rev* **88**, 559 (1952).

The last equality follows from the definition of q in Equation (10.53). A concise way to express this relationship is through use of the hyperbolic tangent function,

$$\tanh x = \frac{1-e^{-2x}}{1+e^{-2x}}. \quad (10.56)$$

Using the hyperbolic tangent, the average magnetic moment is given by

$$\langle \mu \rangle = \mu_0 \tanh\left(\frac{\mu_0 B}{kT}\right). \quad (10.57)$$

In weak magnetic fields or at high temperatures, $\mu_0 B / kT \ll 1$. The Taylor series expansion for exponentials (see Appendix J, Equation (J.1)) gives $1 - e^{-2\mu_0 B / kT} \approx 2\mu_0 B / kT$ and $1 + e^{-2\mu_0 B / kT} \approx 2 - 2\mu_0 B / kT \approx 2$. At high temperatures or in weak fields, the total magnetic moment is inversely proportional to T , and Equation (10.57) becomes **Curie's law**:

$$\langle \mu \rangle = \frac{\mu_0^2 B}{kT}. \quad (10.58)$$

In contrast, at high fields ($B/kT \rightarrow \infty$), Equation (10.55) gives $\langle \mu \rangle = \mu_0$.

Figure 10.11 shows experimental evidence that the magnetization of a material increases with increasing magnetic field. High fields cause all the spins to align with the field. Cooling leads to magnetization (alignment of the spins) and heating leads to demagnetization (random orientations of the spins).

EXAMPLE 10.7 Modeling a protein loop. Let's use Boltzmann's law to model a peptide loop on a protein molecule. You probe the loop with fluorescence spectroscopy, which measures a distance ℓ from a point on the loop to a point on the protein (see Figure 10.12).

You model the loop as having three different conformations: (1) The loop sticks to the side of the protein, $\ell_1 = 1 \text{ \AA}$; define this as the ground state, with energy $\varepsilon_1 = 0$. (2) In a second conformation, the loop is more distant from

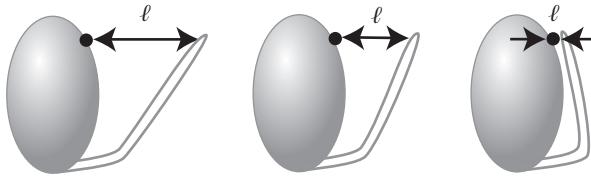


Figure 10.12 Three states of a protein loop in Example 10.7: 4 Å away from the protein, 2 Å away from the protein, or 1 Å away from the protein.

the protein, $\ell_2 = 2 \text{ \AA}$, and the energy is higher by an amount $\varepsilon_2 = 1 \text{ kcal mol}^{-1}$. (3) In the third conformation, the loop flops far away from the protein, $\ell_3 = 4 \text{ \AA}$, and the energy relative to the ground state is $\varepsilon_3 = 1.7 \text{ kcal mol}^{-1}$. At temperature $T = 300 \text{ K}$, the partition function for this model is

$$q = 1 + e^{-\varepsilon_1/RT} + e^{-\varepsilon_2/RT} + e^{-\varepsilon_3/RT} = 1 + 0.19 + 0.06 = 1.25, \quad (10.59)$$

so the populations of the three states are

$$p_1 = \frac{1}{q} = 0.8, \quad p_2 = \frac{e^{-\varepsilon_2/RT}}{q} = 0.15, \quad p_3 = \frac{e^{-\varepsilon_3/RT}}{q} = 0.05. \quad (10.60)$$

The ideal test of your model would be measurements of the three populations in the distribution. However, usually experiments measure only averages, not full distributions. So, to compare with experiments, compute averages such as

$$\langle \ell \rangle = 0.8\ell_1 + 0.15\ell_2 + 0.06\ell_3 = 1.34 \text{ \AA} \quad (10.61)$$

and

$$\langle \varepsilon \rangle = 0.8\varepsilon_1 + 0.15\varepsilon_2 + 0.06\varepsilon_3 = 0.25 \text{ kcal mol}^{-1} \quad (10.62)$$

You can also compute the variance of the loop distance, $\langle \ell^2 \rangle - \langle \ell \rangle^2$, the heat capacity C_V , the entropy and free energy, and other such properties that might be measurable. Increasing temperature will lead to increasing loop opening $\langle \ell \rangle$.

What Is an Ensemble?

A term commonly used in statistical mechanics is *ensemble*. We will use it in the following chapters. The term is usually used in one of two ways. First, it can refer to which set of variables you are controlling: ‘the (U, V, N) ensemble’ or ‘the (T, p, N) ensemble,’ for example. Some ensembles are so common that they have names. In this chapter, we have used the (T, V, N) ensemble, also called the *canonical ensemble*. (U, V, N) is called the *microcanonical ensemble*. (T, p, N) , is called the *isobaric-isothermal ensemble*. Another one that will be important later in Chapters 28 and 29 on ligand binding is (T, V, μ) , called the *grand canonical ensemble*. These are the four main named ensembles.

The term ensemble also has another meaning. An ensemble is the collection of all the possible microstates, or snapshots, of a system. For example, our four-bead polymer chain has an ensemble of five microstates and our six-bead chain has an ensemble of 36 microstates. Below, we briefly note some properties of the microcanonical ensemble.

The Microcanonical Ensemble

The microcanonical ensemble is qualitatively different from the canonical and grand canonical ensembles. In the canonical ensemble, the temperature is fixed, which is equivalent to fixing the average energy $U = \langle E \rangle$. The energy can fluctuate. But in the microcanonical ensemble, every microstate has exactly the same fixed energy, so $U = E$, and there are no fluctuations. For the microcanonical ensemble, it is more useful to focus on the $i = 1, 2, 3, \dots, W(E, V, N)$ different microstates of the system than on t different energy levels, since there is only one energy level. In the microcanonical ensemble, each microstate is equivalent. So you can express the probability that the system is in microstate $i = 1, 2, 3, \dots, W$ as

$$p_i^* = \frac{1}{W}. \quad (10.63)$$

Using the definition of the entropy, $S/k = -\sum_i p_i \ln p_i$, now summed over microstates, you get

$$\begin{aligned} \frac{S}{k} &= -\sum_{i=1}^W p_i \ln p_i = -\sum_{i=1}^W \left(\frac{1}{W}\right) \ln\left(\frac{1}{W}\right) \\ &= \ln W(E, V, N), \end{aligned} \quad (10.64)$$

as you expect from the Boltzmann expression.

Summary

Boltzmann's law gives the equilibrium distribution of atoms and molecules over their energy levels. Starting with a model for the ladder of energies accessible to the system, you can compute the partition function. From the partition function, you can compute the thermodynamic and averaged physical properties. In the next chapter, we illustrate how quantum mechanical models give the energy levels that can be used to predict the properties of gases and simple solids from their atomic structures.

Problems

- 1. Statistical thermodynamics of a cooperative system.** Perhaps the simplest statistical mechanical system having 'cooperativity' is the three-level system in Table 10.2.

Table 10.2

Energies	$2\varepsilon_0$	ε_0	0
Degeneracies	γ	1	1

- (a) Write an expression for the partition function q as a function of energy ε , degeneracy γ (see page 178), and temperature T .
- (b) Write an expression for the average energy $\langle \varepsilon \rangle$ versus T .
- (c) For $\varepsilon_0/kT = 1$ and $\gamma = 1$, compute the equilibrium populations, or probabilities, p_1^*, p_2^*, p_3^* of the three levels of energy 0, ε_0 , $2\varepsilon_0$, respectively.
- (d) Now if $\varepsilon_0 = 2 \text{ kcal mol}^{-1}$ and $\gamma = 1000$, find the temperature T_0 at which $p_1 = p_3$.
- (e) Under condition (d), compute p_1^* , p_2^* , and p_3^* at temperature T_0 .

2. The speed of sound. The speed of sound in air is approximately the average velocity $\langle v_x^2 \rangle^{1/2}$ of the gas molecules. Compute this speed for $T = 0^\circ\text{C}$, assuming that air is mostly nitrogen gas.

3. The properties of a two-state system. Given a two-state system in which the low energy level is 600 cal mol^{-1} , the high energy level is $1800 \text{ cal mol}^{-1}$, and the temperature of the system is 300 K ,

- (a) What is the partition function q ?
- (b) What is the average energy $\langle \varepsilon \rangle$?

4. Binding to a surface. Consider a particle that has two states: bonded to a surface, or non-bonded (released). The non-bonded state is higher in energy by an amount ε_0 .

- (a) Explain how the ability of the particle to bond to the surface contributes to the heat capacity, and why the heat capacity depends on temperature.
- (b) Compute the heat capacity C_V in units of Nk if $T = 300 \text{ K}$ and $\varepsilon_0 = 1.2 \text{ kcal mol}^{-1}$ (which is about the strength of a weak hydrogen bond in water).

5. Entropy depends on distinguishability. Given a system of molecules at $T = 300 \text{ K}$, $q = 1 \times 10^{30}$, and $\Delta U = 3740 \text{ J mol}^{-1}$,

- (a) What is the molar entropy if the molecules are distinguishable?
- (b) What is the molar entropy if the molecules are indistinguishable?

6. The Boltzmann distribution of uniformly spaced energy levels. A system has energy levels uniformly spaced at $3.2 \times 10^{-20} \text{ J}$ apart. The populations of the energy levels are given by the Boltzmann distribution.

What fraction of particles is in the ground state at $T = 300 \text{ K}$?

7. The populations of spins in a magnetic field. The nucleus of a hydrogen atom, a proton, has a magnetic moment. In a magnetic field, the proton has two states of different energy: spin up and spin down. This is the basis of proton NMR. The relative populations can be assumed to be given by the Boltzmann distribution, where the difference in energy between the two states is $\Delta\varepsilon = g\mu B$, $g = 2.79$ for protons, and $\mu = 5.05 \times 10^{-24} \text{ J T}^{-1}$. For a 300 MHz NMR instrument, $B = 7 \text{ T}$.

- (a) Compute the relative population difference, $|N_+ - N_-|/(N_+ + N_-)$, at room temperature for a 300 MHz machine.
- (b) Describe how the population difference changes with temperature.
- (c) What is the partition function?

8. Energy and entropy for indistinguishable particles. Equations (10.36) for $\langle \varepsilon \rangle$ and (10.39) for S apply to distinguishable particles. Compute the corresponding quantities for systems of indistinguishable particles.

9. Computing the Boltzmann distribution. You have a thermodynamic system with three states. You observe the probabilities $p_1 = 0.9$, $p_2 = 0.09$, and $p_3 = 0.01$ at $T = 300 \text{ K}$. What are the energies ε_2 and ε_3 of states 2 and 3 relative to the ground state?

10. The pressure reflects how energy levels change with volume. If energy levels $\varepsilon_i(V)$ depend on the volume of a system, show that the pressure is the average

$$p = -N \left\langle \frac{\partial \varepsilon}{\partial V} \right\rangle.$$

11. The end-to-end distance in polymer collapse. Use the two-dimensional four-bead polymer of Example 10.3. The distance between the chain ends is 1 lattice unit in the compact conformation, 3 lattice units in the extended conformation, and $\sqrt{5}$ lattice units in each of the other three chain conformations. Plot the average end-to-end distance as a function of temperature if the energy is

- (a) $\varepsilon = 1 \text{ kcal mol}^{-1}$;
- (b) $\varepsilon = 3 \text{ kcal mol}^{-1}$.

12. The lattice model of dimerization. Use the lattice model for monomers bonding to form dimers, and assume large volumes $V \gg 1$.

- (a) Derive the partition function.
- (b) Compute $p_1(T)$ and $p_2(T)$, the probabilities of monomers and dimers as functions of temperature, and sketch the dependence on temperature.
- (c) Compute the bond breakage temperature T_0 at which $p_1 = p_2$.

13. Deriving the Boltzmann law two different ways. Use Equation (10.6) to show that the distribution of probabilities p_j^* that minimizes the free energy F at constant

T is the same one you get if instead you maximize the entropy S at constant $U = \langle E \rangle$.

14. Protein conformations. Assume a protein has six different discrete conformations, with energies given in Figure 10.13.

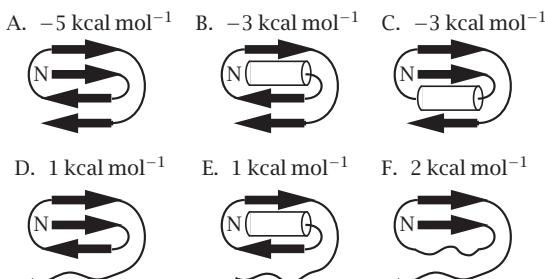


Figure 10.13

- Write an expression for the probability $p(i)$ of finding the protein in conformation i .
- Write an expression for the probability $p(E)$ of finding the protein having energy E .
- Use the expressions you wrote in (a) and (b) to calculate the following probabilities:
 - $p(\text{State B})$.
 - $p(\text{State A})$.
 - $p(\text{State D})$.
 - $p(1 \text{ kcal mol}^{-1})$.
 - $p(-5 \text{ kcal mol}^{-1})$.
- What is the average energy of the ensemble of conformations?

15. Modeling ligand binding. A ligand is bound to a protein with a spring-like square-law energy $\varepsilon(x)$, where x is the distance between the ligand and protein as shown in Figure 10.14.

$$\varepsilon(x) = \frac{1}{2}cx^2$$

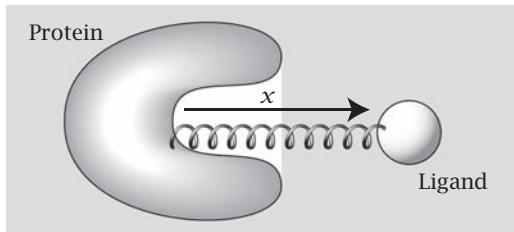


Figure 10.14

- For constant (T, V, N) , write an expression for the probability distribution $p(x)$ of the ligand separation from the protein.
- Sketch a plot of $p(x)$ versus x .
- Write an expression for the average location of the ligand, $\langle x \rangle$.
- Write an expression for the second moment of the location of the ligand, $\langle x^2 \rangle$.
- Calculate the average energy $\langle \varepsilon \rangle$ of the system.

16. Distribution of torsional angles. In a computer simulation that samples a molecular torsion angle θ , you observe a Gaussian distribution $p(\theta)$, shown in Figure 10.15:

$$p(\theta) = p_0 e^{-k_s(\theta-\theta_0)^2}.$$

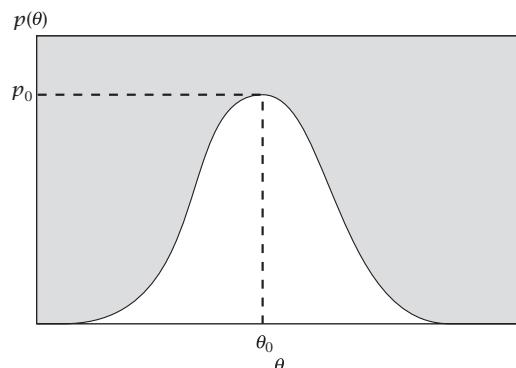


Figure 10.15

What is the underlying energy function $E(\theta)$ that gives rise to it?

17. Three-bead polymer chain model. Consider a three-bead polymer that can undergo conformational change from a nonlinear to a linear form, as shown in Figure 10.16. Both conformations have the same energy. Now suppose the X and Y atoms of the polymer can bind a ligand L (Figure 10.17). Breaking one bond increases the energy by ε and breaking two bonds increases the energy by 2ε . Assume that the ligand-bound conformation has the lowest energy.

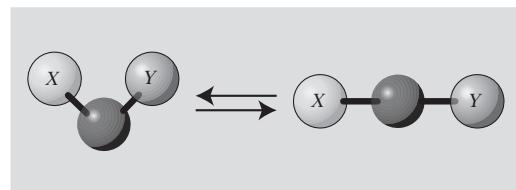


Figure 10.16

- Draw a picture showing the possible binding states.
- Calculate the equilibrium probabilities of these conformations.

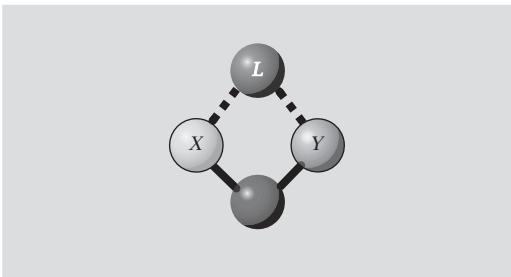


Figure 10.17

- (c) Plot the population distribution of the conformations at temperatures $kT = 0$ and $kT = 1.4\epsilon$. What is the difference in entropy between those two temperatures?
 - (d) Plot the distribution at high temperatures ($T \rightarrow \infty$) and explain its shape.
 - (e) What is the average energy of the system at each temperature in part (c)?
- 18. Protein hinge motions.** A protein has two domains, connected by a flexible hinge as shown in Figure 10.18.

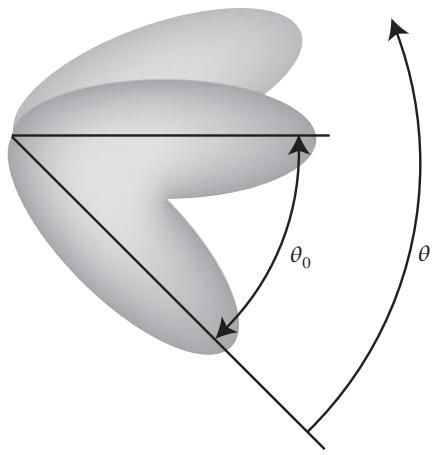


Figure 10.18

The hinge fluctuates around an angle θ_0 . The distribution of the angle θ is Gaussian, around θ_0 , as shown in Figure 10.19:

$$p(\theta) = \left(\frac{1}{\sigma\sqrt{2\pi}} \right) \exp \left[-\frac{(\theta-\theta_0)^2}{2\sigma^2} \right],$$

where σ is the standard deviation.

- (a) Assume a Boltzmann distribution of probabilities: $p(\theta) = b \exp[-c(\theta-\theta_0)^2/kT]$, with energy $\epsilon = c(\theta-\theta_0)^2$. Derive the value of b .
- (b) If you measure a standard deviation $\sigma = 30^\circ$, what is the value of c ? Assume $T = 300\text{ K}$.

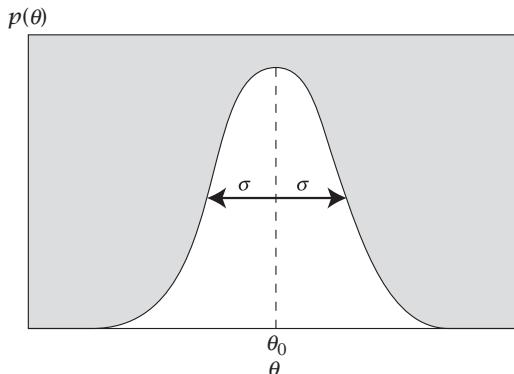


Figure 10.19

- (c) Derive an expression relating $\langle \theta^2 \rangle$ to c .
- (d) Derive an expression for the entropy in terms of kT and c : $S/k = - \int_{-\infty}^{\infty} p(\theta) \ln[p(\theta)] d\theta$.
- (e) If a mutant of the protein has fluctuations that are smaller than the wild type, $\langle \theta^2 \rangle_M = (1/2)\langle \theta^2 \rangle_{WT}$, what is the change in entropy, $\Delta S = S_M - S_{WT}$, where S_M is the entropy of the mutant and S_{WT} is the entropy of the original wild-type protein?

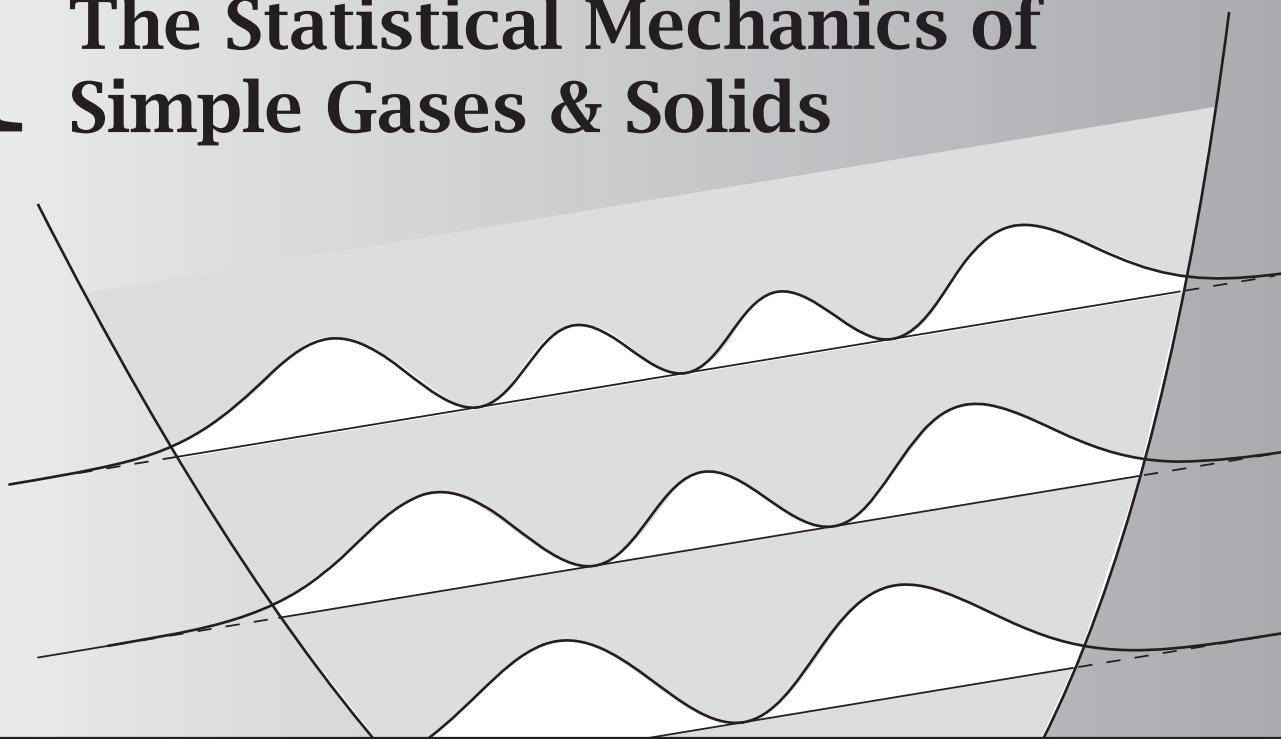
References

- [1] KA Dill, S Bromberg, K Yue, et al., *Prot Sci* 4, 561-602 (1995).

Suggested Reading

- D Chandler, *Introduction to Modern Statistical Mechanics*, Oxford University Press, Oxford, 1987. Concise and insightful.
- HL Friedman, *A Course in Statistical Mechanics*, Prentice-Hall, Englewood Cliffs, NJ, 1985. Good description of the relationship between ensembles in thermodynamics and statistical mechanics.
- TL Hill, *An Introduction to Statistical Thermodynamics*, Addison-Wesley, Reading, MA, 1960 (reprinted by Dover Publications, New York, 1986). A classic with an excellent discussion of ensembles.
- K Huang, *Statistical Mechanics*, 2nd edition, Wiley, New York, 1987. Advanced text.
- R Kubo, with H Ichimara, T Usui, and N Hashitsome, *Statistical Mechanics*, North-Holland, New York, 1965. Advanced text with many excellent worked problems.
- DA McQuarrie, *Statistical Mechanics*, 2nd edition, University Science Books, Sausalito, CA, 2000. A standard and complete text with a good discussion of ensembles.

11 The Statistical Mechanics of Simple Gases & Solids



Statistical Mechanics Predicts Macroscopic Properties from Atomic Structures

To predict the properties of materials from their atomic structures, you need to know *energy ladders*. Energy ladders can be derived from spectroscopy or quantum mechanics. Here we describe some of the quantum mechanics that can predict the properties of ideal gases and simple solids. This will be the foundation for chemical reaction equilibria and kinetics in Chapters 13 and 19. We describe the *particle-in-a-box* model for translational freedom, the harmonic oscillator model for vibrations, and the rigid rotor model for rotations.

Evidence for the quantization of energies comes from atomic spectroscopy. Spectroscopy measures the frequency ν of electromagnetic radiation that is absorbed or emitted by an atom, a molecule, or a material. Absorption of radiation by matter leads to an increase in its energy by an amount $\Delta\varepsilon = h\nu$, where $h = 6.626176 \times 10^{-34}$ J s is Planck's constant. This change $\Delta\varepsilon$ is the difference from one energy level to another on an energy ladder.

Atoms and simple molecules absorb electromagnetic radiation only at certain discrete frequencies, rather than as continuous functions of frequency. For example, Figure 11.1 shows the infrared absorption spectrum of hydrogen bromide. Discrete absorption frequencies imply discrete energy spacings. This is the basis for quantum mechanics.

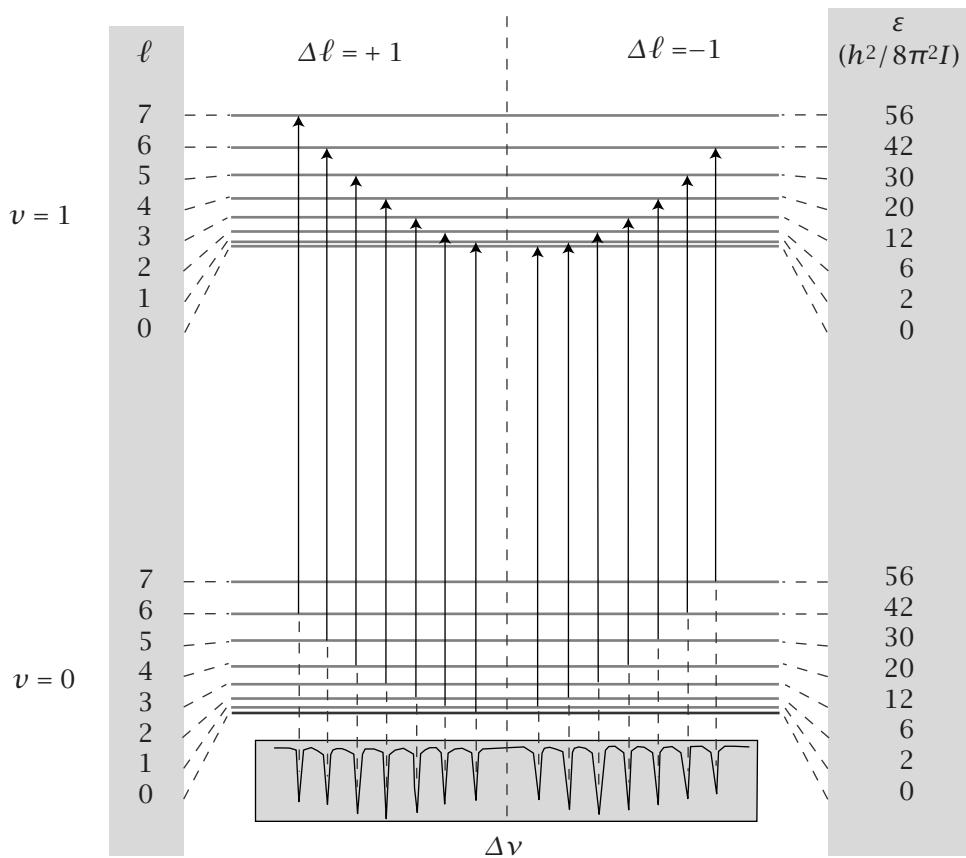


Figure 11.1 Electromagnetic radiation can excite transitions in molecules from one rotational energy level ℓ of a particular vibrational state v to a different level. The frequency of excitation radiation corresponds to the energy change of the transition, through the relation $\Delta\varepsilon = \hbar\nu$. Shown here is the absorption band of gas-phase HBr in the infrared region. Source: GM Barrow, *The Structure of Molecules; an Introduction to Molecular Spectroscopy*, WA Benjamin, New York, 1963.

Energies form a hierarchy. Some energy changes are small, implying closely spaced energy levels, and others are large (see Figure 11.2). Energies fall into classes, including *translations, rotations, vibrations, and electronic excitations*. Each step from one rung to the next on the ladder of electronic excitation energies typically contains many rungs on the vibrational ladder. Each step on the vibrational ladder has many rotational rungs, and each step on the rotational ladder has many translational levels.

The basis for predicting quantum mechanical energy levels is the Schrödinger equation, which can be expressed in compact form:

$$\mathcal{H}\psi = E\psi, \quad (11.1)$$

where the *wavefunction* $\psi(x, y, z)$ is a function of spatial position (x, y, z) (ψ can also be a function of time, but we don't consider that here). Solving Equation (11.1) gives ψ . The square of this function, ψ^2 , is the spatial probability distribution of the particles for the problem of interest.

\mathcal{H} is called the *Hamiltonian operator*. It describes the forces relevant for the problem of interest. In classical mechanics, the sum of kinetic plus potential energies is an invariant conserved quantity, the total energy (see Equation (3.4)). In quantum mechanics, the role of this invariant is played by the Hamiltonian operator. For example, for the one-dimensional translational motion of a particle having mass m and momentum p , the Hamiltonian operator is

$$\mathcal{H} = \frac{p^2}{2m} + \mathcal{V}(x), \quad (11.2)$$

where $p^2/(2m)$ represents the kinetic energy and $\mathcal{V}(x)$ is the potential energy, as a function of the spatial coordinate x .

While classical mechanics regards p^2 and \mathcal{V} as functions of time and spatial position, quantum mechanics regards p^2 and \mathcal{V} as *mathematical operators* that create the right differential equation for the problem at hand. For example, in quantum mechanics, $p^2 = (-\hbar^2/4\pi^2)(d^2/dx^2)$ in one dimension. The expression $\mathcal{H}\psi(x)$ in Equation (11.1) is a shorthand notation for taking the second derivative of ψ , multiplying by a collection of terms and adding $\mathcal{V}\psi$,

$$\mathcal{H}\psi = -\left(\frac{\hbar^2}{8\pi^2 m}\right) \frac{d^2\psi(x)}{dx^2} + \mathcal{V}(x)\psi(x), \quad (11.3)$$

so Equation (11.1) becomes

$$-\left(\frac{\hbar^2}{8\pi^2 m}\right) \frac{d^2\psi(x)}{dx^2} + \mathcal{V}(x)\psi(x) = E\psi(x). \quad (11.4)$$

According to Schrödinger's equation, the operator \mathcal{H} acting on ψ (the left side) equals a constant E multiplied by ψ (the right side). Only certain functions $\psi(x)$ can satisfy Equation (11.4), if E is a constant. There are multiple quantities E_j , called the *eigenvalues* of the equation. They are the discrete energy levels that we seek. The index j for the eigenvalues is called the *quantum number*. Now, we solve the Schrödinger equation for a simple problem to show the quantum mechanical perspective on translational motion.

The Particle in a Box Is the Quantum Mechanical Model for Translational Motion

The particle in a box is a model for the freedom of a particle to move within a confined space, such as an electron contained in an atom or molecule, or a molecule in a test tube. Let's first solve a one-dimensional problem.

A particle is free to move along the x axis over the range $0 < x < L$. The idea of the 'box' is that at the walls ($x = 0$ and $x = L$) the potential is infinite ($\mathcal{V}(0) = \mathcal{V}(L) = \infty$), so the particle cannot escape (see Figure 11.3). But everywhere inside the box the molecule has free motion ($\mathcal{V}(x) = 0$ for $0 < x < L$). So we set $\mathcal{V}(x) = 0$ in Equation (11.4).

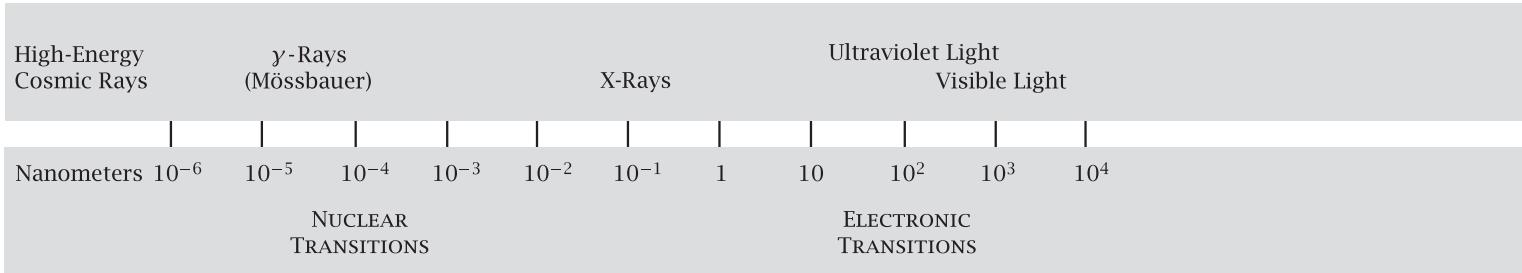


Figure 11.2 Electromagnetic energy spectrum. Rotational and vibrational motions occur in the infrared range. Ionization and covalent bond breakage occur at higher energies in the UV and X-ray range. Nuclear spins are affected at much lower energies. Source: O Howarth, *Theory of Spectroscopy; an Elementary Introduction*, Wiley, New York, 1973.

Equation (11.4) then becomes a linear second-order differential equation

$$\frac{d^2\psi}{dx^2} + K^2\psi = 0, \quad (11.5)$$

where

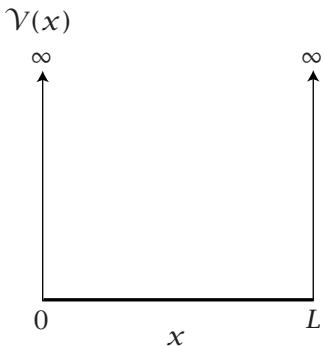


Figure 11.3 The potential energy \mathcal{V} for the particle in a one-dimensional box is infinite at the walls, $x = 0$, and $x = L$. $\mathcal{V}(x) = 0$ inside.

$$K^2 = \frac{8\pi^2 m \varepsilon}{h^2}. \quad (11.6)$$

Because we are dealing with a single particle, we denote the energies by ε rather than E . Equation (11.5) is satisfied by the expression

$$\psi(x) = A \sin Kx + B \cos Kx, \quad (11.7)$$

where A and B are constants.

To determine the constants A and B , use the boundary conditions (the values of \mathcal{V} at the box walls). According to quantum mechanics, the probability of finding the particle in any region of space is the square of the wavefunction $\psi^2(x)$. Because the potential is infinite at the walls, there is no probability of finding the particle exactly at the wall, so

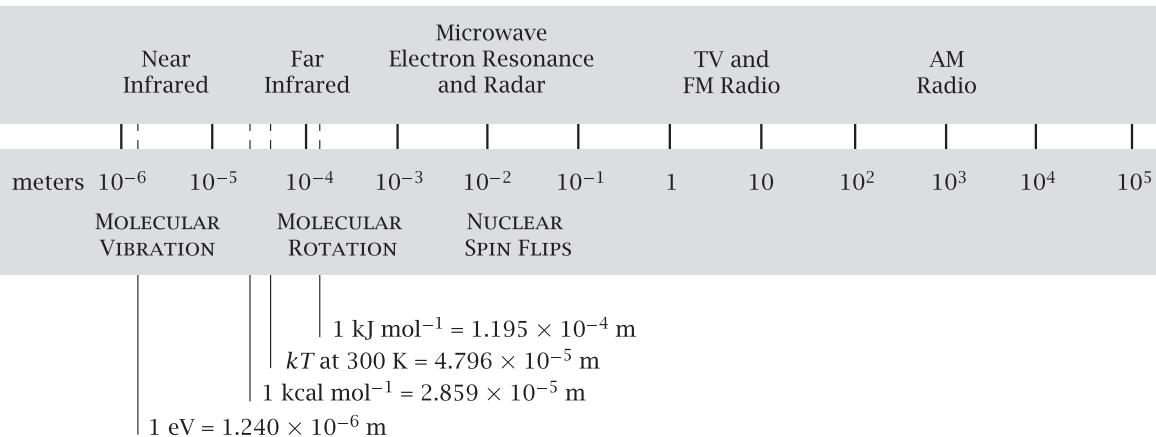
$$\psi^2(0) = 0 \implies \psi(0) = 0. \quad (11.8)$$

When $x = 0$ is substituted into Equation (11.7), you get $A \sin Kx = 0$ and $B \cos Kx = 1$. So you must have $B = 0$ to satisfy the boundary condition (11.8). Equation (11.7) reduces to

$$\psi(x) = A \sin Kx. \quad (11.9)$$

The second boundary condition implies that the wavefunction must equal zero also at $x = L$, the other wall of the box:

$$\psi(L) = A \sin KL = 0. \quad (11.10)$$



Equation (11.10) is satisfied when KL is any integer multiple of π :

$$KL = n\pi \quad \text{for } n = 1, 2, 3, \dots \quad (11.11)$$

$A = 0$ and $n = 0$ aren't useful solutions to Equation (11.10) because they imply that there's no particle in the box.

Substitute Equation (11.6) into Equation (11.11) to find the energy levels ε_n as a function of n :

$$\left(\frac{8\pi^2 m \varepsilon}{h^2}\right)^{1/2} L = n\pi \implies \varepsilon_n = \frac{(nh)^2}{8mL^2}. \quad (11.12)$$

Equation (11.12) is what you need for statistical mechanics—an expression for the energy level spacings ε_n as a function of the index n , the *quantum number*. Figure 11.4 shows the energy ladder for the one-dimensional particle in a box. Because ε_n increases as n^2 , the energy level spacings widen, in this case, with increasing quantum number n .

Before we compute the partition function, let's complete the expression for the wavefunction $\psi(x)$. Probabilities must integrate to 1: $\int_0^L \psi_n^2(x) dx = 1$. The normalized expression for the wavefunction in Equation (11.9) is

$$\psi_n(x) = \left(\frac{2}{L}\right)^{1/2} \sin\left(\frac{n\pi x}{L}\right).$$

Figure 11.5 shows $\psi(x)$, which can be negative, and $\psi^2(x)$, which is always positive. $\psi^2(x)$ represents the spatial distribution of the particle throughout the box. The particle is distributed differently in the ground state ($n = 1$) than in the higher excited states (larger n).

EXAMPLE 11.1 Energy levels for argon in a macroscopic box. What are the energy levels for an argon atom contained in a one-dimensional box of length $L = 1$ cm? The mass of argon is $m = 40 \text{ g mol}^{-1} = (0.040 \text{ kg per mol}^{-1}) / (6.02 \times 10^{23} \text{ atoms mol}^{-1}) = 6.64 \times 10^{-26} \text{ kg per atom}$.

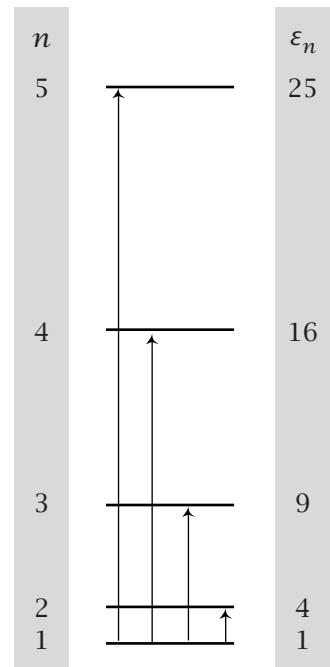


Figure 11.4 Energy ladder for the one-dimensional particle in a box. The energy level ε_n increases with the square of the quantum number n : $\varepsilon_n = n^2 h^2 / (8mL^2)$.

(a) Classical Probability

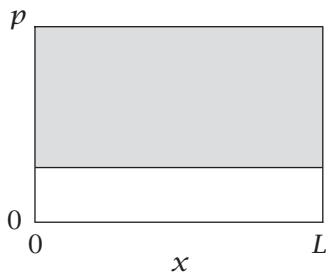
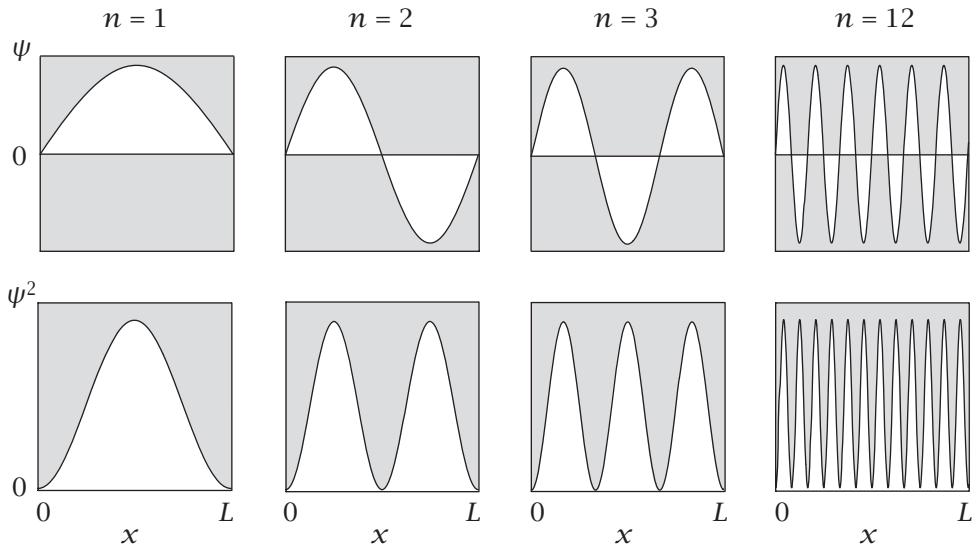
(b) Wavefunction ψ and Probability ψ^2 

Figure 11.5 In contrast to the classical probability (a) of finding a particle distributed uniformly between $x = 0$ and L , the quantum mechanical particle in a box (b) has preferred locations, depending on n . The first row shows the wavefunction $\psi(x)$ and the second row shows $\psi^2(x)$, the probability distribution for the location of the particle.

Equation (11.12) gives the energy spacings as a function of n :

$$\begin{aligned}\varepsilon &= n^2 \frac{(6.626 \times 10^{-34} \text{ Js})^2}{8(6.64 \times 10^{-26} \text{ kg})(10^{-2} \text{ m})^2} \\ &= (8.27 \times 10^{-39} \text{ J}) (6.02 \times 10^{23} \text{ atoms mol}^{-1}) n^2 \approx (5 \times 10^{-15} \text{ J mol}^{-1}) n^2.\end{aligned}$$

The energy spacings are exceedingly small (the smallest one is for $n = 1$), so very many energy levels are populated at room temperature. This is why gases in macroscopic containers can be treated classically, as if they were distributed continuously throughout space. If instead, argon were put in an argon-sized box, say $L = 2 \text{ \AA} = 2 \times 10^{-10} \text{ m}$, the energy would be $12 \text{ J mol}^{-1} n^2$. You need to go to very low temperatures to see quantum effects.

Now that you have the energy ladder, you can compute the translational partition function $q_{\text{translation}}$. Substitute $\varepsilon_n = n^2 h^2 / 8mL^2$ from Equation (11.12) into the definition of the partition function, Equation (10.25). In this case, for each particle,

$$q_{\text{translation}} = \sum_{n=1}^{\infty} e^{-\varepsilon_n/kT} = \sum_{n=1}^{\infty} e^{-n^2 h^2 / (8mL^2 kT)}. \quad (11.13)$$

It is often useful to divide an energy by Boltzmann's constant to express the energy in units of temperature. The translational temperature is $\theta_{\text{trans}} = h^2 / (8mL^2 k)$, so the Boltzmann factors in Equation (11.13) can be expressed as $\exp(-n^2 \theta_{\text{trans}}/T)$. For argon in Example 11.1, $\theta_{\text{trans}} = (8.27 \times 10^{-39} \text{ J}) / (1.38 \times 10^{-23} \text{ JK}^{-1}) \approx 6 \times 10^{-16} \text{ K}$, an extremely small temperature.

When $\theta_{\text{trans}}/T \ll 1$, as it is for argon in a 1 cm one-dimensional box at room temperature, many states are populated and the partition function is large. If the mass of the particle is large, if the size of the box is large, or if the temperature is high (even a few kelvin for most systems) then $\theta_{\text{trans}}/T \ll 1$, and the energy level spacings from Equation (11.12) are so small that the sum in Equation (11.13) can be approximated as an integral:

$$q_{\text{translation}} = \int_0^\infty e^{-(h^2/8mL^2kT)n^2} dn. \quad (11.14)$$

This integral, $\int_0^\infty e^{-ax^2} dx = (1/2)(\pi/a)^{1/2}$, where $a = (h^2/8mL^2kT)$, is given in Equation (K.1) in Appendix K. Therefore the translational partition function for the particle in a one-dimensional box is

$$q_{\text{translation}} = \left(\frac{2\pi mkT}{h^2} \right)^{1/2} L. \quad (11.15)$$

Now let's generalize from one to three dimensions.

A Particle in a Three-Dimensional Box

Suppose the particle is confined within a three-dimensional box with dimensions $0 < x < a$, $0 < y < b$, and $0 < z < c$. The Schrödinger equation is

$$\begin{aligned} -\frac{\hbar^2}{8\pi^2 m} \left(\frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial z^2} \right) \psi(x, y, z) + V(x, y, z) \psi(x, y, z) \\ = \epsilon \psi(x, y, z), \end{aligned} \quad (11.16)$$

where $V(x, y, z) = V(x) + V(y) + V(z)$ and $V(x) = V(y) = V(z) = 0$. Equation (11.16) can be factored into three independent equations:

$$\mathcal{H}_x \psi_x = \epsilon_x \psi_x,$$

$$\mathcal{H}_y \psi_y = \epsilon_y \psi_y,$$

$$\mathcal{H}_z \psi_z = \epsilon_z \psi_z,$$

where $\mathcal{H}_x = -(h^2/8\pi^2 m)(\partial^2/\partial x^2)$, for example. The equations are *separable*—they can be solved separately. Then $\psi(x, y, z) = \psi(x)\psi(y)\psi(z)$ and the energies can be added together. The particle energy ϵ_{n_x, n_y, n_z} is

$$\epsilon_{n_x, n_y, n_z} = \frac{\hbar^2}{8m} \left(\frac{n_x^2}{a^2} + \frac{n_y^2}{b^2} + \frac{n_z^2}{c^2} \right), \quad (11.17)$$

where a , b , and c are the x , y , and z dimensions of the box (see Figure 11.6). Because the three solutions are independent, the partition function for particle translation in the three-dimensional case is just the product of three independent one-dimensional partition functions, each given by Equation (11.15):

$$\begin{aligned} q_{\text{translation}} &= q_x q_y q_z = \left(\frac{2\pi mkT}{h^2} \right)^{3/2} abc = \left(\frac{2\pi mkT}{h^2} \right)^{3/2} V \\ &= \frac{V}{\Lambda^3}, \end{aligned} \quad (11.18)$$

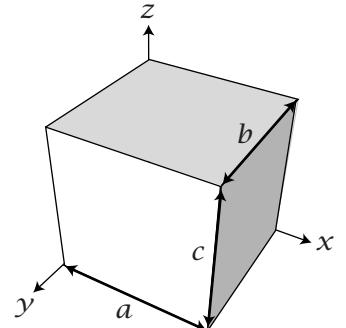


Figure 11.6 a , b , and c are the x , y , and z dimensions for the three-dimensional particle-in-a-box problem.

where V is the volume of the box, and Λ^3 , which has units of volume, is

$$\Lambda^3 = \left(\frac{h^2}{2\pi mkT} \right)^{3/2}. \quad (11.19)$$

This reference volume is of roughly molecular dimensions. For example, $\Lambda = 0.714 \text{ \AA}$ for H_2 at $T = 300 \text{ K}$.

Let's compute the partition function for a monatomic gas.

EXAMPLE 11.2 Calculate the translational partition function for argon. For temperature $T = 273 \text{ K}$, and the volume $V = 2.24 \times 10^{-2} \text{ m}^3$ of one mole of gas at $p = 1 \text{ atm}$, the translational partition function is

$$q_{\text{translation}} = \left(\frac{2\pi mkT}{h^2} \right)^{3/2} V,$$

where

$$\begin{aligned} m \approx 40 \text{ g mol}^{-1} &= \frac{0.04}{6.02 \times 10^{23}} \frac{\text{kg mol}^{-1}}{\text{atoms mol}^{-1}} \\ &= 6.64 \times 10^{-26} \text{ kg per atom.} \end{aligned}$$

$$kT = (1.38 \times 10^{-23} \text{ JK}^{-1})(273 \text{ K}) = 3.77 \times 10^{-21} \text{ J}, \text{ so}$$

$$\begin{aligned} q_{\text{translation}} &= \left(\frac{2\pi(6.64 \times 10^{-26} \text{ kg per atom})(3.77 \times 10^{-21} \text{ J})}{(6.63 \times 10^{-34} \text{ Js})^2} \right)^{3/2} \\ &\quad \times (2.24 \times 10^{-2} \text{ m}^3) \\ &= (2.14 \times 10^{32})V = 4.79 \times 10^{30} \text{ states per atom.} \end{aligned}$$

The partition function $q_{\text{translation}}$ is the number of translational energy levels effectively accessible to an argon atom at this temperature. This number is certainly large enough to justify the replacement of the sum in Equation (11.13) by the integral in Equation (11.14).

Translational motion is only one of the types of degrees of freedom that atoms and molecules use to store and exchange energies. Quantum mechanics goes beyond the kinetic theory of gases in that it treats the internal degrees of freedom as well as translation. If you put all the forces between the nuclei and electrons into a Schrödinger equation to describe a particular atom or molecule, and if you are able to solve it, you can compute the orbitals (wavefunctions) and energy levels for that atom or molecule.

In the same way that we separated translational motion into three independent components, other Schrödinger equations can also be factored into simpler differential equations, which can be solved individually. Such component energies are additive. This is the basis for classifications such as *translations*, *rotations*, and *vibrations*. For example, suppose you are interested in a diatomic gas molecule. The Hamiltonian operator can be expressed in terms of the distance between the two bonded atoms (vibration), their angular distribution (rotation), and the center-of-mass position of the molecule in space (translation). Because the differential operators are additive, you get a Schrödinger equation of the form $(\mathcal{H}_{\text{translation}} + \mathcal{H}_{\text{vibration}} + \mathcal{H}_{\text{rotation}})\psi = \varepsilon\psi$, which can be

further simplified into three independent equations: $\mathcal{H}_{\text{translation}}\psi = \varepsilon_{\text{translation}}\psi$, $\mathcal{H}_{\text{vibration}}\psi = \varepsilon_{\text{vibration}}\psi$, and $\mathcal{H}_{\text{rotation}}\psi = \varepsilon_{\text{rotation}}\psi$.

Although quantum mechanics treats many kinds of problems, our main interest here is in the structures and properties of molecules, as a foundation for chemical reaction equilibria and kinetics. So we will focus on simple models of these types of degrees of freedom. The translational component was treated by the particle-in-a-box model. Now we consider vibrations, rotations, and electronic excitations.

Vibrations Can Be Treated by the Harmonic Oscillator Model

Atoms vibrate inside molecules, in solids, and on surfaces, for example. To a first approximation, such atoms act as if they were connected to other atoms by springs. Here's the simplest model of such bonds.

Consider a particle of mass m that is free to move along the x axis. It is held by a spring-like force around the position $x = 0$ (see Figure 11.7). If the spring constant is k_s , the potential energy is $\mathcal{V}(x) = k_s x^2/2$. This is called a *square-law*, or *parabolic*, potential. Bonding potentials are often well approximated as a polynomial function $\mathcal{V}(x) = V_0 + V_1 x + V_2 x^2 + \dots$. For small motions near the point of minimum energy, where the first derivative is zero ($V_1 = \partial \mathcal{V} / \partial x = 0$), the lowest-order nonzero term is the square-law term, $V_2 x^2$. Therefore parabolic potentials are often good first approximations to bonding energy functions. The corresponding Schrödinger equation for the quantum harmonic oscillator is

$$-\frac{\hbar^2}{8\pi^2 m} \frac{d^2\psi(x)}{dx^2} + \frac{1}{2} k_s x^2 \psi(x) = \varepsilon_v \psi(x), \quad (11.20)$$

where the subscript v indicates that these are vibrational energies. Rearranging gives the differential equation

$$\frac{d^2\psi}{dx^2} - \left(\frac{1}{a^4}\right) x^2 \psi = -\left(\frac{2\varepsilon_v}{a^4 k_s}\right) \psi, \quad (11.21)$$

where $a^4 = \hbar^2 / (4\pi^2 m k_s)$. Wavefunctions $\psi(x)$ that satisfy Equation (11.20) are Gaussian functions multiplied by *Hermite polynomials*. The wavefunctions are given in Table 11.1 and shown in Figure 11.8.

The vibrational energy levels ε_v are

$$\varepsilon_v = \left(v + \frac{1}{2}\right) \hbar\nu, \quad (11.22)$$

where the quantum number is $v = 0, 1, 2, 3, \dots$. Also, $\nu = (1/2\pi)(k_s/m)^{1/2}$ is the vibrational frequency of the harmonic oscillator. (Be careful not to confuse v , the vibrational quantum number, with ν , the vibrational frequency. Unfortunately, this is standard notation.) Notice that the lowest energy level is not zero. The lowest level has energy $\varepsilon_0 = (1/2)\hbar\nu$. This is called the *zero-point energy*. The system vibrates even at $T = 0\text{K}$.

To apply the harmonic oscillator model to diatomic molecules, we need to generalize from a single particle on a spring to two masses m_1 and m_2

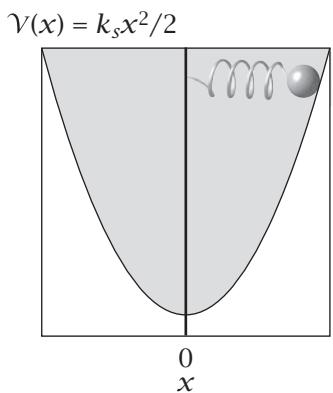


Figure 11.7 The harmonic oscillator model for vibrations is based on a mass on a spring. The spring is attached at $x = 0$. The potential energy \mathcal{V} increases as the square of the displacement x , with spring constant k_s .

Table 11.1 Energy eigenfunctions of the simple harmonic oscillator.

Quantum Number v	Energy Eigenvalue ε_v	Energy Eigenfunction $\psi_v(x) = \left(\frac{1}{v!2^v a\sqrt{\pi}}\right)^{1/2} H_v\left(\frac{x}{a}\right) e^{-x^2/2a^2}$
0	$\frac{1}{2}h\nu$	$\left(\frac{1}{a\sqrt{\pi}}\right)^{1/2} e^{-x^2/2a^2}$
1	$\frac{3}{2}h\nu$	$\left(\frac{1}{2a\sqrt{\pi}}\right)^{1/2} 2\left(\frac{x}{a}\right) e^{-x^2/2a^2}$
2	$\frac{5}{2}h\nu$	$\left(\frac{1}{8a\sqrt{\pi}}\right)^{1/2} \left[2 - 4\left(\frac{x}{a}\right)^2\right] e^{-x^2/2a^2}$
3	$\frac{7}{2}h\nu$	$\left(\frac{1}{48a\sqrt{\pi}}\right)^{1/2} \left[12\left(\frac{x}{a}\right) - 8\left(\frac{x}{a}\right)^3\right] e^{-x^2/2a^2}$
4	$\frac{9}{2}h\nu$	$\left(\frac{1}{348a\sqrt{\pi}}\right)^{1/2} \left[12 - 48\left(\frac{x}{a}\right)^2 + 16\left(\frac{x}{a}\right)^4\right] e^{-x^2/2a^2}$

Source: AP French and EF Taylor, *An Introduction to Quantum Physics*, WW Norton, New York, 1978.

Figure 11.8 The parabola shows the harmonic oscillator potential. The horizontal lines show the energy levels, which are equally spaced for the quantum harmonic oscillator. The lowest level has energy $h\nu/2$. This is called the *zero-point energy*. The curve on each horizontal line shows $\psi^2(x)$, the particle distribution probability for each energy level.

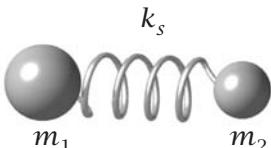
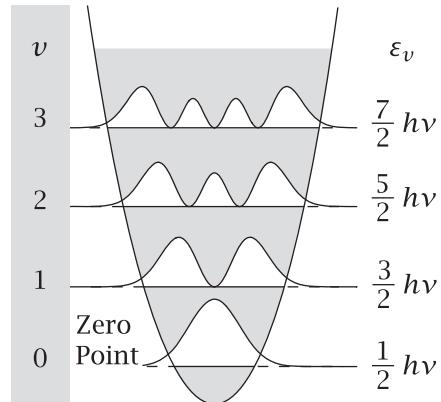


Figure 11.9 Vibrations in a diatomic molecule are modeled as two masses m_1 and m_2 connected by a spring with constant k_s .

connected by a spring (see Figure 11.9). This leads to only a small alteration of the theory. The vibrational energy levels are exactly as given in Equation (11.22), but now the vibrational frequency is

$$\nu = \left(\frac{1}{2\pi}\right) \left(\frac{k_s}{\mu}\right)^{1/2}, \quad (11.23)$$

where

$$\mu = \frac{m_1 m_2}{m_1 + m_2} \quad (11.24)$$

is called the *reduced mass*.

You can get the single-particle partition function by substituting the energy, Equation (11.22), into the partition function, Equation (10.25). Taking energy differences relative to the ground state gives

$$q_{\text{vibration}} = \sum_{v=0}^{\infty} e^{-vh\nu/kT} = 1 + e^{-h\nu/kT} + e^{-2h\nu/kT} + \dots \\ = 1 + x + x^2 + x^3 + \dots, \quad (11.25)$$

where $x = e^{-h\nu/kT}$. A vibrational temperature can be defined as $\theta_{\text{vibration}} = h\nu/k$. Using the series expansion in Equation (J.6) in Appendix J, $(1-x)^{-1} = 1 + x + x^2 + \dots$, for $0 < |x| < 1$, the vibrational partition function can be written more compactly as

$$q_{\text{vibration}} = \frac{1}{1 - e^{-h\nu/kT}}. \quad (11.26)$$

EXAMPLE 11.3 The vibrational partition function of O₂. Oxygen molecules have a vibrational wavenumber of 1580 cm⁻¹. To convert from cm⁻¹ to ν , which has units of s⁻¹, multiply by the speed of electromagnetic radiation (light) to get $\nu = (1580 \text{ cm}^{-1})(2.997 \times 10^{10} \text{ cm s}^{-1}) = 4.737 \times 10^{13} \text{ s}^{-1}$. You have

$$\theta_{\text{vibration}} = \frac{h\nu}{k} = \frac{(6.626 \times 10^{-34} \text{ Js})(4.737 \times 10^{13} \text{ s}^{-1})}{1.38 \times 10^{-23} \text{ JK}^{-1}} = 2274 \text{ K}.$$

At room temperature, $\theta_{\text{vibration}}/T = 2274 \text{ K}/300 \text{ K} = 7.58$. For the partition function at room temperature, Equation (11.26) gives $q_{\text{vibration}} = 1.0005$. Most oxygen molecules are in their vibrational ground states at this temperature. Even at 1000 K, most oxygen molecules are still in their vibrational ground states: $\theta_{\text{vibration}}/T = 2.27$ and $q_{\text{vibration}} = 1.11$.

Rotations Can Be Treated by the Rigid Rotor Model

In classical mechanics, if a mass m orbits at a distance R around a center point and has tangential velocity v_t , its angular momentum is $L = mRv_t$ and its kinetic energy is $K = mv_t^2/2 = L^2/(2mR^2)$. For the corresponding quantum mechanical problem, L is an operator. Our interest here is in a related problem. The rotations of a diatomic molecule can be modeled as masses m_1 and m_2 separated by a rigid connector of length R (see Figure 11.10). The Hamiltonian operator is a function of the two angular degrees of freedom θ and ϕ :

$$\mathcal{H}\psi(\theta, \phi) = \frac{L^2\psi}{2\mu R^2} = \frac{-\hbar^2}{8\pi^2\mu R^2} \left[\frac{1}{\sin\theta} \frac{\partial}{\partial\theta} \left(\sin\theta \frac{\partial}{\partial\theta} \right) + \frac{1}{\sin^2\theta} \frac{\partial^2}{\partial\phi^2} \right] \psi \\ = \varepsilon\psi, \quad (11.27)$$

where μ is the reduced mass given by Equation (11.24), and $I = \mu R^2$ is the *moment of inertia*. The wavefunctions that satisfy Equation (11.27) are called the *spherical harmonics* $Y_{\ell,m}(\theta, \phi)$:

$$\psi(\theta, \phi) = A_{\ell,m} Y_{\ell,m}(\theta, \phi) = A_{\ell,m} P_{\ell,m}(\theta) e^{im\phi},$$

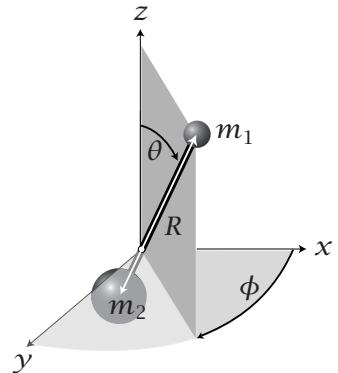


Figure 11.10 In the rigid rotor model of a diatomic molecule, masses m_1 and m_2 are separated by a rigid connector of length R . The origin is at the center of mass, and the angular degrees of freedom are θ and ϕ .

where $\ell = 0, 1, 2, \dots$ and $m = -\ell, -\ell+1, \dots, 0, 1, \dots, \ell-1, \ell$ are the two quantum numbers, the $P(\theta)$'s are called *Legendre polynomials* [1], and the $A_{\ell,m}$'s are constants. The energy levels for the rigid rotor are

$$\varepsilon_\ell = \frac{\ell(\ell+1)h^2}{8\pi^2I}, \quad (11.28)$$

and the rotational temperature is $\theta_{\text{rotation}} = h^2/(8\pi^2Ik)$. Rigid rotors have energy levels at $\ell(\ell+1) = 0, 2, 6, 12, 20, 30, \dots$

For each value of ℓ , there is a degeneracy due to the $2\ell+1$ values of m . So the rotational partition function is

$$q_{\text{rotation}} = \sum_{\ell=0}^{\infty} (2\ell+1)e^{-\varepsilon_\ell/kT}. \quad (11.29)$$

For high temperatures ($T \gg \theta_{\text{rotation}}$), Equation (11.29) can be approximated by an integral to give

$$q_{\text{rotation}} = \frac{T}{\sigma\theta_{\text{rotation}}} = \frac{8\pi^2IkT}{\sigma h^2}, \quad (11.30)$$

where σ is a nuclear and rotation symmetry factor that accounts for the number of equivalent orientations of the molecule ($\sigma = 1$ for heteronuclear diatomic molecules; $\sigma = 2$ for homonuclear diatomic molecules; $\sigma = 2$ for H₂O; $\sigma = 12$ for CH₄, and $\sigma = 12$ for benzene, for example). Think of σ as a correction for overcounting.

For nonlinear molecules with three principal moments of inertia I_a , I_b , and I_c , the rotational partition function is

$$q_{\text{nonlinear rotation}} = \frac{(\pi I_a I_b I_c)^{1/2}}{\sigma} \left(\frac{8\pi^2 k T}{h^2} \right)^{3/2}. \quad (11.31)$$

In general, molecules with large moments of inertia have large rotational partition functions because the spacings between energy levels is small.

EXAMPLE 11.4 The rotational partition function of O₂. The length of the bond in an oxygen molecule is $R = 1.2074 \text{ \AA}$ and the mass of each atom is $m = 16 \text{ g mol}^{-1}$, so the moment of inertia per molecule is

$$\begin{aligned} I &= \mu R^2 = \frac{mR^2}{2} = \frac{\left(0.016 \frac{\text{kg}}{\text{mol}}\right) (1.2074 \times 10^{-10} \text{ m})^2}{2 \left(6.02 \times 10^{23} \frac{\text{molecules}}{\text{mol}}\right)} \\ &= 1.937 \times 10^{-46} \text{ kg m}^2. \end{aligned}$$

The rotational temperature is

$$\begin{aligned} \theta_{\text{rotation}} &= \frac{h^2}{8\pi^2Ik} = \frac{(6.626 \times 10^{-34} \text{ Js})^2}{8\pi^2(1.937 \times 10^{-46} \text{ kg m}^2)(1.38 \times 10^{-23} \text{ JK}^{-1})} \\ &= 2.08 \text{ K}. \end{aligned}$$

Table 11.2 Properties of linear molecules, useful for computing partition functions.

Molecule	R^a (Å)	σ^b	I^a (amu Å ²)	$\theta_{\text{rotation}}^a$ (K)	$\theta_{\text{vibration}}^a$ (K)	$\theta_{\text{electronic}}^c$ (K)	g_0^b	D_0^b (kJ mol ⁻¹)
H ₂	0.7415	2	0.2771	87.53	6338	129,000	1	432.1
HCl	1.2744	1	1.593	15.24	4302	61,800	1	427.8
N ₂	1.0976	2	8.437	2.874	3395	97,400	1	941.4
CO	1.1282	1	8.733	2.777	3122	91,500	1	1070.1
NO	1.1506	1	9.888	2.453	2739	61,700	2	627.7
O ₂	1.2074	2	11.663	2.080	2274	11,100	3	491.9
Cl ₂	1.988	2	70.06	0.346	813	25,700	1	239.2
I ₂	2.666	2	451.0	0.0538	308	16,700	1	148.8

Source: ^aJH Knox, *Molecular Thermodynamics: An Introduction to Statistical Mechanics for Chemists*, Wiley, Chichester, 1978.

^bRA Alberty and RJ Silbey, *Physical Chemistry*, 1st edition, Wiley, New York, 1992; data are from M Chase et al. JANAF Thermochemical Tables, *J Phys Chem Ref Data* 14, Supplement 1 (1985).

^cRS Berry, SA Rice, and J Ross, *Physical Chemistry*, 2nd edition, Oxford University Press, 2000.

Since $\sigma = 2$ for a homonuclear diatomic molecule, the rotational partition function at room temperature is (see Equation (11.30))

$$q_{\text{rotation}} = \frac{300 \text{ K}}{2(2.08 \text{ K})} = 72. \quad (11.32)$$

Therefore many rotational states are populated at room temperature. Table 11.2 gives rotational and vibrational constants for several small molecules.

The Electronic Partition Function

The partition function that accounts for the various states of electronic excitation is

$$q_{\text{electronic}} = g_0 + g_1 e^{-\Delta\varepsilon_1/kT} + g_2 e^{-\Delta\varepsilon_2/kT} + \dots, \quad (11.33)$$

where the g_i 's are the electronic degeneracies and the $\Delta\varepsilon_i$'s are the electronic excited-state energies. Details are given in [1, 2]; we give only an overview here. Table 11.2 shows that the first excited-state energies for diatomic molecules are quite large relative to $T = 300 \text{ K}$. Typically $\theta_e = \Delta\varepsilon_1/k \approx 10^4\text{--}10^5 \text{ K}$. In those cases, the exponential terms approach zero and you can express Equation (11.33) as $q_{\text{electronic}} \approx g_0$. The ground-state electronic degeneracies g_0 are shown in Table 11.2.

Table 11.3 Atomic energy states used to obtain the electronic partition functions of atoms.

Atom	Electron Configuration	Term Symbol	Degeneracy $g = 2J+1$	Energy (eV)
H	1s	$^2S_{1/2}$	2	0
	2p	$^2P_{1/2}$	2	10.20
	2s	$^2S_{1/2}$	2	10.20
	2p	$^2P_{3/2}$	4	10.20
He	1s ²	1S_0	1	0
	1s2s	3S_1	3	19.82
		1S_0	1	20.62
Li	1s ² 2s	$^2S_{1/2}$	2	0
	1s ² 2p	$^2P_{1/2}$	2	1.85
		$^2P_{3/2}$	4	1.85
O	1s ² 3s	$^2S_{1/2}$	2	3.37
	1s ² 2s ² 2p ⁴	3P_2	5	0
		3P_1	3	0.02
		3P_0	1	0.03
		1D_2	5	1.97
F	1s ² 2s ² 2p ⁵	1S_0	1	4.19
		$^2P_{3/2}$	4	0
		$^2P_{1/2}$	2	0.05
	1s ² 2s ² 2p ⁴ 3s	$^4P_{5/2}$	6	12.70
		$^4P_{3/2}$	4	12.73
		$^4P_{1/2}$	2	12.75
		$^2P_{3/2}$	4	13.0
		$^2P_{1/2}$	2	13.0

Source: DA McQuarrie, *Statistical Mechanics*, Harper & Row, New York, 1976, and CE Moore, Atomic Energy States, *Natl Bur Standards Circ* 1, 467 (1949).

For computing chemical equilibria (Chapter 13), you will need $q_{\text{electronic}}$ not just for molecules, but also for individual atoms. For atoms, $q_{\text{electronic}}$ depends on the electronic configuration of the atom (see Table 11.3). For example, the 1s configuration of the H atom has degeneracy $g_0 = 2$, and the 2s configuration has $g_1 = 2$. But the 2s orbital has a high energy, so at temperatures T below about $kT = 10.20$ eV, the only contribution to the electronic partition function is $g_0 = 2$. Table 11.3 shows that for oxygen atoms, $g_0 = 5$, $g_1 = 3$, and $g_2 = 1$. The energies of the first two excited states are small enough for you to retain all three terms except at the very lowest temperatures. Similarly, the first two levels can contribute for fluorine. Atomic degeneracies can be found in [3]. Electronic configurations can be described as *term symbols*, where the lower right subscript (1/2 for H 1s, for example) is J , and the degeneracy is given by $g = 2J+1$.

The Total Partition Function Predicts the Properties of Molecules

Atoms and molecules can store energies in all their various degrees of freedom. To compute thermodynamic properties, sum all the energies, $\varepsilon_{\text{total}} = \varepsilon_{\text{translation}} + \varepsilon_{\text{rotation}} + \varepsilon_{\text{vibration}} + \varepsilon_{\text{electronic}}$ to get the total partition function q , which is the product of terms,

$$q = q_{\text{translation}} q_{\text{rotation}} q_{\text{vibration}} q_{\text{electronic}}. \quad (11.34)$$

For example, for a linear diatomic molecule with $q_{\text{electronic}} = 1$, you will have a total partition function

$$q = \left(\frac{2\pi mkT}{h^2}\right)^{3/2} V \left(\frac{8\pi^2IkT}{\sigma h^2}\right) \left(\frac{1}{1-e^{-hv/kT}}\right). \quad (11.35)$$

Ideal Gas Properties Are Predicted by Quantum Mechanics

Partition functions based on quantum mechanics can provide a more complete description of the ideal gas than our simple lattice model could. In this section, we compute the properties of ideal gases.

Ideal Gas Free Energy

The Helmholtz free energy of an ideal gas is found by substituting $Q = q^N/N!$ into Equation (10.42):

$$\begin{aligned} F &= -kT \ln Q = -kT \ln \left(\frac{q^N}{N!}\right) = -kT \ln \left[\left(\frac{eq}{N}\right)^N\right] \\ &= -NkT \ln \left(\frac{eq}{N}\right), \end{aligned} \quad (11.36)$$

where the factor of e comes from Stirling's approximation for $N!$

Ideal Gas Pressure

We derived the ideal gas law from a lattice model in Example 6.1. Now we derive the gas law from quantum mechanics instead. Equation (11.35) shows that the partition function for a gas is a product of volume V and a volume-independent factor we'll call q_0 . Substituting $q = q_0V$ into Equation (11.36) gives $F = -NkT \ln V - NkT \ln(eq_0/N)$. Taking the derivative that defines pressure in Equation (8.12) gives the ideal gas law:

$$p = -\left(\frac{\partial F}{\partial V}\right)_{T,N} = \frac{NkT}{V}.$$

Ideal Gas Energy

We noted in Example 7.6 that the internal energy of an ideal gas depends only on its temperature and not on its volume: $U = U(T)$. The lattice model doesn't give

a basis for understanding the internal energy or its temperature dependence, but quantum mechanics does. Equation (10.36) expresses the internal energy U in terms of the partition function q and the number of particles N for either distinguishable or indistinguishable particles (see Problem 8 in Chapter 10):

$$U = NkT^2 \left(\frac{\partial \ln q}{\partial T} \right).$$

For the purpose of taking the temperature derivative, express q as a product of its temperature-dependent and temperature-independent parts. For example, for translational freedom, Equation (11.18) would factor into $q = c_0 T^{3/2}$, where $c_0 = (2\pi mk/h^2)^{3/2}V$ is independent of temperature. Substitute this expression for q into Equation (10.36) to get

$$U = \frac{NkT^2}{q} \left(\frac{\partial q}{\partial T} \right) = \frac{NkT^2}{c_0 T^{3/2}} \left(\frac{3}{2} c_0 T^{1/2} \right) = \frac{3}{2} NkT. \quad (11.37)$$

The translational contribution to the average particle energy is $\langle \varepsilon \rangle = U/N = (3/2)kT$. This is the internal energy of an ideal gas.

Because the energy of an ideal gas does not depend on volume, the heat capacity does not either:

$$C_V = \left(\frac{\partial U}{\partial T} \right) = \frac{3}{2} Nk. \quad (11.38)$$

Rotations also contribute to the internal energy. For nonlinear gas molecules,

$$q_{\text{rotation}} = c_1 T^{3/2}, \quad (11.39)$$

where

$$c_1 = \frac{(\pi I_a I_b I_c)^{1/2}}{\sigma} \left(\frac{8\pi^2 k}{h^2} \right)^{3/2}$$

is a temperature-independent constant. Substitute Equation (11.39) into (11.37) to find the component of the internal energy of nonlinear molecules that is due to rotation:

$$U_{\text{rotation(nonlinear)}} = \frac{3}{2} NkT. \quad (11.40)$$

Likewise, for linear gas molecules, the rotational internal energy is

$$U_{\text{rotation(linear)}} = NkT. \quad (11.41)$$

The total internal energy of an ideal gas is given as a sum of terms: $(1/2)kT$ for each of the three translational degrees of freedom and $(1/2)kT$ per rotational degree of freedom. Such $kT/2$ units of energy per degree of freedom are called *equipartition*; see below.

Ideal Gas Entropy

A fundamental validation of the indistinguishability of gas molecules, and of the expression $Q = q^N/N!$ for the system partition function of a gas, is provided by the absolute entropies of gases.

The only contribution to the entropy of an ideal monatomic gas is from its translational freedom. Combining $Q = q^N / N!$, where $q = q_{\text{translation}}$ from Equation (11.18), with $U = (3/2)NkT$ and Stirling's approximation of Equation (10.41) gives the **Sackur-Tetrode equation** for the absolute entropy of a monatomic ideal gas:

$$\begin{aligned} S &= k \ln \left(\frac{q^N}{N!} \right) + \frac{U}{T} = Nk \ln q - k(N \ln N - N) + \frac{3}{2} Nk \\ &= Nk \left(\ln q - \ln N + \frac{5}{2} \right) = Nk \ln \left(\frac{qe^{5/2}}{N} \right) \\ &= Nk \ln \left[\left(\frac{2\pi mkT}{h^2} \right)^{3/2} \left(\frac{e^{5/2}}{N} \right) V \right], \end{aligned} \quad (11.42)$$

where V is the volume. Example 11.5 shows how the Sackur-Tetrode equation is applied.

EXAMPLE 11.5 Calculating the absolute entropy of argon from the Sackur-Tetrode equation. Let's calculate the entropy of argon at $T = 300\text{ K}$ and $p = 1\text{ atm} = 1.01 \times 10^5 \text{ N m}^{-2}$. For argon, the Sackur-Tetrode equation (11.42) is

$$S = Nk \ln \left(\frac{q_{\text{translation}} e^{5/2}}{N} \right). \quad (11.43)$$

We want $q_{\text{translation}}$ at $T = 300\text{ K}$. Use $q_{\text{translation}}$ computed at $T = 273\text{ K}$ from Example 11.2, and correct for the difference in temperature:

$$\begin{aligned} q_{\text{translation}}(300) &= q_{\text{translation}}(273) \left(\frac{300\text{ K}}{273\text{ K}} \right)^{3/2} \\ &= (2.14 \times 10^{32} V)(1.15) \\ &= 2.47 \times 10^{32} V. \end{aligned}$$

Substitute this into Equation (11.43) to get

$$\frac{S}{N} = k \ln \left[2.47 \times 10^{32} \left(\frac{V}{N} \right) e^{5/2} \right]. \quad (11.44)$$

Substitute kT/p from the ideal gas law for V/N :

$$\begin{aligned} \frac{V}{N} &= \frac{kT}{p} = \frac{(1.38 \times 10^{-23} \text{ J K}^{-1} \text{ per molecule})(300\text{ K})}{1.01 \times 10^5 \text{ N m}^2} \\ &= 4.10 \times 10^{-26}, \\ \Rightarrow \frac{S}{N} &= (8.31 \text{ J K}^{-1} \text{ mol}^{-1}) \ln \left[2.47 \times 10^{32} (4.10 \times 10^{-26}) e^{5/2} \right] \\ &= 154.8 \text{ J K}^{-1} \text{ mol}^{-1}. \end{aligned} \quad (11.45)$$

The value of S/N calculated here is likely to be more accurate than the thermodynamically measured value of the entropy for argon ($155 \text{ J K}^{-1} \text{ mol}^{-1}$; see Table 11.4), because of the typical sizes of errors in thermodynamic measurements.

Table 11.4 Gas entropies determined calorimetrically and from statistical mechanics.

Substance	Statistical Entropy ($\text{J K}^{-1} \text{ mol}^{-1}$)	Calorimetric Entropy ($\text{J K}^{-1} \text{ mol}^{-1}$)
Ne ^a	146.22	146.0
Ar ^a	154.74	155.0
N ₂ ^b	191.50	192.05
O ₂ ^b	209.19	205.45
Cl ₂ ^a	222.95	223.1
HCl ^a	186.68	186.0
HBr ^a	198.48	199.0
NH ₃ ^b	192.34	192.09
CO ₂ ^a	213.64	214.00
CH ₃ Cl ^b	234.22	234.05
CH ₃ Br ^a	243.0	242.0
CH ₃ NO ₂ ^b	275.01	275.01
C ₂ H ₄ ^b	219.33	219.58
Cyclopropane ^b	227.02	226.65
Benzene ^b	269.28	269.70
Toluene ^b	320.83	321.21

Source: ^aRS Berry, SA Rice, and J Ross, *Physical Chemistry*, 2nd edition, Oxford University Press, 2000.

^bW Kauzmann, *Thermodynamics and Statistics*, WA Benjamin, New York, 1967. Data are from EF Westrum and JP McCullough, *Physics and Chemistry of the Organic Solid State*, D Fox, MM Labes, and A Weissberger, eds, Chapter 1, Interscience, 1963; DD Wagman et al., *Selected Values of Chemical Thermodynamic Properties*, National Bureau of Standards Technical Note 270-1, Washington, DC, 1965.

The Sackur-Tetrode equation (11.42) also predicts how the entropy changes with volume for an ideal gas at constant temperature:

$$\Delta S = S(V_B) - S(V_A) = Nk \ln\left(\frac{V_B}{V_A}\right), \quad (11.46)$$

which is the same result you would get from the lattice model gas in Example 6.1, page 98.

Ideal Gas Chemical Potential

Chapters 13–16 focus on the chemical potential, the quantity that describes the tendency for molecules to move from one place to another, or from one phase to another. It is the basis for treating boiling and freezing, for describing the partitioning of solute molecules between phases, and for treating chemical reaction equilibria and kinetics. To treat the gas phase in these processes, the key relationship that we will need is $\mu(p)$, the chemical potential as a function of pressure.

Combining the definition of chemical potential $\mu = (\partial F / \partial N)_{T,V}$ with Equation (11.36) gives the chemical potential of an ideal gas in terms of its partition function:

$$\mu = kT \ln\left(\frac{N}{eq}\right) + kT = -kT \ln\left(\frac{q}{N}\right). \quad (11.47)$$

Again we aim to factor the partition function, based on the derivative we will need to take, into a pressure-dependent term and a term for all the pressure-independent quantities. Express the partition function, $q = q_0V$ as a product of the volume V and q_0 , the partition function with the volume factored out. Use the ideal gas law $V = NkT/p$ to get

$$\frac{q}{N} = \frac{q_0V}{N} = \frac{q_0kT}{p}. \quad (11.48)$$

The quantity q_0kT has units of pressure:

$$p_{\text{int}}^\circ = q_0kT = kT \left(\frac{2\pi mkT}{h^2} \right)^{3/2} q_{\text{rotation}} q_{\text{vibration}} q_{\text{electronic}}. \quad (11.49)$$

Our notation indicates that the quantity p_{int}° represents properties *internal* to the molecule. Substituting Equations (11.48) and (11.49) into Equation (11.47) gives

$$\mu = \mu^\circ + kT \ln p = kT \ln \left(\frac{p}{p_{\text{int}}^\circ} \right), \quad (11.50)$$

where $\mu^\circ = -kT \ln p_{\text{int}}^\circ$. Equation (11.50) makes explicit how the chemical potential depends on the applied pressure p . Because p_{int}° has units of pressure, the rightmost side of Equation (11.50) shows that you are taking the logarithm of a quantity that is dimensionless, even though the chemical potential is often written as $\mu^\circ + kT \ln p$, shown as the equality on the left. The quantities p_{int}° and μ° with superscripts \circ are called the *standard-state* pressure and chemical potential. We will use Equation (11.50) throughout Chapters 13–16. (Our chemical potential here is per molecule. If you prefer to express it per mole instead, then replace k with the gas constant R in Equation (11.47).)

EXAMPLE 11.6 The standard-state pressure of helium. To calculate p_{int}° for helium at $T = 300\text{ K}$, first use Equation (11.19) to get

$$\begin{aligned} \Lambda^3 &= \left(\frac{h^2}{2\pi mkT} \right)^{3/2} \\ &= \left[\frac{(6.626 \times 10^{-34}\text{ Js})^2}{2\pi(4 \times 1.66 \times 10^{-27}\text{ kg})(1.38 \times 10^{-23}\text{ JK}^{-1})(300\text{ K})} \right]^{3/2} \\ &= 1.28 \times 10^{-31}\text{ m}^3 = 0.13\text{ \AA}^3. \end{aligned}$$

Then

$$\begin{aligned} p_{\text{int}}^\circ &= \frac{kT}{\Lambda^3} = \frac{(1.38 \times 10^{-23}\text{ JK}^{-1})(300\text{ K})}{1.28 \times 10^{-31}\text{ m}^3} \\ &= 3.2 \times 10^{10}\text{ Pa} = 3.2 \times 10^5\text{ atm.} \end{aligned}$$

The Equipartition Theorem Says That Energies Are Uniformly Distributed to Each Degree of Freedom

On page 208, we noted that the average energies for gases are integral multiples of $kT/2$ for rotational and translational degrees of freedom. This is a manifestation of the principle of equipartition of energy. Where does this principle come from?

When a form of energy depends on a degree of freedom x , the average energy $\langle \varepsilon \rangle$ of that component is

$$\langle \varepsilon \rangle = \frac{\sum_{\text{all } x} \varepsilon(x) e^{-\varepsilon(x)/kT}}{\sum_{\text{all } x} e^{-\varepsilon(x)/kT}}. \quad (11.51)$$

When quantum effects are unimportant, so that a large number of states are populated, the sums in Equation (11.51) can be approximated by integrals and the average energy $\langle \varepsilon \rangle$ is

$$\langle \varepsilon \rangle = \frac{\int_{-\infty}^{\infty} \varepsilon(x) e^{-\varepsilon(x)/kT} dx}{\int_{-\infty}^{\infty} e^{-\varepsilon(x)/kT} dx}, \quad (11.52)$$

for $-\infty \leq x \leq \infty$. For many types of degrees of freedom, the energy is a square-law function: $\varepsilon(x) = cx^2$, where $c > 0$. Perform the integrations using Equations (K.1) and (K.3) from Appendix K. Equation (11.52) becomes

$$\langle \varepsilon \rangle = \frac{\int_{-\infty}^{\infty} cx^2 e^{-cx^2/kT} dx}{\int_{-\infty}^{\infty} e^{-cx^2/kT} dx} = \frac{1}{2} kT, \quad (11.53)$$

which proves the equipartition idea that the energy stored in each such degree of freedom is $kT/2$. The constant c gives the scale for the energy level spacings, but note that $\langle \varepsilon \rangle$ does not depend on c . A useful variant of this expression gives the fluctuations. Because you have $\langle \varepsilon \rangle = c \langle x^2 \rangle = kT/2$, the fluctuations are given by

$$\langle x^2 \rangle = \frac{kT}{2c}. \quad (11.54)$$

Square-law relations hold for many types of degrees of freedom. For translations and rotations, the energy depends on the square of the appropriate quantum number:

$$\varepsilon_n^{\text{translation}} = \left(\frac{\hbar^2}{8mL^2} \right) n^2 \quad \text{and} \quad \varepsilon_{\ell}^{\text{rotation}} = \left(\frac{\hbar^2}{8\pi^2 I} \right) \ell(\ell+1).$$

In classical mechanics, masses on springs, and balls in quadratic valleys have square-law potentials. Even for more complex potentials, the lowest-order contribution to the energy is often a second-order, or square-law, dependence.

When equipartition applies and $kT/2$ is the average energy per degree of freedom, you need only count the number of degrees of freedom to compute the total average energy. Therefore you have energies of $3 \times (1/2)kT$ for translation in three dimensions and $(1/2)kT$ for every rotational degree of freedom. Heat capacities, too, are additive according to equipartition, because for any two types of equipartitional energy ε_A and ε_B ,

$$C_V = N \frac{\partial}{\partial T} (\langle \varepsilon_A \rangle + \langle \varepsilon_B \rangle) = N \left[\left(\frac{\partial \langle \varepsilon_A \rangle}{\partial T} \right) + \left(\frac{\partial \langle \varepsilon_B \rangle}{\partial T} \right) \right] = C_{V_A} + C_{V_B}.$$

The equipartition theorem says that energy in the amount of $kT/2$ partitions into each independent square-law degree of freedom, provided the temperature is high enough for the sum in Equation (11.51) to be suitably approximated by the integral in Equation (11.52).

Equipartition of Vibrations

Quantum mechanical vibrations do not obey a square-law potential. Quantum vibrational energies depend only linearly on the vibrational quantum number: $\varepsilon_v = (v + 1/2)\hbar\nu$. Therefore the equipartition theorem for vibrations is different from Equation (11.53). Using Equation (K.9) in Appendix K, the vibrational equipartition theorem is

$$\langle \varepsilon \rangle = \frac{\int_0^\infty cx e^{-cx/kT} dx}{\int_0^\infty e^{-cx/kT} dx} = kT. \quad (11.55)$$

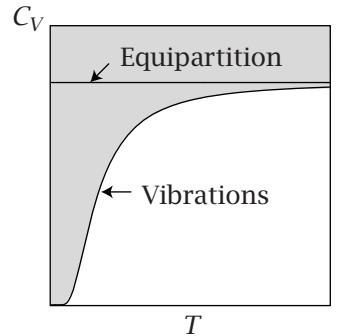
Energy in the amount of kT will partition equally into every vibrational degree of freedom at high temperatures.

Equipartition does not apply when the temperature is low, because then the sums in Equation (11.51) cannot be replaced by the integrals in Equation (11.52). At low temperatures, various degrees of freedom are ‘frozen out,’ as shown in Figures 11.11–11.13. Those degrees of freedom are unable to store energy, because the energy difference from the ground state to the first excited state is too large for thermal energies to induce the system to populate the excited state.

The Einstein Model of Solids

Like gases, solids can lose their ability to absorb energy at low temperatures. According to the equipartition equation (11.55), each vibration contributes kT to the energy. If there are N atoms in a solid, and each atom has three vibrational modes (in the x , y , and z directions), the heat capacity will be $C_V = 3Nk$, independently of temperature. This is called the law of Dulong and Petit, named after the experimentalists who first observed this behavior around 1819. But more recent experimental data, such as that shown in Figures 11.14 and 11.15, indicate that this law does not hold at low temperatures. As the temperature approaches zero, $C_V \rightarrow 0$. The Einstein model, developed in 1907, shows why. This work was among the first evidence for the quantum theory of matter.

(a) Heat Capacity



(b) Average Energy

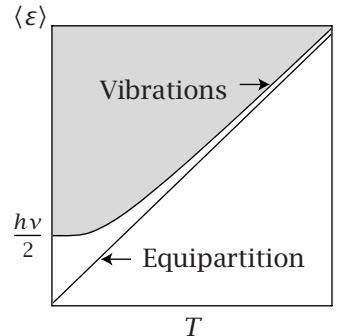


Figure 11.11 (a) Equipartition predicts heat capacities C_V that are constant and independent of temperature T . But at very low temperatures, vibrational heat capacities become small because vibrational degrees of freedom ‘freeze out.’ (b) The corresponding average particle energy $\langle \varepsilon \rangle$ versus temperature.

Figure 11.12 Dependence of the heat capacity on temperature for a linear diatomic ideal gas. Cooling a hot gas leads to freezing out of degrees of freedom: first electronic freedom, then vibrations, then rotations. Source: RS Berry, SA Rice, and J Ross, *Physical Chemistry*, 2nd edition, Oxford University Press, New York, 2000.

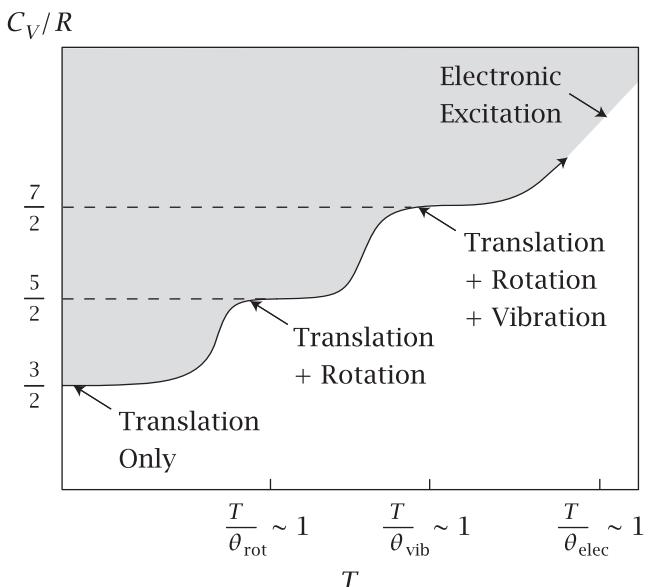


Figure 11.13 Experimental data showing the freezing out of vibrations in diatomic gases, (\bullet) CO, (\circ) N₂, (\square) Cl₂, and (\circ) O₂. Source: R Fowler and EA Guggenheim, *Statistical Thermodynamics: A Version of Statistical Mechanics for Students of Physics and Chemistry*, Cambridge University Press, Cambridge, 1939. Data are from PSH Henry, *Proc R Soc Lond A* **133**, 492 (1931); A Eucken and K Lude, *Z Physikal Chem B* **5**, 413 (1929), GG Sharratt and E Griffiths, *Proc R Soc Lond A* **147**, 292 (1934).

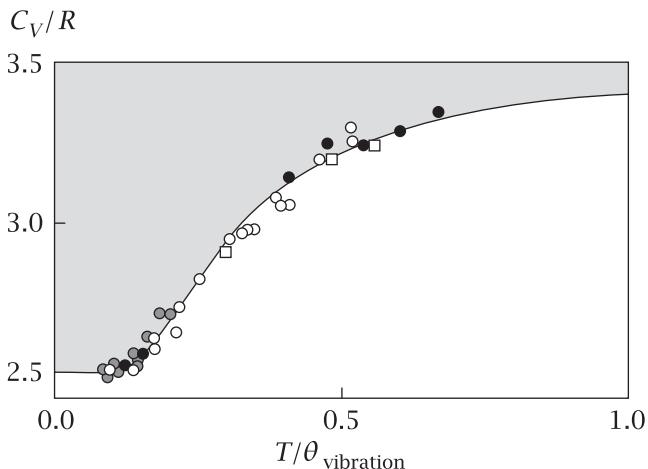
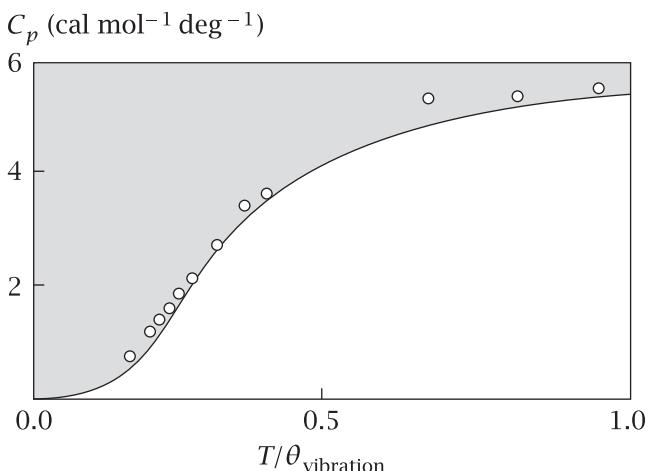


Figure 11.14 Experimental values of the heat capacity of diamond (\circ) compared with values calculated by the Einstein model (—), using the characteristic temperature $\theta_{\text{vibration}} = h\nu/k = 1320$ K. Vibrations are frozen out at low temperatures. Source: RJ Borg and GJ Dienes, *The Physical Chemistry of Solids*, Academic Press, San Diego, 1992. The data are from A Einstein, *Ann Phys* **22**, 180 (1907).



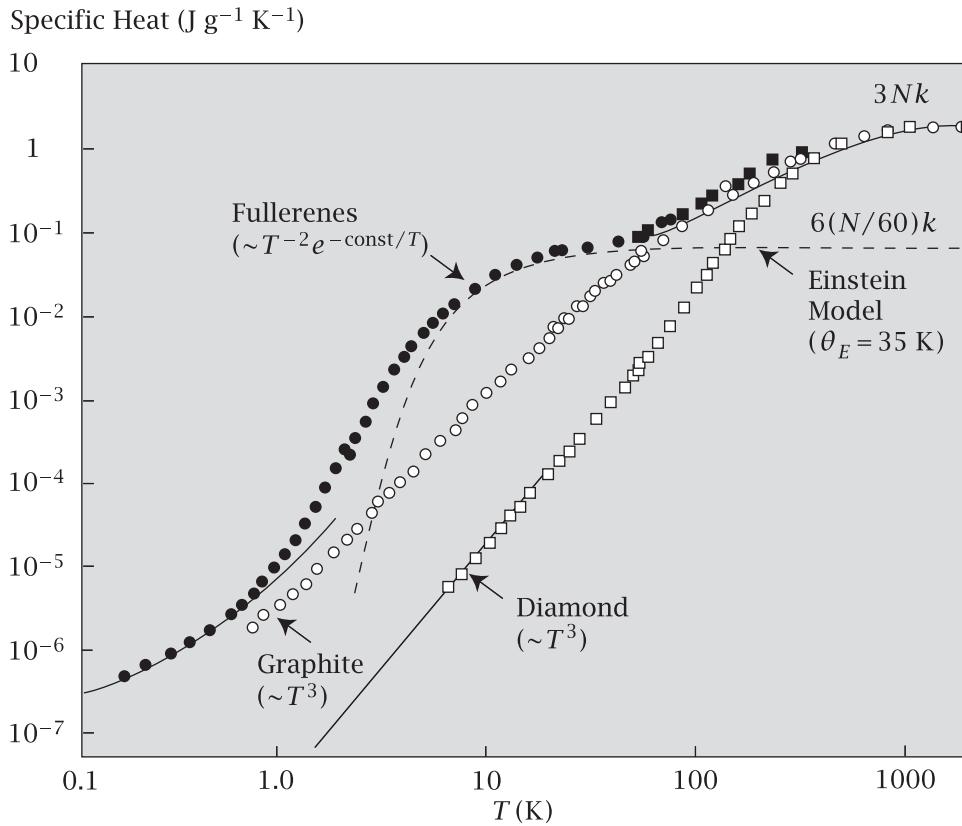


Figure 11.15 Specific heat (heat capacity per gram) versus temperature T for solids: diamond (\square), graphite (\circ), and fullerenes (\bullet). This log–log plot emphasizes the behavior at low temperatures. The Einstein model of independent oscillators ($\sim T^{-2} e^{-\alpha/T}$) characterizes fullerenes from about $T = 5$ K to 100 K, but the more sophisticated Debye model of coupled oscillators $\sim T^3$ is a better model for diamond and graphite at low temperatures. Source: JR Olson, KA Topp, and RO Pohl, *Science* **259**, 1145–1148 (1993).

Einstein assumed that a solid has $3N$ distinguishable oscillators that are independent. The vibrational partition function (Equation (11.26)) is $q = (1 - e^{-\beta h\nu})^{-1}$. The average energy per vibrational mode is given by the derivative in Equation (10.36):

$$\langle \varepsilon \rangle = -\frac{1}{q} \left(\frac{\partial q}{\partial \beta} \right) = h\nu \left(\frac{e^{-\beta h\nu}}{1 - e^{-\beta h\nu}} \right). \quad (11.56)$$

Now take $3N$ times the temperature derivative of Equation (11.56) to get the heat capacity of the solid:

$$C_V = 3N \left(\frac{\partial \langle \varepsilon \rangle}{\partial T} \right) = 3Nk \left(\frac{h\nu}{kT} \right)^2 \frac{e^{-h\nu/kT}}{(1 - e^{-h\nu/kT})^2}. \quad (11.57)$$

(To see the steps of this derivative worked out, look at Equations (10.48)–(10.50) but replace the plus sign in the denominator of Equation (10.49) with a minus sign.)

Figure 11.14 shows that Equation (11.57) captures the freezing out of vibrational modes in diamond at low temperatures. However, very precise measurements on solids show that the Einstein model is not quantitatively correct at extremely low temperatures. As $T \rightarrow 0$, Equation (11.57) predicts the temperature dependence $C_V \sim T^{-2} e^{-hv/kT}$. But very accurate experiments show that $C_V \sim T^3$ at extremely low temperatures. The T^3 dependence is predicted by the more sophisticated Debye theory of solids, which accounts for the coupling between vibrational modes [4]. Figure 11.15 shows that the Debye model predicts the behavior of graphite and diamonds, while the Einstein model works well for buckminsterfullerene, at least from about $T = 1\text{ K}$ to $T = 100\text{ K}$.

EXAMPLE 11.7 Equipartitioning in single-molecule DNA twisting. Equipartition is not limited to simple systems. Here, we apply it to DNA. Twisting a DNA molecule is like twisting an elastic rod. Twisted DNA acts like an angular version of a Hooke's law spring, with a resisting force that is linear in the twisting angle. The DNA will resist the twist with a torque $\tau = -(c_{tw}/L)\theta$, where θ is the angle of twist, L is the length of the DNA or rod, and c_{tw} is an angular spring constant.

Figure 11.16 shows an experiment for twisting a single molecule of DNA [5]. The experiment confirms the linearity of the torque with θ for a DNA molecule of length $L = 5.03\text{ }\mu\text{m}$, and gives a spring constant of $c_{tw} = 420\text{ pN nm}^2$ (Figure 11.16(b)). Since the torque is linear in θ , the twist energy follows a square law, $\varepsilon_{tw} = -\int \tau(\theta) d\theta = (c_{tw}/2L)\theta^2$. So, you can use the equipartition theorem to determine the fluctuations. From the equipartition equation (11.54), you have

$$\langle \theta^2 \rangle = \frac{LkT}{c_{tw}}. \quad (11.58)$$

Measuring the width of the distribution of angles (see Figure 11.16(c)), combined with this equipartition equation, gives you independently that $c_{tw} = 440\text{ pN nm}^2$, quite close to the result measured directly from the torque.

Summary

A major success of statistical mechanics is the ability to predict the thermodynamic properties of gases and simple solids from quantum mechanical energy levels. Monatomic gases have translational freedom, which we treated by using the particle-in-a-box model. Diatomic gases also have vibrational freedom, which we treated by using the harmonic oscillator model, and rotational freedom, for which we used the rigid-rotor model. The atoms in simple solids can be treated by the Einstein model. More complex systems can require more sophisticated treatments of coupled vibrations or internal rotations or electronic excitations. But these simple models provide a microscopic interpretation of temperature and heat capacity in Chapter 12, and they predict chemical reaction equilibria in Chapter 13, and kinetics in Chapter 19.

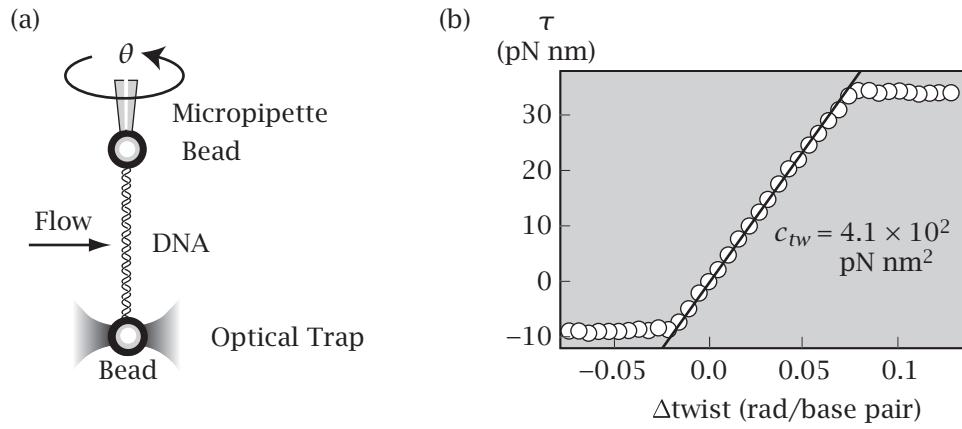


Figure 11.16 (a) A laser optical trap (bottom) grabs a bead at one end of a DNA molecule, while a micropipette attached to a bead at the other end twists the DNA against a torque supplied by a flowing fluid. (b) The measured torque depends linearly on twist angle. (c) The observed distribution of twist angles is Gaussian, indicating that the angular freedom obeys a square law. Source: redrawn from Z Bryant, MD Stone, J Gore, et al. *Nature* **424**, 338–341 (2003).

Problems

1. The heat capacity of an ideal gas. What is the heat capacity C_V of an ideal gas of argon atoms?

2. The statistical mechanics of oxygen gas. Consider a system of one mole of O_2 molecules in the gas phase at $T = 273.15\text{ K}$ in a volume $V = 22.4 \times 10^{-3}\text{ m}^3$. The molecular weight of oxygen is 32.

- Calculate the translational partition function $q_{\text{translation}}$.
- What is the translational component of the internal energy per mole?
- Calculate the constant-volume heat capacity.

3. The statistical mechanics of a basketball. Consider a basketball of mass $m = 1\text{ kg}$ in a basketball hoop. To simplify, suppose the hoop is a cubic box of volume $V = 1\text{ m}^3$.

- Calculate the lowest two energy states using the particle-in-a-box approach.
- Calculate the partition function at $T = 300\text{ K}$. Show whether quantum effects are important or not. (Assume that they are important only if q is smaller than about 10.)

4. The statistical mechanics of an electron. Calculate the two lowest energy levels for an electron in a box of volume $V = 1\text{ \AA}^3$ (this is an approximate model for the hydrogen atom). Calculate the partition function at $T = 300\text{ K}$. Are quantum effects important?

5. The translational partition function in two dimensions. When molecules adsorb on a two-dimensional surface, they have one less degree of freedom than in three dimensions. Write the two-dimensional translational partition function for an otherwise structureless particle.

6. The accessibility of rotational degrees of freedom. Diatomic ideal gases at $T = 300\text{ K}$ have rotational partition functions of approximately $q = 200$. At what temperature would q become small (say $q < 10$) so that quantum effects become important?

7. The statistical thermodynamics of harmonic oscillations. Write the internal energy, entropy, enthalpy, free energy, and pressure for a system of N independent distinguishable harmonic oscillators.

8. Orbital steering in proteins. To prove that proteins do not require ‘orbital steering,’ a process once proposed to orient a substrate with high precision before binding, T Bruice has calculated the dependence of the total energy on the rotational conformation of the hydroxymethylene group of 4-hydroxybutyric acid at $T = 300\text{ K}$. Assume that the curve in Figure 11.17 is approximately parabolic, $\varepsilon = (1/2)k_s(\alpha - \alpha_0)^2$, where α is the dihedral angle of rotation. Use the equipartition theorem.

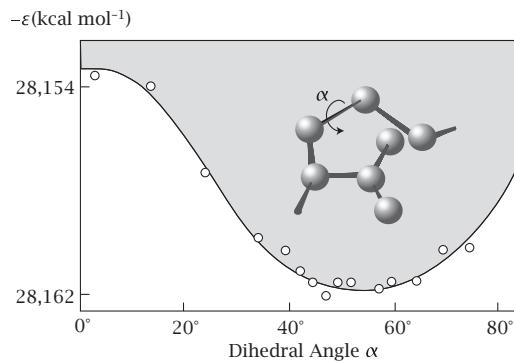
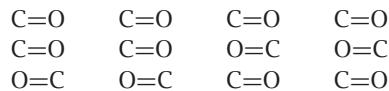


Figure 11.17 Source: TC Bruice, *Cold Spring Harbor Symposia on Quantitative Biology* **36**, 21–27 (1972).

(a) Determine the spring constant k_s .
(b) What is the average energy $\langle \varepsilon \rangle$?
(c) What is the rms dihedral angle $\langle \alpha^2 \rangle^{1/2}$?
9. The entropy of crystalline carbon monoxide at $T = 0\text{ K}$. Carbon monoxide doesn’t obey the ‘Third Law of Thermodynamics’: that is, its entropy is not zero when the temperature is zero. This is because molecules can pack in either the C=O or O=C direction in the crystalline state. For example, one packing arrangement of 12 CO molecules could be:



Calculate the partition function and the entropy of a carbon monoxide crystal per mole at $T = 0\text{ K}$.

10. Temperature-dependent quantities in statistical thermodynamics. Which quantities depend on temperature?

- Planck’s constant h .
- Partition function q .
- Energy levels ε_j .
- Average energy $\langle \varepsilon \rangle$.
- Heat capacity C_V for an ideal gas.

11. Heat capacities of liquids.

- C_V for liquid argon (at $T = 100\text{ K}$) is $18.7\text{ JK}^{-1}\text{ mol}^{-1}$. How much of this heat capacity can you rationalize on the basis of your knowledge of gases?
- C_V for liquid water at $T = 10^\circ\text{C}$ is about $75\text{ JK}^{-1}\text{ mol}^{-1}$. Assuming water has three vibrations, how much of this heat capacity can you rationalize on the basis of gases? What is responsible for the rest?

12. The entropies of CO.

- Calculate the translational entropy for carbon monoxide CO (C has mass $m = 12\text{ amu}$, O has mass $m = 16\text{ amu}$) at $T = 300\text{ K}$, $p = 1\text{ atm}$.

- (b) Calculate the rotational entropy for CO at $T = 300\text{ K}$.
The CO bond has length $R = 1.128 \times 10^{-10}\text{ m}$.

13. Conjugated polymers: why the absorption wavelength increases with chain length. Polyenes are linear double-bonded polymer molecules $(\text{C}=\text{C}-\text{C})_N$, where N is the number of $\text{C}=\text{C}-\text{C}$ monomers. Model a polyene chain as a box in which π -electrons are particles that can move freely. If there are $2N$ carbon atoms each separated by bond length $d = 1.4\text{ \AA}$, and if the ends of the box are a distance d past the end C atoms, then the length of the box is $\ell = (2N+1)d$. An energy level is occupied by two paired electrons. Suppose the N lowest levels are occupied by electrons, so the wavelength absorption of interest involves the excitation from level N to level $N+1$. Compute the absorption energy $\Delta\varepsilon = \varepsilon_{N+1} - \varepsilon_N = hc/\lambda$, where c is the speed of light and λ is the wavelength of absorbed radiation, using the particle-in-a-box model.

14. Why are conjugated bonds so stiff? As in problem 13, model polyene chain boxes of length $\ell \approx 2Nd$, where d is the average length of each carbon–carbon separation, and $2N$ is the number of carbons. There are $2N$ electrons in N energy levels, particles distributed throughout ‘boxes,’ according to the Pauli principle, with at most two electrons per level.

- (a) Compute the total energy.
- (b) Compute the total energy if the chain is ‘bent,’ that is, if there are two boxes, each of length $\ell/2$ containing N electrons each.

15. Electrons flowing in wires carry electrical current. Consider a wire 1 m long and 10^{-4} m^2 in cross-sectional area. Consider the wire to be a box, and use the particle-in-a-box model to compute the translational partition function of an electron at $T = 300\text{ K}$.

16. Fluctuations. A stable state of a thermodynamic system can be described by the free energy $G(x)$ as a function of the degree of freedom x . Suppose G obeys a square law, with spring constant k_s , $G(x)/kT = k_s x^2$, as shown in Figure 11.18.

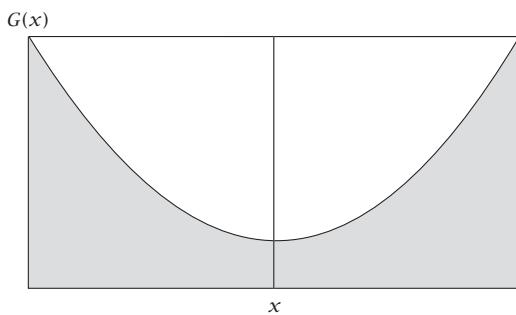


Figure 11.18

- (a) Compute the mean-square thermal fluctuations $\langle x^2 \rangle$ in terms of k_s .

- (b) Some systems have two single minima, with large spring constants k_1 , and others have a single broad minimum with small spring constant k_2 , as shown in Figures 11.19(a) and (b). For example, two-state equilibria may have two single minima, and the free energies near critical points have a single broad minimum. If $k_2 = (1/4)k_1$, what is the ratio of fluctuations $\langle x_2^2 \rangle / \langle x_1^2 \rangle$ for individual energy wells?

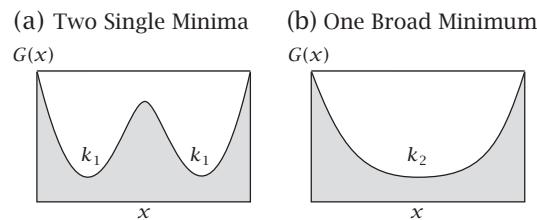


Figure 11.19

17. Heat capacity for Cl_2 . What is C_V at 800 K for Cl_2 treated as an ideal diatomic gas in the high-temperature limit?

18. Protein in a box. Consider a protein of diameter 40 \AA trapped in the pore of a chromatography column. The pore is a cubic box, 100 \AA on a side. The protein mass is 10^4 g mol^{-1} . Assume the box is otherwise empty and $T = 300\text{ K}$.

- (a) Compute the translational partition function. Are quantum effects important?
- (b) If you deuterate all the hydrogens in the protein and increase the protein mass by 10%, does the free energy increase or decrease?
- (c) By how much?

19. Vibrational partition function of iodine. Compute the value of the vibrational partition function for iodine, I_2 , at $T = 308\text{ K}$. (Hint: see Table 11.2.)

20. Electron in a quantum-dot box. An electron moving through the lattice of a semiconductor has less inertia than when it is in a gas. Assume that the effective mass of the electron is only 10% of its actual mass at rest. Calculate the translational partition function of the electron at room temperature (273 K) in a small semiconductor particle of a cubic shape with a side

- (a) 1 mm (10^{-3} m),
- (b) 100 \AA ($100 \cdot 10^{-10}\text{ m}$);
- (c) To which particle would the term ‘quantum dot,’ i.e., a system with quantum mechanical behavior, be applied, and why?

21. A protein, quantum mechanics, and the cell. Assume that a protein of mass $50,000\text{ g mol}^{-1}$ can freely

move in the cell. Approximate the cell as a cubic box $10\text{ }\mu\text{m}$ on a side.

- (a) Compute the translational partition function for the protein in the whole cell. Are quantum effects important?
- (b) The living cell, however, is very crowded with other molecules. Now assume that the protein can freely move only 5 \AA along each x , y , and z direction before it bumps into some other molecule. Compute the translation partition function and conclude whether quantum mechanical effects are important in this case.
- (c) Now assume that we deuterate all the hydrogens in the protein (replace hydrogens with deuterium atoms). If the protein mass is increased by 10%, what happens to the free energy of the modified protein? By how much does it change?

22. Electron in benzene. Consider an electron that can move freely throughout the aromatic orbitals of benzene. Model the electron as a particle in a two-dimensional box $4\text{ \AA} \times 4\text{ \AA}$.

- (a) Compute $\Delta\varepsilon$, the energy change from the ground state to the excited state, $n_x = n_y = 2$.
- (b) Compute the wavelength λ of light that would be absorbed in this transition, if $\Delta\varepsilon = hc/\lambda$, where h is Planck's constant and c is the speed of light.
- (c) Will this transition be in the visible part of the electromagnetic spectrum (i.e., is liquid benzene colored or transparent), according to this simple model?

23. Vibrations in insulin. What is the average energy stored in the vibrational degrees of freedom of one molecule of insulin, a protein with $\text{C}_{256}\text{H}_{381}\text{N}_{65}\text{O}_{76}\text{S}_6$, at room temperature?

24. Escape velocity of gases from the Moon. The escape velocity of an object to leave the Moon is 3.4 km s^{-1} . The temperature on the sunny surface of the Moon is 400 K . What is the weight of the gas that will escape the Moon with an average velocity $\langle V_x^2 \rangle^{1/2}$? What do you conclude about the Moon's atmosphere?

References

- [1] RS Berry, SA Rice, and J Ross, *Physical Chemistry*, 2nd edition, Oxford University Press, New York, 2000.
- [2] RJ Silbey, RA Alberty, and MG Bawendi, *Physical Chemistry*, 4th edition, Wiley, New York, 2005.
- [3] CE Moore, Atomic Energy States, *Natl Bureau Standards Circ* **1**, 467 (1949).

- [4] MW Zemansky and RH Dittman, *Heat and Thermodynamics*, 7th edition, McGraw Hill, New York, 1996.
- [5] Z Bryant, MD Stone, J Gore, et al., *Nature* **424**, 338–341 (2003).

Suggested Reading

PW Atkins and RS Friedman, *Molecular Quantum Mechanics*, 4th edition, Oxford University Press, Oxford, 2005. Graphic and clear discussion of quantum mechanics.

PW Atkins and J de Paula, *Physical Chemistry*, 8th edition, WH Freeman, San Francisco, 2006. Extensive elementary discussion of the principles of quantum mechanics.

HT Davis, *Statistical Mechanics of Phases, Interfaces, and Thin Films*, Wiley-VCH, New York, 1996. Advanced statistical mechanics with a broad range of applications.

AP French and EF Taylor, *An Introduction to Quantum Physics*, WW Norton, New York, 1978. Excellent elementary textbook for the principles of quantum mechanics.

JE House, *Fundamentals of Quantum Mechanics*, 2nd edition, Academic Press, San Diego, 2004. Excellent elementary yet complete discussion of the principles of quantum mechanics including the particle in a box, rotors, vibrations, and simple atoms.

DA McQuarrie, *Statistical Mechanics*, 2nd edition, University Science Books, Sausalito, CA, 2000. Complete, extensive treatment of the subject.

RJ Silbey, RA Alberty, and MG Bawendi, *Physical Chemistry*, 4th edition, Wiley, New York, 2005. Particularly good treatment of the electronic partition function.

Two classic texts on statistical mechanics:

RH Fowler and EA Guggenheim, *Statistical Thermodynamics: A Version of Statistical Mechanics for Students of Physics and Chemistry*, Cambridge University Press, Cambridge, 1939.

GS Rushbrooke, *Introduction to Statistical Mechanics*, Clarendon Press, Oxford, 1949.

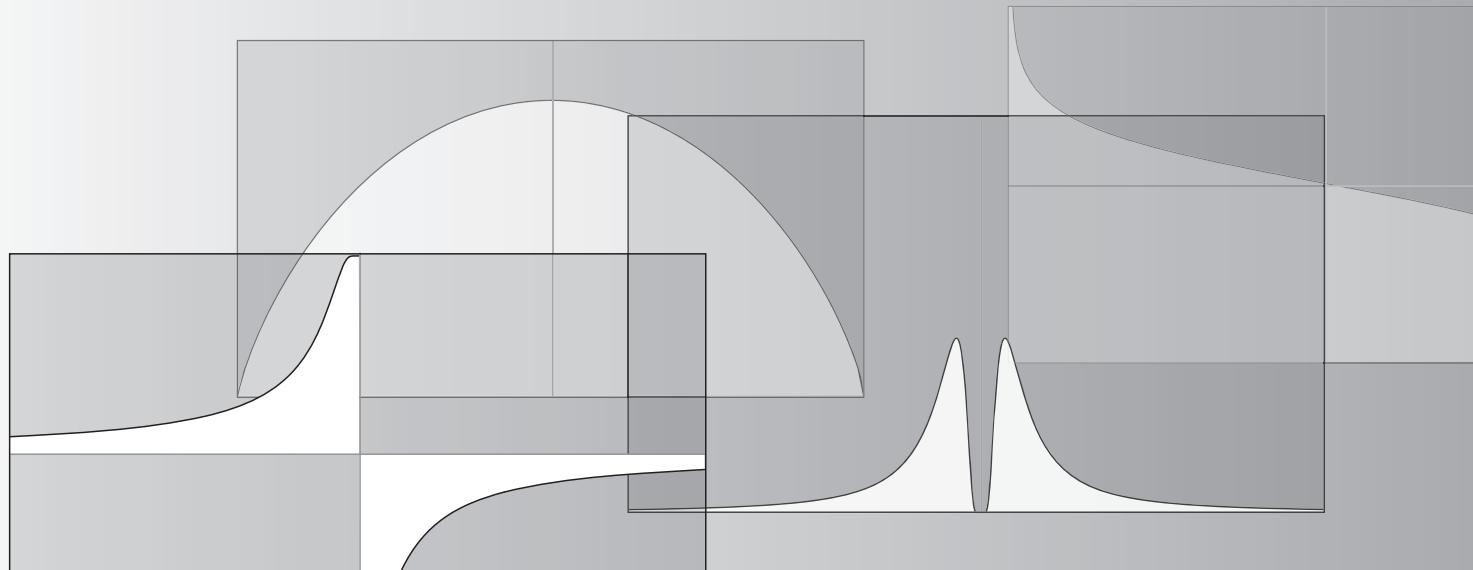
Good elementary treatments of the statistical mechanics of gases:

JH Knox, *Molecular Thermodynamics: An Introduction to Statistical Mechanics for Chemists*, Wiley, Chichester, 1978.

WGV Rosser, *An Introduction to Statistical Physics*, Halsted Press, New York, 1982.

NO Smith, *Elementary Statistical Thermodynamics: A Problems Approach*, Plenum Press, New York, 1982.

12 What Is Temperature? What Is Heat Capacity?



A Microscopic Perspective on Temperature and Heat Capacity

Temperature and heat capacity are two properties that are easy to measure, but not so easy to conceptualize. In this chapter, we develop a conceptual picture of temperature, heat capacity, and related quantities. We use two simple models: the ideal gas, and the two-state Schottky model of Chapter 10. Two-state systems can have *negative* temperatures (lower than $T = 0\text{ K}$). Negative temperatures help to illuminate the meanings of temperature and heat capacity in general.

Temperature is a property of an object. If an object has an entropy $S(U)$ that depends on its energy U , Equation (6.6) defines its temperature in terms of this dependence:

$$\frac{1}{T} = \left(\frac{\partial S}{\partial U} \right)_{V,N}.$$

The Temperature of a Two-State System

Let's use the Schottky two-state model of Chapter 10 to get $S(U)$, to see the meaning of T . The two-state model gives $S(U)$ as $S(W[n(U)])$, described below. In this model, there are two energy levels: the ground state (with energy 0) and the excited state (with energy $\varepsilon_0 > 0$). There are N particles, n of which are in

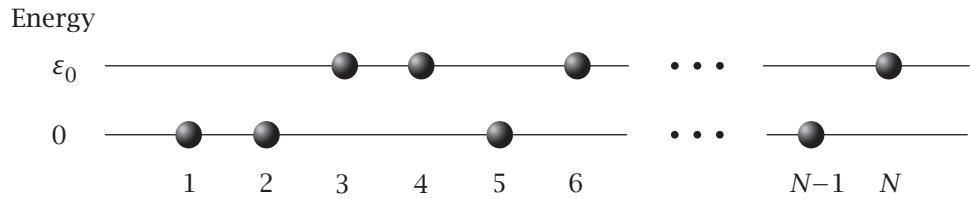


Figure 12.1 A two-state system with N particles, n of which are in the excited state. Each particle has ε_0 or zero energy.

the excited state and $N-n$ of which are in the ground state (see Figure 12.1). When energy enters the system as heat, it excites particles to move from the ground state to the excited state. The energy U of the system is proportional to the number n of molecules in the excited state, so

$$U = n\varepsilon_0 \implies n = \frac{U}{\varepsilon_0}. \quad (12.1)$$

Coin-flip statistics, Equation (1.22), gives the multiplicity of states:

$$W = \frac{N!}{n!(N-n)!},$$

so (see Example 6.1)

$$\frac{S}{k} = \ln W = -n \ln\left(\frac{n}{N}\right) - (N-n) \ln\left(\frac{N-n}{N}\right). \quad (12.2)$$

To get $S(U)$, replace n by U/ε_0 in Equation (12.2). A simple way to get $1/T$ is to express Equation (6.6) as

$$\frac{1}{T} = k \left(\frac{\partial \ln W}{\partial U} \right)_{V,N} = k \left(\frac{\partial \ln W}{\partial n} \right)_{V,N} \left(\frac{dn}{dU} \right). \quad (12.3)$$

Since $dn/dU = 1/\varepsilon_0$, Equation (12.3) becomes

$$\begin{aligned} \frac{1}{T} &= \frac{k}{\varepsilon_0} \left[-1 - \ln\left(\frac{n}{N}\right) + 1 + \ln\left(\frac{N-n}{N}\right) \right] \\ &= -\frac{k}{\varepsilon_0} \ln\left(\frac{n/N}{1-(n/N)}\right) = -\frac{k}{\varepsilon_0} \ln\left(\frac{U/N\varepsilon_0}{1-U/N\varepsilon_0}\right) \\ &= \frac{k}{\varepsilon_0} \ln\left(\frac{f_{\text{ground}}}{f_{\text{excited}}}\right), \end{aligned} \quad (12.4)$$

where $f_{\text{excited}} = (n/N)$ is the fraction of particles in the excited state and $f_{\text{ground}} = 1 - (n/N)$ is the fraction in the ground state.

Equation (12.4) shows that the temperature of a two-state system depends on the energy spacing ε_0 (the property that distinguishes one type of material from another), the number of particles N , and the total energy U through the quantity n .

Figure 12.2 shows $S(U)$ for a two-state system with $N = 3$ particles. For any value of U on the x axis, $1/T$ is the slope of the curve $S(U)$ at that point. Figure 12.2 and Equation (12.4) show how the temperature of a two-state object depends on its total energy. Let's work through the figure from left to right.

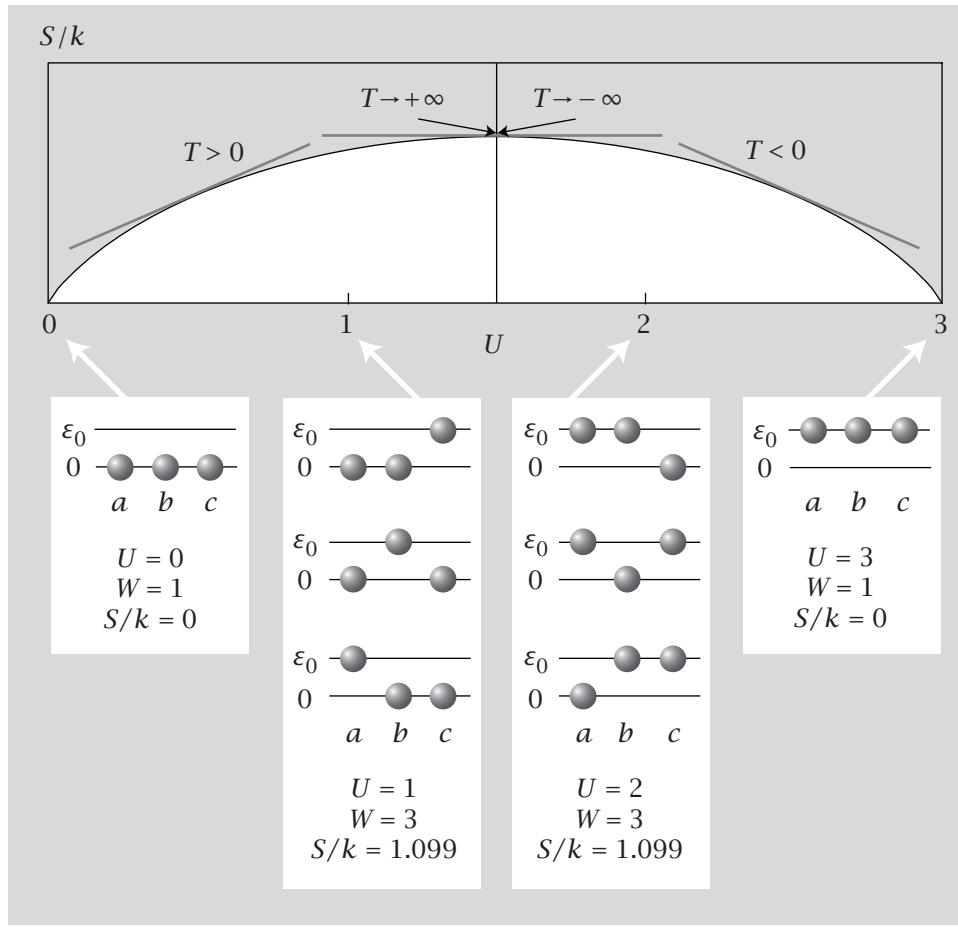


Figure 12.2 The entropy S/k of the two-state system as a function of its energy U . At low U , the multiplicity of states W is small because nearly all the particles are in the ground state. At high U , W is small because nearly all the particles are in the excited state. More states are populated for intermediate U . The slope of $S(U)$ everywhere is $1/T$, where T is the temperature. $T > 0$ on the left, and $T < 0$ on the right.

Positive temperature, $T > 0$. For a two-state system that has low energy U (see the left side of Figure 12.2), most of the particles are in the ground state. As a consequence, $f_{\text{ground}}/f_{\text{excited}} > 1$, so $\ln(f_{\text{ground}}/f_{\text{excited}}) > 0$. According to Equation (12.4), the temperature of such a system is positive. If an external source of energy (a bath) is available, the system will tend to absorb energy from it. Particles will be excited from the ground state to the excited state because of the fundamental principle (the Second Law) that systems tend toward their states of maximum entropy. The system can increase its multiplicity of states by taking up energy. Systems having positive temperature absorb energy to gain entropy.

Our simple model here has only two states. If, instead, we had considered a slightly more complex model—six states instead of two—the distribution function for dice (Figure 5.3(b)) would apply, and populations would follow the Boltzmann law. Then, instead of three particles in the ground state and none in the excited state, we would have had several particles in the ground state, fewer in the first excited state, and even fewer in the second excited state. The

state of positive temperature just means that the higher the energy level, the lower the population.

Infinite temperature, $T = \infty$. For a two-state system having intermediate energy U (middle of Figure 12.2), with equal populations in the ground and excited states, $f_{\text{ground}}/f_{\text{excited}} = 1$, and $\ln(f_{\text{ground}}/f_{\text{excited}}) = 0$. According to Equation (12.4), this implies $1/T = 0$, which means that the temperature is infinite. This is the point at which $S(U)$ is maximal. Just as with coin flips, the maximum multiplicity occurs when half the particles are in each of the two states. Two-state systems at infinite temperature cannot gain additional entropy by absorbing more energy, so they have no tendency to take up energy from a bath.

In a system of six states, infinite temperature would correspond to equal populations of all six states, according to the Boltzmann law. This distribution corresponds to that of unbiased dice (Figure 5.3(a)), in which all outcomes have the same probability.

Negative temperature, $T < 0$. For a two-state system that has high energy U (right side of Figure 12.2), most of the particles are in the excited state, so $f_{\text{ground}}/f_{\text{excited}} < 1$, and therefore $\ln(f_{\text{ground}}/f_{\text{excited}}) < 0$. It follows from Equation (12.4) that $T < 0$. In this condition, if the system were to absorb additional energy, shifting the last few particles from the ground state to the excited state would lead to a *lower* multiplicity of states. The Second Law says that systems tend toward higher entropy, so systems at negative temperatures will tend to *give off* energy, not absorb it. In that regard, a system at negative temperature is *hotter* than a system at positive temperature, since ‘hotness’ corresponds to a tendency to give up energy.

In a six-state system, negative temperatures correspond to the dice distribution shown in Figure 5.3(c). For negative temperatures, populations would *increase* exponentially with increasing energy level, according to the Boltzmann law. Higher energy levels would be more populated than lower levels.

The two-state model shows that the quantity $1/T$ represents the inclination of a system to absorb energy. When $1/T > 0$, a system tends to absorb energy. When $1/T < 0$, it tends to give off energy. When $1/T = 0$, the system has neither tendency. These inclinations result from the drive to maximize entropy.

Why is positive temperature so prevalent in nature and negative temperature so rare? Negative temperatures occur only in saturable systems that have finite numbers of energy levels. Ordinary materials have practically infinite ladders of energy levels and are not saturable, because their particles have translational, rotational, vibrational, and electronic freedom as described in Chapter 11. For such materials, energy absorption always leads to greater entropy. So most materials are energy absorbers and have only positive temperatures. Another reason they are not common is that stable negative-temperature systems require degrees of freedom that are relatively isolated from their surroundings. Otherwise they will spontaneously give off energy.

A system at negative temperature, with more excited than ground-state particles, is said to have a *population inversion*. A population inversion cannot be achieved by equilibration with a normal heat bath, because heat baths invariably have positive temperatures. Using a heat bath that has a positive temperature, the most that increasing the bath temperature can achieve is equal populations

of excited- and ground-state particles of the system. A population inversion can be caused by incident electromagnetic radiation entering the system to excite the particles into a nonequilibrium state. The excited particles will typically then spontaneously emit radiation to give off energy in order to reach a state of higher entropy. Lasers are based on the principle of population inversion.

The Relationship Between Entropy and Heat

What is the microscopic basis for the thermodynamic equation (7.8), $dS = \delta q/T$? Think of it this way. Your system has degrees of freedom (U, V, N). Hold V and N constant, so no work is done on the system. The only way to increase the energy is to take up heat, $dU = \delta q$. A system at a positive temperature will take up heat to gain entropy. The amount of entropy gained by taking up this heat will be $dS = (1/T) dU = \delta q/T$. The system takes up this energy, which is provided to it by a bath at temperature T , because increasing the excited-state population increases the system's multiplicity of states.

Now let's look at the heat capacity and its relationship to temperature.

A Graphical Procedure Shows the Steps from Fundamental Functions to Experimental Measurables

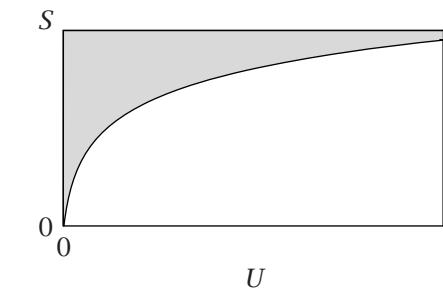
To help untangle the relationships between S, U, T , and C_V , let's follow a series of graphical transformations beginning with $S(U)$, the relationship that is closest to the principle of equilibrium, and move toward $C_V(T)$, the most directly measurable quantity (see Figure 12.3). Let's compare the various thermal properties of an ideal gas and a two-state system, for constant V and N .

First row: S versus U . For an ideal gas at constant volume, $S = (3Nk/2) \ln U$ (Figure 12.3(a)). (To see this, you can invert the logic here, start with the fourth row of this figure, $U(T) = 3NkT/2$, which results from equipartition, and work up. Note, incidentally, that you shouldn't trust this expression for small T or U , near $U = 0$ on the figures, where equipartition doesn't hold.) S is a monotonically increasing function of U because such systems have infinite ladders of energy levels. In contrast, for a two-state system (Figure 12.3(b)), $S(U)$ increases, reaches a maximum, then decreases, as described in Figure 12.2. The first row of Figure 12.3 gives the most fundamental description of either system in terms of the driving force, the entropy.

Second row: $1/T$ versus U . To get the second row of Figure 12.3 from the first row, take the slope $1/T = (\partial S/\partial U)$. This curve represents the driving force to absorb heat. For the ideal gas (Figure 12.3(a)), the driving force to take up heat is always positive: $1/T = 3Nk/(2U)$. For the two-state system (Figure 12.3(b)), the driving force changes sign if the energy U is large.

Third row: T versus U . To get the third row of Figure 12.3 from the second row, take each y -axis value y_0 and invert it to get $1/y_0$. This gives the function $T(U)$. For ideal gases and many other systems, the temperature of the system is proportional to its energy (Figure 12.3(a)): $T = 2U/(3Nk)$. But you can see

(a) Ideal Gas Model



(b) Two-State Model

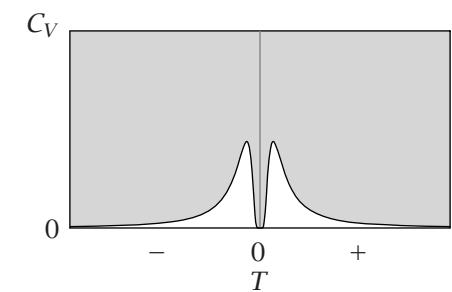
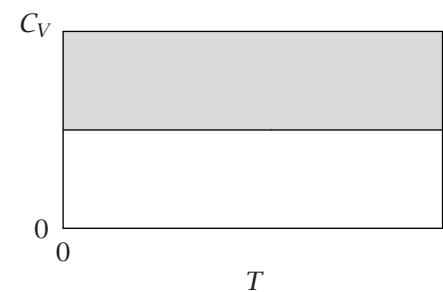
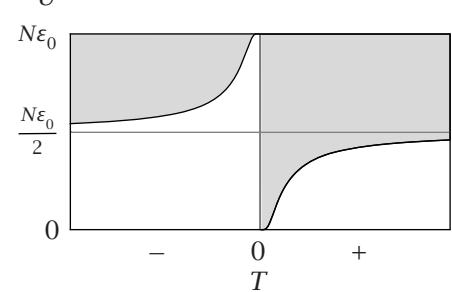
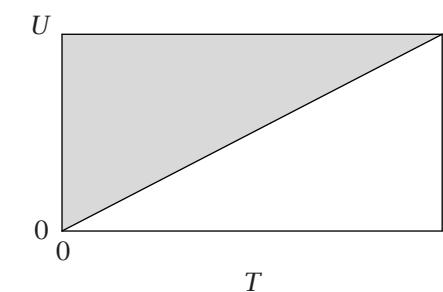
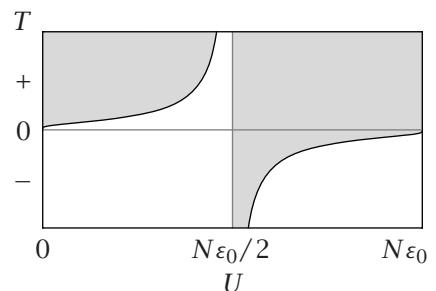
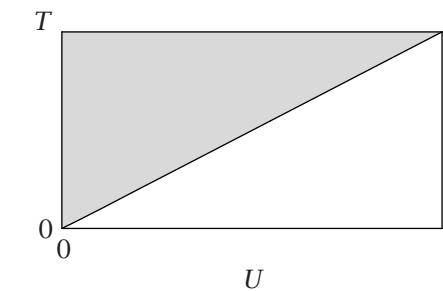
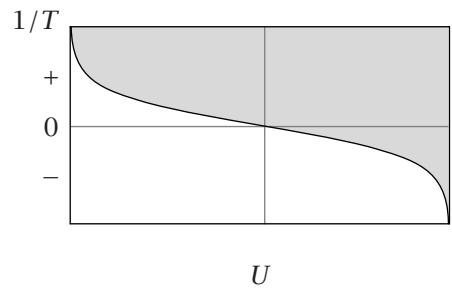
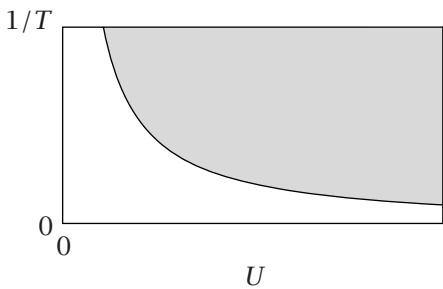
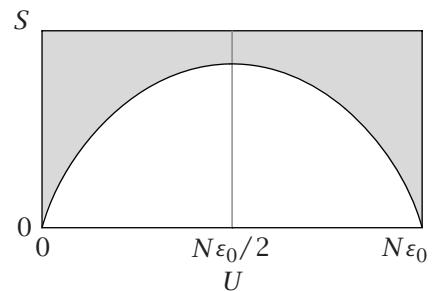


Figure 12.3 A graphical series showing the transformation from $S(U)$ (top row) to the heat capacity $C_V(T)$ (bottom row). The left column (a) describes the ideal gas, which is not saturable by energy because the ladder of energy levels is infinite. The right column (b) describes the two-state model, which is saturable.

from Figure 12.3(b) that in general the temperature need not be proportional to energy. In general, temperature is simply a measure of the relative populations of energy levels: T is inversely proportional to $\ln(f_{\text{ground}}/f_{\text{excited}})$ in the two-state model (Equation (12.4)).

Fourth Row: U versus T . To get the fourth row of Figure 12.3, interchange the x and y axes, to convert $T(U)$ into $U(T)$, a familiar function. For the ideal gas, the energy increases linearly with temperature (see Equation (11.37), $U = (3/2)NkT$). It shows why you can normally speak of ‘putting thermal energy into’ a system. The ideal gas absorbs energy in linear proportion to the temperature of the surrounding bath, at all temperatures.

On the other hand, the two-state system behaves differently: (1) it has negative temperatures and (2) at high temperatures, the two-state system cannot ‘take up’ heat from the bath, because the system is already saturated. The energy of a two-state system can never exceed $N\varepsilon_0/2$, as long as the temperature is positive. At most, the two-state system can reach equal populations of the two states. To reach higher energies requires negative temperatures.

Fifth row: C_V versus T . The fifth row of Figure 12.3 is found by taking the slope of the fourth-row figure to get the heat capacity, following the thermodynamic definition, $C_V = (\partial U / \partial T)$. C_V is the slope of the function $U(T)$. For an ideal gas (Figure 12.3(a)), the heat capacity is constant. For the two-state system (Figure 12.3(b)), there is a peak in $C_V(T)$. No thermal energy is absorbed either at high or low positive temperatures in the two-state system.

The series of steps from the top row to the bottom row of Figure 12.3 gives another example (see also Figure 9.8) of the logic of thermodynamics; it is tortuous, but useful. At the top is the most fundamental expression, $S(U)$, of the drive toward equilibrium, namely how much the entropy increases when energy is taken up. Next is $1/T$ versus U , showing the *driving force* for a system to soak up energy from its surroundings. Systems having little energy are strongly driven to take up energy from the surroundings. The next steps, $T(U)$ and $U(T)$, get us closer to the one quantity—the temperature T —over which we have experimental control. Finally, the last step is $C_V(T)$, a quantity you can measure in a calorimeter, which describes how much energy will be soaked up as you change this driving force (by changing the temperature). In short, by measuring quantities at the bottom of this figure, you can draw inferences about the fundamental quantities at the top.

What Drives Heat Exchange?

Having considered the temperature of a single object, now consider *heat exchange*, a process that involves two objects. In Chapter 6, we used money as a metaphor to illustrate that the tendency of objects in thermal contact to approach the same temperature is not necessarily a tendency to equalize energies. Now let’s use the two-state model to revisit thermal equilibrium. System A has energy U_A , energy spacing ε_A , and particle number N_A . System B has U_B , ε_B , and N_B .

At low positive temperatures, $U/N\varepsilon_0 \ll 1$, and Equation (12.4) becomes

$$\frac{1}{T} = \frac{k}{\varepsilon_0} \ln\left(\frac{N\varepsilon_0}{U}\right).$$

Thermal equilibrium between two two-state systems A and B leads to

$$\frac{1}{T_A} = \frac{1}{T_B} \implies \frac{1}{\varepsilon_A} \ln\left(\frac{N_A \varepsilon_A}{U_A}\right) = \frac{1}{\varepsilon_B} \ln\left(\frac{N_B \varepsilon_B}{U_B}\right). \quad (12.5)$$

Here are some of the implications of Equation (12.5). Sometimes heat exchange equalizes energies. If the two objects are made of the same material ($\varepsilon_A = \varepsilon_B$), and have the same total number of particles ($N_A = N_B$), the tendency to maximize entropy and to equalize temperatures is a tendency to equalize their energies U_A and U_B , according to Equation (12.5).

But thermal equilibrium isn't a tendency of high-energy objects to unload their energy into low-energy objects. A large heat bath can have more energy than a system in a small test tube, but this doesn't force the small system to take up energy from the heat bath. If two two-state objects are made of identical material, $\varepsilon_A = \varepsilon_B$, but their sizes are different, $N_A \neq N_B$, Equation (12.5) tells us that thermal equilibrium will occur when $U_A/N_A = U_B/N_B$, not when $U_A = U_B$. At equilibrium, the objects will share energies in proportion to their sizes. In general, equilibrium is defined by the equality of *intensive* properties (U/N in this case), not extensive properties.

Our treatment of heat capacity in this chapter has been intended to give simple insights. We end this chapter with an additional perspective about the heat capacity, as a measure of the fluctuations in the energy. To see this, we return to the quantum mechanical approach of Chapter 11.

The Heat Capacity Is Proportional to the Energy Fluctuations in a System

Hold a system at constant temperature with a heat bath. Although the temperature is fixed, the energy of the system will undergo fluctuations, usually very small, as heat flows to and from the bath. The heat capacity is a measure of the magnitude of these fluctuations. To derive this relationship, we switch our attention from microstates to energy levels (see Chapter 10), so we need the density of states.

The Density of States

There are often many different ways in which a system can have a given energy. For example, consider a particle in a cubic box of dimensions $a \times a \times a$. Equation (11.17) gives the translational energy as

$$\varepsilon = \varepsilon_{n_x, n_y, n_z} = \frac{\hbar^2}{8ma^2} (n_x^2 + n_y^2 + n_z^2), \quad (12.6)$$

for $n_x, n_y, n_z = 1, 2, 3, \dots$

We want to count the number of states $W(\varepsilon)$ that have an energy between ε and $\varepsilon + \Delta\varepsilon$. You can do this geometrically. Note that n_x, n_y, n_z are integers $1, 2, 3, \dots$ that define a grid of points along the x, y , and z axes in three-dimensional space. Put a dot on each grid point as shown in Figure 12.4. Now, count all the points within a sphere of radius R . The number of integers in a

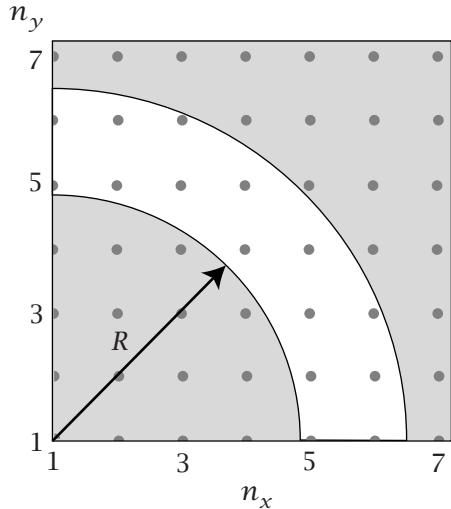


Figure 12.4 In this two-dimensional representation of translational quantum numbers n_x and n_y , the ring or shell of radius R includes all states having approximately equal values of $n_x^2 + n_y^2$ (●). The ring includes the states (points) with energies between ε and $\varepsilon + \Delta\varepsilon$. $W(\varepsilon)$ is the number of points in this shell.

sphere is proportional to the volume, $4\pi R^3/3$. Different combinations of integers n_x , n_y , and n_z can all lead to the same value of the sum $R^2 = n_x^2 + n_y^2 + n_z^2$. Substituting R^2 into Equation (12.6) gives $\varepsilon = (hR)^2/(8ma^2)$. n_x , n_y , and n_z are all positive quantities only for one octant (1/8) of the sphere. So the number of positive integer points M is

$$M = \frac{1}{8} \left(\frac{4\pi R^3}{3} \right) = \frac{\pi}{6} \left(\frac{8m\varepsilon}{h^2} \right)^{3/2} V, \quad (12.7)$$

where $V = a^3$. To get $W(\varepsilon)\Delta\varepsilon$, the number of positive integer points in the shell between energies ε and $\varepsilon + \Delta\varepsilon$, take the derivative:

$$\begin{aligned} W(\varepsilon)\Delta\varepsilon &= M(\varepsilon + \Delta\varepsilon) - M(\varepsilon) \approx \frac{dM}{d\varepsilon}\Delta\varepsilon \\ &= \left(\frac{\pi}{4} \right) \left(\frac{8m}{h^2} \right)^{3/2} V \varepsilon^{1/2} \Delta\varepsilon. \end{aligned} \quad (12.8)$$

$W(\varepsilon)$ is the number of states of a single particle in a box that have energy between ε and $\varepsilon + \Delta\varepsilon$. The following example shows that these degeneracies are very large numbers.

EXAMPLE 12.1 The degeneracy of states for the particle in a box. Compute $W(\varepsilon)$ for an atom of argon, $m = 40 \text{ g mol}^{-1} = 6.64 \times 10^{-26} \text{ kg per atom}$. At $T = 300 \text{ K}$, use $\varepsilon = 3kT/2 = 3(1.38 \times 10^{-23} \text{ J K}^{-1})(300 \text{ K})/2 = 6.21 \times 10^{-21} \text{ J}$. Take $a = 1 \text{ cm}$ and take a 1% band of energy, $\Delta\varepsilon = 0.01\varepsilon$. Equation (12.8) shows that there are a large number of microstates in each energy level:

$$\begin{aligned} W(\varepsilon)\Delta\varepsilon &= \left(\frac{\pi}{4} \right) \left(\frac{8 \times 6.64 \times 10^{-26} \text{ kg per atom}}{(6.63 \times 10^{-34} \text{ J s})^2} \right)^{3/2} \times (10^{-2} \text{ m})^3 \\ &\times (6.21 \times 10^{-21} \text{ J})^{1/2} (0.01) = 5.11 \times 10^{24} \text{ states.} \end{aligned}$$

So far we have focused mostly on the probabilities $p(n_x, n_y, n_z) = q^{-1}e^{-\beta\varepsilon}$ that a particle is in a particular *microstate*, say (n_x, n_y, n_z) for the three-dimensional particle in a box. But now we focus instead on the probability $p(\varepsilon) = q^{-1}W(\varepsilon)e^{-\beta\varepsilon}$ that a system is in a particular *energy level*.

As a check on Equation (12.8), note that the partition function must be the same whether you sum over states or energy levels:

$$q = \sum_{\text{states}} e^{-\beta\varepsilon} = \sum_{\text{energy levels}} W(\varepsilon)e^{-\beta\varepsilon}. \quad (12.9)$$

Substituting Equation (12.8) for $W(\varepsilon)$ into the right-hand side of Equation (12.9) and converting to an integral gives

$$q = \int_0^\infty \left(\frac{\pi}{4}\right) \left(\frac{8m}{h^2}\right)^{3/2} V \varepsilon^{1/2} e^{-\beta\varepsilon} d\varepsilon. \quad (12.10)$$

This integral is of the form $\int_0^\infty x^{1/2} e^{-\beta x} dx$, where $x = \varepsilon$. Using $n = 1/2$ in Equation (K.9) of Appendix K gives this integral as $\Gamma(3/2)/\beta^{3/2}$, where $\Gamma(3/2) = \sqrt{\pi}/2$. Replace the integral with $\sqrt{\pi}(kT)^{3/2}/2$ in Equation (12.10) to get

$$q = \left(\frac{\pi}{4}\right) \left(\frac{8m}{h^2}\right)^{3/2} V \left(\frac{1}{2}\right) (\sqrt{\pi}) (kT)^{3/2} = \left(\frac{2\pi mkT}{h^2}\right)^{3/2} V, \quad (12.11)$$

which is the same result as you get from Equation (11.18), summing over states.

Energy Fluctuations

What is the probability that a particle in a box has a particular energy? The probability of energy ε is $p(\varepsilon) \sim W(\varepsilon)e^{-\beta\varepsilon}$. $p(\varepsilon)$ is a highly peaked function because the factor $W(\varepsilon) \sim \varepsilon^{1/2}$ grows with ε (Equation (12.8)), while the factor $e^{-\beta\varepsilon}$ decreases with ε . If instead of one particle, you have N independent particles in a box with total energy E , $p(E)$ is still a peaked function because $W(E) \sim E^{3N/2-1}$ also grows with E , while $e^{-\beta E}$ decreases (see Figure 12.5).

Now, let's generalize beyond translations, particles in a box, and independent particles. The probability that any arbitrary system is in a particular energy level is $p(E) = Q^{-1}W(E)e^{-\beta E}$. We show that

$$p(E) = Q^{-1}W(E)e^{-\beta E} \quad (12.12)$$

is a highly peaked function with a width that is determined by the heat capacity.

Let E represent one particular energy level of the system and U represent the equilibrium value of the energy, the quantity that appears in thermodynamic expressions. Near the peak, the function $p(E)$ is approximately Gaussian, and its width is proportional to C_V . To see this, express $\ln p(E)$ as a Taylor series (see Appendix A) around the mean value $\langle E \rangle = U$:

$$\begin{aligned} \ln p(E) &= \ln p(U) + \left(\frac{\partial \ln p(E)}{\partial E}\right)_{E=U} (E-U) \\ &\quad + \frac{1}{2} \left(\frac{\partial^2 \ln p(E)}{\partial E^2}\right)_{E=U} (E-U)^2 + \dots \end{aligned} \quad (12.13)$$

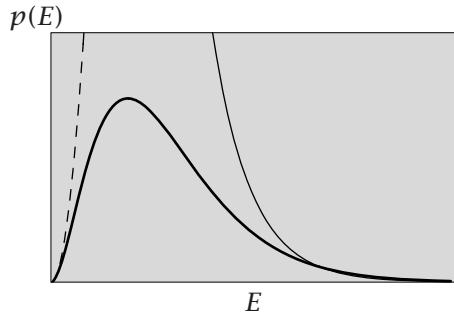


Figure 12.5 The degeneracy of energy states $W(E)$ (---) increases with energy, while the Boltzmann factor $e^{-\beta E}$ (—) decreases. The product is a peaked Gaussian probability distribution function $p(E)$ (—). This figure is computed for translations of N independent particles, $W(E) \sim E^{(3N/2-1)}$ for N independent particles.

The peak of this function defines the equilibrium state: the fluctuations are deviations from it. We are interested in the variations away from equilibrium. At the peak, $E = U$, and the entropy $S(E)$ equals its equilibrium value $S(U)$. At the peak, the temperature $T(E)$ equals its equilibrium value T_0 , which is the temperature of the bath.

In evaluating Equation (12.13), you can use Equation (12.12) to get $\ln p(E) = \ln[Q^{-1}W(E)] - \beta E$. Then take its first and second derivatives. Q is not a function of E (it is a sum over E), so Q does not contribute to such derivatives. Use the expressions $S(E) = k \ln W(E)$ and $(\partial S(E)/\partial E) = 1/T(E)$ to evaluate the terms in Equation (12.13):

$$\begin{aligned} \left(\frac{\partial \ln p(E)}{\partial E} \right) &= \left(\frac{\partial \ln W(E)}{\partial E} \right) - \beta \\ &= \left(\frac{\partial(S/k)}{\partial E} \right) - \beta = \frac{1}{kT(E)} - \frac{1}{kT_0}. \end{aligned} \quad (12.14)$$

At the peak, this first derivative of the Taylor series equals zero, so $T(E) = T_0$. To get the second derivative for the Taylor series, take another derivative of the last term in Equation (12.14) using the relation $(\partial E/\partial T) = C_V$ to get

$$\left(\frac{\partial^2 \ln p(E)}{\partial E^2} \right) = - \left(\frac{1}{kT^2} \right) \left(\frac{\partial T}{\partial E} \right)_{E=U} = - \left(\frac{1}{kT_0^2 C_V} \right). \quad (12.15)$$

Substituting Equation (12.15) into Equation (12.13) and exponentiating shows that $p(E)$ is a Gaussian function (see Equation (1.50)).

$$\begin{aligned} p(E) &= p(U) \exp \left[-\frac{(E-U)^2}{2kT_0^2 C_V} \right] \\ &= \exp[U - T_0 S(U)] \exp \left[-\frac{(E-U)^2}{(2kT_0^2 C_V)} \right]. \end{aligned} \quad (12.16)$$

Equation (12.16) is based on using $S = k \ln W$ and Equation (12.12). Comparing Equation (12.16) with the Gaussian equation (1.50) and the definition of the variance in Equation (1.42) shows that the width of the energy distribution is characterized by its variance σ^2 :

$$\sigma^2 = \langle (E-U)^2 \rangle = \langle E^2 \rangle - U^2 = kT_0^2 C_V. \quad (12.17)$$

Equation (12.17) shows that you can determine the magnitude of the energy fluctuations if you know C_V . Example 12.2 shows that the width of this Gaussian function is usually exceedingly narrow. Sometimes it is useful to express the

variance instead in terms of the partition function (see Equation (C.8)):

$$\sigma^2 = \frac{d^2 \ln Q}{d\beta^2} = \frac{d}{d\beta} \left(\frac{1}{Q} \frac{dQ}{d\beta} \right) = \frac{Q''}{Q} - \frac{Q'}{Q^2}, \quad (12.18)$$

where each ' indicates one derivative of Q with respect to β .

EXAMPLE 12.2 The width of the energy distribution is usually very narrow. A useful dimensionless measure of the width of the distribution in Equation (12.16) is the ratio of the standard deviation to the mean value, σ/U . For an ideal gas, $U = (3/2)NkT$ and $C_V = (3/2)Nk$, so, using Equation (12.17), you have

$$\frac{\sigma}{U} = \frac{(kT^2 C_V)^{1/2}}{\frac{3}{2} N k T} = \frac{\left[\frac{3}{2} N (kT)^2 \right]^{1/2}}{\frac{3}{2} N k T} = \left(\frac{3}{2} N \right)^{-1/2}.$$

If the number of particles in a system is of order 10^{23} , then σ/U is of order 10^{-12} , implying that $p(E)$ has an extremely sharp peak, and the fluctuations are usually exceedingly small. There are exceptions, however. In Chapter 25, we will discuss phase transitions and conditions called critical points where the heat capacity and the fluctuations both become large.

The fact that $p(E)$ is so sharply peaked answers an important question. How do you know whether a system should be considered to be at constant energy (U, V, N) and should be treated by the microcanonical ensemble or whether it should be considered to be at constant temperature (T, V, N) and treated by the canonical ensemble? If $p(E)$ is very sharply peaked, and if you are interested in thermodynamic average properties such as U , then fixing the temperature is tantamount to fixing the energy U to within a very narrow range. For averaged properties such as U , you would make very little error by using the microcanonical ensemble and assuming that the preponderance of accessible states have the same energy. Although the choice of ensemble doesn't matter for averaged properties, it does matter when you are interested in fluctuations. Fluctuations in U where T is fixed (the canonical ensemble) are given by Equation (12.17). There are no fluctuations in U in the microcanonical ensemble.

Summary

$1/T$ is the driving force for the uptake of energy. For two bodies A and B in thermal contact, energy will flow from one to the other until the driving forces are equal: $1/T_A = 1/T_B$. Setting the bath temperature T_B fixes the system temperature $T_A = T_B$ at equilibrium. C_V describes how much energy uptake results from changing the bath temperature, and is a measure of the magnitude of energy fluctuations. But the amount of energy taken up for a given change in temperature depends on the type of system. A two-state system is very different from an equipartitioned system like an ideal gas. The two-state system has a finite number of energy levels, which can be saturated. In such systems, increasing the bath temperature may not substantially increase the energy of the system. In contrast, the ideal gas does not saturate, and tends to absorb more energy at any temperature. Heat capacities also represent the energy fluctuations in a system at constant temperature.

Problems

1. The heat capacity peaks of phase transitions. The peak heat capacity in Figure 12.6 shows that helium gas adsorbed on graphite undergoes a transition from an ordered state at low temperature to a disordered state at high temperature. Use Figure 12.6 to estimate the energy associated with the transition.

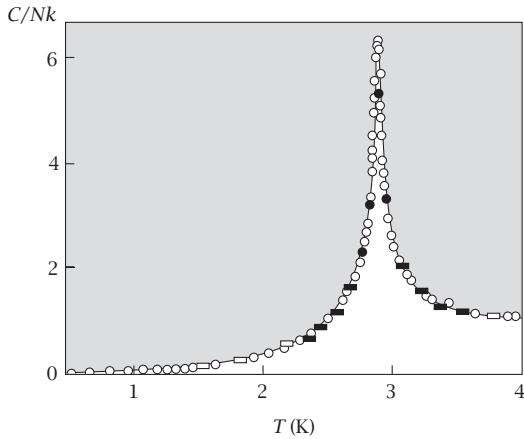


Figure 12.6 Specific heat for helium on graphite in the vicinity of the lattice-gas ordering transition. Source: HJ Kreuzer and ZW Gortel, *Physisorption Kinetics*, Springer-Verlag, Heidelberg, 1986. Data are from RL Elgin and DL Goodstein, *Phys Rev A* **9**, 2657–2675 (1974).

2. A two-state model of a hydrogen bond. Suppose a hydrogen bond in water has an energy of about 2 kcal mol⁻¹. Suppose a ‘made’ bond is the ground state in a two-state model and a ‘broken’ bond is the excited state. At T = 300 K, what fraction of hydrogen bonds are broken?

3. A random energy model of glasses. Glasses are materials that are disordered—and not crystalline—at low temperatures. Here’s a simple model [1]. Consider a system that has a Gaussian distribution of energies E, according to Equation (12.16):

$$p(E) = p_0 \exp\left[-\frac{(E - \langle E \rangle)^2}{2\Delta E^2}\right],$$

where $\langle E \rangle$ is the average energy and ΔE characterizes the magnitude of the fluctuations, that is, the width of the distribution.

- (a) Derive an expression showing that the entropy $S(E)$ is an inverted parabola:

$$S(E) = S_0 - \frac{k(E - \langle E \rangle)^2}{2\Delta E^2}.$$

- (b) An *entropy catastrophe* happens (left side of the curve in Figure 12.7) where $S = 0$. For any physical system, the minimum number of accessible states

is $W = 1$, so $S = k \ln W$ implies that the minimum entropy is $S = 0$. At the point of the entropy catastrophe, the system has no states that are accessible below an energy $E = E_0$. Compute E_0 .

- (c) The *glass transition temperature* T_g is the temperature of the entropy catastrophe. Compute T_g from this model.

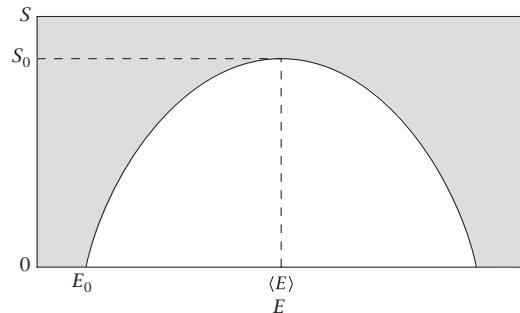


Figure 12.7

4. Fluctuations in enthalpy. The mean-square fluctuations in enthalpy for processes at constant pressure are given by an expression similar to Equation (12.17) for processes at constant pressure:

$$\langle \delta H^2 \rangle = \langle (H - \langle H \rangle)^2 \rangle = kT^2 C_p.$$

What is the root-mean-square enthalpy fluctuation $\langle \delta H^2 \rangle^{1/2}$ for water at T = 300 K?

5. Oxygen vibrations. Oxygen molecules have a vibrational wavenumber of 1580 cm⁻¹ (see Example 11.3). If the relative populations are $f_{\text{ground}}/f_{\text{excited}} = 100$ in the Schottky model, what is the temperature of the system?

6. Modeling a population inversion. Population inversion, in which more particles of a system are in an excited state than a ground state, is used to produce laser action. Consider a system of atoms at 300 K with three energy levels: 0 kcal mol⁻¹, 0.5 kcal mol⁻¹, and 1.0 kcal mol⁻¹.

- (a) Compute the probabilities p_1^* , p_2^* , and p_3^* that an atom is in each energy level.
 (b) What is the average energy $\langle \varepsilon \rangle$ of the system? What is the entropy (in cal mol⁻¹ K⁻¹)? Now suppose you excite the system to cause a population inversion, resulting in a new population distribution p^{**} as follows: $p_1^{**} = p_3^*$, $p_2^{**} = p_2^*$, $p_3^{**} = p_1^*$.
 (c) What is the average energy $\langle \varepsilon \rangle$ of the system after the population inversion? What is the entropy (in cal mol⁻¹ K⁻¹) of the system after the population inversion?
 (d) What is the temperature of the system after the population inversion?

7. Schottky model for a two-level transition in hydrogen molecules. Hydrogen molecules have a vibrational

wavenumber of 4300 cm^{-1} . If 0.5% of the molecules are in the excited state, what is the temperature of the system?

8. Finding the C_V of a gas. A gas is placed in an airtight container. A piston decreases the volume by 10% in an adiabatic process, and the temperature is observed to increase by 10%. What is the constant-volume heat capacity of the gas?

References

- [1] JN Onuchic, Z Luthey-Schulten, and PG Wolynes, *Ann Rev Phys Chem* **48**, 545–600 (1997).

Suggested Reading

P Hakonen and OV Lounasmaa, *Science* **265**, 1821–1825 (1994). Describes experiments on spins in metals showing negative temperatures.

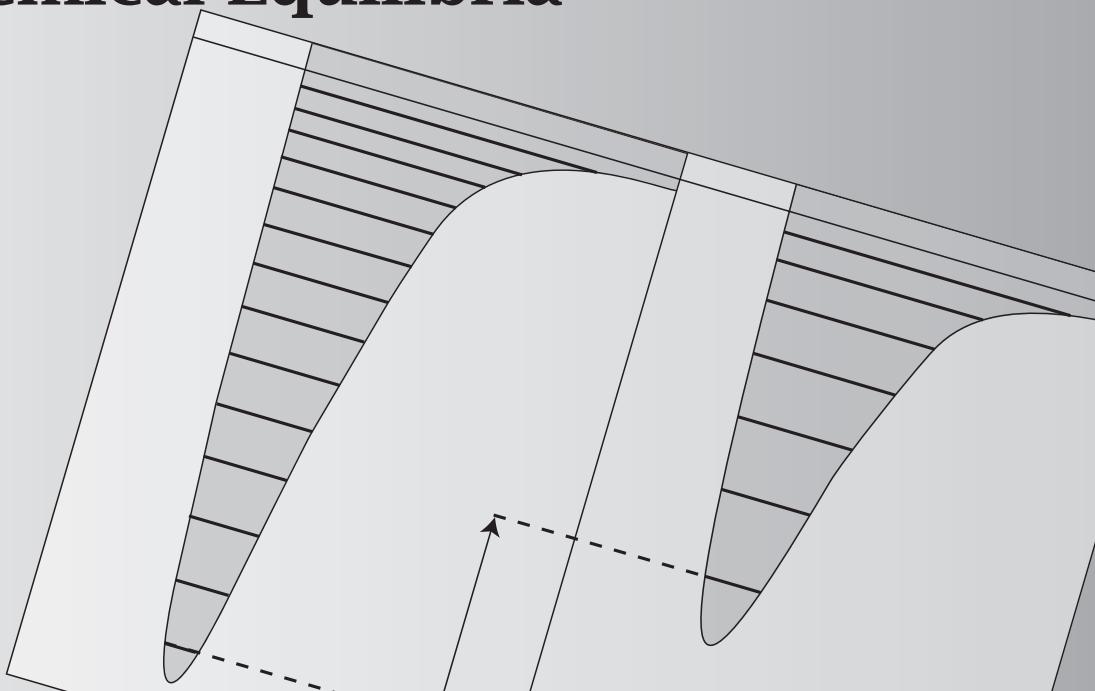
TL Hill, *An Introduction to Statistical Thermodynamics*, Addison-Wesley, Reading, MA, 1960 (reprinted by Dover Publications, New York, 1986). The classic text for discussion of ensembles and fluctuations, and many applications of statistical mechanics.

C Kittel and H Kroemer, *Thermal Physics*, 2nd edition, WH Freeman, New York, 1980. Outstanding general text on statistical thermodynamics. Good discussion of negative temperatures.

RK Pathria, *Statistical Mechanics*, 2nd edition, Butterworth-Heinemann, Oxford, 1996. Excellent advanced treatment of statistical mechanics, including a good description of fluctuations.

JR Waldram, *The Theory of Thermodynamics*, Cambridge University Press, Cambridge, 1985. Excellent discussion of fluctuations.

13 Chemical Equilibria

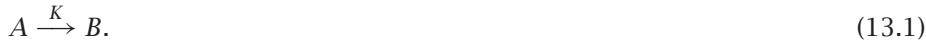


Simple Chemical Equilibria Can Be Predicted from Atomic Structures

A major goal in chemistry is to predict the equilibria of chemical reactions, the relative amounts of the reactants and products, from their atomic masses, bond lengths, moments of inertia, and other structural properties. In this chapter, we focus on chemical equilibria in the gas phase, to illustrate the principles, since they are simpler than reactions in liquid phases. In Chapter 16, we will consider solvation effects.

The Condition for Chemical Equilibrium

First consider the simplest equilibrium between two states, *A* and *B*:



Two-state equilibria include chemical isomerization, the folding of a biopolymer from an open to a compact state, the binding of a ligand to a surface or a molecule, the condensation of a vapor to a liquid, and the freezing of a liquid to a solid. The *equilibrium constant K* is the ratio of the numbers or concentrations of the two species at equilibrium. In this chapter, we suppose that you are able to measure the concentrations of reactants and products. Our goal is to predict those equilibria from the molecular structures of the products and reactants and from temperature and pressure.

To make it unambiguous whether K is the ratio of the amount of A divided by the amount of B or B divided by A , the arrow in Equation (13.1) has a direction. It points to the *final state* from the *initial state*. The terms ‘initial’ and ‘final’ have no implication about time or kinetics. They imply only which is the numerator and which is the denominator in the equilibrium constant K . With the direction shown in Equation (13.1), K is the ratio B/A . You can point the arrow in either direction. Once you choose the direction, the sign of every thermodynamic quantity is determined.

The quantity that predicts chemical equilibria is the chemical potential. A chemical potential represents the tendency of a molecule or particle to move or convert or escape from the state it is in. For equilibria at fixed temperature and pressure, the appropriate extremum function is the Gibbs free energy: $dG = -S \, dT + V \, dp + \mu_A \, dN_A + \mu_B \, dN_B$, where N_A and N_B are the numbers of particles in the two states and μ_A and μ_B are their chemical potentials. At constant T and p , the condition for equilibrium is

$$dG = \mu_A \, dN_A + \mu_B \, dN_B = 0. \quad (13.2)$$

If every molecule is in either state A or B , the total number of molecules N_{total} is constant:

$$N_A + N_B = N_{\text{total}} = \text{constant} \implies dN_A + dN_B = 0. \quad (13.3)$$

A molecule lost as A is converted to B , so $dN_A = -dN_B$, and the condition for equilibrium, Equation (13.2), can be written in terms of dN_A :

$$(\mu_A - \mu_B) \, dN_A = 0. \quad (13.4)$$

Equation (13.4) must hold for any nonzero variation dN_A , so the condition for equilibrium is that the chemical potentials of species A and B are the same:

$$\mu_A = \mu_B. \quad (13.5)$$

To proceed further, we need to relate each chemical potential to its partition function.

Partition Functions for Chemical Reactions

First, express the partition function, Equation (10.19), in terms of q' , where the added ‘ \prime ’ indicates that we include the ground state explicitly:

$$q' = \sum_{j=0}^t e^{-\varepsilon_j/kT} = e^{-\varepsilon_0/kT} + e^{-\varepsilon_1/kT} + \dots + e^{-\varepsilon_t/kT}. \quad (13.6)$$

We need to retain the ground-state terms for chemical equilibria because the species A and B do not necessarily have the same ground states. Now we redefine q , without the ‘ \prime ’, as a reduced partition function with the ground-state term

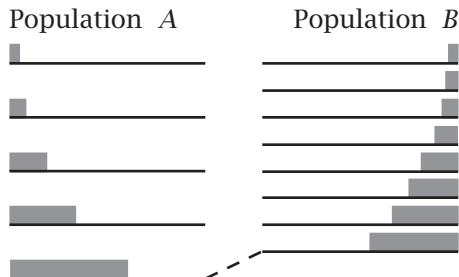


Figure 13.1 Hypothetical energy ladders for the two-state equilibrium $A \rightarrow B$. Two factors contribute to the equilibrium populations of A and B : low energies and high densities of states. A is shown here to have the lowest-energy ground state. B has the higher density of states.

factored out:

$$\begin{aligned} q &= e^{\varepsilon_0/kT} q' \\ &= 1 + e^{-(\varepsilon_1 - \varepsilon_0)/kT} + e^{-(\varepsilon_2 - \varepsilon_0)/kT} + \dots + e^{-(\varepsilon_t - \varepsilon_0)/kT}. \end{aligned} \quad (13.7)$$

In terms of q' , the chemical potential μ_A for species A is given by Equation (11.47):

$$\mu_A = -kT \ln \left(\frac{q'_A}{N_A} \right). \quad (13.8)$$

Similarly for B ,

$$\mu_B = -kT \ln \left(\frac{q'_B}{N_B} \right). \quad (13.9)$$

At equilibrium, all the energy levels of both species A and B are accessible to any molecule. A given molecule may be species A for a while at one energy level, then it may change to species B , at some other energy level. At equilibrium, the individual molecules of species A and B will be distributed according to the Boltzmann distribution, with the largest number of molecules in the lowest energy state. The number of molecules populating a given state is determined only by the energy, and not by whether the molecule is in the form of A or B .

Figure 13.1 shows a pair of possible energy ladders for A and B , illustrating a case in which the state of lowest energy (and therefore highest population) is the ground state of A . This does not mean that more particles will be A 's than B 's. The relative numbers of A 's and B 's is a sum of Boltzmann factors over all the energy levels. If the energy level spacings are closer for B than for A , B may be the species with the greater total population.

Substituting Equations (13.9) and (13.8) into the equilibrium condition $\mu_A = \mu_B$ (Equation (13.5)) gives $q'_B/N_B = q'_A/N_A$. Defining the **equilibrium constant** K as the ratio N_B/N_A gives

$$K = \frac{N_B}{N_A} = \left(\frac{q'_B}{q'_A} \right) = \left(\frac{q_B}{q_A} \right) e^{-(\varepsilon_{0B} - \varepsilon_{0A})/kT}. \quad (13.10)$$

Equation (13.10) gives a way to compute K from the atomic properties of A and B through their partition functions. Equation (13.10) takes no account of the interactions of one molecule with another, and therefore it applies only to isolated particles such as those in the gas phase.

More Complex Equilibria

Let's generalize to the reaction



where a , b , and c indicate the stoichiometries of species A , B , and C . What is the appropriate way to define an equilibrium constant for this reaction? At constant T and p , the condition for equilibrium is

$$dG = \mu_A dN_A + \mu_B dN_B + \mu_C dN_C = 0. \quad (13.12)$$

Now, the stoichiometry defines some constraint relationships among the quantities N_A , N_B , and N_C . These constraints are most simply expressed in terms of a *progress variable* or *extent of reaction* ξ . Every unit reaction produces c product molecules and uses up a molecules of species A and b molecules of species B . So, you have

$$dN_C = c d\xi,$$

$$dN_A = -a d\xi,$$

$$dN_B = -b d\xi. \quad (13.13)$$

These relations are illustrated in Example 13.1.

EXAMPLE 13.1 Stoichiometry. To construct one car C , suppose you need four tires T and six windows W . The ‘equilibrium’ for the ‘reaction’ that transforms tires and windows into cars can be written as



In this case, $a = 4$, $b = 6$, and $c = 1$. Making one car is represented by $d\xi = 1$. So, one reaction transformation gives $c d\xi = +1$ new car and it uses up $-a d\xi = -4$ tires and $-b d\xi = -6$ windows.

Substitute the stoichiometry constraint equations (13.13) into Equation (13.12) to get

$$(c\mu_C - a\mu_A - b\mu_B) d\xi = 0, \quad (13.14)$$

so, at equilibrium,

$$c\mu_C = a\mu_A + b\mu_B. \quad (13.15)$$

Substituting the definitions of chemical potential from Equations (13.8) and (13.9) into the equilibrium equation (13.15) gives

$$\begin{aligned} c \left[-kT \ln \left(\frac{q'_C}{N_C} \right) \right] &= a \left[-kT \ln \left(\frac{q'_A}{N_A} \right) \right] + b \left[-kT \ln \left(\frac{q'_B}{N_B} \right) \right], \\ \Rightarrow \quad \left(\frac{q'_C}{N_C} \right)^c &= \left(\frac{q'_A}{N_A} \right)^a \left(\frac{q'_B}{N_B} \right)^b. \end{aligned} \quad (13.16)$$

To express the relative numbers of particles of each species present at equilibrium, a natural definition of the equilibrium constant K arises from rearranging Equation (13.16):

$$K = \frac{N_C^c}{N_A^a N_B^b} = \left(\frac{(q'_C)^c}{(q'_A)^a (q'_B)^b} \right) = \left(\frac{q_C^c}{q_A^a q_B^b} \right) e^{-(c\varepsilon_{0C} - a\varepsilon_{0A} - b\varepsilon_{0B})/kT}. \quad (13.17)$$

Equation (13.17) allows you to predict chemical equilibria from atomic structures.

In this chapter, we apply quantum mechanical models to predicting chemical equilibria of small molecules, such as diatomics, in the gas phase. But first, let's illustrate the main ideas in a simple model of binding.

EXAMPLE 13.2 A lattice model of a binding equilibrium. To illustrate chemical equilibria, let's model a ligand L binding to a receptor R to form a complex C (see Figure 13.2). There are N_R , N_L , and N_C molecules each of R , L , and C on a lattice. The reaction takes place in a volume V . The molecules are at low concentration. When a ligand molecule orients correctly in a receptor site, it binds with energy $\varepsilon_0 < 0$. Neglecting incorrectly bound ligands, the partition functions are

$$\begin{aligned} q_R &= Vz, \\ q_L &= Vz, \\ q_C &= Vze^{-\varepsilon_0/kT}. \end{aligned} \quad (13.18)$$

In this model, V is the translational partition function. V is the number of different lattice sites on which each L , R , or C molecule can be centered (we neglect overlaps because the molecules are dilute). z is the rotational partition function. z is the number of different directions that an arrow on a given lattice site can point ($z = 6$ on a three-dimensional cubic lattice). The binding energy advantage only occurs when the ligand and receptor arrows are correctly aligned. Use Equation (13.17) to calculate the equilibrium binding affinity K :

$$K = \frac{N_C}{N_R N_L} = \frac{q_C}{q_R q_L} = \frac{e^{-\varepsilon_0/kT}}{Vz}. \quad (13.19)$$

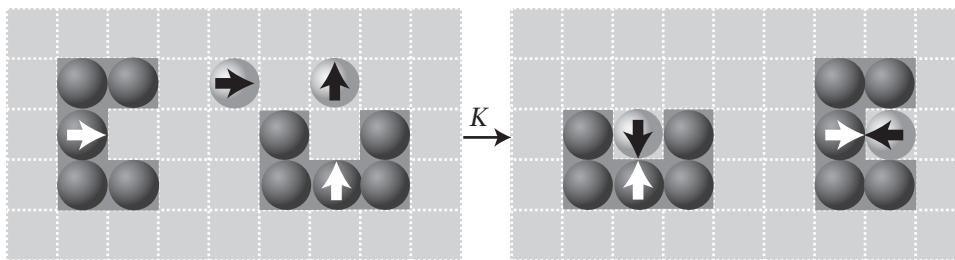


Figure 13.2 Each ligand molecule L is represented as a sphere with a black arrow. Each receptor molecule R is an object with a white arrow. When R and L come together in a proper orientation (arrows lined up), they form a complex C having a favorable binding energy $\varepsilon_0 < 0$.

This calculation shows the general features of chemical equilibria: (1) The binding affinity K is greater for ligands having tighter binding energies (ε_0 more negative). (2) Greater dilution (larger V) reduces the amount of complex C for given amounts of R and L . (3) Larger values of z lead to smaller affinities K . That is, if the system requires a high degree of orientational alignment to acquire the binding energy, it is entropically costly, as noted below.

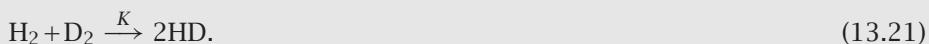
This model predicts a temperature dependence. Take the logarithm of Equation (13.19) to get $\ln K = -\varepsilon_0/(kT) - \ln(Vz)$. Then take the derivative to get

$$\frac{d \ln K}{dT} = \frac{\varepsilon_0}{kT^2}. \quad (13.20)$$

Since $\varepsilon_0 < 0$, this says that binding becomes weaker at higher temperatures. We will see below (in Equations (13.33)–(13.35)) how to compute the enthalpic and entropic components. The result is that the binding enthalpy for the present model is $\Delta h = \varepsilon_0$ and the binding entropy is $\Delta s = \ln(Vz)$. Binding is driven by the enthalpy. Binding is opposed by the translational and rotational entropies of complex formation.

Now, let's compute chemical equilibria for small-molecule reactions in the gas phase. To compute equilibrium constants using Equation (13.17), you need to know $\Delta\varepsilon_0 = c\varepsilon_{0C} - a\varepsilon_{0A} - b\varepsilon_{0B}$, the difference in ground-state energies. For $A \rightarrow B$, both states A and B must contain the same atoms, so the fully dissociated state of A must be identical to the fully dissociated state of B . Each quantity ε_0 is negative and represents the depth of the ground state relative to zero, the fully dissociated state. It is common to express each ground-state energy instead as a *dissociation energy* $D = -\varepsilon_0$, a positive quantity representing how much energy is required to break the bond, going from the ground state to the dissociated state. In the examples below, we use the dissociation energies from Table 11.2 and we use $\Delta\varepsilon_0 = -\Delta D$.

EXAMPLE 13.3 Predicting K in a change-of-symmetry reaction. Consider the gas-phase exchange reaction



The equilibrium constant is (see Equation (13.17))

$$K = \left(\frac{q_{\text{HD}}^2}{q_{\text{H}_2} q_{\text{D}_2}} \right) e^{\Delta D/RT}, \quad (13.22)$$

where ΔD is the difference between the molar dissociation energies of all the products and all the reactants. Because these energies are in molar units, divide them by RT rather than kT to get Boltzmann factors with dimensionless exponents. The dissociation energies are 431.8 kJ mol⁻¹ for H₂, 439.2 kJ mol⁻¹ for D₂, and 435.2 kJ mol⁻¹ for HD. Therefore,

$$\Delta D = 2D_{\text{HD}} - D_{\text{H}_2} - D_{\text{D}_2} = -0.6 \text{ kJ mol}^{-1}, \quad (13.23)$$

so $e^{\Delta D/RT} = 0.79$ at $T = 300 \text{ K}$.

You can write the equilibrium constant as a product of its component translational (t), rotational (r), and vibrational (v) factors, $K = K_t K_r K_v e^{\Delta D/RT}$, where

$$K_t = \frac{\left[(2\pi m_{\text{HD}} k T h^{-2})^{3/2}\right]^2}{(2\pi m_{\text{H}_2} k T h^{-2})^{3/2} (2\pi m_{\text{D}_2} k T h^{-2})^{3/2}}$$

$$= \left(\frac{m_{\text{HD}}^2}{m_{\text{H}_2} m_{\text{D}_2}}\right)^{3/2} = \left(\frac{3^2}{2 \times 4}\right)^{3/2} = 1.19 \quad (13.24)$$

and

$$K_r = \frac{[(8\pi^2 I_{\text{HD}} k T) / (\sigma_{\text{HD}} h^2)]^2}{[(8\pi^2 I_{\text{H}_2} k T) / (\sigma_{\text{H}_2} h^2)] [(8\pi^2 I_{\text{D}_2} k T) / (\sigma_{\text{D}_2} h^2)]}$$

$$= \left(\frac{\sigma_{\text{H}_2} \sigma_{\text{D}_2}}{\sigma_{\text{HD}}^2}\right) \left(\frac{I_{\text{HD}}^2}{I_{\text{H}_2} I_{\text{D}_2}}\right) = 4 \left(\frac{(6.13)^2}{(4.60)(9.19)}\right) = 3.56. \quad (13.25)$$

K_t is the ratio of the appropriate powers of translational partition functions from Equation (11.18), and K_r is the ratio of rotational partition functions from Equation (11.30), where $\sigma_{\text{H}_2} = \sigma_{\text{D}_2} = 2$ and $\sigma_{\text{HD}} = 1$. Because the vibrational energies are large for all three species at room temperature, $(h\nu)/kT \gg 1$ and $q_{\text{vibration}} = 1$,

$$K_v = \frac{(1 - e^{-h\nu_{\text{HD}}/kT})^{-2}}{(1 - e^{-h\nu_{\text{H}_2}/kT})^{-1} (1 - e^{-h\nu_{\text{D}_2}/kT})^{-1}} = 1 \quad (13.26)$$

is the ratio of vibrational partition functions $q_{\text{vibration}}$, from Equation (11.26). Combining the factors $K = K_t K_r K_v e^{\Delta D/RT}$ gives $K = (1.19)(3.56)(1)(0.79) = 3.35$. As $T \rightarrow \infty$, $K \rightarrow (1.19)(3.56) = 4.23$. The change in rotational symmetry is the main contributor to this equilibrium. The reaction is driven by the gain in entropy due to the rotational asymmetry of the products.

Pressure-Based Equilibrium Constants

Because pressures are easier to measure than particle numbers for gases, it is often more convenient to use equilibrium constants K_p based on pressures rather than equilibrium constants K based on the numbers of molecules. Combining Equation (13.17) with the ideal gas law $N = pV/kT$ gives

$$K = \frac{N_C^c}{N_A^a N_B^b} = \frac{(p_C V / kT)^c}{(p_A V / kT)^a (p_B V / kT)^b} = \frac{q_C^c}{q_A^a q_B^b} e^{\Delta D / kT}. \quad (13.27)$$

To get the pressure-based equilibrium constant K_p , multiply the two terms on the right side of Equation (13.27) by $[(V/kT)^a (V/kT)^b] / [(V/kT)^c]$:

$$K_p = \frac{p_C^c}{p_A^a p_B^b} = (kT)^{c-a-b} \frac{(q_{0C})^c}{(q_{0A})^a (q_{0B})^b} e^{\Delta D / kT}, \quad (13.28)$$

where $q_0 = q/V = [(2\pi m k T) / h^2]^{3/2} q_r q_v \dots$ is the partition function with the volume V factored out.

You can also express chemical potentials in terms of the partial pressures of each gas. Use $\mu = -kT \ln(q'/N)$ from Equation (13.8), where q' is the partition function that includes the ground-state energy term. Now, factor out the volume: $q' = q'_0 V$. Substitute the ideal gas law $V = NkT/p$ to get $q'/N = q'_0 kT/p$.

Now you have

$$\mu = -kT \ln\left(\frac{q'_0 kT}{p}\right) = kT \ln\left(\frac{p}{p_{\text{int}}^\circ}\right) = \mu^\circ + kT \ln p, \quad (13.29)$$

where we have collected together constants that have units of pressure, $p_{\text{int}}^\circ = q'_0 kT$. Or, you can express this constant instead in units of chemical potential, $\mu^\circ = -kT \ln q'_0 kT$.

Equation (13.29) will be useful in the following chapters. It divides the chemical potential into a part that depends on pressure ($kT \ln p$) and a part that does not (μ°). μ° is called the *standard-state* chemical potential. Although the rightmost expression ($\mu^\circ + kT \ln p$) would appear to imply that you should take the logarithm of a quantity that has dimensions, the preceding expression ($kT \ln(p/p_{\text{int}}^\circ)$) shows that the argument of the logarithm is indeed dimensionless. Operationally, you simply treat the term $\ln p$ as if it had no units.

The pressure-based equilibrium constant K_p is computed for a dissociation reaction in Example 13.4.

EXAMPLE 13.4 Dissociation reaction. For the dissociation of iodine, $I_2 \rightarrow 2I$, compute the equilibrium constant at $T = 1000\text{ K}$. The mass of an iodine atom is $m_I = 0.127\text{ kg mol}^{-1}/(6.02 \times 10^{23} \text{ molecules mol}^{-1}) = 2.109 \times 10^{-25} \text{ kg}$. Table 11.2 and the associated discussion give the ground-state electronic degeneracies as $q_{\text{elec},I} = 4$, $q_{\text{elec},I_2} = 1$, $\theta_{\text{rotation}} = 0.0537\text{ K}$, $\theta_{\text{vibration}} = 308\text{ K}$, and $\Delta D = -35.6\text{ kcal mol}^{-1}$. From Equation (13.28), the pressure-based equilibrium constant is

$$K_p = kT \frac{(q_{0I})^2}{q_{0I_2}} e^{\Delta D/RT}. \quad (13.30)$$

You have

$$\begin{aligned} kT = RT/\mathcal{N} &= \frac{(8.206 \times 10^{-5} \text{ m}^3 \text{ atm K}^{-1} \text{ mol}^{-1})(1000 \text{ K})}{6.02 \times 10^{23} \text{ molecules mol}^{-1}} \\ &= 1.363 \times 10^{-25} \text{ m}^3 \text{ atm}, \end{aligned}$$

and

$$e^{\Delta D/RT} = e^{-35,600/(1.987)(1000)} = 1.66 \times 10^{-8}.$$

Factor the partition functions q_0 into translational (q_t), rotational (q_r), vibrational ($q_{\text{vibration}}$), and electronic (q_e) components:

$$\frac{q_{0I}^2}{q_{0I_2}} = \left(\frac{q_{tI}^2}{q_{tI_2}}\right) \left(\frac{1}{q_{rI_2}}\right) \left(\frac{1}{q_{\text{vibration}I_2}}\right) \left(\frac{q_{eI}^2}{q_{eI_2}}\right). \quad (13.31)$$

For the rotations, $\sigma = 2$, and

$$q_{rI_2} = \frac{T}{2\theta_{\text{rotation}}} = \frac{1000}{2(0.0537)} = 9310.$$

For the vibrations (Equation (11.26)),

$$\begin{aligned} q_{vI_2} &= \frac{1}{1 - e^{-\theta_{\text{vibration}}/T}} = \frac{1}{1 - e^{-308/1000}} \\ &= 3.77. \end{aligned}$$

For the electronic terms, $q_{eI}^2/q_{eI_2} = 16$. For the translations (which have units of m^{-3} because volume has been factored out), $m_{I_2} = 2m_I$, and

$$\begin{aligned}\frac{q_{tI}^2}{q_{tI_2}} &= \frac{[(2\pi m_I kT/h^2)^{3/2}]^2}{(2\pi m_{I_2} kT/h^2)^{3/2}} = \left(\frac{\pi m_I kT}{h^2}\right)^{3/2} \\ &= \left[\frac{\pi(2.109 \times 10^{-25} \text{ kg})(1.38 \times 10^{-23} \text{ JK}^{-1})(1000 \text{ K})}{(6.626 \times 10^{-34} \text{ Js})^2}\right]^{3/2} \\ &= 3.01 \times 10^{33} \text{ m}^{-3}.\end{aligned}$$

Combine all these terms using Equations (13.30) and (13.31) to get

$$\begin{aligned}K_p &= (1.363 \times 10^{-25} \text{ m}^3 \text{ atm}) (1.66 \times 10^{-8}) \left(\frac{1}{9310}\right) \left(\frac{1}{3.77}\right) (16) \\ &\quad \times (3.01 \times 10^{33} \text{ m}^{-3}) = 3.1 \times 10^{-3} \text{ atm}.\end{aligned}\quad (16)$$

Figure 13.3 shows that this prediction is quite good. The measured dissociation constant of $I_2 \rightarrow 2I$ shown in Figure 13.3 is highly dependent on temperature. This is mainly because of the exponential dependence of the Boltzmann factor $e^{\Delta D/RT}$ on temperature. Increasing the temperature dissociates I_2 into iodine atoms.

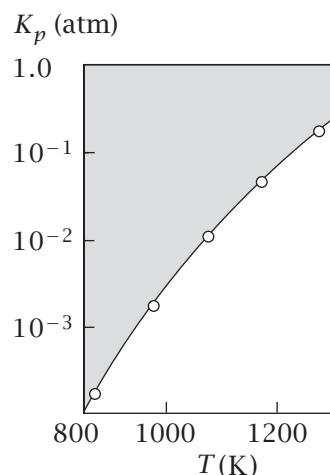
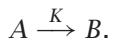


Figure 13.3 Theoretical curve and experimental points for the dissociation of iodine as a function of temperature.
Source: ML Perlman and GK Rollefson, *J Chem Phys* **9**, 362–369 (1941).

Le Chatelier's Principle Describes the Response to Perturbation from Equilibrium

Any perturbation away from a stable equilibrium state will increase the free energy of the system. The system will respond by moving back toward the state of equilibrium. To illustrate, let's return to the general two-state equilibrium,



Suppose a fluctuation changes the number of B molecules by an amount dN_B . The resulting change dG in free energy is given by Equations (13.2) and (13.4):

$$dG = (\mu_B - \mu_A) d\xi, \quad (13.32)$$

where ξ is the *reaction coordinate*; see Equation (13.13) and Figure 13.4. To move toward equilibrium, $dG \leq 0$, implying that the quantities $(\mu_B - \mu_A)$ and $d\xi$ must have opposite signs. If the system fluctuates into a state in which B happens to have the higher chemical potential, $\mu_B > \mu_A$, then the direction toward equilibrium is $d\xi < 0$, meaning that N_B will decrease. The chemical potential μ is sometimes called the **escaping tendency** because the higher the value of μ_B , the greater is the tendency of the system to escape from the state B to enter the state A . *Le Chatelier's principle* is the term that refers to the tendency of a system to return to equilibrium by moving in a direction opposite to that caused by an external perturbation.

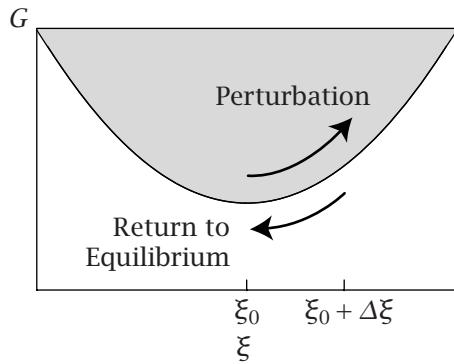
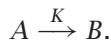


Figure 13.4 For the reaction $A \rightarrow B$, the extent of reaction is ξ . At equilibrium, $\xi = \xi_0$. A fluctuation that increases B leads to $\xi = \xi_0 + \Delta\xi$. A stable system returns to equilibrium by reducing the amount of B .

The Temperature Dependence of Equilibrium Is Described by the van't Hoff Equation

If you measure an equilibrium constant $K(T)$ at different temperatures, you can learn the enthalpy and entropy of the reaction, which are often useful for constructing or testing microscopic models. To illustrate, we return to the two-state equilibrium



At constant T and p , the condition for equilibrium is $\mu_A = \mu_B$. The pressure-based equilibrium constant is $K_p = p_B/p_A$. Using Equation (13.29), you have, at equilibrium,

$$\begin{aligned} \mu_A^\circ + kT \ln p_A &= \mu_B^\circ + kT \ln p_B \\ \Rightarrow \ln K_p &= \ln\left(\frac{p_B}{p_A}\right) = \frac{-(\mu_B^\circ - \mu_A^\circ)}{kT} = -\frac{\Delta\mu^\circ}{kT}. \end{aligned} \quad (13.33)$$

For gases, $\Delta\mu^\circ$ depends only on temperature (see Equation (13.29)) and not on pressure (for ideal gases). Express the chemical potential in terms of its thermal components (see Equation (9.32)):

$$\Delta\mu^\circ = \Delta h^\circ - T \Delta s^\circ,$$

where the lower-case h and s indicate an enthalpy and entropy *per molecule* or *per mole* (in the same way that μ expresses a free energy per molecule or per mole). The symbol $^\circ$ means that these quantities pertain to equilibria at $p = 1$ atm (see Equation (13.29)). Take the derivative of both sides of Equation (13.33) with respect to temperature, and substitute Equation (9.32) to get

$$\left(\frac{\partial \ln K_p}{\partial T} \right) = -\frac{\partial}{\partial T} \left(\frac{\Delta\mu^\circ}{kT} \right) = -\frac{\partial}{\partial T} \left(\frac{\Delta h^\circ - T \Delta s^\circ}{kT} \right). \quad (13.34)$$

If $\Delta h^\circ = h_B^\circ - h_A^\circ$ and $\Delta s^\circ = s_B^\circ - s_A^\circ$ are independent of temperature, Equation (13.34) gives

$$\left(\frac{\partial \ln K_p}{\partial T} \right) = \frac{\Delta h^\circ}{kT^2}. \quad (13.35)$$

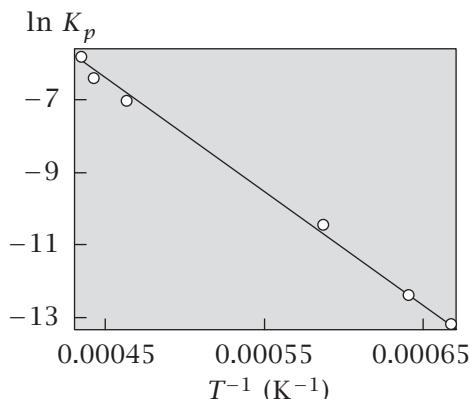


Figure 13.5 van't Hoff plot for the dissociation of water in the gas phase, $H_2O \rightarrow H_2 + (1/2)O_2$. Higher temperatures cause more dissociation than lower temperatures. Source: MW Zemansky and RH Dittman, *Heat and Thermodynamics: an Intermediate Textbook*, 6th edition, McGraw-Hill, New York, 1981.

An alternative form of Equation (13.35), called the **van't Hoff relation**, provides a useful way to plot data to obtain Δh° . Use $d(1/T) = -(1/T^2) dT$:

$$\left(\frac{\partial \ln K_p}{\partial (1/T)} \right) = -\frac{\Delta h^\circ}{k}. \quad (13.36)$$

van't Hoff plots show $\ln K$ versus $1/T$: the slope is $-\Delta h^\circ/k$. Figure 13.5 shows a van't Hoff plot for the dissociation of water vapor, $H_2O \rightarrow H_2 + (1/2)O_2$. Water is more dissociated at high temperatures than at lower temperatures. For dissociation, $\Delta h^\circ > 0$, which is characteristic of bond-breaking processes. Because this enthalpy change is positive, dissociation must be driven by entropy. Figure 13.5 illustrates a common, but not universal, feature of van't Hoff plots: they are often linear, implying that Δh° is independent of temperature.

When Δh° is independent of temperature, integration of Equation (13.36) gives

$$\ln \left[\frac{K_p(T_2)}{K_p(T_1)} \right] = \frac{-\Delta h^\circ}{k} \left(\frac{1}{T_2} - \frac{1}{T_1} \right). \quad (13.37)$$

Equation (13.37) can be used to find how K_p depends on temperature if you know Δh° , or to determine Δh° if you measure $K_p(T)$.

EXAMPLE 13.5 Getting Δh° from a van't Hoff plot. To get the enthalpy of dissociation of water, take two points $(1/T, \ln K)$ from Figure 13.5. At temperatures $T = 1500\text{ K}$ and 2257 K , you have $1/T = 6.66 \times 10^{-4}$ and $1/T = 4.43 \times 10^{-4}$, respectively. The corresponding points are $(6.66 \times 10^{-4}, -13.147)$ and $(4.43 \times 10^{-4}, -6.4)$. Use Equation (13.37):

$$\begin{aligned} \Delta h^\circ &= \frac{-R[\ln K_p(T_2) - \ln K_p(T_1)]}{(1/T_2) - (1/T_1)} \\ &= -(8.314 \text{ J K}^{-1} \text{ mol}^{-1}) \frac{[-13.1 - (-6.4)]}{(6.66 \times 10^{-4} - 4.43 \times 10^{-4}) \text{ K}^{-1}} \\ &= 249 \text{ kJ mol}^{-1}. \end{aligned}$$

Other thermodynamic quantities are also readily obtained from Figure 13.5. For example, at $T = 1500\text{ K}$,

$\Delta\mu^\circ = -RT \ln K_p = -(8.314 \text{ J K}^{-1} \text{ mol}^{-1})(1500 \text{ K})(-13.147) = 164 \text{ kJ mol}^{-1}$
and

$$\Delta s^\circ = \frac{\Delta h^\circ - \Delta\mu^\circ}{T} = \frac{(249 - 164) \text{ kJ mol}^{-1}}{1500 \text{ K}} = 56.7 \text{ J K}^{-1} \text{ mol}^{-1}.$$

Equations (13.35)–(13.37) are quite general and apply to any two-state equilibrium, not just to those that are in the gas phase or to chemical reactions. Here is a more general derivation showing that Equations (13.35) and (13.36) hold even when Δh° is dependent on temperature.

The Gibbs–Helmholtz Equation for Temperature–Dependent Equilibria

Let's generalize beyond chemical equilibria and beyond gases. How does the free energy $G(T)$ of any system depend on temperature? Rearrange the definition of free energy $G = H - TS$:

$$H = G + TS. \quad (13.38)$$

Substitute Equation (8.20), $S = -(\partial G / \partial T)_p$, into Equation (13.38) to get

$$H = G - T \left(\frac{\partial G}{\partial T} \right)_p. \quad (13.39)$$

Use $d(u/v) = (vu' - uv')/v^2$ with $v = T$ and $u = G$ to express the temperature derivative of G/T as

$$\left(\frac{\partial(G/T)}{\partial T} \right)_p = \frac{1}{T} \left(\frac{\partial G}{\partial T} \right)_p - \frac{G}{T^2} = -\frac{1}{T^2} \left[G - T \left(\frac{\partial G}{\partial T} \right)_p \right]. \quad (13.40)$$

Substituting Equation (13.39) into the square brackets in Equation (13.40) leads to the **Gibbs–Helmholtz equation**:

$$\left(\frac{\partial(G/T)}{\partial T} \right)_p = -\frac{H(T)}{T^2}. \quad (13.41)$$

Similarly, for the Helmholtz free energy, you have:

$$\left(\frac{\partial(F/T)}{\partial T} \right)_V = -\frac{U(T)}{T^2}. \quad (13.42)$$

Here's an illustration of the Gibbs–Helmholtz equation.

EXAMPLE 13.6 Obtaining $S(T)$ and $H(T)$ from $G(T)$. Suppose the free energy has the temperature dependence $G(T) = \alpha T^3$, where α is a constant. Equation (8.20) then gives $S(T) = -3\alpha T^2$. From Equation (13.39), you get $H(T) = \alpha T^3 - 3\alpha T^3 = -2\alpha T^3$. This is consistent with the result from Equation (13.41): $d(\alpha T^2)/dT = 2\alpha T = -H(T)/T^2$. This shows that the Gibbs–Helmholtz equation does *not* mean that the temperature dependence of the Gibbs free energy is solely due to the enthalpy.

Pressure Dependence of the Equilibrium Constant

Applying pressure can shift an equilibrium. The pressure dependence $K(p)$ indicates a difference in volume on the two sides of the equilibrium. Consider the two-state equilibrium $A \rightleftharpoons B$. The slope of the equilibrium constant with pressure is

$$\frac{\partial \ln K(p)}{\partial p} = \frac{\partial}{\partial p} \left[-\frac{(\mu_B^\circ - \mu_A^\circ)}{kT} \right] = -\frac{1}{kT} \left(\frac{\partial \Delta\mu^\circ}{\partial p} \right). \quad (13.43)$$

For ideal gases, this quantity is zero, but for other systems, it may be nonzero. To determine the pressure dependence $\Delta\mu(p)$, use Maxwell's relation

$$\left(\frac{\partial \mu}{\partial p} \right)_{T,N} = \left(\frac{\partial V}{\partial N} \right)_{T,p} = v, \quad (13.44)$$

where v is the volume per molecule, or per mole, depending on the units of μ . Substituting Equation (13.44) into (13.43) gives

$$\left(\frac{\partial (\mu_B^\circ - \mu_A^\circ)}{\partial p} \right)_T = v_B^\circ - v_A^\circ = \Delta v^\circ. \quad (13.45)$$

where v° refers to the volume of the component A or B when $p = 1$ bar. Substituting Equation (13.45) into (13.43) gives

$$\left(\frac{\partial \ln K}{\partial p} \right)_T = -\frac{\Delta v^\circ}{kT}. \quad (13.46)$$

Therefore, if B is the state with the smaller volume, $v_B - v_A < 0$, then increasing the pressure will shift the equilibrium from A toward B . Two examples follow below.

EXAMPLE 13.7 Pressure drives equilibria toward dense states. To see how pressure affects equilibria, combine Equation (13.46) with the model of ligand binding in Example 13.2. Figure 13.6 shows two different cases.

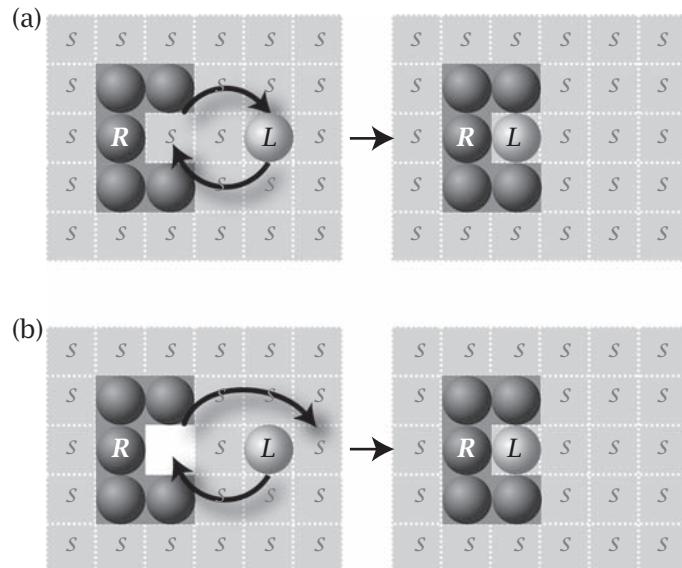


Figure 13.6 Pressure can shift equilibria. (a) Binding a ligand displaces a water molecule, so the volumes of products and reactants are about the same: $\Delta v = 0$. Pressure does not affect this equilibrium. (b) Binding a ligand fills a previously empty space, reducing the volume of the system: $\Delta v < 0$. In this case, increasing pressure increases binding.

In Figure 13.6(a), a ligand displaces a water molecule from the receptor site. The net change in volume, Δv , of the system (receptor + displaced water + ligand) is approximately zero because this condensed phase is relatively incompressible. So K does not change with applied pressure. In contrast, in Figure 13.6(b), the ligand fills a space that was previously empty, so the volume decreases upon binding: $\Delta v < 0$. So, Equation (13.46) says that applied pressure will shift this equilibrium to the right, driving the system to form more receptor-ligand complex C , i.e., stabilizing the state that has the smaller volume. Pressure shifts equilibria in whatever direction crunches the total system into a smaller volume.

Here's an application of Equations (13.35) and (13.46) to pressure effects on anesthesia.

EXAMPLE 13.8 Pressure affects a two-state equilibrium. A common anesthetic drug molecule is halothane (2-bromo-2-chloro-1,1,1-trifluoroethane). Its mode of action is presumed to involve partitioning from water (state A) into lipid bilayer membranes (state B). Figure 13.7 shows how the partitioning equilibrium depends on pressure. Putting values of $(p, \ln K) = (0, 7.84)$ and $(280, 7.6)$ into Equation (13.46) gives $\Delta v = v_{\text{bilayer}} - v_{\text{water}}$:

$$\begin{aligned}\Delta v &= -RT \frac{(\ln K_2 - \ln K_1)}{p_2 - p_1} \\ &= \frac{(8.205 \times 10^{-5} \text{ m}^3 \text{ atm K}^{-1} \text{ mol}^{-1})(300 \text{ K})(7.84 - 7.6)}{280 \text{ atm}} \left(\frac{10^2 \text{ cm}}{\text{m}} \right)^3 \\ &= 21 \text{ cm}^3 \text{ mol}^{-1}.\end{aligned}$$

Multiplying by $(10^8 \text{ \AA}/\text{cm})^3$ and dividing by Avogadro's number gives a volume change of 35 \AA^3 per molecule. The anesthetic occupies more volume in the membrane phase than in water. Pressure forces the anesthetic molecules to go into the water, where the volumes they occupy are smaller.

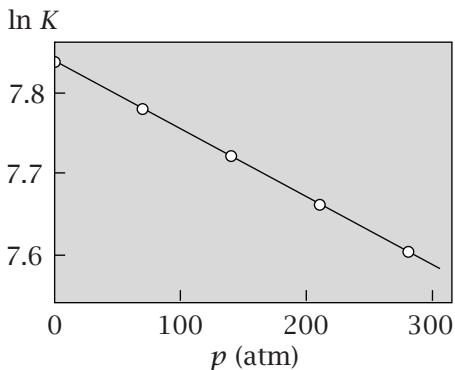


Figure 13.7 Applying pressure drives anesthetic molecules into water from lipid bilayer membranes. K is the partition coefficient from water into the bilayer. Source: S Kaneshina, H Kamaya, I Ueda, *J Coll Interf Sci* **93**, 215–224 (1983).

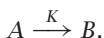
Summary

A remarkable achievement of statistical mechanics is the accurate prediction of gas-phase chemical reaction equilibria from atomic structures. From atomic masses, moments of inertia, bond lengths, and bond strengths, you can calculate partition functions. You can then calculate equilibrium constants and their dependence on temperature and pressure. In Chapter 19, we will apply these ideas to chemical kinetics, which pertains to the *rates* of reactions. Reactions can be affected by the medium they are in. Next, in Chapter 14, we will develop models of liquids and other condensed phases.

Problems

1. Iodine dissociation. Compute the dissociation constant K_p for iodine, $I_2 \rightarrow 2I$, at $T = 300\text{ K}$.

2. Temperature dependence of a simple equilibrium. In a reaction



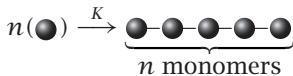
the equilibrium constant is $K_p = 10$ at $T = 300\text{ K}$.

- (a) What is $\Delta\mu^\circ$?
- (b) If $\Delta h^\circ = 10\text{ kcal mol}^{-1}$, what is K_p at $T = 310\text{ K}$?
- (c) What is Δs° at $T = 300\text{ K}$?

3. Dissociation of oxygen, $O_2 \rightarrow 2O$. Compute K_p , the pressure-based equilibrium constant for this dissociation reaction at $T = 3000\text{ K}$. The electronic ground-state degeneracy for O is $g_0(0) = 9$.

4. Temperature dependence of K_p . For the dissociation of O_2 , derive an expression for $d \ln K_p / dT$ near $T = 300\text{ K}$ from the expression for K_p that you used in Problem 3.

5. Polymerization. Consider a polymerization reaction in the gas phase in which n moles of identical monomers are in equilibrium with one mole of chains of n -mers:



- (a) Do you expect typical polymerization processes to be driven or opposed by enthalpy? By entropy? What are the physical origins of these enthalpies and entropies?
- (b) Do you expect polymerizations to be exothermic (giving off heat) or endothermic (taking up heat)? Explain the explosions in some polymerization processes.
- (c) Are polymerization equilibrium constants for long chains more or less sensitive to temperature than those for shorter chains?

6. Hydrogen ionization. A hydrogen atom can ionize in the gas phase:



Calculate the equilibrium constant K for temperature $T = 5000\text{ K}$. There is no rotational or vibrational partition function for H, H^+ , or e^- , but there are spin partition functions: $q_s = 1$ for H^+ and $q_s = 2$ for e^- . $\Delta D = -311\text{ kcal mol}^{-1}$.

7. Free energy along the reaction coordinate. Consider the ideal gas-phase equilibrium $2A \rightarrow B$, at constant temperature and a constant pressure of 1 atm. Assume that there is initially 1 mole of A and no B present, and that $\mu_A^\circ = 5\text{ kcal mol}^{-1}$ and $\mu_B^\circ = 10\text{ kcal mol}^{-1}$ at this temperature.

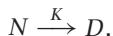
- (a) Show that G , at any time during the reaction, can be written as

$$G = \left(5 + RT \ln \left[\left(\frac{1-2\xi}{1-\xi} \right)^{1-2\xi} \left(\frac{\xi}{1-\xi} \right)^\xi \right] \right) \text{ kcal mol}^{-1},$$

where ξ is the extent of reaction.

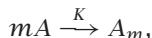
- (b) What is the value of the extent of reaction ξ at equilibrium?

8. Pressure denaturation of proteins. For a typical protein, folding can be regarded as involving two states, native (N) and denatured (D):



At $T = 300\text{ K}$, $\Delta\mu^\circ = 10\text{ kcal mol}^{-1}$. Applying about 10,000 atm of pressure can denature a protein at $T = 300\text{ K}$. What is the volume change Δv for the unfolding process?

9. Clusters. Sometimes isolated molecules of type A can be in a two-state equilibrium with a cluster of m monomers:



where A_m represents an m -mer cluster.

- (a) At equilibrium, what is the relationship between μ_1 , the chemical potential of the monomer, and μ_m , the chemical potential of A in the cluster?
- (b) Express the equilibrium constant K in terms of the partition functions.

10. Two-state protein folding. In Example 8.2, we considered a two-state model of proteins: the folded state has energy $-\varepsilon_0$ (where $\varepsilon_0 > 0$) and the unfolded state has $\gamma = 4$ open conformations. The partition function is

$$q = 4 + e^{\varepsilon_0/kT}.$$

- (a) Write an expression for the equilibrium constant for folding, $K_f = (\text{population}_{\text{folded}})/(\text{population}_{\text{unfolded}})$, in terms of ε_0 and temperature T .
- (b) Using the van't Hoff expression $(\partial \ln K) / (\partial (1/T))$ derive the enthalpy of folding, ΔH_f , for this model.
- (c) What is the free energy of folding versus temperature T : $\Delta F_f(T) = F_{\text{folded}} - F_{\text{unfolded}}$?
- (d) This model predicts that $\Delta F_f(T)$ is a linearly increasing function of T . This is often a good approximation near the denaturation midpoint T_m . For myoglobin, which has about 150 amino acids,

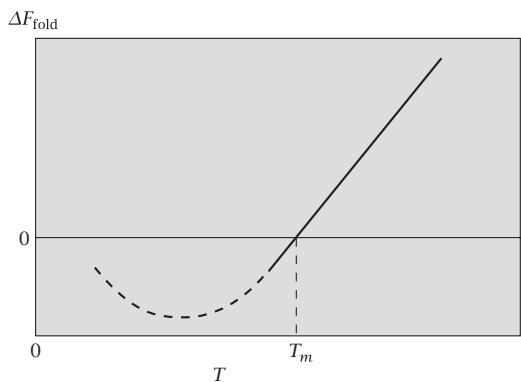


Figure 13.8

the slope is $0.5 \text{ kcal mol}^{-1} \text{ deg}^{-1}$) at $T_m = 70^\circ\text{C}$. If instead of choosing $\gamma = 4$, as in our simple model, you chose γ to fit the experimental slope in Figure 13.8, what value of γ is needed?

Suggested Reading

Elementary and detailed discussions of chemical equilibria:

JH Knox, *Molecular Thermodynamics: An Introduction to Statistical Mechanics for Chemists*, Wiley, Chichester, 1978.

NO Smith, *Elementary Statistical Thermodynamics: A Problems Approach*, Plenum Press, New York, 1982.

Excellent advanced texts with worked examples:

TL Hill, *Introduction to Statistical Thermodynamics*, Addison-Wesley, Reading, MA, 1960 (reprinted by Dover Publications, New York, 1986).

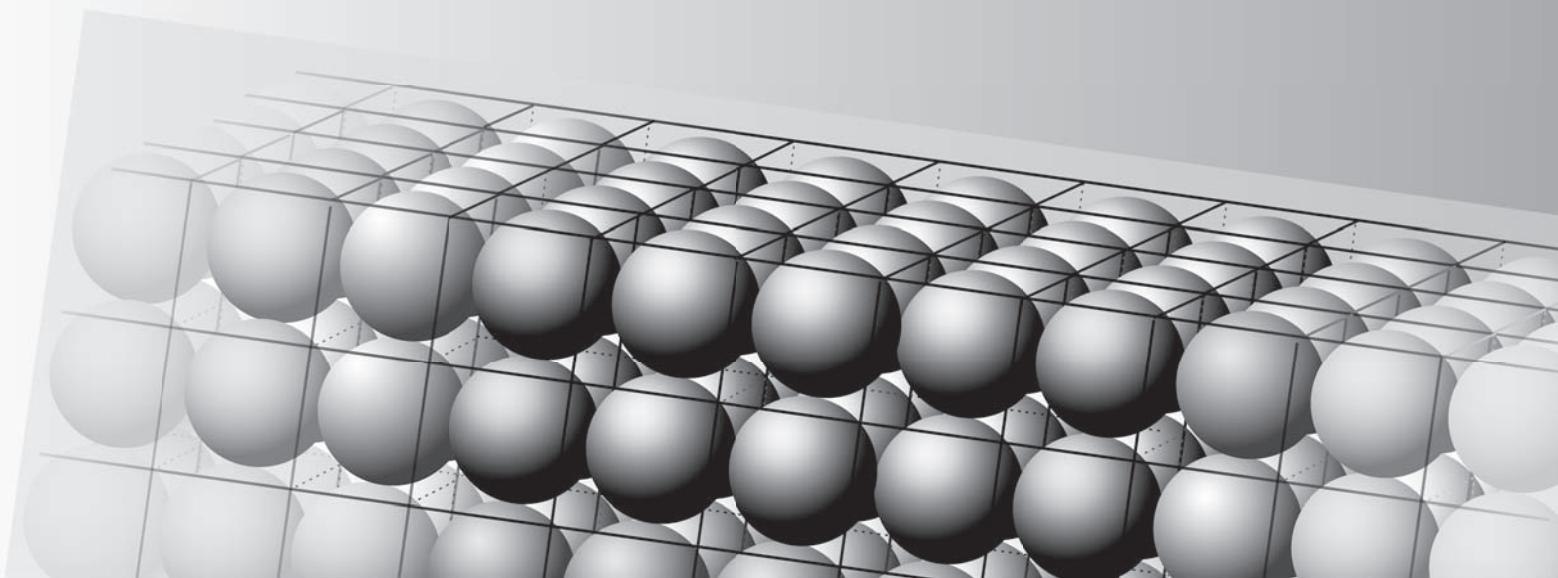
R Kubo, with H Ichimura, T Usui, and N Hashitsome, *Statistical Mechanics*, North-Holland, New York, 1965.

DA McQuarrie, *Statistical Mechanics*, 2nd edition, University Science Books, Sausalito, CA, 2000.

CL Tien and JH Lienhard, *Statistical Thermodynamics*, Hemisphere, New York, 1979.

This page is intentionally left blank.

14 Equilibria Between Liquids, Solids, & Gases



Phase Equilibria Are Described by the Chemical Potential

In this chapter, we explore some properties of pure liquids and solids. We model the vapor pressures over liquids, the processes of boiling and sublimation, and surface tension. This is the foundation for treating solvation and desolvation in chemical reactions and in binding and conformational equilibria, in Chapters 15 and 16.

In Chapter 13, we found that the chemical potential describes the tendency for particles to interconvert from one chemical species to another. Here we find that the chemical potential also describes the tendency for particles to move between two phases, between a liquid and a vapor phase, for example. The surface tension of a liquid is described in terms of the tendency for particles to move between the bulk (interior) and the surface of a material.

Why Do Liquids Boil?

At room temperature and pressure, water is mostly liquid, but some water molecules escape to the vapor phase. In contrast, above the boiling point, water vaporizes. (*Vapor* usually refers to the gas phase of a material that is mostly liquid or solid at room temperature, such as water, while *gas* refers to a material that is mostly gaseous at room temperature, such as helium.) How is vaporization controlled by temperature and pressure? The essence is a balance of

two forces. Attractions hold the molecules together in the liquid phase. On the other hand, molecules gain translational entropy when they escape to the vapor phase. Increasing the temperature boils water because the entropy gain at high temperatures is greater than the loss of bond energies, and therefore dominates the free energy $\Delta F = \Delta U - T\Delta S$. At lower temperatures, energies dominate, forming a stable liquid state.

Consider a beaker of liquid (the condensed phase is indicated with the subscript c) in equilibrium with its vapor (indicated with the subscript v). Hold the temperature T constant. Also, hold the pressure p constant and don't allow molecules to escape; you can do this with a piston or balloon over the vapor phase. The degree of freedom is the number of particles N_c in the condensed phase, or the number N_v in the vapor phase. The free energy change of the combined system (vapor plus condensed phase) is expressed in terms of the chemical potentials μ_c and μ_v . Because $dT = dp = 0$, the free energy depends only on the chemical potentials and the numbers of particles in the two phases:

$$dG = -S dT + V dp + \sum_{j=1}^2 \mu_j dN_j = \mu_v dN_v + \mu_c dN_c. \quad (14.1)$$

If the total number of molecules is conserved in the exchange between vapor and condensed phases, the constraint is

$$\begin{aligned} N_v + N_c &= N = \text{constant} \\ \Rightarrow dN_v + dN_c &= dN = 0 \Rightarrow dN_c = -dN_v, \end{aligned} \quad (14.2)$$

so you can express the change in free energy as

$$dG = (\mu_v - \mu_c) dN_v. \quad (14.3)$$

At constant T and p , the condition for equilibrium is $dG = 0$, or, in a more useful form,

$$\mu_v = \mu_c. \quad (14.4)$$

To compute the vaporization equilibrium from Equation (14.4), you need μ_c and μ_v . If the vapor is dilute enough to be an ideal gas, then, according to Equations (11.47) and (11.50), you have

$$\mu_v = kT \ln\left(\frac{p}{p_{\text{int}}^\circ}\right), \quad (14.5)$$

where $p_{\text{int}}^\circ = q_0^\circ kT$ has units of pressure, describes the internal degrees of freedom in the molecule, and is a constant that does not depend on the vapor pressure p .

Now you need a model for μ_c , the chemical potential for the molecules in the condensed phase. The particles in liquids and solids are close together, so they interact significantly with each other. Predicting the properties of condensed phases can be complex. Instead of proceeding toward sophisticated theory, we return to the lattice model, which gives simple insights.

Let's model a liquid or solid as a lattice of particles of a single type (see Figure 14.1). Because gases are roughly a thousandfold less dense than liquids,

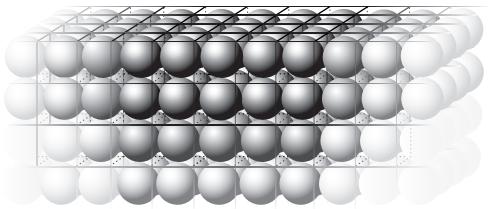


Figure 14.1 Lattice model of liquids and solids. For practical purposes, the lattice is infinite.

and liquids are only about 10% less dense than solids, we will neglect the distinction between liquids and solids here. We will model a liquid as if its particles occupied a crystalline lattice, with every site occupied by one particle.

This is an approximation in two respects: liquids are slightly less dense than solids, and solids have longer-ranged repeating periodicities, called *long-range order*. Solids are more regular. Every atom or molecule in a crystalline solid has the same number of nearest-neighbor atoms or molecules as every other. But in a liquid, there is greater variation—one molecule may have five nearest neighbors for a while, then six for a while, etc. The main insight represented by the lattice liquid model is that the most important energetic interactions for holding liquids together are the *short-ranged* interactions of each particle with its nearest neighbors, and that the number of such neighbors has a relatively well-defined average (see Chapter 24). A particle interacts only very weakly with other particles that are more distant.

Because liquids and solids are much less compressible than gases, it won't matter whether we treat the liquid as being at constant pressure (T, p, N) or constant volume (T, V, N). When the pressure is held constant, the volumes of liquids and solids don't change much. In microscopic models, it is often simpler to treat constant V than constant p . Here we use constant V , and we compute the free energy $F = U - TS$ to determine which phase—vapor or liquid—is most stable under various conditions.

The Lattice Model for the Entropy and Energy of a Condensed Phase

The translational entropy of the lattice of particles shown in Figure 14.1 is zero. $W = 1$ and $S = k \ln W = 0$, because if pairs of particles trade positions, the rearrangement can't be distinguished from the original arrangement.

Now we need the energy of the lattice model liquid. At equilibrium, the exchange of particles between the liquid and the vapor phases involves little or no change in the internal quantum mechanical state of the particles—their rotations, vibrations, and electronic states do not change, to a first approximation. So the energies that are relevant for vaporization are those between pairs of particles, not within each particle. In the gas phase, if it is ideal, the particles do not interact with each other.

Uncharged particles are attracted to each other (see Chapter 24). Let's represent the attraction between two particles of type A in terms of a 'bond' energy: $w_{AA} < 0$. This energy applies to every pair of particles that occupy neighboring lattice sites. Each such pair has one bond, or contact. w_{AA} is negative, indicating that the particles attract. The energies that hold uncharged atoms and molecules together in liquids and solids are less than a few kilocalories per mole of bonds. They are much weaker than the covalent and ionic bonds that hold

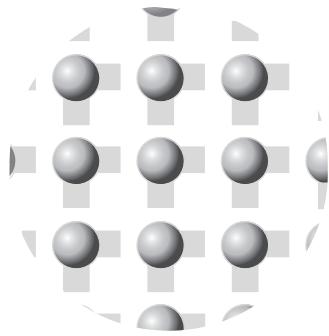


Figure 14.2 Focus on an interior particle in an infinite lattice of particles. Each particle has z neighbors ($z = 4$ in this two-dimensional case). To avoid double-counting, assign $z/2 = 2$ bonds (shown as shaded bars) to each particle. For N particles, the total number of bonds is $m = Nz/2$.

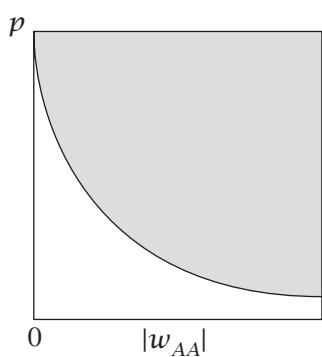


Figure 14.3 The vapor pressure p of A molecules versus the strength of the AA bond, $|w_{AA}|$.

atoms in molecules (tens to hundreds of kilocalories per mole). A reasonable first approximation is that the bond energies are independent of temperature.

To compute the interaction energy of the lattice model liquid, we need to count the number of bonds among the N particles and multiply by the energy per bond, w_{AA} . Each particle on a lattice has z nearest neighbors. z is called the *coordination number* of the lattice. Even though each particle has z neighbors, and therefore z interactions, the total energy is *not* $U = Nz w_{AA}$, because that would count every bond twice. Figure 14.2 shows that you can avoid double-counting by assigning $z/2$ of the interactions to each particle. So, because the outer edges contribute negligibly to an infinite lattice, the number of bonds is $m = Nz/2$, and the total bond energy U is

$$U = mw_{AA} = \left(\frac{Nzw_{AA}}{2} \right). \quad (14.6)$$

Free Energy and Chemical Potential

The free energy of the lattice model liquid is

$$F = U - TS = U = N \left(\frac{zw_{AA}}{2} \right) \quad (14.7)$$

and the chemical potential μ_c is

$$\mu_c = \left(\frac{\partial F}{\partial N} \right)_{T,V} = \frac{zw_{AA}}{2}. \quad (14.8)$$

The Vapor Pressure

For the condensed-phase chemical potential μ_c , you can use the lattice model equation (14.8). Use Equation (14.5) for μ_v and set the chemical potentials equal, $\mu_c = \mu_v$, to get the condition for equilibrium in terms of the vapor pressure p :

$$kT \ln \left(\frac{p}{p_{\text{int}}^{\circ}} \right) = \frac{zw_{AA}}{2} \quad \Rightarrow \quad p = p_{\text{int}}^{\circ} e^{zw_{AA}/2kT}. \quad (14.9)$$

Equation (14.9) describes the vapor pressure p of the A molecules that escape the liquid. The vapor pressure is a measure of the density of the vapor-phase molecules (because $p = (N/V)kT$ from the ideal gas law).

Our model makes two predictions. First, stronger bonding drives molecules into the liquid phase. In a series of different types of molecules A , those having stronger AA bonds (w_{AA} more negative, toward the right on Figure 14.3), have lower vapor pressures. You can smell benzene and gasoline and ‘aromatic’ compounds. They are only weakly attracted to each other in the liquid, so they have relatively high vapor pressures and your nose detects them. But you can’t smell water or liquid mercury, partly because they are strongly bonded together in the liquid, so they have low vapor pressures (but also partly because your nose does not detect all molecules equally).

Second, vapor pressures increase with temperature. As a lake becomes warm, water evaporates, leaving a haze that you can sometimes see. Or, when wet air is cooled, the equilibrium reverses, and water condenses, resulting in rain or fog. Figure 14.4 shows how increasing the temperature increases the vapor

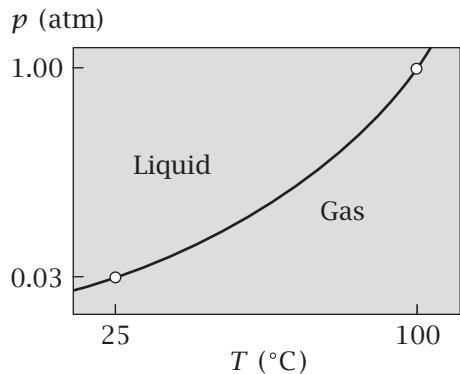


Figure 14.4 Gas–liquid phase boundary for water. Water’s vapor pressure is 1 atm at $T = 100^\circ\text{C}$ and is 0.03 atm at $T = 25^\circ\text{C}$. At the phase boundary, the applied pressure equals the vapor pressure.

pressure according to Equation (14.9). In particular, it shows the line of pressures and temperatures at which water boils. Water boils at $T = 100^\circ\text{C}$ and $p = 1 \text{ atm}$, but you can also boil water at room temperature ($T = 25^\circ\text{C}$) if you reduce the applied pressure to $p = 0.03 \text{ atm}$. This model shows how you can use that information to learn about the strength of bonding in the liquid state.

EXAMPLE 14.1 Computing molecular attractions from boiling experiments. Given that water’s vapor pressure is 1 atm at $T = 100^\circ\text{C}$ and 0.03 atm at $T = 25^\circ\text{C}$, let’s find the value of zw_{AA} . Take the logarithm of Equation (14.9) to get the boiling pressure p_1 at temperature T_1 as $\ln(p_1/p_{\text{int}}^\circ) = zw_{AA}/RT_1$ and the boiling pressure p_2 at temperature T_2 as $\ln(p_2/p_{\text{int}}^\circ) = zw_{AA}/RT_2$. Subtract the first equation from the second to get

$$\ln\left(\frac{p_2}{p_1}\right) = \frac{zw_{AA}}{2R} \left(\frac{1}{T_2} - \frac{1}{T_1} \right). \quad (14.10)$$

(We neglect the small temperature dependence in p_{int}° .) Substituting these values for (T_1, p_1) and (T_2, p_2) gives $zw_{AA}/2 = 40.66 \text{ kJ mol}^{-1}$. Because water molecules have about $z = 4$ neighbors in the liquid state, you can estimate that $w_{AA} \approx 20.3 \text{ kJ mol}^{-1}$.

Now let’s switch from this simple lattice model to a more general thermodynamic treatment of phase equilibria.

The Clapeyron Equation Describes $p(T)$ at Phase Equilibrium

In this section, we describe phase equilibria such as boiling, and how they change with pressure and temperature. Figure 14.5 shows a phase diagram. Each line on the diagram (called a phase boundary) represents a set of (p, T) points at which two phases are equally stable: solid–gas (sublimation or deposition), solid–liquid (melting or freezing), or liquid–gas (boiling or condensation). The liquid–gas line shows that increasing the pressure increases the boiling temperature. When (p, T) is above the line, the system is a liquid, and when (p, T) is below the line, the system is a gas. For example, under conditions $p = 1 \text{ atm}$ and $T = 25^\circ\text{C}$, water is liquid.

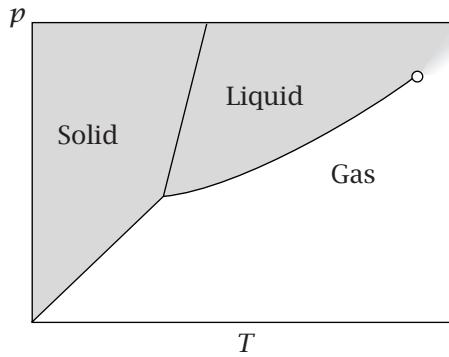


Figure 14.5 A (p, T) phase diagram. The value of applied pressure p and temperature T determines whether the system is gas, liquid, or solid at equilibrium. The phase boundary lines represent boiling-condensation (liquid-gas), sublimation-deposition (solid-gas), and melting-freezing (liquid-solid).

In Figures 14.4 and 14.5, p is the applied pressure, which need not equal the vapor pressure. In a series of equilibrium experiments, when the applied pressure equals the vapor pressure, the system changes phase. When the applied pressure is greater (say 1 atm) than the vapor pressure (say 0.03 atm at $T = 25^\circ\text{C}$), water is a liquid, but some molecules are in the vapor phase. The positive slope of $p(T)$ is a general feature of many types of phase equilibria.

The slope of a phase boundary gives useful information. Figure 14.6 shows the equilibrium between gas (G) and liquid (L). We want to compute how the boiling pressure changes with boiling temperature. We have two points, (p_1, T_1) and (p_2, T_2) , at which the liquid and the vapor are in equilibrium, so

$$\mu_L(T_1, p_1) = \mu_G(T_1, p_1) \quad \text{and} \quad \mu_L(T_2, p_2) = \mu_G(T_2, p_2). \quad (14.11)$$

The chemical potential at point 2 involves a small perturbation from its value at point 1,

$$\mu_L(T_2, p_2) = \mu_L(T_1, p_1) + d\mu_L(T, p)$$

and

$$\mu_G(T_2, p_2) = \mu_G(T_1, p_1) + d\mu_G(T, p). \quad (14.12)$$

Substituting Equations (14.11) into Equations (14.12) gives

$$d\mu_G(T, p) = d\mu_L(T, p). \quad (14.13)$$

We are regarding μ as a function of (T, p) , so

$$d\mu(T, p) = \left(\frac{\partial\mu}{\partial T}\right)_{p,N} dT + \left(\frac{\partial\mu}{\partial p}\right)_{T,N} dp. \quad (14.14)$$

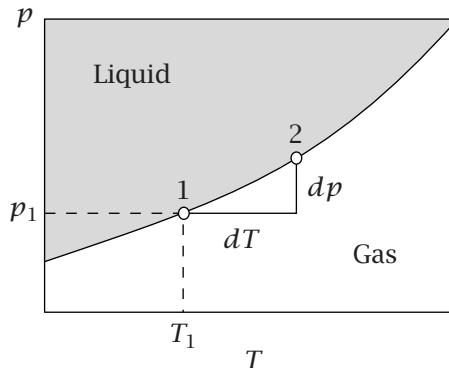


Figure 14.6 The slope of the boiling equilibrium line is given by the Clapeyron equation.

If you use free energies *per mole* (rather than *per molecule*) as units for μ , the derivatives in Equation (14.14) are the partial molar entropy s and partial molar volume v (see the Maxwell relations in Table 9.1),

$$\left(\frac{\partial \mu}{\partial T}\right)_{p,N} = -\left(\frac{\partial S}{\partial N}\right)_{T,p} = -s$$

and

$$\left(\frac{\partial \mu}{\partial p}\right)_{T,N} = \left(\frac{\partial V}{\partial N}\right)_{T,p} = v. \quad (14.15)$$

Substitute Equations (14.15) into Equations (14.14) and (14.13) to get

$$d\mu_G = -s_G dT + v_G dp = d\mu_L = -s_L dT + v_L dp. \quad (14.16)$$

Rearranging Equation (14.16) gives

$$\frac{dp}{dT} = \frac{s_G - s_L}{v_G - v_L} = \frac{\Delta s}{\Delta v}, \quad (14.17)$$

where $\Delta s = s_G - s_L$ is the partial molar change of entropy and $\Delta v = v_G - v_L$ is the partial molar change of volume from liquid to gas. At the phase transition, $\Delta\mu = \Delta h - T \Delta s = 0$, so

$$\Delta h = T \Delta s, \quad (14.18)$$

where $\Delta h = h_G - h_L$ is the enthalpy difference between the two phases. Substituting Equation (14.18) into Equation (14.17) yields the **Clapeyron equation**:

$$\frac{dp}{dT} = \frac{\Delta h}{T \Delta v}. \quad (14.19)$$

The molar volume of the gas phase is much larger than the molar volume of the liquid or solid phase, so $\Delta v = v_{\text{gas}} - v_{\text{condensed}} \approx v_{\text{gas}} = RT/p$,¹ if the ideal gas law applies. Substituting this expression into Equation (14.19) replaces the right-hand side with $p\Delta h/(RT^2)$. Combining this form of Equation (14.19) with $dp/p = d \ln p$ gives the **Clausius-Clapeyron equation**,

$$\frac{d \ln p}{dT} = \frac{\Delta h}{RT^2}. \quad (14.20)$$

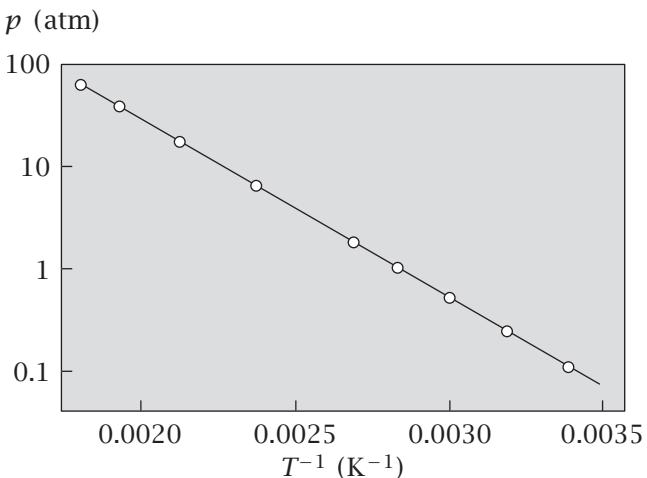
To put Equation (14.20) into a more useful form, integrate it. When Δh is independent of p and T ,

$$\begin{aligned} \int_{p_1}^{p_2} d \ln p &= \int_{T_1}^{T_2} \frac{\Delta h}{RT^2} dT \\ \Rightarrow \ln\left(\frac{p_2}{p_1}\right) &= -\frac{\Delta h}{R} \left(\frac{1}{T_2} - \frac{1}{T_1}\right). \end{aligned} \quad (14.21)$$

Figure 14.7 shows the vapor pressure of benzene versus $1/T$, illustrating that $\ln p$ is linearly proportional to $1/T$, and therefore that Δh is independent of T .

¹Volumes per mole are given by RT/p , while volumes per molecule are given by kT/p . Use kT to express energies per molecule, or multiply by Avogadro's number ($N = 6.022 \times 10^{23}$ molecules mol⁻¹) to get $RT = NkT$ if you want energies per mole.

Figure 14.7 Vapor pressure of benzene versus $1/T$. When $\ln p$ is a linear function of $1/T$, as it is here, Equation (14.21) shows that Δh is independent of temperature. Source: JS Rowlinson and FL Swinton, *Liquids and Liquid Mixtures*, 3rd edition, Butterworth, London, 1982.



Values of Δh can be obtained from tables (see Table 14.1) and used to predict a boiling point (p_2, T_2) if another boiling point (p_1, T_1) is known, as shown in Example 14.2.

EXAMPLE 14.2 Boiling water at high altitudes. Water boils at $T_1 = 373\text{ K}$ and $p_1 = 1\text{ atm}$. At a high altitude where $p_2 = 0.5\text{ atm}$, what is the boiling temperature T_2 ? Substituting the enthalpy of vaporization for water, $\Delta h_{\text{vap}} = 40.66\text{ kJ mol}^{-1}$, from Table 14.1 into Equation (14.21) gives

$$\begin{aligned}\frac{1}{T_2} &= \frac{1}{T_1} - \frac{R}{\Delta h_{\text{vap}}} \ln\left(\frac{p_2}{p_1}\right) \\ \Rightarrow \quad \frac{1}{T_2} &= \frac{1}{373} - \left(\frac{8.314\text{ J K}^{-1}\text{ mol}^{-1}}{40,660\text{ J mol}^{-1}}\right) \ln\left(\frac{1}{2}\right) \\ \Rightarrow \quad T_2 &= 354\text{ K} = 81^\circ\text{C}.\end{aligned}$$

Water boils at lower temperatures at higher altitudes.

Equation (14.17) says that if the slope dp/dT of the phase boundary is positive, the phase with the greater entropy has the greater volume per molecule. Because the vapor has higher entropy than the liquid or solid, the vapor has the greater volume per molecule. Similarly, a typical liquid has more volume than its solid. The correlation of increasing volume with increasing entropy is a general property of materials, although an interesting exception is water (see Chapter 30). For water, the solid-liquid phase boundary has a negative slope, $dp/dT < 0$. Because liquid water has greater entropy than solid water (ice), liquid water must have a *smaller* volume per molecule than ice. Indeed, ice floats on water.

Comparing Equation (14.9) with Equation (14.21) shows that the vaporization enthalpy is related to the lattice model pair interaction energy w_{AA} through

$$\Delta h_{\text{vap}} = -\frac{zw_{AA}}{2}. \quad (14.22)$$

Figure 14.8 illustrates that the enthalpy of vaporization comes from breaking bonds, removing a particle, and closing a cavity in a liquid.

Table 14.1 Enthalpies of fusion (Δh_f) and boiling (Δh_{vap}) and the corresponding transition temperatures (T_f and T_b) at various pressures.

Substance	Fusion ^a		Evaporation ^b	
	T_f (K)	Δh_f (kJ mol ⁻¹)	T_b (K)	Δh_{vap} (kJ mol ⁻¹)
He	3.5	0.021	4.22	0.084
Ar	83.81	1.188	87.29	6.506
H ₂	13.96	0.117	20.38	0.916
N ₂	63.15	0.719	77.35	5.586
O ₂	54.36	0.444	90.18	6.820
Cl ₂	172.1	6.41	239.1	20.41
Br ₂	265.9	10.57	332.4	29.45
I ₂	386.8	15.52	458.4	41.80
Hg	234.3	2.292	629.7	59.30
Ag	1234	11.30	2436	250.6
Na	371.0	2.601	1156	98.01
CO ₂	217.0	8.33	194.6	25.23 ^c
H ₂ O	273.15	6.008	373.15	40.656 ^d
NH ₃	195.4	5.652	239.7	23.35
H ₂ S	187.6	2.377	212.8	18.67
CH ₄	90.68	0.941	111.7	8.18
C ₂ H ₆	89.85	2.86	184.6	14.7
C ₆ H ₆	278.61	10.59	353.2	30.8
CH ₃ OH	175.2	3.16	337.2	35.27

^aVarious pressures; ^bat 1 atm; ^csublimation; ^d 44.016 at 298.15 K.

Source: PW Atkins, *Physical Chemistry*, 6th edition, WH Freeman, San Francisco, 1997.

Cavities in Liquids and Solids

Molecular processes in liquids and solids often involve creating or filling cavities or changing a state of *solvation*. When drugs bind to proteins, water must first be stripped away from the drug and the protein. When a toxin molecule partitions from a watery to an oily environment, the process often involves first removing water molecules adjacent to the toxin. Figure 14.8(a) shows that the energy cost ΔU_{remove} of removing one particle, leaving behind an open cavity, is

$$\Delta U_{\text{remove}} = -zw_{AA}, \quad (14.23)$$

if the cavity doesn't distort or adjust after the particle is removed. The energy goes up when the particle is removed ($\Delta U_{\text{remove}} > 0$) because this process involves bond breaking. Of the z bonds that are broken, $z/2$ are associated with the molecule that leaves (all its bonds are broken), and $z/2$ are associated with the molecules that line the cavity that is left behind.

Another process of interest is the removal of one particle and the subsequent closure of the cavity. $\Delta U_{\text{remove+close}}$ must equal $U(N-1) - U(N)$, the

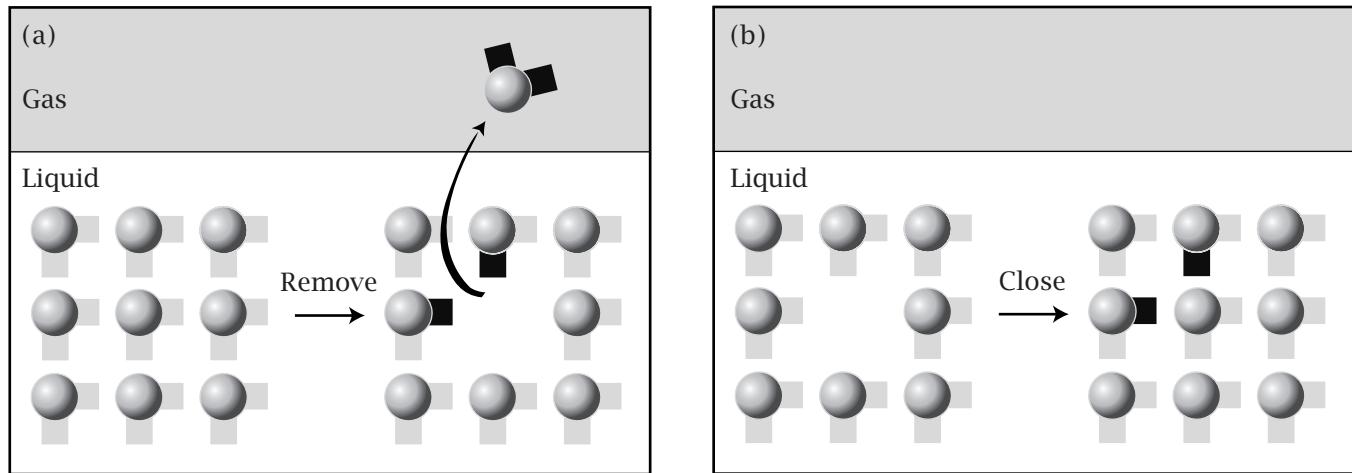


Figure 14.8 (a) Removing a particle and leaving a cavity behind (reducing the particle number by one) breaks the $z/2$ bonds of the escaping particle, and the $z/2$ bonds of the cavity molecules, so $\Delta U_{\text{remove}} = -(z/2 + z/2)w_{AA} = -zw_{AA}$. (b) Closing a cavity (at fixed particle number) makes $z/2$ bonds, so $\Delta U_{\text{close}} = zw_{AA}/2$.

energy difference between a system of $N-1$ particles and a system of N particles. Using Equation (14.6) for $U(N)$ gives

$$\Delta U_{\text{remove+close}} = U(N-1) - U(N) = -\frac{zw_{AA}}{2}. \quad (14.24)$$

By taking the difference between Equations (14.24) and (14.23), you can also see that the energy costs of opening and closing a cavity are

$$\begin{aligned}\Delta U_{\text{close}} &= \frac{zw_{AA}}{2}, \\ \Delta U_{\text{open}} &= -\frac{zw_{AA}}{2}\end{aligned} \quad (14.25)$$

(see Figure 14.8(b)).

Is it better to think of particles escaping from the surface or from the interior of the liquid? It doesn't matter, because the chemical potential of a molecule must be the same at the surface or in the bulk. Otherwise the system and its surface would not be in equilibrium. In Chapters 15 and 16, we will find Equations (14.23)–(14.25) useful for interpreting solvation and partitioning processes.

Vapor–liquid equilibria are often of practical importance. Let's look at their role in refrigerators and heat pumps.

How Do Refrigerators and Heat Pumps Work?

A refrigerator uses external energy to absorb heat from a cold place (inside the refrigerator box) and to dump the heat in a warmer place (outside the refrigerator). A heat pump does the same. It pumps heat from a cold place (outside a house in winter, for example) to a warmer place (inside the house). In both types of device, a 'working fluid' is pumped around a system of tubes and undergoes repeated thermodynamic cycles of vaporization and condensation.

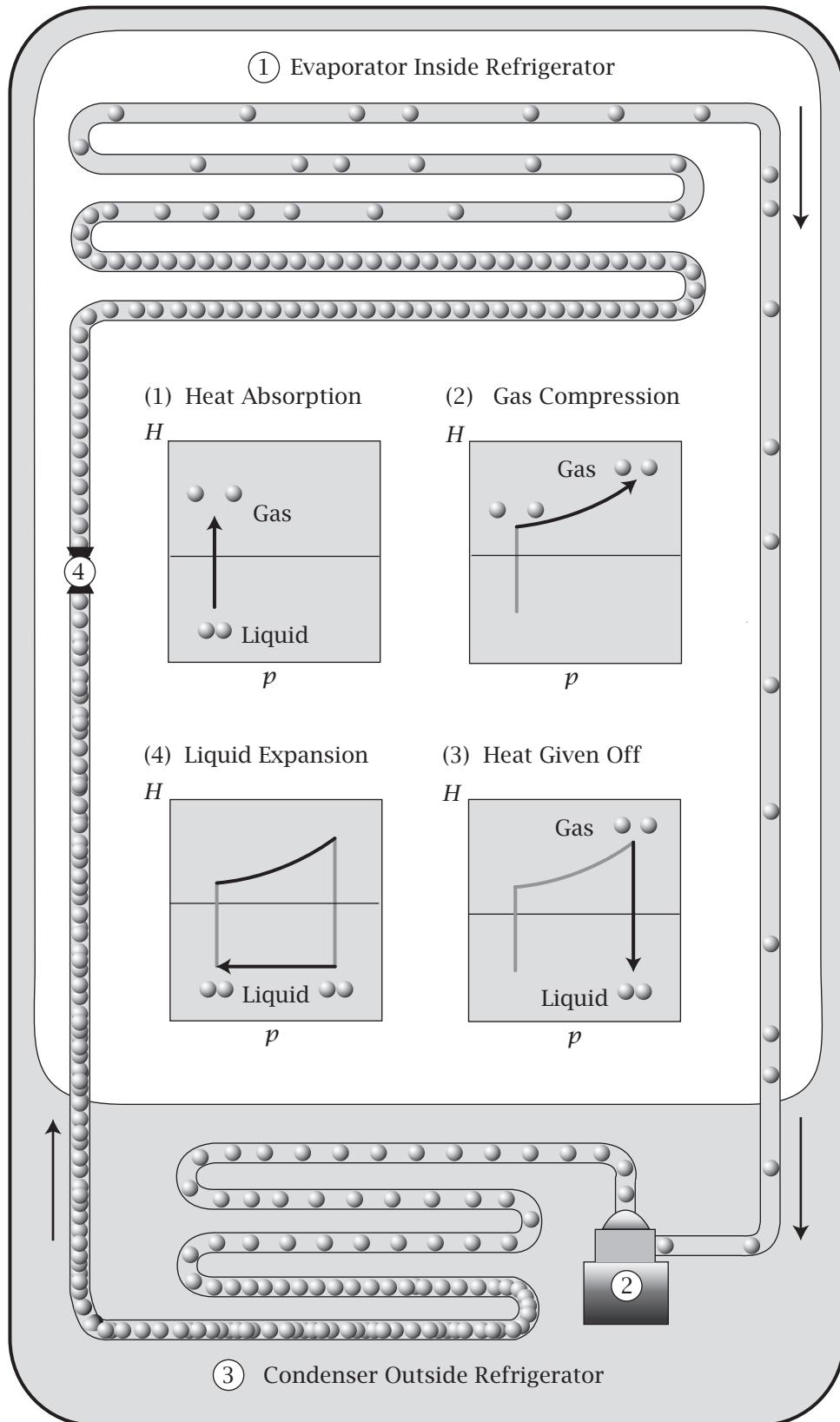
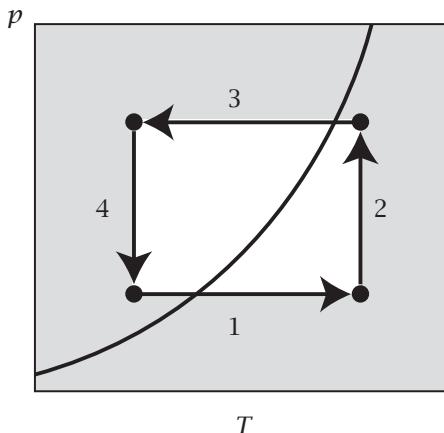


Figure 14.9 The thermodynamic cycle of a refrigerator. The spheres indicate the relative densities of the refrigerant molecules as they move through the system. There are four steps in the cycle: (1) Heat is absorbed from inside the refrigerator compartment to boil the working fluid. (2) The fluid is compressed and pumped. (3) Heat is dumped outside. The gas refrigerant re-condenses into a liquid. (4) The fluid is expanded until it is ready to boil.

Figure 14.10 The p - T cycle of a refrigerator. The fluid is liquid above the curved line and vapor below it. (1) The working fluid takes up heat and boils, converting from liquid to gas. (2) The compressor compresses the working-fluid gas. (3) The working fluid dumps heat behind the refrigerator, lowering its temperature, and condensing back to liquid. (4) An expansion valve reduces the pressure of the liquid working fluid, poised it for the next boiling cycle.



Refrigerators and heat pumps are based on two principles: (1) that boiling stores energy by breaking noncovalent bonds and condensation gets that energy back and (2) that a fluid can be boiled at a low temperature and recondensed at a high temperature by controlling the pressure. Figure 14.9 shows the four steps in an idealized refrigeration cycle. The graphs of enthalpy versus pressure show the stages where the noncovalent bonds of the working fluid break (high enthalpy) and form (low enthalpy). Figure 14.10 shows the same four steps on a pressure–temperature diagram instead, showing how you can pump heat uphill by cyclic pressure changes. Alternating steps of constant temperature and constant pressure are called *Ericsson cycles*.

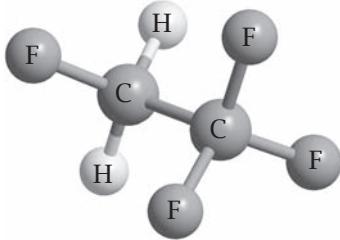


Figure 14.11
1,1,1,2-Tetrafluoroethane, a modern refrigerant, circulates through refrigerators, undergoing repeated cycles of boiling and condensation.

(1) Absorbing heat from inside the cold refrigerator. Before step 1, the working fluid is in the liquid state at low pressure, poised to boil. In step 1, the working fluid flows through the coils inside the refrigeration box, absorbing heat from the food compartment and cooling the inside of the refrigerator. This heat uptake boils the fluid, breaking the liquid–liquid bonds, increasing the enthalpy. The choice of fluid is based on its ability to boil at an appropriately low temperature and pressure. A typical refrigerant (working fluid) is 1,1,1,2-tetrafluoroethane, which has the molecular structure shown in Figure 14.11 and boils at $T = -26.1^\circ\text{C}$. Older refrigerants such as Freons contained chlorines, which react and deplete the ozone layer of the Earth, so they are now less common.

(2) Compressing the working fluid. The low-pressure gas is then compressed to a gas at higher pressure. The gas is now under sufficiently high pressure that it is poised to condense to the liquid state.

(3) Dumping the heat outside the refrigerator. The working gas now flows through coils mounted on the outside of the refrigeration box. Heat is given off to the surroundings, condensing the working gas into a liquid under high pressure. This step results in a decrease in enthalpy, because the liquefaction reforms liquid–liquid bonds. The heat captured from inside the refrigerator in step 1 is now released to the surrounding room in step 3.

(4) Reducing the pressure of the working fluid. The fluid flows through an expansion valve, where its pressure is reduced without much enthalpy change. The liquid is again at low pressure, poised for vaporization, and the cycle repeats. This is how vaporization cycles can be used to pump heat from cold places to hot places.

In this chapter, we have focused on liquid–vapor equilibria. The same ideas can be applied to other phase equilibria. For example, the sublimation pressure of a gas over a solid can be computed from Equation (14.21) by replacing the enthalpy of vaporization with the enthalpy of sublimation. Now we consider another process of particle exchange, not between condensed phase and vapor, but between the interior and the surface of a liquid.

Surface Tension Describes the Equilibrium Between Molecules at the Surface and in the Bulk

A *surface* is defined as the boundary between a condensed phase and a gas or vapor. More generally, an *interface* is defined as the boundary between any two media. *Surface tension* is the free energy cost of increasing the surface area of the system. For example, when a water droplet is spherical, it has the smallest possible ratio of surface to volume. When the droplet changes shape, its surface gets larger relative to its volume. Water tends to form spherical droplets because deviations away from spherical shapes are opposed by the surface tension. Here is a model.

Consider a lattice condensed phase of N molecules. If n of the molecules are at the surface, then $N-n$ are in the ‘bulk’ (the interior of the phase, away from the surface; see Figure 14.12). Bulk particles have z nearest neighbors. Surface particles have only $z-1$ nearest neighbors because they have one side exposed at the surface. Taking this difference into account, the total energy of the surface plus the bulk can be computed according to Equation (14.6):

$$U = \left(\frac{zw_{AA}}{2} \right) (N-n) + \left(\frac{(z-1)w_{AA}}{2} \right) n = \frac{w_{AA}}{2} (Nz - n). \quad (14.26)$$

The surface tension is defined as the derivative of the free energy with respect to the total area \mathcal{A} of the surface: $\gamma = (\partial F / \partial \mathcal{A})_{T,V,N}$. Because the lattice liquid has $S = 0$ and $F = U$,

$$\gamma = \left(\frac{\partial F}{\partial \mathcal{A}} \right)_{T,V,N} = \left(\frac{\partial F}{\partial n} \right)_{T,V,N} \left(\frac{dn}{d\mathcal{A}} \right) = \left(\frac{\partial U}{\partial n} \right)_{T,V,N} \left(\frac{dn}{d\mathcal{A}} \right). \quad (14.27)$$

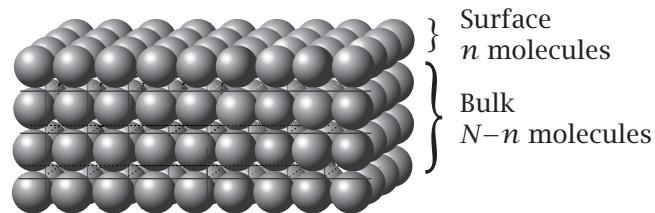


Figure 14.12 Lattice model of surface tension. Of N total molecules, n are at the surface. The surface tension is the free energy of moving molecules from the bulk to the surface, per unit area.

Use the definition of U in Equation (14.26) to take the derivative:

$$\left(\frac{\partial U}{\partial n} \right) = \frac{-w_{AA}}{2}. \quad (14.28)$$

The total area of the surface is $\mathcal{A} = na$, where a is the area per particle. So the second part of the derivative that you need for Equation (14.27) is

$$\left(\frac{dn}{d\mathcal{A}} \right) = \frac{1}{a}.$$

Therefore, the lattice model gives

$$\gamma = \frac{-w_{AA}}{2a} \quad (14.29)$$

for the surface tension.

The surface tension is a positive quantity (w_{AA} is negative). It describes the free energy cost of transferring particles from the bulk to the surface to increase the surface area. Surface tensions are greatest for the materials with the strongest intermolecular attractive forces. Table 14.2 shows that surface tensions are high for liquid metals, such as silver, iron, and mercury, in which free-flowing electrons bind atoms tightly together. This is why a mercury

Table 14.2 The surface tensions γ of some solids and liquids.

Material	γ (erg cm ⁻²)	T (°C)
W (solid) ^a	2900	1727
Au (solid) ^a	1410	1027
Ag (solid) ^a	1140	907
Ag (liquid) ^b	879	1100
Fe (solid) ^a	2150	1400
Fe (liquid) ^b	1880	1535
Pt (solid) ^a	2340	1311
Hg (liquid) ^b	487	16.5
NaCl (solid) ^c	227	25
KCl (solid) ^c	110	25
CaF ₂ (solid) ^c	450	-195
He (liquid) ^b	0.308	-270.5
Ethanol (liquid) ^b	22.75	20
Water (liquid) ^b	72.75	20
Benzene (liquid) ^b	28.88	20
<i>n</i> -Octane (liquid) ^b	21.80	20
Carbon tetrachloride ^b (liquid)	26.95	20
Nitrobenzene ^b (liquid)	25.2	20

Source: G Somorjai, *Introduction to Surface Chemistry and Catalysis*, Wiley, New York, 1994.

^aJM Blakeley and PS Maiya, *Surfaces and Interfaces*, JJ Burke et al., eds, Syracuse University Press, Syracuse, NY, 1967.

^bAW Adamson, *Physical Chemistry of Surfaces*, Wiley, New York, 1967.

^cGC Benson and RS Yuen, *The Solid-Gas Interface*, EA Flood, ed., Marcel Dekker, New York, 1967.

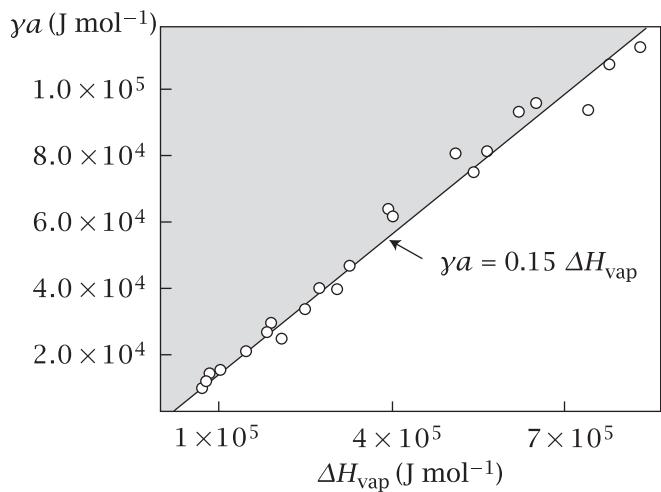


Figure 14.13 Vaporization and surface tension γ both involve breaking bonds, modeled through the quantity w_{AA} . The slope, $\gamma \times \text{area } (a)$, versus ΔH_{vap} indicates that 6.6-fold more bonds are broken per atom in vaporization than in creating surface, over a range of liquid metals. Source: redrawn from SH Overbury, PA Bertrand, and GA Somorjai, *Chem Rev* **75**, 547–560 (1975).

droplet is so spherical: it strongly seeks the shape of minimum possible surface. For water, which is hydrogen-bonded, the surface tension is 72.8 erg cm^{-2} at 20°C . For *n*-hexane, which is bound together only by dispersion forces, the surface tension is 18.4 erg cm^{-2} at 20°C . Adding surfactants to water decreases the surface tension, resulting in the increased surface area that you see in the foams on beer and soapy water.

Measuring various equilibria can give information about the intermolecular interaction energies w_{AA} . Different measures are related to each other. For example, Δh_{vap} and γ are both linear functions of w_{AA} , so Δh_{vap} and γ should be linearly related to each other. From the lattice model, the surface tension of a liquid is predicted to be $1/z$ times the energy of vaporization per unit area of the particle. Figure 14.13 confirms this approximate relationship between surface tensions and vaporization enthalpies for liquid metals.

Summary

Vaporization equilibria involve a balance of forces. Particles stick together at low temperatures, but they vaporize at high temperatures to gain translational entropy. The chemical potential describes the escaping tendency: μ_v from the vapor phase, and μ_c from the condensed phase. When these escaping tendencies are equal, the system is at equilibrium and no net particle exchange takes place. We used the statistical mechanical model of Chapter 11 for the chemical potential of the gas phase. For the liquid or solid phase, we used a lattice model. Vapor pressure and surface tension equilibria, combined with models, give information about intermolecular interactions. In the next chapter, we will go beyond pure phases and consider the equilibria of mixtures.

Problems

1. Applying the Clausius–Clapeyron equation.

- (a) The vapor pressure of water is 23 mmHg at $T = 300\text{ K}$ and 760 mmHg at $T = 373\text{ K}$. Calculate the enthalpy of vaporization Δh_{vap} .
- (b) Assuming that each water has $z = 4$ nearest neighbors, calculate the interaction energy w_{AA} .

2. How does surface tension depend on temperature?

If the surface tension of a pure liquid is due entirely to energy (and no entropy), will the surface tension *increase*, *decrease*, or *not change* with increasing temperature?

3. Why do spray cans get cold?

Explain why an aerosol spray can gets cold when you spray the contents.

4. The surface tension of water.

The surface tension of water $\gamma(T)$ decreases with temperature as shown in Figure 9.13.

- (a) As a first-order approximation, estimate the water–water pair attraction energy w_{AA} , assuming that the interaction is temperature-independent.
- (b) Explain the basis for this attraction.
- (c) Explain why γ decreases with T .

5. Applying the Clausius–Clapeyron equation again.

Suppose you want to add a perfuming agent to a liquid aerosol spray. You want its vapor pressure to double if the temperature increases from 25°C to 35°C . Calculate the pair interaction energy that this agent should have in the pure liquid state. Assume that the number of nearest-neighbor molecules in the liquid is $z = 6$.

6. How does the boiling temperature depend on the enthalpy of vaporization?

A fluid has a boiling temperature $T = 300\text{ K}$ at $p = 1\text{ atm}$; $\Delta h_{\text{vap}} = 10\text{ kcal mol}^{-1}$. Suppose you make a molecular modification of the fluid that adds a hydrogen bond so that the new enthalpy is $\Delta h_{\text{vap}} = 15\text{ kcal mol}^{-1}$ at $p = 1\text{ atm}$. What is the new boiling temperature?

7. Trouton's rule.

- (a) Using Table 14.1, show that the entropy of vaporization $\Delta s_{\text{vap}} = \Delta h_{\text{vap}}/T_b$ is relatively constant for a broad range of different substances. Δh_{vap} is the enthalpy of vaporization and T_b is the boiling temperature. This is called Trouton's rule.

- (b) Why are the values of Δs_{vap} for water and CO_2 in Table 14.1 larger than for other substances?

8. Sublimation of graphite.

The heat of sublimation of graphite is $\Delta h_{\text{sub}} = 716.7\text{ kJ mol}^{-1}$. Use this number to estimate the strength of a carbon–carbon bond.

9. Surface tension of mercury.

The surface tension of water is 72 erg cm^{-2} and that of liquid mercury is 487 erg cm^{-2} . If the water–water attraction is about 5 kcal mol^{-1} , what is the mercury–mercury attraction?

10. Squeezing ice.

Use the Clapeyron relation to compute the pressure that is required to lower the melting temperature of ice by 10 K . For water, $\Delta h = 6.008\text{ kJ mol}^{-1}$ and $\Delta v = -1.64\text{ cm}^3\text{ mol}^{-1}$.

Suggested Reading

RS Berry, SA Rice, and J Ross, *Physical Chemistry*, 2nd edition, Oxford University Press, New York, 2000. A complete and advanced text on physical chemistry, including thermodynamics.

TD Eastop and A McConkey, *Applied Thermodynamics for Engineering Technologists*, 5th edition, Longman, Harlow, Essex, UK, 1993. An engineering perspective on refrigerators and heat pumps.

Good treatments of phase boundaries and the Clausius–Clapeyron equation:

PW Atkins and J de Paula, *Physical Chemistry*, 8th edition, WH Freeman, San Francisco, 2006.

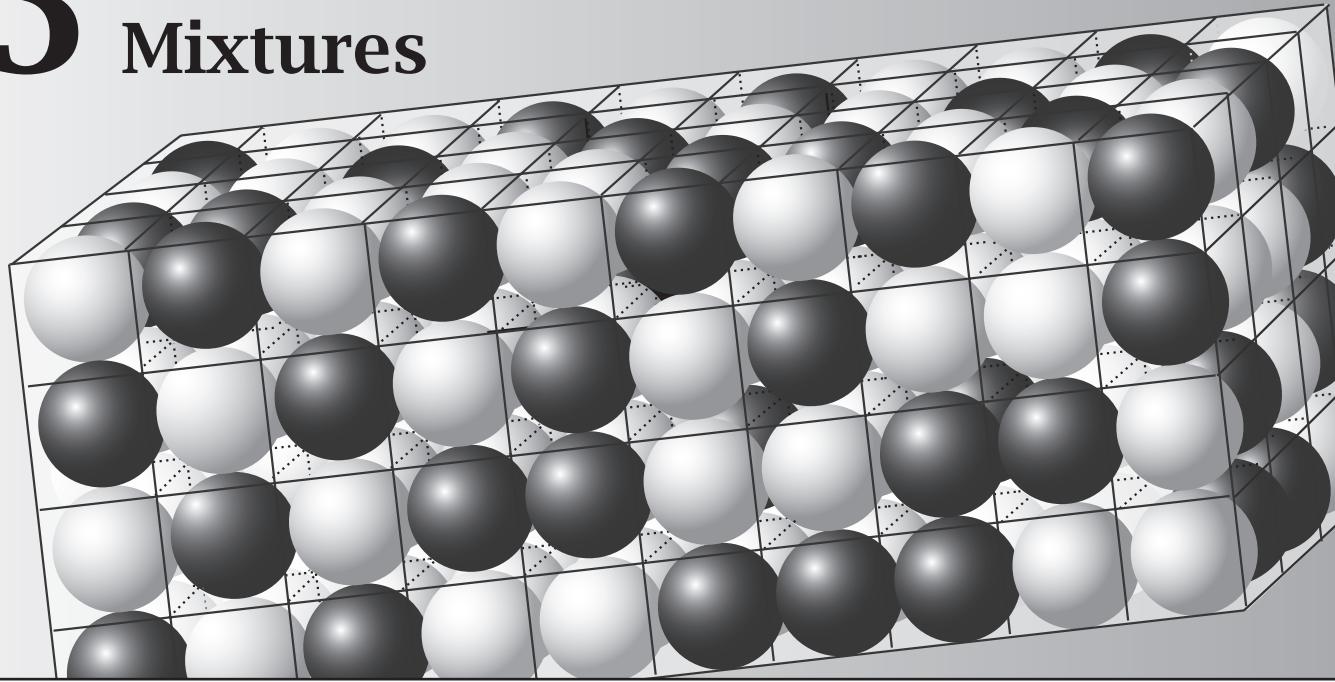
DA McQuarrie and JD Simon, *Physical Chemistry: A Molecular Approach*, University Science Books, Sausalito, CA, 1997.

Surface tension is discussed in:

PC Hiemenz, *Principles of Colloid and Surface Chemistry*, 3rd edition, Marcel Dekker, New York, 1997.

J Lyklema, *Fundamentals of Interface and Colloid Science, Volume I: Fundamentals*, Academic Press, San Diego, 1991.

15 Solutions & Mixtures



A Lattice Model Describes Mixtures

In Chapter 14, we considered pure liquids or solids composed of a single chemical species. Here we consider *solutions*, i.e., homogeneous mixtures of more than one component. The fundamental result that we derive in this chapter, $\mu = \mu^\circ + kT \ln x$, is a relationship between a molecule's chemical potential μ and its concentration x in the solution. This relationship will help address questions in Chapters 16 and 25–31: When does one chemical species dissolve in another? When is it insoluble? How do solutes lower the freezing point of a liquid, elevate the boiling temperature, and cause osmotic pressure? What forces drive molecules to partition differently into different phases? We continue with the lattice model because it gives simple insights and because it gives the foundation for treatments of polymers, colloids, and biomolecules.

We use the (T, V, N) ensemble, rather than (T, p, N) , because it allows us to work with the simplest possible lattice model that captures the principles of solution theory. The appropriate extremum principle is based on the Helmholtz free energy $F = U - TS$, where S is the entropy of solution and U accounts for the interaction energies between the lattice particles.

The Entropy of Solution

Suppose there are N_A molecules of species *A* and N_B molecules of species *B*. Particles of *A* and *B* are the same size—each occupies one lattice site—and

A B

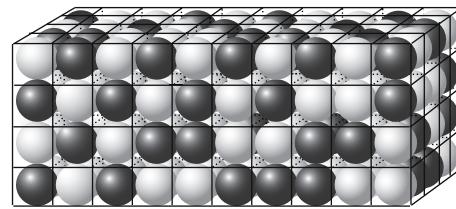


Figure 15.1 A lattice mixture of two components *A* and *B*. The number of *A*'s is N_A and the number of *B*'s is N_B . The total number of lattice sites is $N = N_A + N_B$. All sites are filled.

together they completely fill a lattice of N lattice sites (see Figure 15.1):

$$N = N_A + N_B. \quad (15.1)$$

(See Chapter 32 for the Flory–Huggins treatment for particles of different sizes, such as polymers in simple solvents.)

The multiplicity of states is the number of spatial arrangements of the molecules:

$$W = \frac{N!}{N_A! N_B!}. \quad (15.2)$$

The translational entropy of the mixed system can be computed by using the Boltzmann equation (5.1), $S = k \ln W$, and Stirling's approximation, Equation (B.3):

$$\begin{aligned} \Delta S_{\text{solution}} &= k(N \ln N - N_A \ln N_A - N_B \ln N_B) \\ &= k(N_A \ln N + N_B \ln N - N_A \ln N_A - N_B \ln N_B) \\ &= -Nk \left[\frac{N_A}{N} \ln \left(\frac{N_A}{N} \right) + \frac{N_B}{N} \ln \left(\frac{N_B}{N} \right) \right] \\ &= -k(N_A \ln x_A + N_B \ln x_B). \end{aligned} \quad (15.3)$$

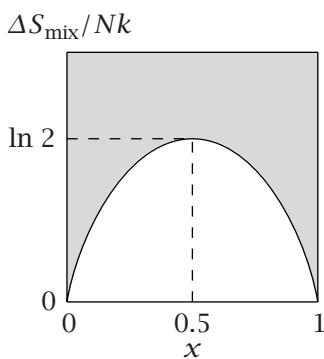


Figure 15.2 The entropy of solution as a function of the mole fraction x .

This entropy can be expressed in terms of the relative *concentrations* of *A* and *B*. We will express concentrations as *mole fractions* $x = N_A/N$ and $(1-x) = N_B/N$. Equation (15.3) gives the entropy of solution of a binary (two-component) solution as

$$\frac{\Delta S_{\text{solution}}}{Nk} = -x \ln x - (1-x) \ln(1-x). \quad (15.4)$$

Figure 15.2 shows how the entropy of solution $\Delta S_{\text{solution}}/Nk$ depends on x , the mole fraction of *A*, according to Equation (15.4).¹ The process is illustrated in Figure 15.3. Put N_A molecules of *A* into a container that has N_B molecules of *B*. The composition x is now fixed. If the two materials did not mix with each other, there would be no change in the entropy, relative to the two pure fluids: $\Delta S_{\text{solution}} = 0$. If the two materials do mix in the random way that we have assumed, the entropy will increase from 0, for the unmixed state, to $\Delta S_{\text{solution}}(x)$, the value shown for that particular composition x in Figure 15.2.

¹ $\Delta S_{\text{solution}}$ is often called the 'mixing entropy,' but the entropy actually arises from the greater volume available to each particle type [1]: N_A particles of *A* begin in N_A sites and end up distributed in $N_A + N_B$ sites, for example.

The components mix because of the multiplicity of ways of intermingling the A 's with the B 's. This is the driving force that causes atoms and molecules to mix.

Don't confuse x with a degree of freedom. The system doesn't change its composition toward $x = 1/2$ to reach the highest entropy. x is the composition that is fixed by the number of A and B molecules put into the solution. The solution entropy is maximal at $x = 1/2$ because this is the solution composition that gives the highest multiplicity of all possible compositions. Also, note that this entropy arises from arrangements of particles at fixed overall density and not from the mixing of holes and particles that we used to describe gas pressures.

EXAMPLE 15.1 Mixing entropy (entropy of solution). For a solution containing a mole fraction of 20% methanol in water, compute the entropy of solution. Equation (15.4) gives

$$\begin{aligned}\frac{\Delta S_{\text{solution}}}{N} &= R(-0.2 \ln 0.2 - 0.8 \ln 0.8) \\ &= (1.987 \text{ cal K}^{-1} \text{ mol}^{-1})(0.5) \approx 1.0 \text{ cal K}^{-1} \text{ mol}^{-1}\end{aligned}$$

If there were no interaction energy, the free energy of solution at $T = 300 \text{ K}$ would be

$$\begin{aligned}\frac{\Delta F_{\text{solution}}}{N} &= \frac{-T \Delta S_{\text{solution}}}{N} \\ &= -300 \text{ cal mol}^{-1}.\end{aligned}$$

Ideal Solutions

A solution is called *ideal* if its free energy of solution is given by $\Delta F_{\text{solution}} = -T \Delta S_{\text{solution}}$, with $\Delta S_{\text{solution}}$ taken from Equation (15.4). Mixing an ideal solution involves no change in energy. And it involves no other entropies due to changes in volume, or structuring, or ordering in the solution. Chapter 25 describes solubility and insolubility in more detail. Now we move to a model of mixtures that is more realistic than the ideal solution model.

The Energy of Solution

In practice, few solutions are truly ideal. Real solutions involve energies of solution. In the lattice model, the total energy of solution is the sum of the contact interactions of noncovalent bonds of all the pairs of nearest neighbors in the mixture. For a lattice solution of components A and B , Figure 15.4 shows the three possible types of contact: an AA bond, a BB bond, or an AB bond. There are no other options, because the lattice is completely filled by A 's and B 's.

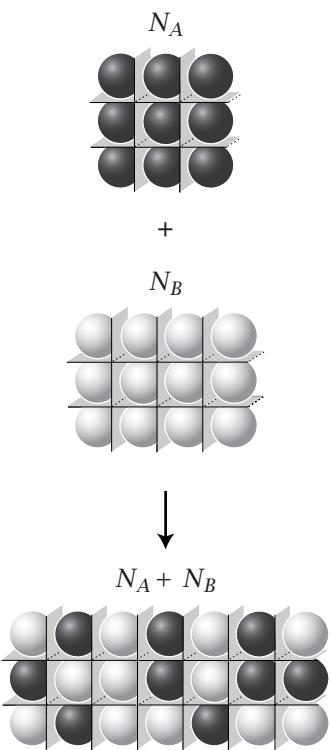


Figure 15.3 Mixing is a process that begins with N_A molecules of pure A and N_B molecules of pure B , and combines them into a solution of $N_A + N_B$ molecules.

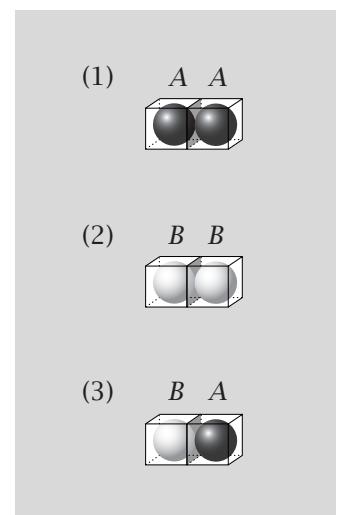


Figure 15.4 Three types of contact (or bond) occur in a lattice mixture of components A and B .

The total energy of the system is the sum of the individual contact energies over the three types of contact:

$$U = m_{AA}w_{AA} + m_{BB}w_{BB} + m_{AB}w_{AB}, \quad (15.5)$$

where m_{AA} is the number of AA bonds, m_{BB} is the number of BB bonds, m_{AB} is the number of AB bonds, and w_{AA}, w_{BB}, w_{AB} are the corresponding contact energies. As noted in Chapter 14, the quantities w are negative.

In general, the numbers of contacts, m_{AA} , m_{BB} , and m_{AB} , are not known. To put Equation (15.5) into a more useful form, you can express the quantities m in terms of N_A and N_B , the known numbers of A's and B's. Each lattice site has z 'sides,' just as in Chapter 14. Figure 15.5 shows that every contact involves two sides. The total number of sides of type A particles is zN_A , which can be expressed in terms of the numbers of contacts as

$$zN_A = 2m_{AA} + m_{AB}, \quad (15.6)$$

because the total number of A sides equals

$$(\text{number of AA bonds}) \times \left(\frac{2 \text{ A sides}}{\text{AA bond}} \right) + (\text{number of AB bonds}) \times \left(\frac{1 \text{ A side}}{\text{AB bond}} \right).$$

Similarly, for type B particles,

$$zN_B = 2m_{BB} + m_{AB}. \quad (15.7)$$

Solve Equations (15.6) and (15.7) for the number of AA bonds m_{AA} and for the number of BB bonds m_{BB} :

$$m_{AA} = \frac{zN_A - m_{AB}}{2} \quad \text{and} \quad m_{BB} = \frac{zN_B - m_{AB}}{2}. \quad (15.8)$$

Substitute Equations (15.8) into Equation (15.5) to arrive at an expression for the total interaction energy, in which the only unknown term is now m_{AB} , the number of AB contacts:

$$\begin{aligned} U &= \left(\frac{zN_A - m_{AB}}{2} \right) w_{AA} + \left(\frac{zN_B - m_{AB}}{2} \right) w_{BB} + m_{AB}w_{AB} \\ &= \left(\frac{zw_{AA}}{2} \right) N_A + \left(\frac{zw_{BB}}{2} \right) N_B + \left(w_{AB} - \frac{w_{AA} + w_{BB}}{2} \right) m_{AB}. \end{aligned} \quad (15.9)$$

Now we use the Bragg-Williams, or *mean-field*, approximation to evaluate m_{AB} [2–4].

The Mean-Field Approximation

Different arrangements of the system's particles will have different values of m_{AB} . In principle, we should consider each configuration of the system, and we should account for its appropriate Boltzmann weight (for more discussion, see page 278). This would lead to sophisticated models. Here we explore a much simpler approach that gives many of the same insights. We make an assumption, called the mean-field approximation, that for any given numbers N_A and N_B , the particles are mixed as randomly and uniformly as possible. This gives us a way to estimate m_{AB} .

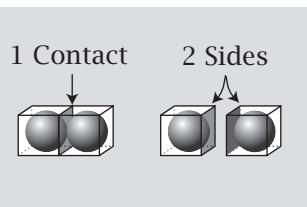


Figure 15.5 One contact between lattice particles involves two lattice site sides.

Consider a specific site next to an *A* molecule. What is the probability that a *B* occupies that neighboring site? In the Bragg–Williams approximation, you assume that the *B*'s are randomly distributed throughout all the sites. The probability p_B that any site is occupied by *B* equals the fraction of all sites that are occupied by *B*'s:

$$p_B = \frac{N_B}{N} = x_B = 1 - x. \quad (15.10)$$

Because there are z nearest-neighbor sites for each molecule of *A*, the average number of *AB* contacts made by that particular molecule of *A* is $zN_B/N = z(1-x)$. The total number of molecules of *A* is N_A , so

$$m_{AB} \approx \frac{zN_A N_B}{N} = zN_x(1-x). \quad (15.11)$$

Now compute the total contact energy of the mixture from the known quantities N_A and N_B by substituting Equation (15.11) into Equation (15.9):

$$\begin{aligned} U &= \left(\frac{zw_{AA}}{2}\right)N_A + \left(\frac{zw_{BB}}{2}\right)N_B + z\left(w_{AB} - \frac{w_{AA} + w_{BB}}{2}\right)\frac{N_A N_B}{N} \\ &= \left(\frac{zw_{AA}}{2}\right)N_A + \left(\frac{zw_{BB}}{2}\right)N_B + kT\chi_{AB}\frac{N_A N_B}{N}, \end{aligned} \quad (15.12)$$

where we define a dimensionless quantity called the **exchange parameter** χ_{AB} :

$$\chi_{AB} = \frac{z}{kT} \left(w_{AB} - \frac{w_{AA} + w_{BB}}{2} \right) \quad (15.13)$$

(see also page 275).

How does the Bragg–Williams approximation err? If *AB* interactions are more favorable than *AA* and *BB* interactions, then *B*'s will prefer to sit next to *A*'s more often than the mean-field assumption predicts. Or, if the self-attractions are stronger than the attractions of *A*'s for *B*'s, then *A*'s will tend to cluster together, and *B*'s will cluster together, more than the random mixing assumption predicts. Nevertheless, the Bragg–Williams mean-field expression is often a reasonable first approximation.

The Free Energy of Solution

Now combine terms to form the free energy $F = U - TS$, using Equation (15.3) for the entropy and Equation (15.12) for the energy:

$$\begin{aligned} \frac{F(N_A, N_B)}{kT} &= N_A \ln\left(\frac{N_A}{N}\right) + N_B \ln\left(\frac{N_B}{N}\right) \\ &\quad + \left(\frac{zw_{AA}}{2kT}\right)N_A + \left(\frac{zw_{BB}}{2kT}\right)N_B + \chi_{AB}\frac{N_A N_B}{N}. \end{aligned} \quad (15.14)$$

$F(N_A, N_B)$ is the free energy of a mixed solution of N_A *A*'s and N_B *B*'s, totaling $N = N_A + N_B$ particles.

Typically, we are interested in the free energy *difference* between the mixed final state and the initial pure states of *A* and *B*, $\Delta F_{\text{solution}}$ (see Figure 15.3):

$$\Delta F_{\text{solution}} = F(N_A, N_B) - F(N_A, 0) - F(0, N_B). \quad (15.15)$$

$F(N_A, 0) = zw_{AA}N_A/2$ is the free energy of a pure system of N_A A's, which is found by substituting $N = N_A$ and $N_B = 0$ into Equation (15.14). Similarly, $F(0, N_B) = zw_{BB}N_B/2$ is the free energy of a pure system of N_B B's. $F(N_A, N_B)$, the free energy of the mixed final state, is given by Equation (15.14). Substitute these three free energy expressions into Equation (15.15) and divide by N to get the free energy of solution in terms of the mole fraction x and the interaction parameter χ_{AB} :

$$\frac{\Delta F_{\text{solution}}}{NkT} = x \ln x + (1 - x) \ln(1 - x) + \chi_{AB}x(1 - x). \quad (15.16)$$

This model was first described by JH Hildebrand in 1929, and is called the *regular solution model* [5]. While ideal solutions are driven only by the entropy of solution of particles of roughly equal size, regular solutions are driven also by the energy of the mean-field form described above. Solutions having free energies of solution of the form in Equation (15.16) are called regular solutions. Sometimes energies are more complicated than in this simple model: in electrolytes, for example; or when molecules are rod-shaped, as in solutions of liquid crystals; or when A has a very different size than B, as in polymer solutions. Polymers are described in the last chapters of this book. Here's an application of Equation (15.16).

EXAMPLE 15.2 Oil and water don't mix. Show that oil and water don't mix unless one component is very dilute. A typical value is $\chi \approx 5$ for hydrocarbon/water interactions. If $x_{\text{oil}} = 0.3$ and $T = 300\text{ K}$, Equation (15.16) gives the free energy of solution as

$$\begin{aligned} \frac{\Delta F_{\text{solution}}}{N} &= [0.3 \ln 0.3 + 0.7 \ln 0.7 + 5(0.3)(0.7)](8.314\text{ J K}^{-1}\text{ mol}^{-1})(300\text{ K}) \\ &= 1.1\text{ kJ mol}^{-1}. \end{aligned}$$

Because this value is positive, it predicts that oil and water won't mix to form a random solution of this composition. If the oil is very dilute, $x_{\text{oil}} = 10^{-4}$, then

$$\begin{aligned} \frac{\Delta F_{\text{solution}}}{N} &= [10^{-4} \ln 10^{-4} + 0.9999 \ln 0.9999 + 5(10^{-4})(0.9999)] \\ &\quad \times (8.314\text{ J K}^{-1}\text{ mol}^{-1})(300\text{ K}) = -1.3\text{ J mol}^{-1}, \end{aligned}$$

which is negative, so mixing is favorable.

In general, determining solubility is not as simple as these calculations imply, because another option is available to the system—it may separate into phases with different compositions. Phase separation, for which this model is a starting point, is described in Chapter 25.

The Chemical Potentials

The chemical potential for A in this lattice model of a two-component mixture is found by taking the derivative $\mu_A = (\partial F / \partial N_A)_{T, N_B}$ (see Equation (9.31)) of F

(Equation (15.14)) with respect to N_A , holding N_B (not N) constant:

$$\begin{aligned}\frac{\mu_A}{kT} &= \left[\frac{\partial}{\partial N_A} \left(\frac{F}{kT} \right) \right]_{T, N_B} \\ &= \ln \left(\frac{N_A}{N} \right) + 1 - \frac{N_A}{N} - \frac{N_B}{N} + \frac{zw_{AA}}{2kT} + \chi_{AB} \frac{(N_A + N_B)N_B - N_A N_B}{(N_A + N_B)^2} \\ &= \ln x_A + \frac{zw_{AA}}{2kT} + \chi_{AB}(1 - x_A)^2.\end{aligned}\quad (15.17)$$

Similarly, the chemical potential for B is

$$\frac{\mu_B}{kT} = \left[\frac{\partial}{\partial N_B} \left(\frac{F}{kT} \right) \right]_{T, N_A} = \ln x_B + \frac{zw_{BB}}{2kT} + \chi_{AB}(1 - x_B)^2. \quad (15.18)$$

The main result is an equation of the form

$$\mu = \mu^\circ + kT \ln \gamma x, \quad (15.19)$$

where γ (not to be confused with the surface tension) is called the *activity coefficient*.

The lattice model leads to a small ambiguity. In principle, our degrees of freedom are (T, V, N_A, N_B) . But the volume is not independent of the total particle number because the lattice is fully filled and constrained by $N_A + N_B = N$. Because the lattice contains no empty sites, this model does not treat pV effects. We regard the relevant constraints as (T, N_A, N_B) and neglect the pV term. This is why you hold N_B constant, rather than N , when taking the derivative with respect to N_A .

Equations (15.17)–(15.19) provide the foundation for our treatment of mixing, solubility, partitioning, solvation, and colligative properties in Chapter 16. We will find that the chemical potential describes the escaping tendency for particles to move from one phase or one state to another. Particles are driven by at least two tendencies. First particles of A tend to leave regions of high concentration of A and move toward regions of low concentration of A to gain solution entropy (this is described by the term $kT \ln x$). Second, particles of A are attracted to regions or phases for which they have high chemical affinity, described in the next chapter by the quantities μ° and χ_{AB} .

What's the physical interpretation of the exchange parameter χ_{AB} ? From the steps leading to Equation (15.13), you can determine that χ_{AB} describes the energetic cost of beginning with the pure states of A and B and transferring one B into a medium of pure A 's and one A into a medium of pure B 's (see Figure 15.6):



where K_{exch} is the equilibrium constant for the exchange process. Think of the A 's as little boys at a dance and the B 's as little girls at a dance. When $\chi_{AB} > 0$, it means the boys and girls don't want to mix. The larger the value of χ_{AB} , the more they don't want to mix. On the other hand, *older* boys and girls at a dance *do* want to mix, so in those cases $\chi_{AB} < 0$. When $\chi_{AB} = 0$, it means there is no preference either way, and mixing happens freely. χ_{AB} does not include the translational entropy, so imagine beginning this exchange by choosing a particular A and B located at fixed spatial positions in their respective media.

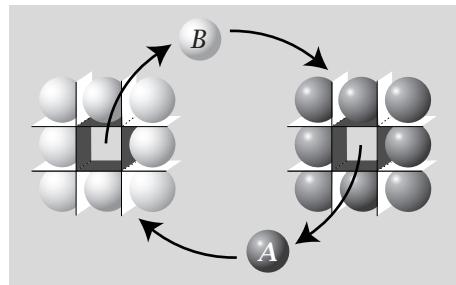


Figure 15.6 The quantity $2\chi_{AB}$ is the energy divided by kT for the process that begins with pure components A and B , and exchanges an A for a B .

If this process is favorable in the direction of the arrow in Equation (15.20) and Figure 15.6, $K_{\text{exch}} > 1$. In that case, χ_{AB} , which has units of energy divided by kT , is negative. According to *Hildebrand's principle*, for most systems the AB affinity is weaker than the AA and BB affinities, so usually $\chi_{AB} > 0$. The quantity χ_{AB} also contributes to the interfacial free energy between two materials, which we take up in the next section.

In the next few chapters, we apply this theory of mixtures to liquids. However, this lattice theory is also widely used for studying properties of solids. For example, metal alloys are mixtures of metals: *brass* is copper and zinc; *bronze* is tin and copper; *solder*, used in electronics, is tin and lead; and *sterling silver* is mostly silver with a little copper. Metals are driven by the solution entropy to form these alloys.

Interfacial Tension Describes the Free Energy of Creating Surface Area

The boundary between two condensed phases is an *interface*. The interfacial tension γ_{AB} is the free energy cost of increasing the interfacial area between phases A and B . If γ_{AB} is large, the two media will tend to minimize their interfacial contact. Let's determine γ_{AB} by using the lattice model for molecules of types A and B that are identical in size.

Suppose there are N_A molecules of A , n of which are at the interface, and there are N_B molecules of B , n of which are at the interface in contact with A (see Figure 15.7). (Because the particles have the same size in this model, there will be the same number n of each type for a given area of interfacial contact). The total energy of the system is treated as it was for surface tension

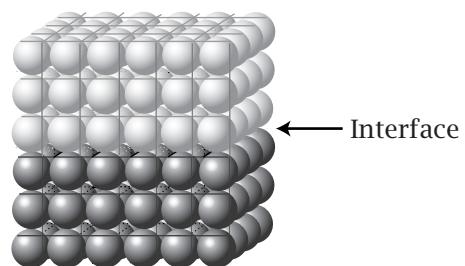


Figure 15.7 Lattice model of interfacial tension.

(see Equation (14.26)), with the addition of n AB contacts at the interface:

$$U = (N_A - n) \left(\frac{zw_{AA}}{2} \right) + n \left(\frac{(z-1)w_{AA}}{2} \right) + nw_{AB} \\ + (N_B - n) \left(\frac{zw_{BB}}{2} \right) + n \left(\frac{(z-1)w_{BB}}{2} \right). \quad (15.21)$$

Because the entropy of each bulk phase is zero according to the lattice model, the interfacial tension is defined by

$$\gamma_{AB} = \left(\frac{\partial F}{\partial \mathcal{A}} \right)_{N_A, N_B, T} = \left(\frac{\partial U}{\partial \mathcal{A}} \right)_{N_A, N_B, T} = \left(\frac{\partial U}{\partial n} \right)_{N_A, N_B, T} \left(\frac{dn}{d\mathcal{A}} \right), \quad (15.22)$$

where \mathcal{A} is the total area of the surface in lattice units. You have $dn/d\mathcal{A} = 1/a$, where a is the area per molecule exposed at the surface. Take the derivative ($\partial U/\partial n$) using Equation (15.21):

$$\left(\frac{\partial U}{\partial n} \right)_{N_A, N_B, T} = w_{AB} - \frac{w_{AA} + w_{BB}}{2}. \quad (15.23)$$

You can assemble an expression for γ_{AB} from Equations (15.13), (15.22), and (15.23):

$$\gamma_{AB} = \frac{1}{a} \left(w_{AB} - \frac{w_{AA} + w_{BB}}{2} \right) = \left(\frac{kT}{za} \right) \chi_{AB}. \quad (15.24)$$

When there are no molecules of B , $w_{AB} = w_{BB} = 0$, and γ_{AB} reduces to the surface tension given by Equation (14.29), $-w_{AA}/2a$.

EXAMPLE 15.3 Estimating χ_{AB} from interfacial tension experiments. The interfacial tension between water and benzene at 20°C is 35 dyn cm⁻¹. Suppose $z = 6$ and the interfacial contact area of a benzene with water is about 3 Å × 3 Å = 9.0 Å². Determine χ_{AB} at 20°C. Equation (15.24) gives

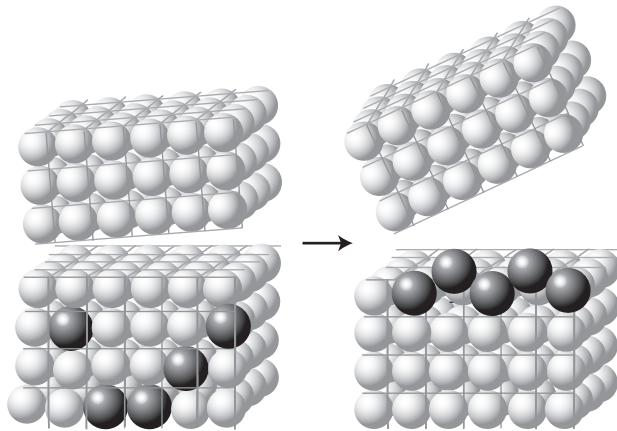
$$\chi_{AB} = \frac{zay_{AB}}{kT} = \frac{(6)(9 \text{ Å}^2)(35 \times 10^{-7} \text{ J cm}^{-2}) \left(\frac{1 \text{ cm}}{10^8 \text{ Å}} \right)^2}{(1.38 \times 10^{-23} \text{ JK}^{-1})(293 \text{ K})} = 4.7.$$

χ_{AB} is a dimensionless quantity, and $RT\chi_{AB} = 11.3 \text{ kJ mol}^{-1} = 2.7 \text{ kcal mol}^{-1}$ is the exchange energy.

EXAMPLE 15.4 Where does the interfacial tension matter? Interfacial tension—from either covalent or noncovalent bonding—holds paint on walls and barnacles on ships. It holds your cells together so your body doesn't fall apart, and it helps snails and other slimy critters to crawl. Cell-adhesion proteins help hold cells to surfaces and other cells; they are critical in cancer. Water on your waxed car shapes itself into spherical beads to minimize its contact with the wax, because χ_{AB} between water and wax is large. Water beads are flatter on a rusty hood than a waxed hood, since χ_{AB} is smaller between water and rusty metal. Figure 15.8 shows how impurities in metals and other solids

can migrate to *grain boundaries* and *dislocations*, which are interfaces between different microparticles of solid materials. This can lead to weakening or breakage of solid materials, sometimes causing catastrophic structural failures of materials.

Figure 15.8 Impurity atoms (large dark spheres) in metals and solids (atoms shown as small spheres) can migrate to boundaries between individual crystals. This is a major source of the microscopic cracks and fractures that cause the weakening and breakage of solid materials.



What Have We Left Out?

There are two main ways in which our lattice model is a simplification. First, the partition function Q should be a sum over all the possible states of the system. It should be a sum of the number of states with small numbers of AB contacts, m_{AB} , and states with large m_{AB} , rather than a mean-field estimate of the number of uniformly mixed conformations (see Figure 15.9). In this way, we have made the approximation that

$$Q = \sum_{m_{AB}} W(N_A, N_B, m_{AB}) e^{-E(N_A, N_B, m_{AB})/kT} \\ \approx \frac{N!}{N_A! N_B!} e^{-U/kT}, \quad (15.25)$$

where E is the energy of a configuration having m_{AB} contacts and W is the density of states, the number of configurations having the given value of m_{AB} (see page 177). U is the mean-field average energy from Equation (15.12).

The second approximation that we made was leaving out the quantum mechanical degrees of freedom: rotations, vibrations, and electronic configurations. Those degrees of freedom normally are the same in the pure phase as in the mixed phase. In *differences* of thermodynamic quantities, like $\Delta F_{\text{solution}}$, such terms cancel, so you are at liberty to leave them out in the first place. It is only the quantities that *change* in a process that need to be taken into account. Only the intermolecular interactions and the translational entropy change in simple mixing processes. But for chemical reactions in solution,

quantum mechanics contributes too (see Chapter 16). Taking internal degrees of freedom into account gives

$$Q = q_A^{N_A} q_B^{N_B} \sum_{m_{AB}} W(N_A, N_B, m_{AB}) e^{-E(N_A, N_B, m_{AB})/kT}$$

$$\approx q_A^{N_A} q_B^{N_B} \frac{N!}{N_A! N_B!} e^{-U/kT}, \quad (15.26)$$

where q_A and q_B are the partition functions for the rotations, vibrations, and electronic states of molecules A and B , respectively. Using $F = -kT \ln Q$ (Equation (10.42)) with Equation (15.26), you get a general expression for the free energy of solution that includes the quantum mechanical contributions in the mean-field lattice model:

$$\frac{F}{kT} = N_A \ln \left(\frac{N_A}{N} \right) + N_B \ln \left(\frac{N_B}{N} \right) + \left(\frac{zw_{AA}}{2kT} \right) N_A + \left(\frac{zw_{BB}}{2kT} \right) N_B$$

$$+ \chi_{AB} \frac{N_A N_B}{N} - N_A \ln q_A - N_B \ln q_B. \quad (15.27)$$

Similarly, the generalizations of Equation (15.17) for the chemical potential are

$$\frac{\mu_A}{kT} = \ln \left(\frac{x_A}{q_A} \right) + \frac{zw_{AA}}{2kT} + \chi_{AB}(1-x_A)^2,$$

and

$$\frac{\mu_B}{kT} = \ln \left(\frac{x_B}{q_B} \right) + \frac{zw_{BB}}{2kT} + \chi_{AB}(1-x_B)^2. \quad (15.28)$$

The quantum mechanical components cancel in quantities such as $\Delta F_{\text{solution}}$ when the internal degrees of freedom are unaffected by the mixing process. To see this, use Equation (15.27) to get the pure state components for Equation (15.15),

$$F(N_A, 0) = N_A [(zw_{AA}/2) - \ln q_A] \text{ and } F(0, N_B) = N_B [(zw_{BB}/2) - \ln q_B],$$

and subtract. You get the same result as in Equation (15.16):

$$\frac{\Delta F_{\text{solution}}}{NkT} = x \ln x + (1-x) \ln(1-x) + \chi_{AB}x(1-x).$$

Summary

We have developed a model for the thermodynamic properties of ideal and regular solutions. Two components A and B will tend to mix because of the favorable entropy resulting from the many different ways of interspersing A and B particles. The degree of mixing also depends on whether the AB attractions are stronger or weaker than the AA and BB attractions. In the next chapters, we will apply this model to the properties of solutions.

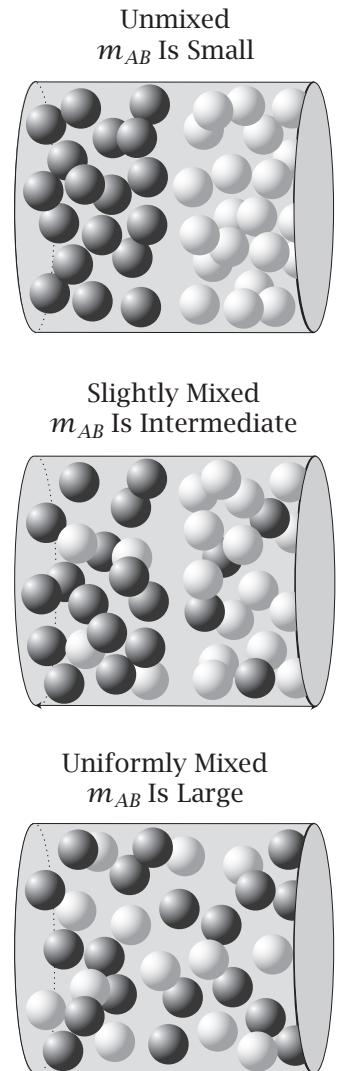


Figure 15.9 Particles of A and B can mix to different degrees to have different numbers m_{AB} of AB contacts. The partition function Q is a sum over all these states, but the mean-field model assumes uniform mixing.

Problems

1. Ternary lattice mixtures. Consider a lattice model liquid mixture of three species of spherical particles: *A*, *B*, and *C*. As with binary mixtures, assume that all $N = n_A + n_B + n_C$ sites are filled.

- (a) Write an expression for the entropy of solution.
- (b) Using the Bragg–Williams approximation, write an expression for the energy *U* of solution in terms of the binary interaction parameters χ .
- (c) Write an expression for the chemical potential μ_A of *A*.

2. Enthalpy of solution. For a mixture of benzene and *n*-heptane having equal mole fractions $x = 1/2$ and temperature $T = 300\text{ K}$, the enthalpy of solution is $\Delta H_{\text{solution}} = 220\text{ cal mol}^{-1}$. Compute χ_{AB} .

3. Plot $\mu(x)$. Plot the chemical potential versus x for:

- (a) $\chi_{AB} = 0$,
- (b) $\chi_{AB} = 2$,
- (c) $\chi_{AB} = 4$.

4. Limit of $x \ln x$ terms in solution entropies and free energies. What is the value of $x \ln x$ as $x \rightarrow 0$? What is the implication for the entropy of solution?

5. Hydrophobic entropy. The experimentally determined total entropy of dissolving benzene in water at high dilution is approximately $14\text{ cal K}^{-1}\text{ mol}^{-1}$ at $T = 15^\circ\text{C}$.

- (a) How does this total entropy compare with the solution entropy component?
- (b) Speculate on the origin of this entropy.

6. Solubility parameters. The quantity χ_{AB} describes *AB* interactions relative to *AA* and *BB* interactions. If instead of a two-component mixture, you had a mixture of *N* different species *A*, *B*, ..., then, to obtain the pairwise quantities χ_{ij} for all of them, you would need $\sim N^2$ experiments involving mixtures of all the components $i = 1, 2, \dots, N$ with all the components $j = 1, 2, \dots, N$. However, sometimes this can be much simpler. If all the species are nonpolar, you can make the approximation $w_{AB} \approx \sqrt{w_{AA} w_{BB}}$ (see Chapter 24). Show how this reduces the necessary number of experiments to only $\sim N$.

7. Self-assembly of functional electrical circuits. G Whitesides and colleagues have pioneered the formation of nano- and mesoscale structures based on the self-assembly of patterned components [6]. Consider a circuit made from the self-assembly of *N* different building block components. When shaken up together, each component must find its proper place in a two-dimensional square lattice for the circuit to function correctly. Using Figure 15.10, suppose that if the letters are correctly ordered as on the left, each unit interacts pairwise with its neighbor through an interaction energy w_{match} . However, if a letter is surrounded by an incorrect neighbor, it interacts

more weakly, with an energy w_{mismatch} . Find the contact energy difference

$$\Delta w = w_{\text{match}} - w_{\text{mismatch}},$$

necessary to ensure that the circuit can be reliably self-assembled. That is, the fraction of working circuits should be higher than δ , where δ is some small value. Ignore shape and surface effects (assume each of the *N* components makes $z = 4$ contacts) and only consider the equilibrium between a perfectly formed circuit and all possible misformed circuits in which two components have swapped places. Assume that *N* is large enough that the defective components are not in contact with one another.

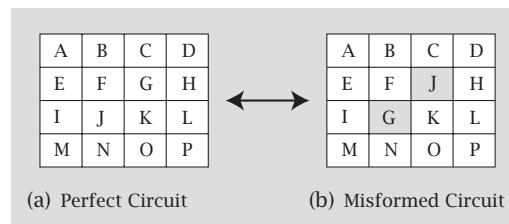


Figure 15.10 Comparing a correctly self-assembled two-dimensional circuit of parts with an incorrectly assembled circuit [6]. (a) A perfect Whitesides circuit. (b) One of the possible misformed Whitesides circuits.

References

- [1] A Ben-Naim, *A Farewell to Entropy: Statistical Thermodynamics Based on Information*, World Scientific, Singapore, 2008.
- [2] PJ Flory, *Principles of Polymer Chemistry*, Cornell University Press, Ithaca, NY, 1953.
- [3] TL Hill, *An Introduction to Statistical Thermodynamics*. Addison-Wesley, Reading, MA, 1960 (reprinted by Dover Publications, New York, 1986).
- [4] R Kubo, *Statistical Mechanics*, 2nd edition, North-Holland, New York, 1987.
- [5] JH Hildebrand and RL Scott. *Regular Solutions*. Prentice-Hall, Englewood Cliffs, 1962.
- [6] DH Gracias, J Tien, TL Breen, et al., *Science* **289**, 1170–1172 (2000).

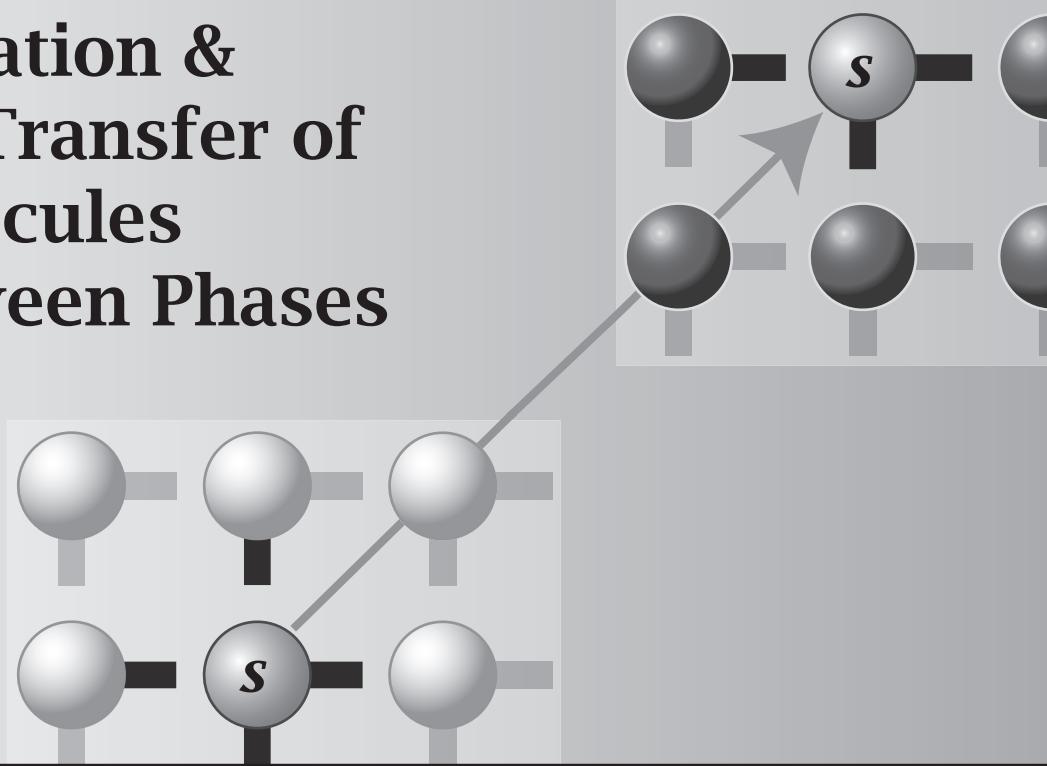
Suggested Reading

Excellent discussions of the lattice model of mixing:
PJ Flory, *Principles of Polymer Chemistry*, Cornell University Press, Ithaca, NY, 1953.

- TL Hill, *Introduction to Statistical Thermodynamics*, Reading, MA, 1960.
- R Kubo with H Ichimura, T Usui, and N Hashitsume, *Statistical Mechanics*, North-Holland, New York, 1965.
- Good treatments of liquids, regular solutions, and surfaces:
- JS Rowlinson and FL Swinton, *Liquids and Liquid Mixtures*, 3rd edition, Butterworth, London, 1982. A complete and extensive discussion of mixing, including much experimental data.
- JH Hildebrand and RL Scott, *Regular Solutions*, Prentice-Hall, Englewood Cliffs, 1962. The regular solution model is described in detail.
- JS Rowlinson and B Widom, *Molecular Theory of Capillarity*, Clarendon Press, Oxford, 1982 (reprinted by Dover Publications, New York, 2002). An excellent advanced discussion of surfaces and interfacial tensions.
- WE Acree, Jr, *Thermodynamic Properties of Nonelectrolyte Solutions*, Academic Press, San Diego, 1984.

This page is intentionally left blank.

16 Solvation & the Transfer of Molecules Between Phases



The Chemical Potential Describes the Tendency of Molecules to Exchange and Partition

We now consider the solvation, partitioning, and transfer of molecules from one medium to another. For example, toxins dissolve in one medium, such as water, and *partition* or *transfer* into another medium, such as the oil components of fish. A central type of action in biology—the binding of a drug or metabolite to a protein—requires first *desolvation*, i.e., the stripping away of water molecules from the ligand and the receptor. Drugs that act on the brain must be designed not only for their ability to bind their biological targets but also for their ability to transport across an interface called the blood-brain barrier. Related to partitioning and solvation are *colligative properties*: vapor pressure depression by solutes, boiling temperature elevation, freezing temperature depression, and osmotic pressure. For example, salt dumped on icy roads will melt the ice. Salt added to a boiling pot of water will reduce the boiling. Reverse osmosis is a way to purify salt water to make it drinkable. In this chapter, we use the liquid mixture lattice model to illustrate, in a simple approximate way, the molecular basis for solvation and colligative processes.

What drives transfer and partitioning processes? On one hand, atoms and molecules move from regions of high concentration to regions of low concentrations, to increase the entropy of the system. On the other hand, molecules also move into regions for which they have high chemical affinity. In this chapter, we explore the chemical potential, an expression of these driving forces.

While the free energy describes the tendency of a whole system toward equilibrium, the chemical potential describes the tendency toward equilibrium for each individual chemical component. In a mixture, the low-concentration component is called the *solute* and the high-concentration component is the *solvent*.

Solvation Is the Transfer of Molecules Between Vapor and Liquid Phases

Recall from Chapter 14 that liquids have a vapor pressure: some water molecules will escape to the vapor phase over a container of pure liquid water. Now add a little salt. Interestingly, water's vapor pressure over salty water is lower than over pure water. Why does adding *salt* reduce the vapor pressure of the *water*? While Chapter 14 focused on pure liquids, this chapter focuses on mixtures and solutions.

Consider a liquid mixture of two components, *A* and *B*. Suppose *B* is volatile and *A* is not. That is, *B* exchanges freely between the liquid and the vapor phase, but *A* stays in the liquid phase. Think about salt water. Because water is much more volatile than salt, we label water as *B* and salt as *A*. The degree of freedom is the number of *B* molecules in the gas phase, $N(\text{gas})$, or the number of *B* molecules in the liquid phase, $N(\text{liquid})$. The total number of *B* molecules is fixed: $N(\text{gas}) + N(\text{liquid}) = \text{constant}$. Component *A* does not exchange, so you don't need to write an equilibrium for it.

At constant temperature and pressure, the *B* molecules will be at equilibrium when they have the same chemical potential in the gas phase as in the solution phase:

$$\mu_B(\text{gas}) = \mu_B(\text{liquid}). \quad (16.1)$$

The Lattice Solution Model for Solvation and Vapor Pressure Depression

To find the relationship between the vapor pressure p_B of the molecules of *B* and the solution concentration x_B , use Equation (11.50), $\mu_B(\text{gas}) = kT \ln(p_B / p_{B,\text{int}}^\circ)$, for the chemical potential of *B* in the vapor phase. For the chemical potential of *B* in the solution, use the lattice model equation (15.17), $\mu_B(\text{liquid}) = kT \ln x_B + z w_{BB} / 2 + kT \chi_{AB}(1 - x_B)^2$. Substitute these equations into the equilibrium equation (16.1) and exponentiate to get

$$\frac{p_B}{p_{B,\text{int}}^\circ} = x_B \exp \left[\chi_{AB}(1 - x_B)^2 + \frac{z w_{BB}}{2kT} \right]. \quad (16.2)$$

Equation (16.2) can be expressed more compactly as

$$p_B = p_{B,\text{int}}^\circ x_B \exp \left[\chi_{AB}(1 - x_B)^2 \right], \quad (16.3)$$

where $p_{B,\text{int}}^\circ$ is the vapor pressure of *B* over a pure solvent *B*:

$$p_{B,\text{int}}^\circ = p_{B,\text{int}}^\circ \exp \left(\frac{z w_{BB}}{2kT} \right). \quad (16.4)$$

Figure 16.1 plots Equation (16.3): the vapor pressure of *B* versus the concentration x_B of the volatile component in solution. You can use this in

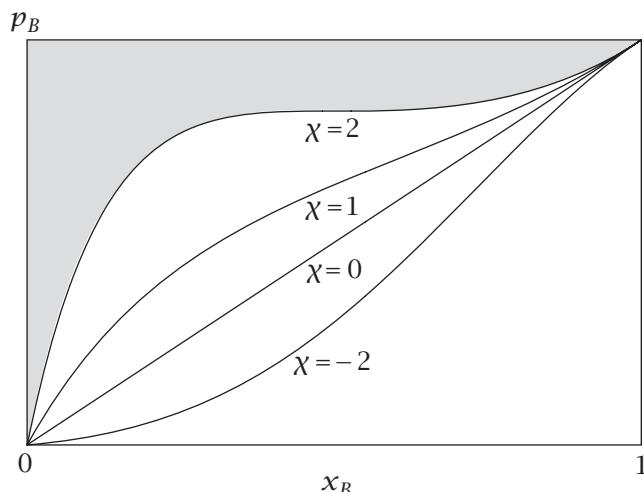


Figure 16.1 The vapor pressure p_B of molecules of B over a liquid solution having composition x_B , according to the lattice model. Increasing the concentration x_B of B in the liquid increases the vapor pressure of B . $\chi = 0$ represents an ideal solution. B has more escaping tendency when $\chi > 0$ and less when $\chi < 0$, for a given x_B .

two ways: (1) Put a concentration x_B of molecules of volatile B in solution, and Equation (16.3) will tell you the vapor pressure of B over the solution. Or, (2) apply a vapor pressure p_B , and Equation (16.3) will tell you what concentration of B dissolves in the liquid. Look at the line indicated by $\chi_{AB} = 0$. The linear dependence, $p_B = p_B^\circ x_B$, defines an *ideal solution*, also called *Raoult's law*. Start at the right side of the figure: $x_B = 1$ describes the state of pure water with no salt. Moving slightly to the left on the figure, $x_B < 1$ describes a system of slightly salty water. Adding salt reduces the vapor pressure of the water. This is not a consequence of energetic preferences, because $\chi_{AB} = 0$ along this line. Rather, you can trace this reduced vapor pressure to the solution entropy (see Figure 16.2). The number of distinguishable arrangements of water molecules is $W = 1$ in pure water, whereas the number of arrangements of water molecules in the salt water mixture is $W > 1$. A mixture of A 's and B 's has more solution entropy than a liquid of pure B 's, so a molecule of B would rather be in an AB mixture than in a liquid of pure B . The exchangeable component B is drawn to 'dilute out' the non-exchangeable component A .

So far, we have described ideal solutions, $\chi_{AB} = 0$. Now consider nonideal solutions, $\chi_{AB} \neq 0$. When A and B energetically 'dislike' each other ($\chi_{AB} > 0$), the molecules of B will have a higher vapor pressure than they would have had over an ideal solution, in order to escape contact with the molecules of A in the liquid; see Figure 16.1. But when A and B 'like' each other ($\chi_{AB} < 0$), the molecules of B will have a lower vapor pressure than over an ideal solution because molecules of B prefer to contact molecules of A in the liquid. Typically, nonpolar molecules have more affinity for molecules that are like themselves than for molecules that are different, so usually $\chi_{AB} > 0$. This is captured in the general rule that 'like dissolves like' (see Chapter 24).

So far, we looked at the right side of Figure 16.1 as a model for mixtures such as salty water. Now look at the left side of the figure, where the volatile species B is now the minority component ($x_B \rightarrow 0$). That could describe wet salt, i.e., mostly salt and very little water, but such problems are rarely of interest. For some problems, you will find the right side of Figure 16.1 to be useful (whenever the *solvent* is the exchangeable component). For other problems, you will find the left side of Figure 16.1 more useful (whenever the *solute* is the exchangeable

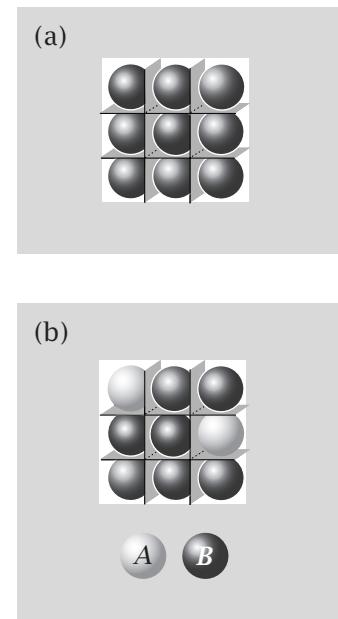


Figure 16.2 (a) For pure B , the number of arrangements is $W = 1$. (b) Mixing in A 's gives $W > 1$, increasing the entropy and reducing the escaping tendency μ of B .

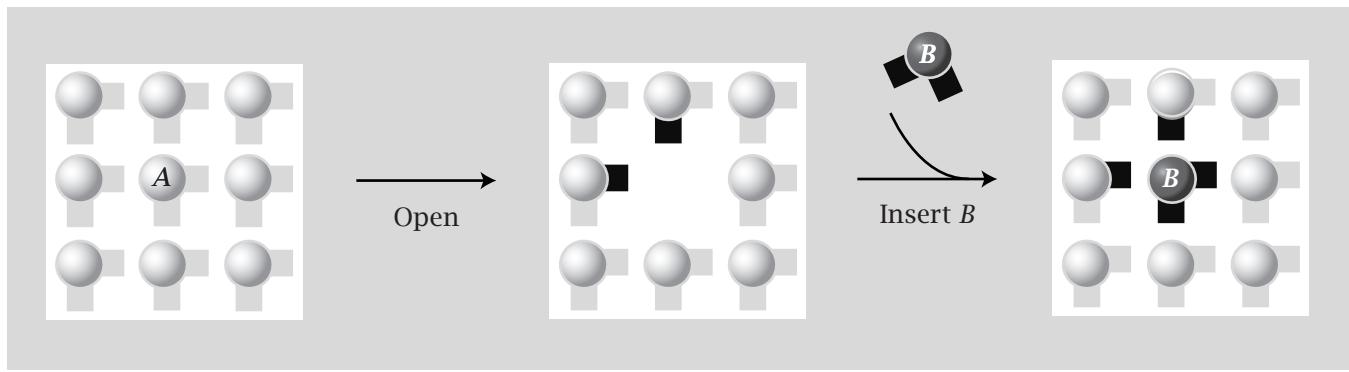


Figure 16.3 The microscopic process underlying the Henry's law coefficient k_H and μ_B° in the solute convention: opening a cavity in pure A (breaking $(z/2)$ AA contacts), then inserting a B (making z AB contacts).

component). For example, the left side of Figure 16.1 applies when a small amount of gas B is dissolved in a large amount of water A . In this case, the gas is much more volatile than the water, so now you would label the gas as the exchangeable component B and the water A as the nonvolatile component. For instance, high pressures of CO_2 gas (the volatile component B) over water (which is now the nonvolatile component) create carbonated beverages. To get the vapor pressure of B at low concentrations, substitute Equation (16.4) into (16.3) and let $x_B \rightarrow 0$. The result predicts that p_B is linearly proportional to the composition x_B :

$$\frac{p_B}{p_{B,\text{int}}^\circ} = x_B \exp\left(\chi_{AB} + \frac{zw_{BB}}{2kT}\right) = x_B \exp\left[\frac{z}{kT}\left(w_{AB} - \frac{w_{AA}}{2}\right)\right], \quad (16.5)$$

Table 16.1 Henry's law constants for gases in water at 25°C.

Gas	k_H (atm)
He	131×10^3
N_2	86×10^3
CO	57×10^3
O_2	43×10^3
CH_4	41×10^3
Ar	40×10^3
CO_2	1.6×10^3
C_2H_2	1.3×10^3

Source: I Tinoco, K Sauer, and JC Wang, *Physical Chemistry: Principles and Applications in Biological Sciences*, Prentice-Hall, Englewood Cliffs, 1978.

which can be expressed in terms of a slope k_H ,

$$p_B = k_H x_B. \quad (16.6)$$

k_H , which has units of pressure, is the *Henry's law constant*, given in our model by

$$k_H = p_{B,\text{int}}^\circ \exp\left[\frac{z}{kT}\left(w_{AB} - \frac{w_{AA}}{2}\right)\right] = p_{B,\text{int}}^\circ \exp\left(\frac{\Delta h_{\text{solution}}^\circ}{kT}\right). \quad (16.7)$$

The quantity $\Delta h_{\text{solution}}^\circ = [z(w_{AB} - w_{AA}/2)]$ is the enthalpy of solution and represents the process shown in Figure 16.3: a cavity is opened in A and a molecule of B is inserted. The linear regime on the left side of Figure 16.1 is known as the Henry's law region. Henry's law constants for gases in aqueous solutions at 25°C are given in Table 16.1. In the Henry's law equation (16.6), Equation (16.5) shows that B 's will concentrate in liquid A if either the AB affinity is strong or the AA affinity is weak. A weak AA affinity means that the medium A readily accommodates a molecule of B because it costs little energy to open a cavity in A .

EXAMPLE 16.1 Henry's law predicts effects of global warming on marine life. Marine organisms breathe oxygen that is dissolved in oceans. The eastern tropical North Atlantic ocean is warming at the rate of 0.9°C per century. Increased temperature will reduce the dissolved oxygen concentration in the ocean. To see this, first compute the oxygen concentration, x_B in pure water, approximating ocean water, at temperature $T_1 = 300\text{ K}$. Air is 21% oxygen, so the oxygen vapor pressure is 0.21 atm. Use Equation (16.6) and the value of the Henry's law constant for oxygen in water from Table 16.1:

$$x_B = \frac{p_B}{k_H} = \left(\frac{0.21 \text{ atm}}{43,000 \text{ atm}} \right) \left(\frac{1 \text{ mol H}_2\text{O}}{18 \text{ g}} \right) \left(\frac{1000 \text{ g}}{1 \text{ kg}} \right) = 270 \mu\text{mol kg}^{-1}. \quad (16.8)$$

The enthalpy of dissolution of oxygen gas in water is $\Delta h_{\text{solution}}^\circ = -3.4 \text{ kcal mol}^{-1}$, so $\Delta h_{\text{solution}}^\circ/R = -1700 \text{ K}$. Equation (16.8) shows that x_B is inversely proportional to k_H for fixed p_B . So, the change in dissolved oxygen concentration, from x_1 to x_2 , upon changing the temperature from T_1 to T_2 is

$$\begin{aligned} \ln\left(\frac{x_2}{x_1}\right) &= \ln\left(\frac{k_{H,1}}{k_{H,2}}\right) = \left(\frac{\Delta h_{\text{solution}}^\circ}{R}\right)\left(\frac{1}{T_1} - \frac{1}{T_2}\right) \\ &= -1700 \text{ K} \left(\frac{1}{300 \text{ K}} - \frac{1}{300.9 \text{ K}}\right) = -0.17. \end{aligned} \quad (16.9)$$

So, we estimate that the oxygen concentration in water at the higher temperature is $x_2 = \exp(-0.17)x_1 = 0.84x_1$. The loss in dissolved oxygen due to warming is $x_2 - x_1 = (0.84 - 1)(x_1) = -0.16x_1 = -0.16 \times 270 = -42 \mu\text{mol kg}^{-1}$. This is close to the value of $-34 \mu\text{mol kg}^{-1}$ that has been measured experimentally in that ocean [1].

A Thermodynamic Model Defines the Activity and Activity Coefficient

The models above—of ideal solutions and lattice mixtures—give basic general insights. But some mixtures are more complex. How can you get deeper insights from your data? First, measure the vapor pressure $p_B(x_B)$ as a function of the solution concentration of B , the volatile component. Then plot the ratio p_B/x_B over the range of interest of x_B . Equations (16.2) and (16.3) show that if this ratio is constant, independent of x_B , then your solution is ideal, and the constant is p_B° , and you can get insights about BB interactions using Equation (16.4). However, if your solution is more complex, the ratio p_B/x_B will be a more complicated function of x_B . In general, you can use the following thermodynamic model.

Express the gas-phase chemical potential of the exchangeable component B as

$$\mu_B(\text{gas}) = \mu_B^\circ(\text{gas}) + kT \ln p_B, \quad (16.10)$$

and the liquid-phase chemical potential as

$$\mu_B(\text{liquid}) = \mu_B^\circ(\text{liquid}) + kT \ln[\gamma_B(x_B)x_B]. \quad (16.11)$$

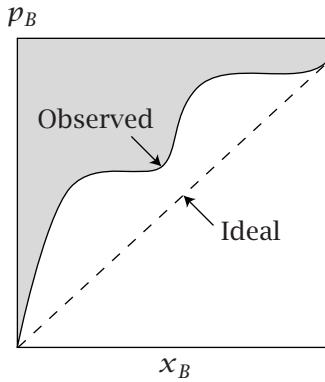


Figure 16.4 In an ideal solution, p_B is a linear function of x_B . A real solution can be more complex. The degree of nonideality is determined by the ratio p_B/x_B .

In equations such as (16.10), you must take the logarithm of a dimensionless quantity, so if $p_B = 5.2$ atm, for example, it implies there is a denominator of 1 atm. The function γ_B of x_B is called the *activity coefficient*. The product $\gamma(x_B) \times x_B$ is called the *activity*. The quantity $\Delta\mu_B^\circ = \mu_B^\circ(\text{liquid}) - \mu_B^\circ(\text{gas})$ is called the *standard-state chemical potential*. Setting equal the gas-phase and liquid-phase chemical potentials gives a general expression for the equilibrium vapor pressure,

$$\frac{p_B}{x_B} = \gamma_B(x_B) \exp\left(\frac{\Delta\mu_B^\circ}{kT}\right). \quad (16.12)$$

The strategy embodied in Equation (16.12) is to measure p_B/x_B versus x_B (as shown in Figure 16.4) and then to curve-fit Equation (16.12) using some function $\gamma_B(x_B)$ and a constant $\Delta\mu_B^\circ$. In some cases, you might obtain the function $\gamma_B(x_B)$ from a microscopic model; in other cases, you might just pick a simple math function.

In order to use this thermodynamic approach, you must first resolve one ambiguity. Equation (16.12) gives the experimental quantity p_B/x_B as a product of a function $\gamma_B(x_B)$ and a term containing $\Delta\mu_B^\circ$. However, any product can be factored in different ways (10 is either 5×2 or 10×1 , for example.) To resolve this ambiguity, we choose a *convention*. First, decide whether your problem is more focused on the right side of Figure 16.1, where the volatile species B is the majority (solvent) component, or the left side, where the volatile species is the minority (solute) component. If your exchangeable component is the solvent, then, by convention, you choose a function (the **solvent convention**) that satisfies the limit $\gamma_B(x_B) \rightarrow 1$ as $x_B \rightarrow 1$. If instead your exchangeable component is a solute, then you choose a function (the **solute convention**) having the property that $\gamma_B(x_B) \rightarrow 1$ as $x_B \rightarrow 0$. These choices remove any ambiguities for fitting experimental data. These choices also give you useful microscopic insights, as shown below.

Interpreting the Thermodynamic Quantities by Using the Lattice Model

To see the logic of these two conventions, here's a microscopic interpretation of μ_B° and γ_B , using the lattice model. Comparing Equation (15.19) with Equation (15.18) shows that

$$\frac{\mu_B^\circ}{kT} + \ln \gamma_B = \frac{zw_{BB}}{2kT} + \chi_{AB}(1-x_B)^2. \quad (16.13)$$

SOLVENT CONVENTION. To find the correspondence between the lattice and thermodynamic models in the solvent convention, let $\gamma_B(x_B) \rightarrow 1$ as $x_B \rightarrow 1$ to reduce Equation (16.13) to

$$\frac{\mu_B^\circ}{kT} = \frac{zw_{BB}}{2kT}. \quad (16.14)$$

When the solvent is volatile, you see that μ_B° represents the free energy cost of opening a cavity in pure B , and inserting a B .

To interpret the activity coefficient, substitute Equation (16.14) into Equation (16.13) and exponentiate:

$$\gamma_B = \exp[\chi_{AB}(1-x_B)^2]. \quad (16.15)$$

In this lattice model, the nonideal concentration dependence is due to the energies of the AB interactions.

SOLUTE CONVENTION. If B is the solute, substitute $\gamma_B \rightarrow 1$ as $x_B \rightarrow 0$ into Equation (16.13) to get

$$\begin{aligned} \frac{\mu_B^\circ}{kT} &= \frac{zw_{BB}}{2kT} + \chi_{AB} \\ &= \frac{z}{kT} \left(w_{AB} - \frac{w_{AA}}{2} \right), \end{aligned} \quad (16.16)$$

which describes the process shown in Figure 16.3. When the solute is the exchangeable component, μ_B° represents the free energy cost of inserting a B molecule at infinite dilution into otherwise pure solvent A . Substitute Equation (16.16) into Equation (16.13) to find the definition of γ_B in the solute convention:

$$\gamma_B = \exp[\chi_{AB}x_B(x_B-2)]. \quad (16.17)$$

These lattice-model comparisons show how the solute and solvent conventions give a quantity μ_B° that conveys insight about the molecular process involved in either limit $x_B \rightarrow 1$ or $x_B \rightarrow 0$. Any other definition of μ_B° would not have been as useful for microscopic insights, because it would have described mixed interactions of B with both A 's and B 's.

Now let's consider a different problem. How can a solute A change the phase equilibrium of B ?

Adding Solute Can Raise the Boiling Temperature of the Solvent

At $p = 1$ atm, pure water boils at $T_{b0} = 100^\circ\text{C}$. But salty water boils at a higher temperature, call it T_{b1} . Adding salt changes water's boiling temperature. This is another manifestation of the principle that salt reduces the vapor pressure of water. Think of it in steps: (1) pure water boils at 100°C , then (2) adding salt reduces water's vapor pressure so at 100°C water is no longer boiling, then (3) raising the temperature increases water's vapor pressure, compensating for the effects of the salt, returning the water to boiling. To see this, first substitute $p_B = 1$ atm and $T = T_{b0}$ into Equation (16.2) to treat the boiling of pure water (no salt):

$$p_B = 1 \text{ atm} = p_{B,\text{int}}^\circ \exp\left(\frac{zw_{BB}}{2RT_{b0}}\right). \quad (16.18)$$

Now, consider the boiling of salty water: x_B is slightly less than 1. Substitute $p_B = 1$ atm, x_B , and the new boiling temperature T_{b1} into Equation (16.2) to get

$$p_B = 1 \text{ atm} = p_{B,\text{int}}^\circ x_B \exp\left(\frac{zw_{BB}}{2RT_{b1}}\right) \quad (16.19)$$

(since $(1 - x_B)^2 \approx 0$). Setting Equations (16.18) and (16.19) equal and taking the logarithm of both sides gives

$$\ln x_B = \frac{zw_{BB}}{2R} \left(\frac{1}{T_{b0}} - \frac{1}{T_{b1}} \right). \quad (16.20)$$

Equation (16.20) gives the boiling temperature T_{b1} of salt water if you know the boiling temperature T_{b0} of pure water and the amount x_B of salt in the solution. Let's make a simplifying approximation. For small solute concentrations $x_A \ll 1$, the increase in boiling temperature, $\Delta T = T_{b1} - T_{b0}$, due to added salt, will be small, so

$$\frac{1}{T_{b0}} - \frac{1}{T_{b1}} = \frac{T_{b1} - T_{b0}}{T_{b0}T_{b1}} \approx \frac{\Delta T}{T_{b0}^2} \quad (16.21)$$

Now, expand the logarithm (see Appendix J, Equation (J.4)), $\ln x_B = \ln(1 - x_A) \approx -x_A - x_A^2/2 - x_A^3/3 - \dots$, substitute Equation (16.21) into Equation (16.20), and use $\Delta h_{\text{vap}}^\circ = h_{\text{gas}}^\circ - h_{\text{liq}}^\circ = -zw_{BB}/2$ to get

$$\Delta T = \frac{RT_{b0}^2 x_A}{\Delta h_{\text{vap}}^\circ}. \quad (16.22)$$

Equation (16.22) predicts that the boiling temperature elevation, like the vapor pressure depression, depends linearly on the concentration x_A of the nonvolatile solute A . Equation (16.22) says that the only property of the solute you need to know is its concentration x_A ; all the other information you need is for pure B : $\Delta h_{\text{vap}}^\circ$, the enthalpy of vaporization of pure water in this case, and T_{b0} , the boiling temperature of pure water. Because ideal solution properties depend only on the solute concentration, and not on its chemical character, the colligative laws are analogous to the ideal gas law for dilute solutions.

Units of Concentration

You may sometimes prefer concentration units other than mole fractions. The *molality* m_A of A is the number of moles of solute A dissolved in 1000 g of solvent. The *molarity* c_A is the number of moles of A per liter of solution. To compute ΔT from Equation (16.22), let's convert x_A to other units. If n_A is the number of moles of the solute A and n_B is the number of moles of the solvent B , then, in a dilute solution,

$$x_A = \frac{n_A}{n_A + n_B} \approx \frac{n_A}{n_B}. \quad (16.23)$$

In a solution containing 1000 g of solvent B , having solute molality m_A , n_A is given by

$$n_A \text{ moles of } A = \left(\frac{m_A \text{ moles of } A}{1000 \text{ g of } B} \right) (1000 \text{ g of } B). \quad (16.24)$$

Because the solution is nearly pure B , the number of moles n_B of solvent in this system can be computed simply from its molecular mass. If M_B is the molecular

mass of B in grams per mole, then $1/M_B$ is the number of moles of B contained in 1 g of solvent. So, in 1000 g of solvent B , the number of moles is

$$n_B \text{ moles of } B = \left(\frac{1}{M_B} \right) (1000 \text{ g of } B). \quad (16.25)$$

Substituting Equations (16.24) and (16.25) into Equation (16.23) gives

$$x_A \approx m_A M_B. \quad (16.26)$$

Expressed in terms of the molality of A in a dilute solution, the boiling point elevation is

$$T - T_b = K_b m_A, \quad (16.27)$$

where K_b is a constant for boiling that depends only on the pure solvent (e.g., $K_b = 0.51$ for water; see Example 16.2).

EXAMPLE 16.2 Compute the boiling point elevation constant K_b for water. Combine Equations (16.22) and (16.27) with Equation (16.26), and take $\Delta h_{\text{vap}}^\circ = 40,656 \text{ J mol}^{-1}$ and $T_b = 373.15 \text{ K}$ from Table 14.1, to get

$$\begin{aligned} K_b &= \frac{RT_b^2 M_B}{\Delta h_{\text{vap}}^\circ} \\ &= \frac{(8.314 \text{ JK}^{-1} \text{ mol}^{-1})(373.15 \text{ K})^2 (18 \text{ g mol}^{-1})}{(40,656 \text{ J mol}^{-1})} \\ &= 0.513 \text{ K kg mol}^{-1}. \end{aligned}$$

If you put 2 mol of sucrose (molecular mass 342 g mol^{-1}), which is 684 g , into 1000 g of water, it has a concentration of $m_A = 2 \text{ molal}$. Equation (16.27) gives the boiling point of this solution, $100^\circ\text{C} + (0.513)(2) = 101^\circ\text{C}$.

For high-precision measurements, the molality is often the preferred measure of concentration because it is expressed in terms of weight, which is independent of temperature. Molarities are expressed in terms of volumes, which depend on temperature. In an ideal solution, the solute contributes to colligative properties only through its concentration, and not through its interaction with the solvent. Solutes affect freezing points in much the same way that they affect boiling points.

Adding Solute Can Lower the Freezing Temperature of a Solvent

Adding salt lowers the freezing temperature of water below 0°C , following the same principles as boiling point elevation. The melting process is an equilibrium of an exchangeable component, such as water, between its liquid and solid phases. The melting temperature of pure water is the point at which the escaping tendency of water molecules from the liquid to the solid is the same as the escaping tendency from the solid to the liquid. Salt preferentially dissolves in the liquid, so, to a first approximation, salt is not an exchangeable component

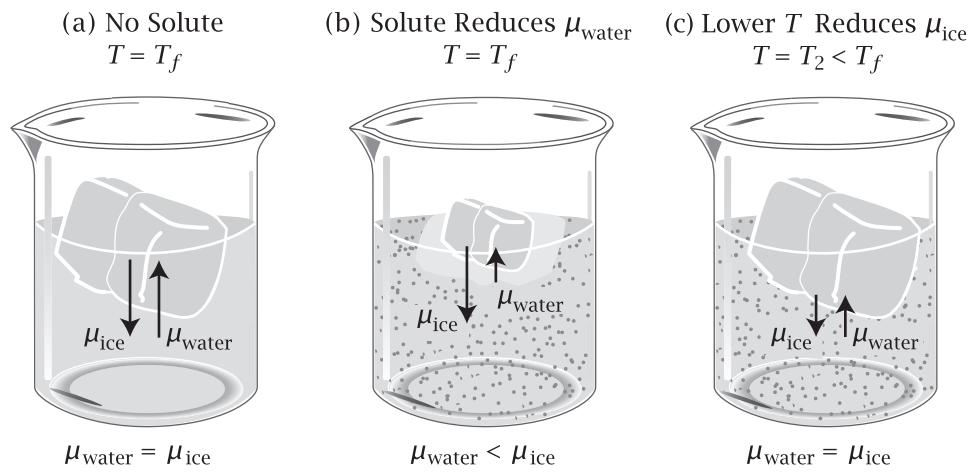


Figure 16.5 The freezing temperature is lowered by adding solute. (a) The freezing temperature of pure solvent is T_f . (b) Adding solute reduces the tendency of the water to escape from the liquid mixture to ice. At T_f , this melts the ice. (c) Lowering the temperature to $T < T_f$ shifts the equilibrium back towards the solid, to reach a new freezing point.

in this equilibrium. Salt reduces the escaping tendency of the water from the liquid to the solid, on balance drawing water from the pure phase (ice) to the mixture (liquid), and melting the ice (see Figure 16.5).

The lowering of the freezing temperature ΔT_f by solute can be treated in the same way as the boiling point elevation in Equations (16.20)–(16.22), giving

$$\Delta T_f = T_f - T = \frac{kT_f^2 x_A}{\Delta h_{\text{fus}}^\circ}, \quad (16.28)$$

where $T_f > T$. T_f is the freezing temperature of the pure liquid B and $\Delta h_{\text{fus}}^\circ = h_{\text{liquid}}^\circ - h_{\text{solid}}^\circ$ is the enthalpy of *fusion* (which means *melting*). Concentration units other than mole fraction x_A may be more convenient. In terms of molality,

$$T_f - T = K_f m_A, \quad (16.29)$$

where K_f is defined by the freezing properties of the solvent (e.g., $K_f = 1.86 \text{ K kg mol}^{-1}$ for water). Figure 16.6 shows that the freezing temperature of water is depressed in linear proportion to the concentration of glycerol and some sugars, in good agreement with this ideal solution model.

In a few instances, a solute preferentially dissolves in the solid, rather than in the liquid. For example, some long-chain alkanes, such as decane, dissolve in solid-phase lipid bilayer membranes but not in liquid-phase bilayers. Such solutes must increase, rather than decrease, the freezing temperature of the system.

The next section treats another colligative effect, the osmotic pressure.

Freezing Point Lowering (°C)

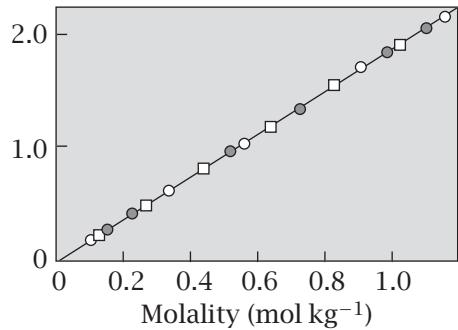


Figure 16.6 The depression of the freezing temperature of water is a linear function of the concentration of glycerol (○), glucose (●), and sucrose (□). Source: CH Langford and RA Beebe, *The Development of Chemical Principles*, Dover Publications, New York, 1969.

Adding Solute on One Side of a Semipermeable Membrane Causes an Osmotic Pressure

Suppose you have a pure liquid *B* on the right side of a container and a mixture of *B* with a solute *A* on the left side. The two liquids are separated by a *semipermeable membrane*, through which molecules of *B* can pass freely, but those of *A* cannot (see Figure 16.7).

The colligative principle described above applies to this situation. Molecules of *B* will be drawn from the right (pure liquid) into the left compartment (the mixture) because of the favorable entropy of mixing *B* with the *A* component. This flow of *B* toward the left can be halted by applying an opposing pressure to the left-hand compartment. The extra pressure that is required on the left to hold the system in equilibrium is called the *osmotic pressure*.

What is the relationship between the composition x_B of the solution and the osmotic pressure π that is required to hold the solvent flow in equilibrium? In our colligative modeling above, we sought a relationship between composition x_B and temperature. Now, we want a relationship between x_B and the extra pressure π applied to the left compartment. At equilibrium, the chemical potential for the exchangeable component *B* must be the same on both sides of the membrane:

$$\mu_B(\text{pure}, p) = \mu_B(\text{mixture}, p + \pi, x_B). \quad (16.30)$$

To find the relationship between π and x_B , use the thermodynamic cycle shown in Figure 16.8. For the left compartment, imagine starting with pure liquid *B* at pressure p , then increasing its pressure to $p + \pi$ in step 1, then adding the solute at constant pressure $p + \pi$ in step 2. How does μ_B change in these two steps?

For the first step, how does μ_B depend on pressure in a pure liquid *B* at constant *T*? Integrate $(\partial \mu_B / \partial p)$:

$$\mu_B(\text{pure}, p + \pi) = \mu_B(\text{pure}, p) + \int_p^{p+\pi} \frac{\partial \mu_B}{\partial p} dp. \quad (16.31)$$

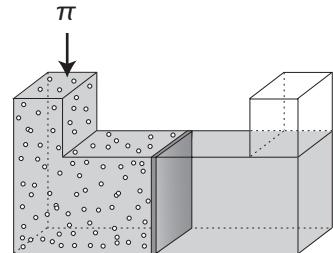


Figure 16.7 Osmotic pressure. The right compartment contains pure liquid *B*. The left compartment contains a mixture of *A* and *B*. *B* can exchange freely across the membrane, but *A* cannot. At equilibrium, the net flow of *B* to the left is halted by a pressure π applied to the left compartment to hold the system in equilibrium. The pressure π that is needed to halt the net flow increases linearly with the concentration of solute *A* in the left compartment.

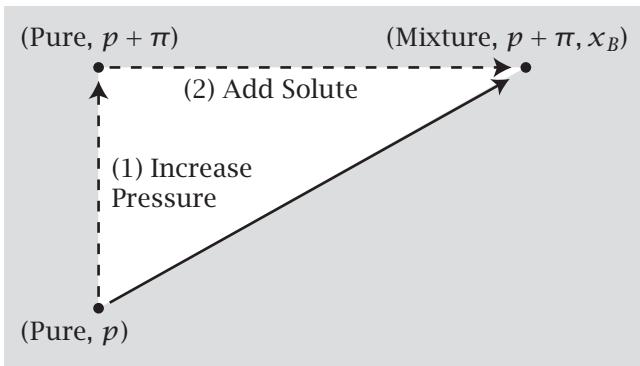


Figure 16.8 Thermodynamic cycle for osmotic pressure in the left compartment. Begin with pure solvent B at pressure p . Step 1: increase its pressure to $p + \pi$. Step 2: add solute A until the composition x_B of B is reached.

Using the Maxwell relation $\partial\mu_B/\partial p = \partial V/\partial N_B = v_B$, where v_B is the molar volume of B , you get

$$\begin{aligned}\mu_B(\text{pure}, p + \pi) &= \mu_B(\text{pure}, p) + \int_p^{p+\pi} v_B dp \\ &= \mu_B(\text{pure}, p) + \pi v_B,\end{aligned}\quad (16.32)$$

since v_B is typically independent of pressure.

Now, adding solute to a solvent at any pressure p_0 gives

$$\mu_B(\text{mixture}, p_0, x_B) = \mu_B(\text{pure}, p_0) + RT \ln(y_B x_B). \quad (16.33)$$

So, for the second step—adding the solute—use Equation (16.32) to replace $\mu_B(\text{pure}, p + \pi)$ in Equation (16.33) to get

$$\mu_B(\text{mixture}, p + \pi, x_B) = \mu_B(\text{pure}, p) + \pi v_B + RT \ln(y_B x_B). \quad (16.34)$$

Substitute Equation (16.30) into Equation (16.34) to get

$$-\pi v_B = RT \ln(y_B x_B). \quad (16.35)$$

To get the osmotic pressure π in terms of the composition of a dilute solution ($y_B \approx 1$), use $x_B = 1 - x_A$ and expand the logarithm in Equation (16.35) to first order, to get the ideal solution result

$$\pi = \frac{RTx_A}{v_B}. \quad (16.36)$$

To convert to molarity c_A , use $x_A \approx n_A/n_B$ and note that $n_B v_B \approx V$ is approximately the volume of the solution. Then Equation (16.36) becomes

$$\pi \approx \frac{n_A RT}{n_B v_B} = \frac{n_A RT}{V} = c_A RT. \quad (16.37)$$

EXAMPLE 16.3 Determining molecular masses from osmotic pressure. Find the molecular mass of a compound that has a concentration $w = 1.2 \text{ g L}^{-1}$ in solution, and an osmotic pressure of $\pi = 0.20 \text{ atm}$ at $T = 300 \text{ K}$. Combine Equation (16.37), $\pi = cRT$, with $c = w/M$ to get $\pi = wRT/M$. Now rearrange

to get

$$M = \frac{wRT}{\pi} = \frac{(1.2 \text{ g L}^{-1})(8.2 \times 10^{-2} \text{ Latm K}^{-1} \text{ mol}^{-1})(300 \text{ K})}{0.20 \text{ atm}} \\ = 147 \text{ g mol}^{-1}.$$

The osmotic pressure is a useful surrogate for the chemical potential of mixing. To see this, combine Equations (16.33) and (16.35):

$$\pi v_B = -\Delta\mu_B(\text{mix}), \quad (16.38)$$

where $\Delta\mu_B(\text{mix}) = \mu_B(\text{mixture}, p_0, x_B) - \mu_B(\text{pure}, p_0)$ is the change in chemical potential upon adding the solute to the solvent at pressure p_0 .

Solutes Can Transfer and Partition from One Medium to Another

In many processes, including chromatographic separations, the adsorption of drugs to surfaces, the equilibration of solutes between oil and water phases, and diffusion across polymeric or biological membranes, molecules start in one environment and end in another. Partitioning processes have been used to model the formation of aggregates or droplets, the self-assembly of membranes and micelles, and protein folding.

So far we've concentrated on colligative processes where the exchangeable component is the *solvent*, i.e., the majority species. Now focus on solute transfer processes, in which the exchangeable component is the minority species, so we use the solute convention (see page 289).

You have two immiscible solvents *A* and *B*, say oil and water. You also have a third component, a solute species *s*. *s* dissolves in both *A* and *B* and exchanges between them. At equilibrium, if the concentration of *s* in *A* is x_{sA} and the concentration of *s* in *B* is x_{sB} , then the **partition coefficient** is defined as $K_A^B = x_{sB}/x_{sA}$. (Don't confuse a partition coefficient, which is this ratio of concentrations, with a partition function, which is a statistical mechanical sum.) Partition coefficients can also be expressed in other concentration units.

The degree of freedom in this problem is the number of solute molecules in each medium. The total number of solute molecules is fixed: The chemical potential for *s* in *A* is $\mu_s(A)$ and the chemical potential for *s* in *B* is $\mu_s(B)$. To determine the partition coefficient, use the equilibrium condition that the chemical potential of the solute must be the same in both phases:

$$\mu_s(A) = \mu_s(B). \quad (16.39)$$

To compute the partition coefficient, use the lattice model for the chemical potential of *s* in each liquid. According to Equations (15.17) and (15.18),

$$\frac{\mu_s(A)}{kT} = \frac{zw_{ss}}{2kT} + \ln x_{sA} + \chi_{sA}(1-x_{sA})^2 \quad (16.40)$$

and

$$\frac{\mu_s(B)}{kT} = \frac{zw_{ss}}{2kT} + \ln x_{sB} + \chi_{sB}(1-x_{sB})^2. \quad (16.41)$$

Setting the chemical potential for s in A equal to the chemical potential of s in B gives the partition coefficient K_A^B :

$$\ln K_A^B = \ln\left(\frac{x_{sB}}{x_{sA}}\right) = \chi_{sA}(1-x_{sA})^2 - \chi_{sB}(1-x_{sB})^2. \quad (16.42)$$

The corresponding result for the thermodynamic model comes from combining Equations (16.39) and (16.11):

$$\ln K_A^B = \ln\left(\frac{x_{sB}}{x_{sA}}\right) = -\frac{\mu_s^\circ(B) - \mu_s^\circ(A)}{kT} - \ln\left(\frac{\gamma_{sB}}{\gamma_{sA}}\right). \quad (16.43)$$

The quantity $\Delta\mu^\circ = \mu_s^\circ(B) - \mu_s^\circ(A)$ is called the *free energy of transfer*. At infinite dilution of solute in both phases, $x_{sA}, x_{sB} \ll 1$. In the solute convention ($\gamma_{sA} \rightarrow 1$ as $x_{sA} \rightarrow 0$ and $\gamma_{sB} \rightarrow 1$ as $x_{sB} \rightarrow 0$), you have

$$\ln K_A^B = \frac{\mu_s^\circ(A) - \mu_s^\circ(B)}{kT} = \chi_{sA} - \chi_{sB}. \quad (16.44)$$

EXAMPLE 16.4 Partition coefficient for transferring butane into water. Figure 16.9 shows $\Delta\mu^\circ = \mu_{\text{water}}^\circ - \mu_{\text{hydrocarbon}}^\circ = 6 \text{ kcal mol}^{-1}$ for butane. To compute the partition coefficient for this transfer, $K_{\text{hc}}^w = x_{s,w}/x_{s,\text{hc}}$ at $T = 300 \text{ K}$, use Equation (16.43):

$$K_{\text{hc}}^w = e^{-\Delta\mu^\circ/RT} = \exp\left[\frac{-6000}{(1.987)(300)}\right] \\ = 4.3 \times 10^{-5}.$$

Partitioning experiments provide direct information about the binary interaction quantities χ_{AB} . Here, $-kT\chi_{\text{butane,water}} = -6 \text{ kcal mol}^{-1}$, so at $T = 300 \text{ K}$, $\chi_{\text{butane,water}} \approx 10$.

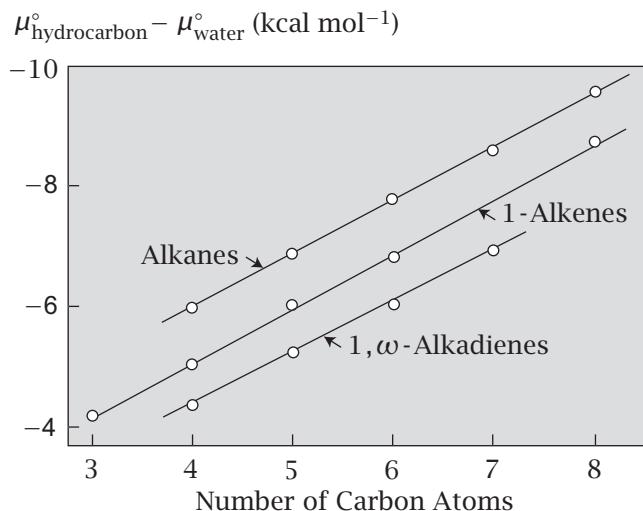


Figure 16.9 $\Delta\mu^\circ$ for the transfer of hydrocarbons from aqueous solution to pure liquid hydrocarbon at 25°C , based on solubility measurements of C McAuliffe, *J Phys Chem* **70**, 1267 (1966). The transfer free energy is linearly proportional to the number of solute–solvent contacts, represented by z in the lattice model. Source: C Tanford, *The Hydrophobic Effect: Formation of Micelles and Biological Membranes*, 2nd edition, Wiley, New York, 1980.

EXAMPLE 16.5 Oil/water partitioning predicts the potencies of general anesthetic drugs. In 1846, it was discovered that surgery could be made tolerable by giving patients ether. Ether anesthetizes patients, temporarily removing their pain. What is the biological action of ether? Around 1900, HH Meyer and E Overton discovered that many different simple nonpolar molecules act as anesthetics. Remarkably, they found that the degree to which a molecule anesthetizes a patient correlates with its oil/water or gas/water partition coefficient; see Figure 16.10. While the mechanism of anesthetic drugs is still not fully understood, the current view is that anesthetics partition into lipid bilayer membranes where they affect membrane proteins, either directly or through indirect interactions with the membrane.

Minimum Alveolar Partial Pressure

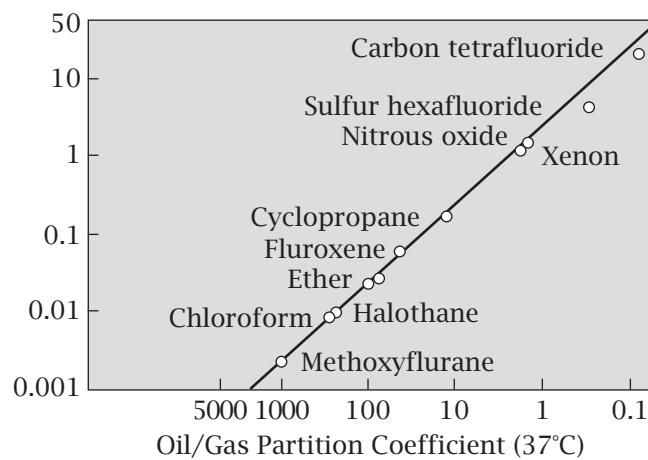


Figure 16.10 Potencies of anesthetic drugs correlate with their coefficients of partitioning from the vapor phase into oil. Molecules having greater potency can cause anesthesia at lower minimum partial pressures in the alveoli. Source: MJ Halsey et al., in *General Anesthesia*, JF Nunn, JE Utting, and BR Brown, eds, Butterworth, London, 1989.

Equation (16.44) shows how to compute the partition coefficient if you know $\chi_{sB} - \chi_{sA}$. Using the definition of χ , you have

$$\begin{aligned}\chi_{sB} - \chi_{sA} &= \left(\frac{z}{kT}\right) \left[\left(w_{sB} - \frac{w_{ss} + w_{BB}}{2}\right) - \left(w_{sA} - \frac{w_{ss} + w_{AA}}{2}\right) \right] \\ &= \left(\frac{z}{kT}\right) \left(w_{sB} - w_{sA} + \frac{w_{AA}}{2} - \frac{w_{BB}}{2} \right).\end{aligned}\quad (16.45)$$

Figure 16.11 shows that partitioning is driven by the energetic cost of opening a cavity in B (step 1, $-w_{BB}/2$), transferring a solute molecule from medium A into the cavity in B (step 2, $w_{sB} - w_{sA}$), then closing the cavity in A (step 3, $w_{AA}/2$).

Let's generalize our partitioning model. You may prefer some other measure of concentration c_{sA} of solute s in medium A than mole fraction x_{sA} . So, express the chemical potential as

$$\mu_s(A) = \mu_s^\circ(A) + kT \ln(\gamma c_{sA}) \approx \mu_s^\circ(A) + kT \ln c_{sA}. \quad (16.46)$$

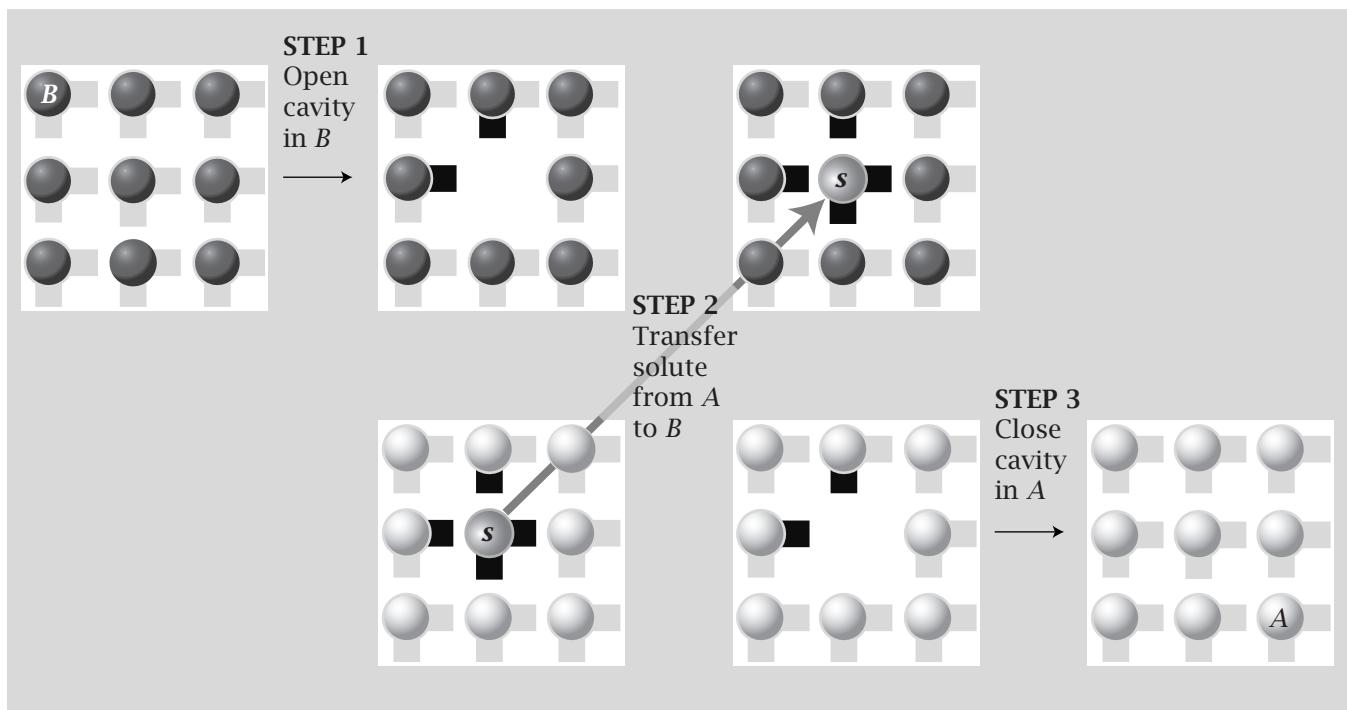


Figure 16.11 The solute partitioning process: (1) open a cavity in *B* (breaking $z/2$ *BB* contacts), (2) transfer solute *s* from medium *A* to *B* (breaking z *sA* contacts and making z *sB* contacts), then (3) close the cavity in *A* (making $z/2$ *AA* contacts).

(Note that c_{sA} will usually have units, say molarity, moles per liter. But, you can only take a logarithm of a dimensionless quantity. This is handled within the companion quantity, $\mu_s^\circ(A)$. That is, if $c_{sA} = 0.2$ M, then operationally you take $\ln(0.2\text{ M}/1\text{ M})$ in Equation (16.46).)

Here is the logic of solution chemistry. From experiments, you want quantities such as μ_s° . So perform experiments in dilute solutions, where c_{sA} is as small as possible. As indicated in Equation (16.46), this eliminates the solution complexities described by γ . For example, in dilute solutions, Equation (16.44) will convert your measurements of K_A^B to $[\mu_s^\circ(A) - \mu_s^\circ(B)]/kT$ or $\chi_{sA} - \chi_{sB}$, quantities that give you direct information about how *s* interacts with pure solvation shells of *A* and *B*. Higher concentrations would reflect a process in which *s* begins as solvated by some molecules of *A* and some molecules of *s* and ends as solvated by some molecules of *B* and some molecules of *s* (see Equations (16.42) and (16.43)), and would be hard to interpret.

Common to the past few chapters is the exchange-energy quantity χ_{AB} . It appears in colligative, partitioning, transfer, and interfacial properties. So, not surprisingly, various solution properties correlate with others. For example, χ_{AB} for the transfer of a solute *B* into a liquid *A* is often proportional to the interfacial tension between those two materials; see Figure 16.12.

One basic process of partitioning is the transfer of *s* from pure liquid *s* to a mixture with *A*. By definition, $\chi_{ss} = 0$ (see Equation (15.13)). In that case,

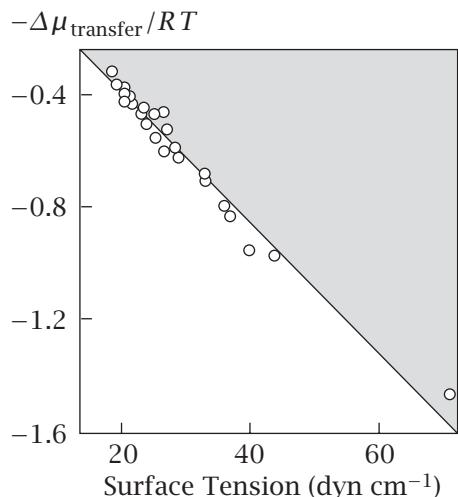


Figure 16.12 The free energy of transferring argon into various hydrocarbons at 25°C correlates with surface tension, indicating how various solution properties are related to each other through χ_{AB} . Source: A Ben-Naim, *Hydrophobic Interactions*, Plenum Press, New York, 1980. Data are reproduced with changes from R Battino and HL Clever, *Chem Rev* **66**, 395 (1966).

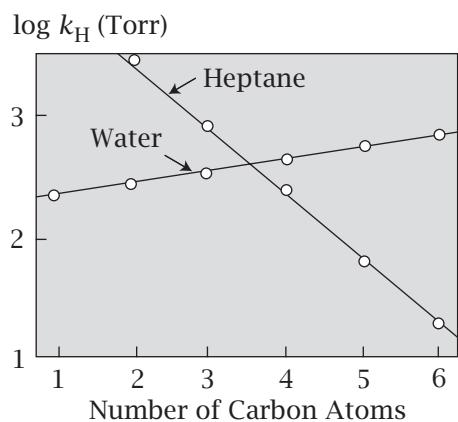


Figure 16.13 The logarithm of the Henry's law constants k_H at 25°C versus carbon number for straight-chain alcohols in water and heptane. Increasing the chain length of alcohols increases their escaping tendency from water, but decreases their escaping tendency from heptane. Source: JH Ryting, LP Huston, and T Higuchi, *J Pharm Sci* **67**, 615–618 (1978).

Equation (16.42) gives

$$\ln K_{SA} = -\chi_{SA}. \quad (16.47)$$

Larger solute molecules typically have larger partition coefficients. In the lattice model, the driving force for partitioning depends on the lattice coordination number z , which serves as a measure of the surface area of contact between the solute molecule and surrounding medium. It follows that $\mu_s^\circ(A) - \mu_s^\circ(B)$ is proportional to the sizes of solutes; see Figure 16.9. Similarly, the Henry's law constant should increase with solute size; see Figure 16.13.

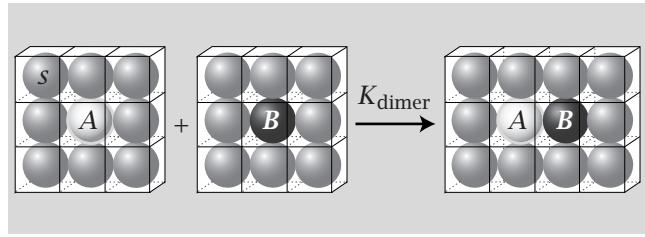
Now, let's look at other solution processes: association and dimerization.

Dimerization in Solution Involves Desolvation

In this section, we compute the equilibrium constant for the association of a molecule of type A with a molecule of type B to form species AB (see Figure 16.14) in a solvent s .

The degrees of freedom are the numbers N_A of monomer particles of type A and N_B of type B , and the number N_{AB} of dimers of type AB . Pressure and

Figure 16.14 Two molecules, (A) and (B), associate in solvent (s).



temperature are constant. Changes in free energy are given by

$$dG = \mu_A dN_A + \mu_B dN_B + \mu_{AB} dN_{AB}. \quad (16.48)$$

Conservation of A and B particles leads to the constraints

$$N_A + N_{AB} = N_{A,\text{total}} = \text{constant}$$

and

$$N_B + N_{AB} = N_{B,\text{total}} = \text{constant}$$

$$\Rightarrow dN_A = dN_B = -dN_{AB}. \quad (16.49)$$

The equilibrium condition $dG = 0$ can be expressed as

$$\begin{aligned} (-\mu_A - \mu_B + \mu_{AB}) dN_{AB} &= 0 \\ \Rightarrow \mu_{AB} &= \mu_A + \mu_B. \end{aligned} \quad (16.50)$$

Assume A , B , and AB are at infinite dilution in the solvent s . Substituting the definition of μ from Equation (16.11) into Equation (16.50), you have

$$\frac{\mu_{AB}^\circ}{kT} + \ln x_{AB} = \left(\frac{\mu_A^\circ}{kT} + \ln x_A \right) + \left(\frac{\mu_B^\circ}{kT} + \ln x_B \right). \quad (16.51)$$

From Equation (16.51), the dimerization equilibrium constant K_{dimer} is

$$\ln K_{\text{dimer}} = \ln \left(\frac{x_{AB}}{x_A x_B} \right) = -\frac{1}{kT} (\mu_{AB}^\circ - \mu_A^\circ - \mu_B^\circ). \quad (16.52)$$

The lattice model gives μ_A° and μ_B° (see Equation (15.28)):

$$\frac{\mu_A^\circ}{kT} = \frac{zw_{AA}}{2kT} + \chi_{sA} - \ln q_A = \frac{z}{kT} \left(w_{sA} - \frac{w_{ss}}{2} \right) - \ln q_A \quad (16.53)$$

and

$$\frac{\mu_B^\circ}{kT} = \frac{zw_{BB}}{2kT} + \chi_{sB} - \ln q_B = \frac{z}{kT} \left(w_{sB} - \frac{w_{ss}}{2} \right) - \ln q_B. \quad (16.54)$$

Now you need to include the internal partition functions, q_A , q_B , and q_{AB} because the internal degrees of freedom, such as the rotational symmetry, can change upon dimerization (see Equation (15.28)).

You also need μ_{AB}° , the free energy cost of opening a two-site cavity in s and inserting an A and a B adjacent to each other (see Chapter 15). The simplest way to get this quantity is to consider a two-site cavity surrounded by a first shell of

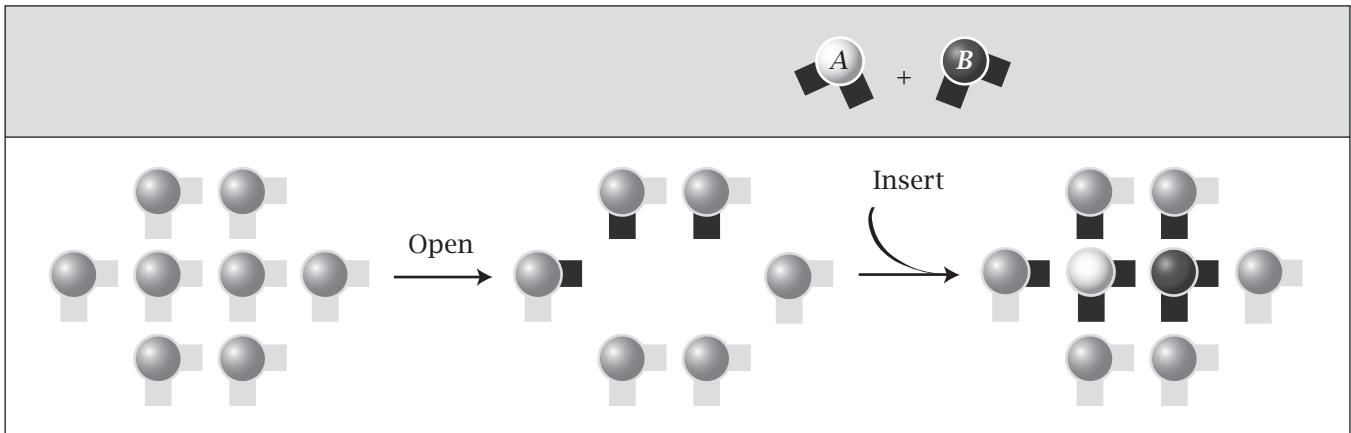


Figure 16.15 For dimerization, a two-site cavity opens in solvent s . There are $2(z-1)$ first-neighbor s molecules (six first-shell neighbors in this two-dimensional example in which $z = 4$). The number of ss bonds broken in this process is half this number ($z-1 = 3$, indicated by the three bonds pointed inward in this case). Inserting one A makes $z-1$ contacts between s and A . Inserting one B makes $z-1$ contacts between s and B . There is one AB contact.

$2(z-1)$ solvent molecules. The number of bonds that are broken upon opening the cavity is half the number of first-shell neighbors, $2(z-1)/2 = z-1$ (see Figure 16.15). So the energy cost of opening a two-site cavity is $(-\omega_{ss})(z-1)$. The energy gained upon inserting one A into the cavity is $(z-1)\omega_{sA}$, and upon inserting one B the energy gain is $(z-1)\omega_{sB}$. The energy gained in forming the one AB contact is ω_{AB} , so

$$\frac{\mu_{AB}^\circ}{kT} = \left(\frac{z-1}{kT}\right)(\omega_{sA} + \omega_{sB} - \omega_{ss}) + \frac{\omega_{AB}}{kT} - \ln q_{AB}, \quad (16.55)$$

where q_{AB} is the partition function for the dimer. To find K_{dimer} , substitute Equations (16.53)–(16.55) into Equation (16.52):

$$\begin{aligned} \ln K_{\text{dimer}} &= -\frac{\Delta\mu^\circ}{kT} = \frac{\omega_{sA} + \omega_{sB} - \omega_{ss} - \omega_{AB}}{kT} + \ln\left(\frac{q_{AB}}{q_A q_B}\right) \\ &= \left(\frac{1}{z}\right)(\chi_{sA} + \chi_{sB} - \chi_{AB}) + \ln\left(\frac{q_{AB}}{q_A q_B}\right). \end{aligned} \quad (16.56)$$

Dimerization in solution can be described by the thermodynamic cycle shown in either Figure 16.16(a) or (b).

AB association in solution can be driven by the affinity of A for B , or by the affinity of s for s ($\omega_{ss} \ll 0$), or by the lack of affinity of s for A or B . Oil and water don't mix, mainly because the water-water attractions are so strong that water squeezes out the oil (although oil-water interactions are complicated by additional entropic factors; see Chapter 31).

If you are interested in AA dimerization, then substitute A for B in Equation (16.56):

$$\ln\left(\frac{x_{AA}}{x_A^2}\right) = \frac{2\omega_{sA} - \omega_{ss} - \omega_{AA}}{kT} + \ln\left(\frac{q_{AA}}{q_A^2}\right) = \frac{2}{z}\chi_{sA} + \ln\left(\frac{q_{AA}}{q_A^2}\right). \quad (16.57)$$

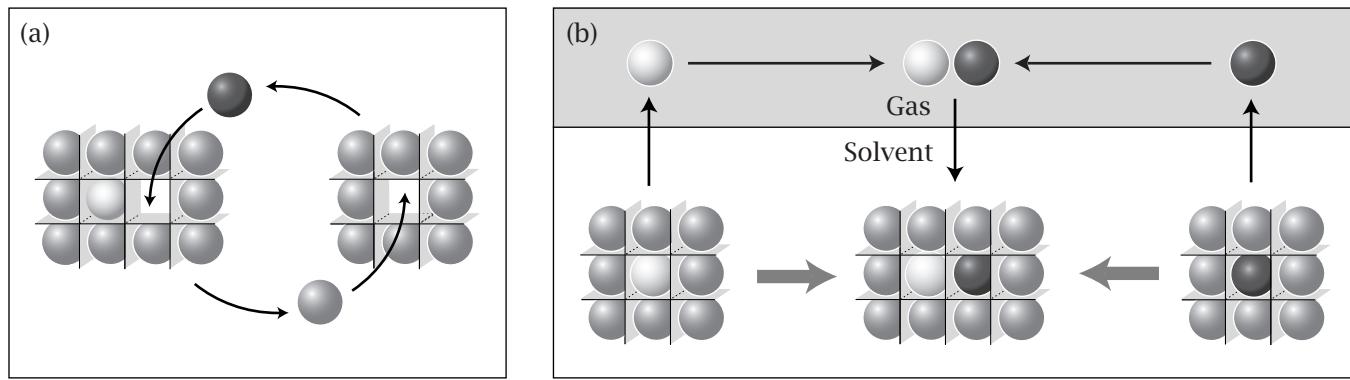


Figure 16.16 (a) The association of two molecules ($\textcircled{1}$) and ($\textcircled{2}$) in solvent ($\textcircled{3}$) can be represented as an exchange process, or (b) as a series of steps: first A and B are transferred from solvent to the vapor phase, then A associates with B , then the AB dimer is solvated.

As for dimerization in the gas phase, the rotational partition function favors the dimer, because there are more distinguishable orientations of AB than of A or B spheres.

EXAMPLE 16.6 Ligand binding to a protein involves desolvation. When a ligand L binds to a protein receptor R , water must first be stripped away. Figure 16.17 shows the binding cycle: (a) ligand L desolvates; (b) solvent molecules s leave the R site, desolvating the protein; (c) desolvated L binds to desolvated R ; (d) the molecules of s that are released enter the bulk liquid. The graphical cycle in Figure 16.17 shows that the binding affinity K can be computed as

$$\ln K = \ln \left(\frac{x_{RL}}{x_R x_L} \right) = \frac{\mathcal{A}}{kT} (-w_{RL} - w_{ss} + w_{Rs} + w_{Ls}) \quad (16.58)$$

where $\mathcal{A} = 3$ is the contact area (in this two-dimensional lattice model, we use units of $z = 3$ lattice contact sites); x_{RL} is the concentration of the bound complex (RL), x_L is the concentration of unbound ligand, and x_R is the concentration of unbound receptor. The w 's are the corresponding interaction energies. So, binding affinities arise not only from the direct interaction, w_{RL} , but also from stripping off waters and moving them to the bulk.

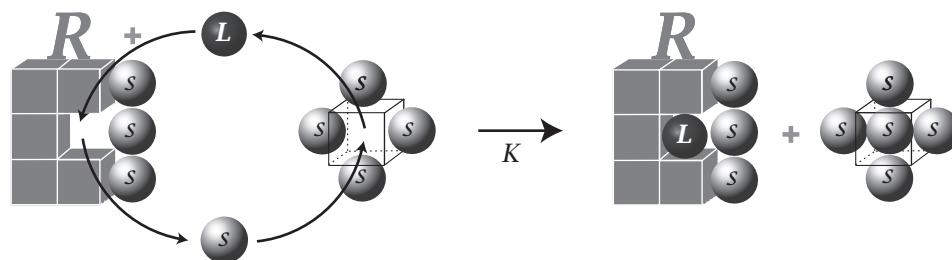


Figure 16.17 When a ligand molecule binds to a protein, the ligand desolvates, the receptor site desolvates, and there is a gain of water–water contacts.

Summary

We have considered mixtures of two components in which one can exchange into the gas phase or into other liquids or solids. The degree to which a species exchanges depends on its concentration in the solution. Such processes can be described by thermodynamic models, which do not describe the microscopic details, or statistical mechanical models such as the lattice model. Chapters 27 and 28 treat dimerization and other binding processes in more detail.

Problems

1. The mechanism of anesthetic drugs. Anesthetic drug action is thought to involve the solubility of the anesthetic in the hydrocarbon region of the lipid bilayer of biological membranes. According to the classical Meyer-Overton hypothesis, anesthesia occurs whenever the concentration of drug is greater than 0.03 mol kg^{-1} membrane, no matter what the anesthetic.

- (a) For gaseous anesthetics like nitrous oxide or ether, how would you determine what gas pressure of anesthetic to use in the inhalation mix for a patient in order to achieve this membrane concentration?
- (b) Lipid bilayers ‘melt’ from a solid-like state to a liquid-like state. Do you expect introduction of the anesthetic to increase, decrease, or not change the melting temperature? If the melting temperature changes, how would you predict the change?

2. Divers get the bends. Divers returning from deep dives can get the bends from nitrogen gas bubbles in their blood. Assume that blood is largely water. The Henry’s law constant for N_2 in water at 25°C is 86,000 atm. The hydrostatic pressure is 1 atm at the surface of a body of water and increases by approximately 1 atm for every 33 feet in depth. Calculate the N_2 solubility in the blood as a function of depth in the water, and explain why the bends occur.

3. Hydrophobic interactions. Two terms describe the hydrophobic effect: (i) hydrophobic hydration, the process of transferring a hydrophobic solute from the vapor phase into a very dilute solution in which the solvent is water, and (ii) hydrophobic interaction, the process of dimer formation from two identical hydrophobic molecules in a water solvent.

- (a) Using the lattice model chemical potentials, and the solute convention, write the standard state chemical potential differences for each of these processes, assuming that these binary mixtures obey the regular solution theory.
- (b) Describe physically, or in terms of diagrams, the driving forces and how these two processes are similar or different.

4. Solutes partition into lipid bilayers. Robinson Crusoe and his friend Friday are stranded on a desert island with no whiskey, only pure alcohol. Crusoe, an anesthesiologist, realizes that the effect of alcohol, as with other anesthetics, is felt when the alcohol reaches a particular critical concentration c_0 in the membranes of neurons. Crusoe and Friday have only a supply of propanol, butanol, and pentanol, and a table of their free energies of transfer for $T = 300 \text{ K}$.

- (a) If a concentration of c_1 of ethanol in water is needed to produce a concentration c_0 in the membrane, a hydrocarbon-like environment, use Table 16.2 to

predict what concentrations in water of the other alcohols would produce the same degree of anesthesia.

Table 16.2 Partitioning quantities in cal mol^{-1} .

	$\mu_w^\circ - \mu_{hc}^\circ$	$h_w^\circ - h_{hc}^\circ$	$s_w^\circ - s_{hc}^\circ$
Ethanol	760	-2430	-10.7
Propanol	1580	-2420	-13.4
Butanol	2400	-2250	-15.6
Pentanol	3222	-1870	-17.1

- (b) Mythical cannibals find Crusoe and Friday and throw them into a pot of boiling water. Will the alcohol stay in their membranes and keep them drunk at 100°C ?
- (c) Which alcohol has the greatest tendency to partition into the membrane per degree of temperature rise?

5. Global warming. CO_2 in the Earth’s atmosphere prevents heat from escaping, and is responsible for roughly half of the greenhouse effect that causes global warming. Would global warming cause a further increase in atmospheric CO_2 through vaporization from the oceans? Assume that the ocean is a two-component solution of water plus CO_2 , and that CO_2 is much more volatile than water. Give an algebraic expression for the full temperature dependence of the Henry’s law constant k_H for the CO_2 in water; that is, derive an equation for $\partial k_H / \partial T$.

6. Modeling cavities in liquids. Assume you have a spherical cavity of radius r in a liquid. The surface tension of the liquid is γ , in units of energy per unit area.

- (a) Write an expression for the energy $\varepsilon(r)$ required to create a cavity of radius r .
- (b) Write an expression for the probability $p(r)$ of finding a cavity of radius r .
- (c) What is the average energy $\langle \varepsilon \rangle$ for cavities in the liquid?
- (d) Write an expression for the average cavity radius.
- (e) If $kT = 600 \text{ cal mol}^{-1}$ and $\gamma = 25 \text{ cal \AA}^{-2} \text{ mol}^{-1}$, compute $\langle r \rangle$.

7. Sparkling drinks. CO_2 in water has a Henry’s law constant $k_H = 1.25 \times 10^6 \text{ mmHg}$. What mole fraction of CO_2 in water will lead to ‘bubbling up’ and a vapor pressure equal to 1 atm?

8. Modeling binding sites. You have a two-dimensional molecular lock and key in solvent s , as shown in Figure 16.18. Different parts of each molecule have different chemical characters A , B , and C .

- (a) In terms of the different pair interactions, (w_{AB} , w_{AC} , w_{AS} , etc.) write an expression for the binding constant K (i.e., for association).
- (b) Which type of pair interaction (AB , AC , BC) will dominate the attraction?

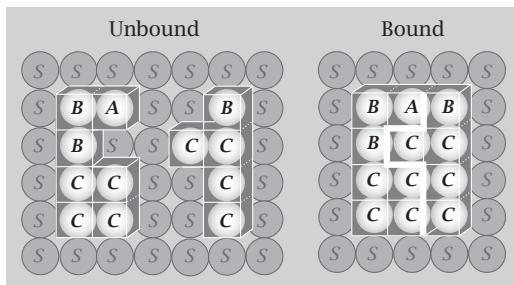


Figure 16.18

9. Oil/water partitioning of drugs. In the process of partitioning of a drug from oil into water, $\Delta s^\circ = -50 \text{ cal deg}^{-1} \text{ mol}^{-1}$ and $\Delta h^\circ = 0$ at $T = 300 \text{ K}$.

- What is $\Delta\mu^\circ$ at $T = 300 \text{ K}$?
- What is the partition coefficient from oil to water, $K_{\text{oil}}^{\text{water}}$ at $T = 300 \text{ K}$?
- Assume that Δs° and Δh° are independent of temperature. Calculate K_0^w at $T = 320 \text{ K}$.

10. Another oil/water partitioning problem. Assume that a drug molecule partitions from water to oil with partition coefficients $K_1 = 1000$ at $T_1 = 300 \text{ K}$, and $K_2 = 1400$ at $T_2 = 320 \text{ K}$.

- Calculate the free energy of transfer $\Delta\mu^\circ$ at $T_1 = 300 \text{ K}$.
- Calculate the enthalpy of transfer Δh° (water to oil).
- Calculate the entropy of transfer Δs° (water to oil).

11. Oil/water partitioning of benzene. You put the solute benzene into a mixture containing the two solvents: oil and water. You observe the benzene concentration in water to be $x_w = 2.0 \times 10^{-6} \text{ M}$ and that in oil to be $x_o = 5.08 \times 10^{-3} \text{ M}$.

- What is the partition coefficient $K_{\text{water}}^{\text{oil}}$ (from water into oil)?
- What is $\Delta\mu^\circ$ for the transfer of benzene from water into oil at $T = 300 \text{ K}$?

12. Balancing osmotic pressures. Consider a membrane-enclosed vesicle that contains within it a protein species *A* that cannot exchange across the membrane. This causes an osmotic flow of water into the cell. Could you reverse the osmotic flow by a sufficient concentration of a different nonexchangeable protein species *B* in the external medium?

13. Vapor pressures of large molecules. Why do large molecules, such as polymers, proteins, and DNA, have very small vapor pressures?

14. Osmosis in plants. Consider the problem of how plants might lift water from ground level to their leaves. Assume that there is a semipermeable membrane at the roots, with pure water on the outside, and an ideal solution inside a small cylindrical capillary inside the plant. The solute mole fraction inside the capillary is $x = 0.001$. The radius of the capillary is 10^{-2} cm . The gravitational

potential energy that must be overcome is mgh , where m is the mass of the solution, and g is the gravitational acceleration constant, 980 cm s^{-2} . The density of the solution is 1 g cm^{-3} . What is the height of the solution at room temperature? Can osmotic pressure account for raising this water?

15. Polymerization in solution. Using the lattice dimerization model (see Figure 16.19), derive the equilibrium constant for creating a chain of m monomers of type *A* in a solvent *s*.

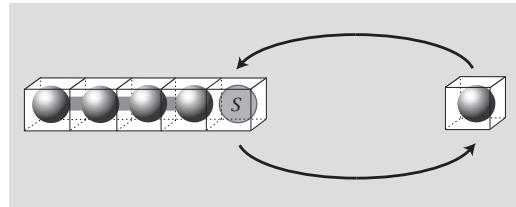


Figure 16.19

16. Ethane association in water. The association of ethane molecules in water is accompanied by an enthalpy of $2.5 \text{ kcal mol}^{-1}$ at $T = 25^\circ \text{C}$. Calculate $(\partial \ln K_{\text{assoc}} / \partial T)$ at this temperature. Does the “hydrophobic effect” get stronger or weaker as temperature is increased?

17. Freezing point depression by large molecules. Freezing temperature depression is a useful way of measuring the molecular masses of some molecules. Is it useful for proteins? One gram of protein of molecular mass $100,000 \text{ g mol}^{-1}$ is in 1 g of water. Calculate the freezing temperature of the solution.

18. Partitioning into small droplets is opposed by interfacial tension. When a solute molecule (*s*) partitions from water (*A*) into a small oil droplet (*B*), the droplet will grow larger, creating a larger surface of contact between the droplet and the water. Thus, partitioning the solute into the droplet will be opposed by the droplet’s interfacial tension with water. Derive an expression to show how much the partition coefficient will be reduced as a function of the interfacial tension and radius of the droplet.

19. Alternative description of Henry’s law. Show that an alternative way to express Henry’s law of gas solubility is to say that the volume of gas that dissolves in a fixed volume of solution is independent of pressure at a given temperature.

20. Benzene transfer into water. At $T = 25^\circ \text{C}$, benzene dissolves in water up to a mole fraction of 4.07×10^{-4} (its solubility limit).

- Compute $\Delta\mu^\circ$ for transferring benzene to water.
- Compute $\chi_{\text{benzene}, \text{water}}$.
- Write an expression for the temperature dependence of $\Delta\mu^\circ$ as a function of Δh° , Δs° (the molar enthalpy and entropy at 25°C), and ΔC_p .

- (d) Using the expression you wrote for (c), calculate $\Delta\mu^\circ$ for transferring benzene to water at $T = 100^\circ\text{C}$, if $\Delta h^\circ = 2 \text{ kJ mol}^{-1}$, $\Delta s^\circ = -58 \text{ JK}^{-1} \text{ mol}^{-1}$, and $\Delta C_p = 225 \text{ JK}^{-1} \text{ mol}^{-1}$.

21. Raoult's law for methanol-water mixtures. Figure 16.20 plots models for the vapor pressure p_{methanol} versus methanol composition x_{methanol} in a binary solution with water at 300 K. Raoult's law behavior is indicated by the straight line.

- Using the lattice model, calculate the contact energy w for methanol. To simplify, neglect the internal degrees of freedom for methanol, and assume a lattice coordination number $z = 6$.
- From the graph on the right of Figure 16.20, estimate the value of the exchange parameter χ for the methanol-water mixture.
- Using the lattice model expression for the activity coefficient, show that p_{methanol} approaches p_{methanol}^* (the equilibrium vapor pressure above pure

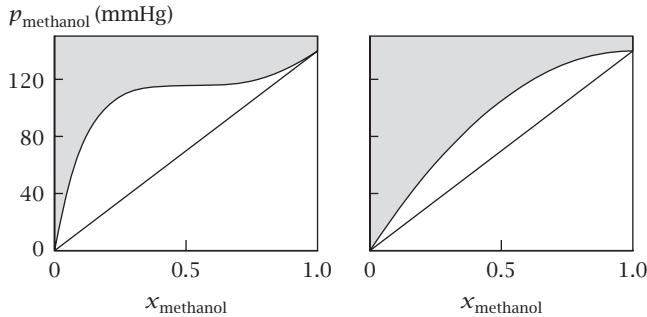


Figure 16.20 Two possible models for the vapor pressure of methanol over methanol–water mixtures, as a function of methanol mole fraction.

methanol) with the same slope as Raoult's law. That is, say why the plot for p_{methanol} should resemble the right-hand graph in Figure 16.20, and not the left.

22. Vapor pressure of water.

- At a vapor pressure $p = 0.03 \text{ atm}$, and a temperature $T = 25^\circ\text{C}$, what is the density of water vapor? Assume an ideal gas.
- What is the corresponding density of liquid water under the same conditions?
- If the vapor pressure of water is 0.03 atm at $T = 25^\circ\text{C}$ and 1 atm at $T = 100^\circ\text{C}$, what is the pairwise interaction energy w_{AA} if each water molecule has four nearest neighbors?

23. Osmotic pressure of sucrose. What is the osmotic pressure of a 0.1 molar solution of sucrose in water at $T = 300 \text{ K}$?

References

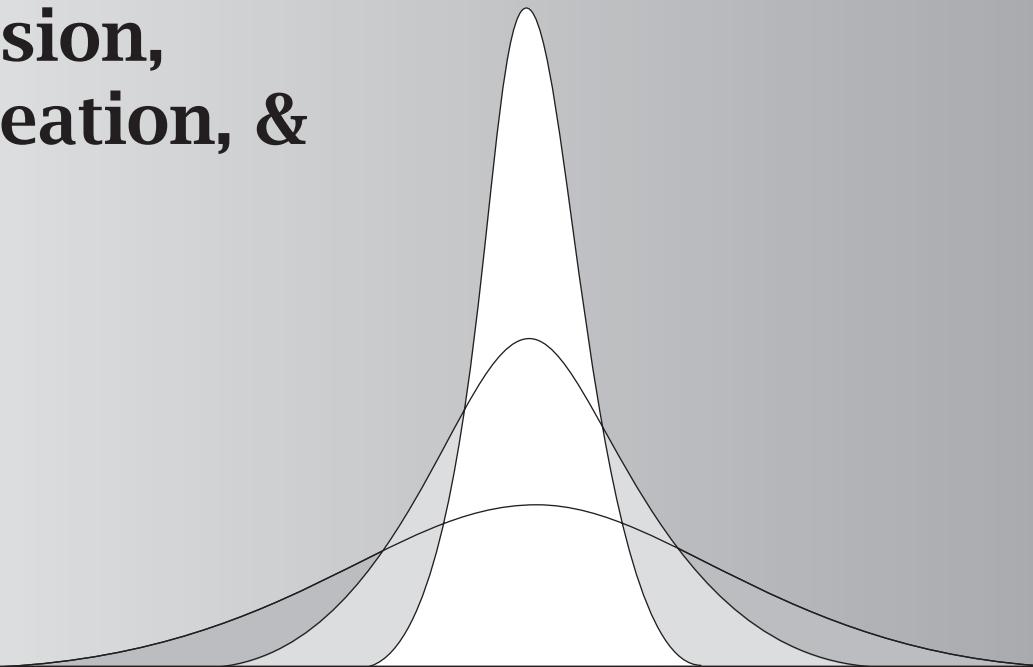
- [1] L Stramma, GC Johnson, J Sprintall, and V Mohrholz, *Science* **320**, 655–658 (2008).

Suggested Reading

PW Atkins and J de Paula, *Physical Chemistry*, 8th edition, WH Freeman, San Francisco, 2006. Good general physical chemistry text.

IM Klotz and RM Rosenberg, *Chemical Thermodynamics: Basic Concepts and Methods*, 7th edition, Wiley, New York, 2008. Excellent discussion of standard states and activities.

17 Physical Kinetics: Diffusion, Permeation, & Flow



Applied Forces Cause Liquids to Flow

Molecules diffuse, transporting matter and heat. For example, soft drinks lose their fizz when CO_2 diffuses out of their plastic containers. Drugs diffuse out of clever encapsulation devices into the body according to tailored time schedules. Metabolites flow in and out of cells through biological membranes. The electrical currents that power household appliances, modern electronics, and neurons result from the flows of ions and electrons that are driven by gradients of electrical potentials. These processes are described by laws of kinetics; they are outside the scope of equilibrium thermodynamics. What are the rates at which molecules flow from one place to another? What forces drive them? In this chapter, we discuss the phenomenology; in the following chapter, we take a more microscopic view.

Defining the Flux

A flowing fluid is characterized by its *flux*. To avoid vector arithmetic (see Appendix G for vector methods), let's just consider a flow along a single direction, which we'll choose to be the x axis. In this chapter, flux J is defined as the amount of material passing through a unit area per unit time. (Elsewhere, flux is sometimes not divided by an area.) You are free to define the amount of material in terms of either the number of particles, or their mass, or their volume, whichever is most convenient for the problem at hand.

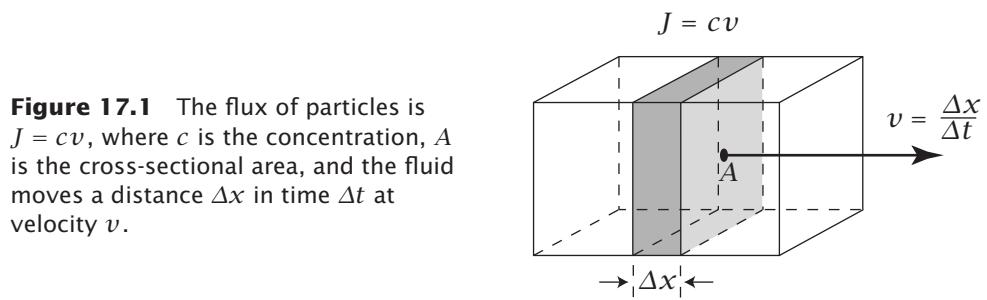


Figure 17.1 The flux of particles is $J = cv$, where c is the concentration, A is the cross-sectional area, and the fluid moves a distance Δx in time Δt at velocity v .

Think of a garden hose. The flux of water can be expressed as the number of gallons of water coming out of the hose per minute, divided by the cross-sectional area of the hose. The flux is proportional to the velocity of flow.

Figure 17.1 shows a rectangular element of a fluid, Δx units long, with cross-sectional area A , having a volume $A \Delta x$. If the fluid carries particles having concentration $c = (\text{number of particles})/(\text{unit volume})$, then the total number of particles in the volume element is $cA \Delta x$. The flux J is the number of particles, $cA \Delta x$, divided by the unit area A per unit time Δt . That is, $J = c \Delta x / \Delta t$. If the flow velocity is $v = \Delta x / \Delta t$, the flux is

$$J = \frac{c \Delta x}{\Delta t} = cv. \quad (17.1)$$

This relationship will be useful throughout this chapter.

Flows can arise from applied forces. For example, gravity draws particles downward in a beaker of water. Or applied voltages drive charged particles such as electrons or ions to flow as an electrical current, as in *Ohm's law*. Forces accelerate particles according to Newton's laws. But for microscopic particles in liquids, the acceleration lasts only nanoseconds; thereafter particles move with constant velocities. So, our interest in this chapter is mostly the linear relationships between forces and velocities. The velocity v times a proportionality constant called the *friction coefficient* ξ equals the applied force f :

$$f = \xi v. \quad (17.2)$$

Substituting Equation (17.2) into Equation (17.1) gives the flux of particles J that results from an applied force f :

$$J = cv = \frac{cf}{\xi}. \quad (17.3)$$

In this case, and many others, fluxes are linearly proportional to the forces that cause them,

$$J = Lf, \quad (17.4)$$

where L is a proportionality constant. While some flows are caused by external forces, other flows arise from gradients.

Fick's Law and Fourier's Law: Gradients of Particles or Heat Cause Flows

A single-phase system that is in thermodynamic equilibrium and not subject to external forces will have particles uniformly distributed in space. But systems

not at equilibrium may have *gradients*. The thermodynamic tendency toward equilibrium is a tendency toward uniformity of particle concentrations or temperatures.

Dynamical processes are governed by fundamental empirical laws. Among them are *Fick's law* relating the forces and flows of particles and *Fourier's law* relating the forces and flows of heat. **Fick's first law** captures the general observation that the flux J of particles is proportional to the gradient of concentration, dc/dx :

$$J = -D \frac{dc}{dx}, \quad (17.5)$$

if the flow is one-dimensional. In three dimensions, the flux \mathbf{J} is a vector (c is always a scalar), and flows can have components in all directions, so

$$\mathbf{J} = -D \nabla c. \quad (17.6)$$

The proportionality constant D is called the *diffusion coefficient* and the minus sign indicates that particles flow *down* their concentration gradients, from regions of high concentration to regions of low concentration (see Figure 17.2). The flow rate of particles is proportional to the gradient of particle concentrations. Steep concentration gradients cause large fluxes of particles. Small gradients cause small fluxes. Also, the larger the diffusion constant, the higher the flux, for a given gradient. Table 17.1 lists some molecular diffusion constants.

While Fick's law describes how particle flow is driven by concentration gradients, **Fourier's law** describes how heat flow J_q is driven by temperature gradients dT/dx in materials:

$$J_q = -\kappa \frac{dT}{dx}, \quad (17.7)$$

where κ is called the *thermal conductivity* of the material. Metals and diamond have greater thermal conductivities than glass, water, and wood. In metals, heat is conducted rapidly away from regions of high temperatures; in glasses, heat is conducted away more slowly.

EXAMPLE 17.1 Diffusion can be seen as an 'effective force.' Because particles are driven to flow by concentration gradients, think of gradients as creating an effective force that moves the particles. If particle concentration $c(x)$ varies with spatial position x , the drive toward equalizing the chemical potential everywhere in space is a tendency to flatten concentration gradients. This is not a force that acts on each particle directly. It results from an entropy. Just as Newton's laws give forces as derivatives of energies, the force from a concentration gradient can be expressed as a derivative of the chemical potential. If c depends on x , then the chemical potential can be expressed as $\mu(x) = \mu^\circ + kT \ln c(x)$ and the average force f from a concentration gradient is

$$f = -\frac{d\mu}{dx} = -kT \frac{d \ln c}{dx} = -\frac{kT}{c} \frac{dc}{dx}. \quad (17.8)$$

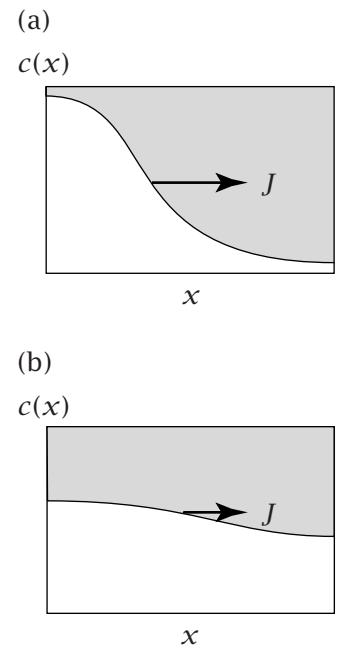


Figure 17.2 Flow results from concentration gradients. (a) A steep gradient causes large flux, while (b) a shallow gradient causes a small flux.

Table 17.1 Diffusion constants D for various molecules in air or water, as a function of their molecular masses M .

Molecule	Medium	T (°C)	M (g mol $^{-1}$)	D (cm 2 s $^{-1}$)
Hydrogen	Air	0	2	6.11×10^{-1}
Helium	Air	3	4	6.24×10^{-1}
Oxygen	Air	0	32	1.78×10^{-1}
Benzene	Air	25	78	9.60×10^{-2}
Hydrogen	Water	25	2	4.50×10^{-5}
Helium	Water	25	4	6.28×10^{-5}
Water (self-diffusion)	Water	25	18	2.13×10^{-5}
Oxygen	Water	25	32	2.10×10^{-5}
Urea	Water	25	60	1.38×10^{-5}
Benzene	Water	25	78	1.02×10^{-5}
Sucrose	Water	25	342	5.23×10^{-6}
Ribonuclease	Water	20	13,683	1.19×10^{-6}
Green fluorescent protein (GFP)	Water	25	31,000	8.7×10^{-7}
Hemoglobin	Water	20	68,000	6.90×10^{-7}
Catalase	Water	20	250,000	4.10×10^{-7}
Myosin	Water	20	493,000	1.16×10^{-7}
DNA	Water	20	6,000,000	1.30×10^{-8}
Tobacco mosaic virus (TMV)	Water	20	50,000,000	3.00×10^{-8}

Source: RW Baker, *Controlled Release of Biologically Active Agents*, Wiley, New York, 1987.

Substituting Equation (17.8) into Equation (17.4) gives

$$J = -\frac{LkT}{c} \frac{dc}{dx}. \quad (17.9)$$

Comparing Equation (17.9) with Equation (17.5) gives a useful relationship: $D = u k T$, where $u = L/c$ is called the *mobility* and $1/u$ is called the *friction coefficient*.

The next section describes how particle concentrations vary in time and space.

The Diffusion Equation Describes How Concentration Gradients Change Over Time and Space

Here, we derive the *diffusion equation*, which describes how particles flow due to concentration gradients. To keep it simple, consider diffusion in the x direction. Figure 17.3 shows a small volume element of flow. There are two ways to

calculate how the number of particles in this slab changes during a time interval Δt . First, if the flux $J_{\text{in}} = J(x, t)$ into this volume element is greater than the flux $J_{\text{out}} = J(x + \Delta x, t)$ out, then the number of particles in the slab will increase during Δt by $(J_{\text{in}} - J_{\text{out}})A\Delta t$, where A is the area of the slab, according to the definition of J . Second, if the concentration $c(t + \Delta t)$ at the end of the time interval is greater than the concentration $c(t)$ at the beginning, then you can also express the number of particles that accumulate during the time interval Δt as $[c(t + \Delta t) - c(t)]A\Delta x$, where Δx is the thickness of the slab, c is the number of particles per unit volume, and $A\Delta x$ is the volume element. Equating $\Delta J A \Delta t$ with $\Delta c A \Delta x$ for the increased number of particles in the volume element during the time interval Δt gives the relation

$$\frac{\Delta c}{\Delta t} = -\frac{\Delta J}{\Delta x}. \quad (17.10)$$

Take the limits $\Delta x \rightarrow dx$ and $\Delta t \rightarrow dt$ to get

$$\left(\frac{\partial c}{\partial t}\right) = -\left(\frac{\partial J}{\partial x}\right). \quad (17.11)$$

Substituting Equation (17.5) into Equation (17.11) gives **Fick's second law**, also called the **diffusion equation**:

$$\left(\frac{\partial c}{\partial t}\right) = \left(\frac{\partial}{\partial x} \left[D \left(\frac{\partial c}{\partial x}\right)\right]\right) = D \left(\frac{\partial^2 c}{\partial x^2}\right), \quad (17.12)$$

where the last equality applies if the diffusion coefficient D does not vary with x . The generalization to three dimensions is

$$\left(\frac{\partial c}{\partial t}\right) = -\nabla \cdot \mathbf{J} = D \nabla^2 c. \quad (17.13)$$

The diffusion equation is a partial differential equation. You solve it to find $c(x, y, z, t)$, the particle concentration as a function of spatial position and time. To solve it, you need to know two *boundary conditions* and one *initial condition*, that is, two pieces of information about the concentration at particular points in space and one piece of information about the concentration distribution at some particular time. Example 17.2 shows how to solve the diffusion equation for the permeation of particles through a planar film or membrane.

EXAMPLE 17.2 Diffusion through a slab or membrane. Drugs and metabolites pass through cell membranes or skin. In other instances, gases are forced through thin polymer membranes in industrial processes of gas purification or recovery. For example, forcing air through polymeric films separates O₂ from N₂ gases, because O₂ permeates faster.

Here's a model for the permeation of atoms and molecules through films and membranes. Figure 17.4 shows a slab of material having thickness h , through which particles flow. To the left of the membrane is a higher concentration c_l of particles and to the right is a lower concentration c_r . The difference in particle concentrations, $\Delta c = c_l - c_r$, drives particles to flow from left to right through the slab. We want to know the concentration profile $c(x, t)$ of particles inside the slab.

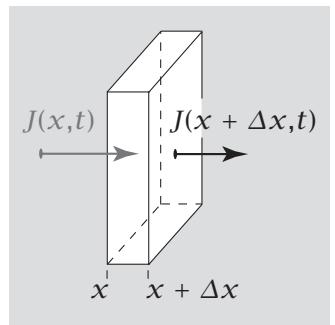
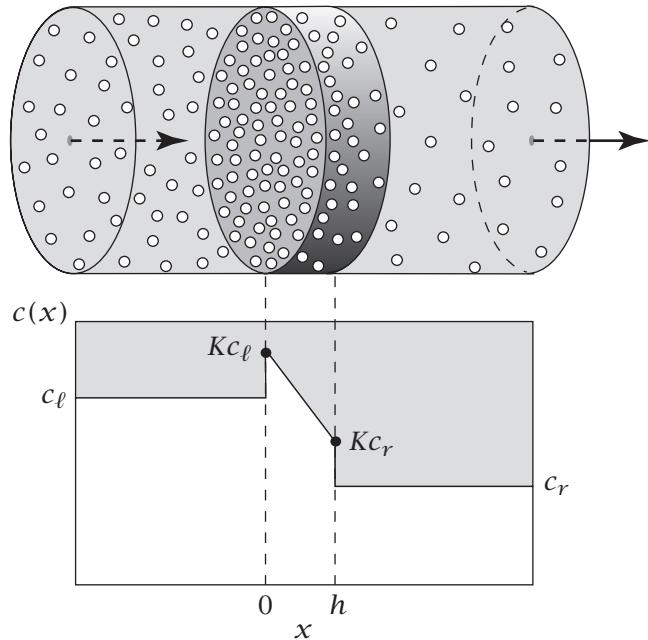


Figure 17.3 The flux $J(x + \Delta x, t)$ out of a volume element at $x + \Delta x$ at time t can be different from the flux $J(x, t)$ into the element because the material can be depleted or accumulated in the volume element.

Figure 17.4 Particles flow from a reservoir at concentration c_ℓ on the left, through a slab of material of thickness h , to a reservoir having a lower concentration c_r on the right. The partition coefficient from the solutions into the slab is K . At steady state, there is a linear gradient of particle concentration along the x axis through the slab.



For simplicity, consider *steady-state* flow, $(\partial c / \partial t) = 0$. This occurs, for example, if the particle reservoir on the left is large, so that c_ℓ does not change during the experiment. Then the diffusion equation (17.12) reduces to

$$\left(\frac{\partial^2 c}{\partial x^2} \right) = 0. \quad (17.14)$$

Integrating Equation (17.14) once gives

$$\left(\frac{\partial c}{\partial x} \right) = A_1, \quad (17.15)$$

where A_1 is a constant of integration. Integrating again gives

$$c(x) = A_1 x + A_2, \quad (17.16)$$

where A_2 is the second constant of integration.

Equation (17.16) says that the concentration profile in the slab is linear in x . You can get A_1 and A_2 if you know the partition coefficient K . K is the particle concentration in the slab divided by the particle concentration in the liquid; it is a property that depends on the two materials (see Chapter 16). Just inside the left wall of the slab at $x = 0$, the concentration is $c(0) = Kc_\ell$. Substituting this condition and $x = 0$ into Equation (17.16) gives

$$A_2 = Kc_\ell. \quad (17.17)$$

Inside the right wall at $x = h$, the concentration is $c(h) = Kc_r$. Substituting these boundary conditions into Equation (17.16) gives

$$A_1 = \frac{K(c_r - c_\ell)}{h}. \quad (17.18)$$

With these two boundary conditions, Equation (17.16) becomes

$$c(x) = \frac{K(c_r - c_\ell)}{h}x + Kc_\ell. \quad (17.19)$$

The gradient of the concentration in the slab is shown in Figure 17.4.

What is the flux J of particles through the slab? Take the derivative of Equation (17.19), as prescribed by Fick's law, Equation (17.5), to get

$$J = \frac{KD}{h}(c_\ell - c_r) = \frac{KD}{h}\Delta c. \quad (17.20)$$

The flow of particles through the membrane is driven by the difference in concentration on the two sides. Fluxes are highest for particles having large diffusion coefficients or large partition coefficients. The thicker the slab, the smaller the flux.

The *permeability* P of a membrane is defined as the flux divided by the driving force Δc :

$$P = \frac{J}{\Delta c} = \frac{KD}{h}. \quad (17.21)$$

The *resistance* of the membrane to the particle flux is defined as $1/P$, the inverse of the permeability. Figure 17.5 shows that the permeabilities of various solutes through lipid bilayer membranes increase approximately linearly with the partition coefficient, as predicted. Flow into biological cells is called *passive transport* if it is driven by such concentration gradients. *Active transport* is flow that goes against concentration gradients, driven by some energy source.

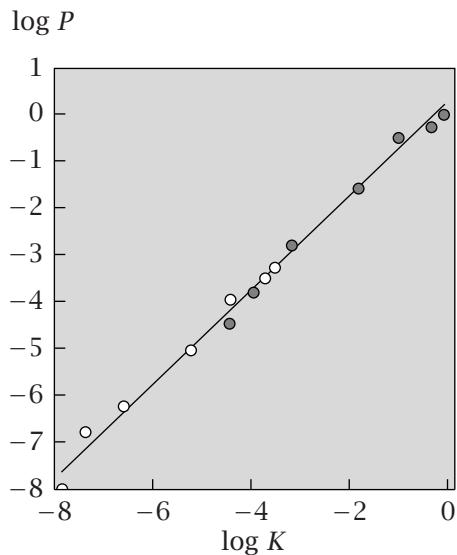


Figure 17.5 The permeability P is proportional to the oil/water partition coefficient K for *p*-methylhippuric acids (○) and *p*-toluic acids (●) through lipid bilayer membranes. Source: P Mayer, T-X Xiang, and BD Anderson, *AAPS PharmSci* **2**, article 14 (2000).

EXAMPLE 17.3 Oxygen permeability through a contact lens. Contact lenses must be highly permeable to oxygen so that your eyes can breathe. Assume a contact lens is a slab of material of thickness $h = 0.1$ mm. If the diffusion constant of oxygen through the material is $D = 2 \times 10^{-5}$ cm 2 s $^{-1}$ and if the partition coefficient is $K = 2.5 \times 10^4$, then the permeability P is

$$P = \frac{KD}{h} = \frac{(2.5 \times 10^4)(2 \times 10^{-5} \text{ cm}^2 \text{ s}^{-1})}{(10^{-2} \text{ cm})} = 50.0 \text{ cm s}^{-1} \quad (17.22)$$

So, if the concentration difference of oxygen across the lens is $\Delta c = 10^{-3}$ mol L $^{-1}$, then the flux is

$$\begin{aligned} J &= P\Delta c = (50 \text{ cm s}^{-1})(10^{-3} \text{ mol L}^{-1})\left(\frac{1 \text{ L}}{1000 \text{ cm}^3}\right) \\ &= 50 \mu\text{mol cm}^{-2} \text{ s}^{-1}. \end{aligned} \quad (17.23)$$

Now we switch from one-dimensional diffusion to a three-dimensional problem.

Diffusion of Particles Toward a Sphere

Consider a spherical particle of radius a such as a protein or micelle, toward which small molecules or ligands diffuse (see Figure 17.6). At what rate do the small molecules collide with the sphere? To compute the rate, solve the diffusion equation (17.13) in spherical coordinates (see Table G.1 in Appendix G). At steady state, $\partial c/\partial t = 0$. Assume that the flow depends only on the radial distance away from the sphere, and not on the angular variables. Then you can compute $c(r)$ using

$$\nabla^2 c = \frac{1}{r} \frac{d^2(r c)}{dr^2} = 0. \quad (17.24)$$

To solve for $c(r)$, integrate Equation (17.24) once to get

$$\frac{d(r c)}{dr} = A_1, \quad (17.25)$$

where A_1 is the first constant of integration. Integrate again to get

$$rc = A_1 r + A_2, \quad (17.26)$$

where A_2 is the second constant of integration. Rearranging Equation (17.26) gives $c(r)$:

$$c(r) = A_1 + \frac{A_2}{r}. \quad (17.27)$$

Two boundary conditions are needed to determine A_1 and A_2 . First, you know the concentration c_∞ of small molecules at a large distance from the sphere, $r \rightarrow \infty$, so Equation (17.27) must have $A_1 = c_\infty$. Second, assume that when each small molecule contacts the sphere, it is absorbed, dissipated, or transformed (for example, by a chemical reaction and dispersal of the products). This boundary condition gives an upper limit on the rate at which

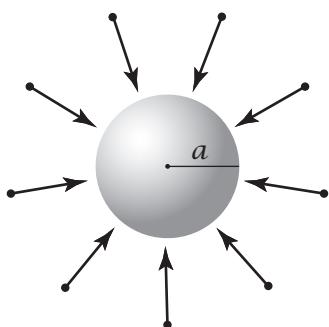


Figure 17.6 Small molecules (●) diffuse toward a sphere of radius a .

collisions lead to conversion to product. This condition, called an *absorbing boundary condition*, gives $c(a) = 0$ as the concentration at the surface of the sphere. Substituting $c(a) = 0$ and $A_1 = c_\infty$ into Equation (17.27) gives $A_2 = -c_\infty a$. With these boundary conditions, Equation (17.27) becomes

$$c(r) = c_\infty \left(1 - \frac{a}{r}\right). \quad (17.28)$$

Figure 17.7 shows this distribution of small molecules near the sphere. To get the flux, use

$$J(r) = -D \frac{dc}{dr} = \frac{-Dc_\infty a}{r^2}. \quad (17.29)$$

To get the number of collisions per second at $r = a$, called the *current* $I(a)$, multiply the flux (the number of particles colliding per unit area per second, from Equation (17.29)) by the area of the sphere's surface:

$$I(a) = J(a)4\pi a^2 = -4\pi Dc_\infty a. \quad (17.30)$$

The minus sign in Equation (17.30) for $I(a)$ indicates that the current is in the direction $-r$, toward the sphere. The current is large if D or c_∞ or a are large.

For spherical particles, **diffusion-controlled** reaction rates are given by Equation (17.30). Diffusion control implies that reactions are fast. Other reactions involve additional rate-limiting steps that occur after the reacting molecules have come into contact with the sphere. Diffusion control defines an upper limit on the speed of reactions. Any other kind of process must be slower than the diffusion-controlled process because the reaction must take additional time to complete after contact. Association rates are often expressed in terms of a rate coefficient k_a defined by $I(a) = -k_a c_\infty$, where

$$k_a = 4\pi Da. \quad (17.31)$$

EXAMPLE 17.4 Diffusion-controlled kinetics. Compute the diffusion-controlled rate coefficient for benzene, which has a diffusion constant $D = 1 \times 10^{-5} \text{ cm}^2 \text{ sec}^{-1}$, colliding with a protein, modeled as a sphere of radius $a = 10 \text{ \AA}$. Equation (17.31) gives the rate coefficient

$$\begin{aligned} k_a &= 4\pi \left(1 \times 10^{-5} \text{ cm}^2 \text{ s}^{-1}\right) \left(10^{-7} \text{ cm}\right) \left(6.02 \times 10^{23} \text{ molecules mol}^{-1}\right) \\ &\times \left(\frac{1}{1000 \text{ cm}^3}\right) = 7.6 \times 10^9 \text{ M}^{-1} \text{ s}^{-1}, \end{aligned}$$

where M is moles per liter. These are standard units for rate coefficients; this is the number of molecules per second hitting the sphere, divided by the bulk concentration, in molarity units.

Now, let's look at diffusion from a point source.

EXAMPLE 17.5 Diffusion from a point source. Put a drop of dye in water and observe how the dye spreads. To make the math simple, consider the spreading in one dimension. Begin by putting the dye at $x = 0$. To determine the dye

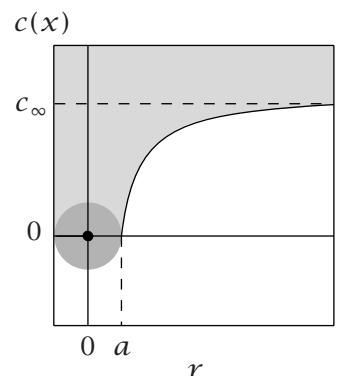


Figure 17.7
Concentration profile $c(r) = c_\infty(1 - a/r)$ of molecules as a function of distance r away from an absorbing sphere of radius a .

profile $c(x, t)$, you can solve Equation (17.12) by various methods. The solutions of this equation for many different geometries and boundary conditions are given in two classic texts on diffusion [1] and on heat conduction [2].

Heat conduction obeys the same differential equation as diffusion: if you replace particle concentration $c(x, y, z, t)$ with temperature $T(x, y, z, t)$, and replace the diffusion constant D with the *thermal diffusivity* κ , which expresses how the flow of heat depends on the temperature gradient, then the distribution of temperature in space and time is given by

$$\frac{\partial T}{\partial t} = \kappa \nabla^2 T. \quad (17.32)$$

So, one way you can handle a diffusion equation is to solve instead the corresponding heat flow equation having the same boundary conditions, and replace T with c , and κ with D . You can find solutions to diffusion problems in books on either diffusion or heat conduction for the given boundary conditions. It is sometimes useful to solve diffusion or heat flow problems by looking up the solutions in this way, particularly for systems having complicated geometries. For the present problem, see Equation (2.6) in [1]. The solution to this one-dimensional problem is

$$c(x, t) = \frac{n_0}{\sqrt{4\pi Dt}} \exp\left(-\frac{x^2}{4Dt}\right), \quad (17.33)$$

where n_0 is the initial amount of dye at the point $x = 0$. You can verify this solution by substituting Equation (17.33) into Equation (17.12) and taking the appropriate derivatives. Figure 17.8 shows how the dye spreads out with time. The larger the diffusion constant D , the faster the dye diffuses away from the site of the initial droplet. To generalize to two or three dimensions, you can regard diffusion along the x , y , and z axes as independent of each other. For any dimensionality, $d = 1, 2, 3, \dots$, Equation (17.33) becomes

$$c(r, t) = \frac{n_0}{(4\pi Dt)^{d/2}} \exp\left(-\frac{r^2}{4Dt}\right), \quad (17.34)$$

where r represents the distance in d -dimensional space. In three dimensions, for example, $r^2 = x^2 + y^2 + z^2$.

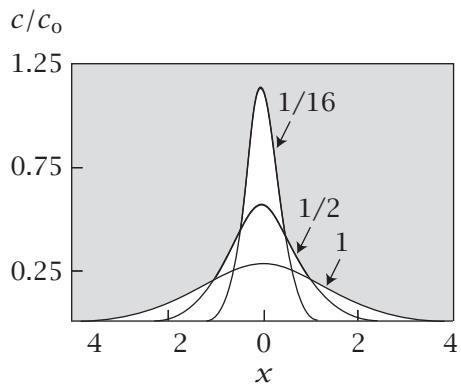


Figure 17.8 Concentration–distance curves for molecules that begin at $x = 0$ and diffuse in one dimension along x . The numbers on the curves are the values of Dt . Source: J Crank, *The Mathematics of Diffusion*, 2nd edition, Clarendon Press, Oxford, 1975.

So far, we have considered particles diffusing freely, driven only by spatial concentration gradients. Let's now consider particles that are also subjected to external applied forces.

The Smoluchowski Equation Describes Particles Driven by Both Applied Forces and Diffusion

Sometimes particles are driven both by diffusion, due to gradients, and by additional applied forces. To model their flows, you sum the component fluxes. To generalize Fick's first law (Equation (17.5)) add a net flux $c\nu$ (Equation (17.3)):

$$J = -D \left(\frac{\partial c}{\partial x} \right) + c\nu = -D \left(\frac{\partial c}{\partial x} \right) + \frac{cf}{\xi}. \quad (17.35)$$

Combine Equations (17.11) and (17.35) to get the **Smoluchowski equation**:

$$\left(\frac{\partial c}{\partial t} \right) = D \left(\frac{\partial^2 c}{\partial x^2} \right) - \frac{f}{\xi} \left(\frac{\partial c}{\partial x} \right). \quad (17.36)$$

EXAMPLE 17.6 The Smoluchowski equation applies to horse races, nuclear technology, biomolecular separations, and ion flows through membrane channel proteins. Let's model a horse race using the Smoluchowski equation. All the horses start at position $x = 0$ at time $t = 0$. The horses move toward the finish line with average velocity ν . Substituting $\nu = f/\xi$ into Equation (17.36) and solving gives [3]:

$$c(x, t) = \frac{n_0}{\sqrt{4\pi Dt}} \exp \left[-\frac{(x - \nu t)^2}{4Dt} \right], \quad (17.37)$$

where n_0 represents the number of horses at the starting line at time 0 (see Figure 17.9). The distribution $c(x, t)$ broadens with time, with variance $2Dt$. This treatment is a model for diffusing particles that are driven by an applied force to have an average velocity. It is only approximate for horse races, which usually have only small numbers of horses.

The Smoluchowski equation also plays a prominent role in separation technology. Take two types of particles, having masses m_1 and m_2 . Spin them in a centrifuge. The centrifugal forces on them will be different: $f_1 = m_1 \omega^2 x$ and $f_2 = m_2 \omega^2 x$. So, the particles will migrate toward the bottom of the centrifuge tubes with different average velocities $\nu = f/\xi$, in proportion to their masses. Centrifuges are used to separate biomolecules by molecular weight and to measure their molecular weights. T Svedberg won the 1926 Nobel Prize in Chemistry for the invention of the *ultracentrifuge*, which is a centrifuge that spins at very high speeds. With it, he showed in 1926 that protein molecules are single long-chain covalently bonded macromolecules, and not noncovalent assemblies of smaller molecules. In 1936, JW Beams and FB Haynes showed that centrifuges could separate isotopes of chlorine gas [4]. This was the precursor of the modern method of separating the uranium isotopes ^{238}U and ^{235}U in uranium hexafluoride gas to obtain ^{235}U for use in nuclear reactors and

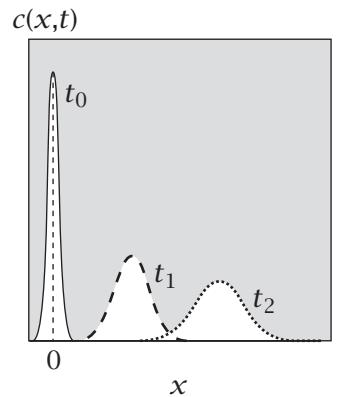


Figure 17.9
Smoluchowski model of a horse race. The horses start at time $t = 0$ at the starting gate $x = 0$. The peak of the distribution moves toward the finish line with velocity ν . The distribution is Gaussian; see Equation (17.37). Its variance increases with time as $2Dt$.

nuclear weapons. Another example of the Smoluchowski equation is the *Nernst–Planck* equation (22.40), which describes how ions diffuse through membrane channel proteins, driven by applied electrical forces.

Now we use Equation (17.36) to derive a relationship between the frictional drag experienced by a particle moving through a fluid, and the particle's diffusion constant.

The Einstein–Smoluchowski Equation Relates Diffusion and Friction

The Einstein–Smoluchowski equation says that the diffusion constant D is related to the friction coefficient ξ . To derive this relationship, consider the equilibrium illustrated in Figure 17.10. Put some molecules in a beaker of liquid. Gravity forces the molecules toward the bottom. Particles begin to concentrate at the bottom, forming a concentration gradient. This concentration gradient then acts to drive particles back upward, opposing the force of gravity. At equilibrium, the two fluxes (from gravity, forcing the particles downward, and from the concentration gradient, forcing the particles upward) are balanced so $J = 0$ in Equation (17.35). At equilibrium, c does not depend on time and you have

$$D \left(\frac{dc}{dx} \right) = \frac{cf}{\xi}. \quad (17.38)$$

Rearranging Equation (17.38) to put the c 's on one side and x on the other gives

$$D \frac{dc}{c} = \frac{f dx}{\xi}. \quad (17.39)$$

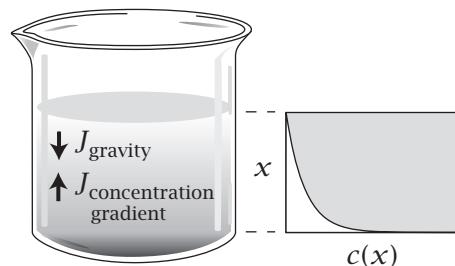
Integrate both sides and use Equation (3.9) for the reversible work $w = - \int f dx$. The minus sign on the work applies because gravity acts downwards while the positive x direction is upwards. This gives

$$D \ln \left[\frac{c(x)}{c(0)} \right] = \frac{-w}{\xi}. \quad (17.40)$$

Exponentiate Equation (17.40) to get

$$\frac{c(x)}{c(0)} = e^{-w(x)/\xi D}. \quad (17.41)$$

Figure 17.10 Particles in a beaker flow downward due to gravity, building up a high concentration at the bottom, leading to an opposing flux upward at equilibrium.



Because the system is in equilibrium, the Boltzmann distribution law also applies, and the quantity on the right side of Equation (17.41) must also equal $e^{-w/kT}$. Equating the exponents of the Boltzmann law and Equation (17.41) gives the **Einstein-Smoluchowski equation**:

$$D = \frac{kT}{\xi}. \quad (17.42)$$

This relationship shows how to determine the friction coefficient ξ if you know D . Because ξ depends on the sizes and shapes of particles, measuring the diffusion constant D gives some information about the structures of particles. For example, solving fluid flow equations (which we don't do here; see [5]) gives *Stokes's law* for spherical particles:

$$\xi = 6\pi\eta a, \quad (17.43)$$

where η is the viscosity of the solvent and a is the radius of the sphere.

Stokes's law is known from the following type of experiment. Put a ball of radius a in a liquid. The gravitational force is $(\Delta m)g$, where Δm is the mass difference between the object and displaced liquid. Measure its velocity. The ratio of gravitational force to velocity is ξ . Figure 17.11 gives friction factors for particles that aren't spherical. Combining Equations (17.42) and (17.43) gives

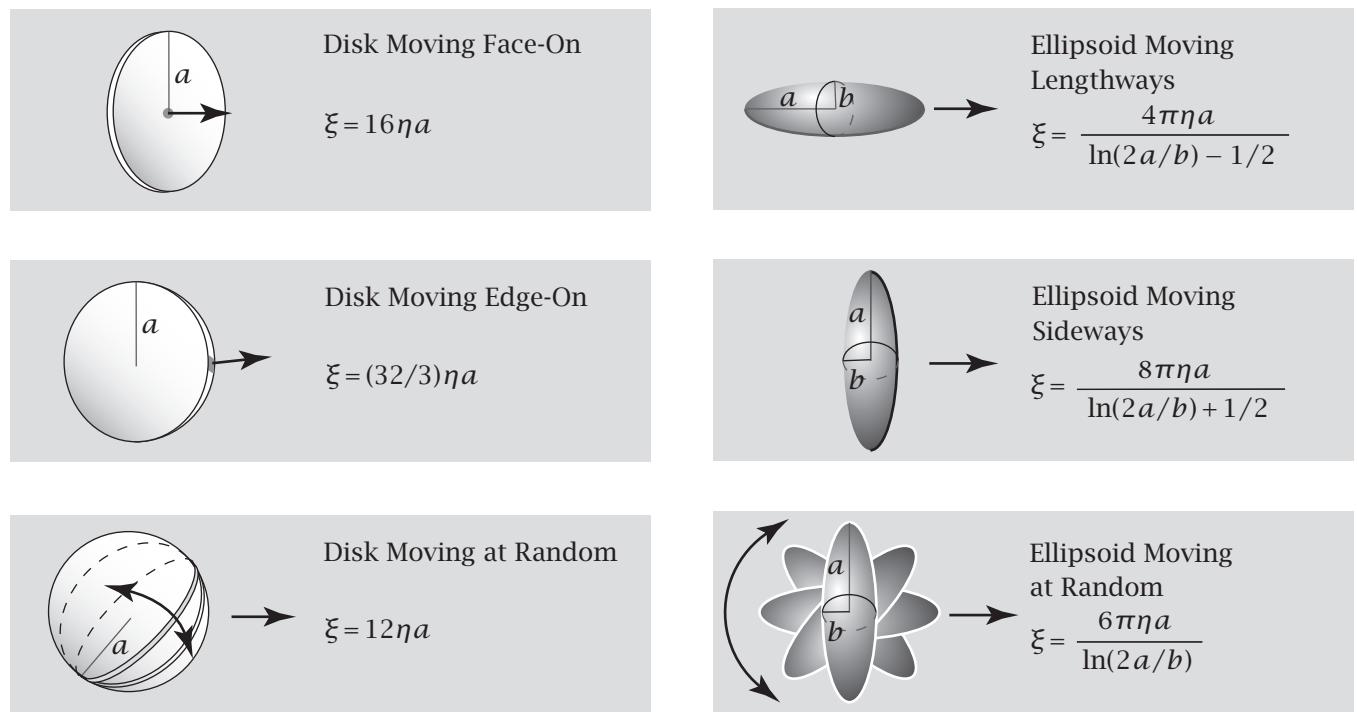
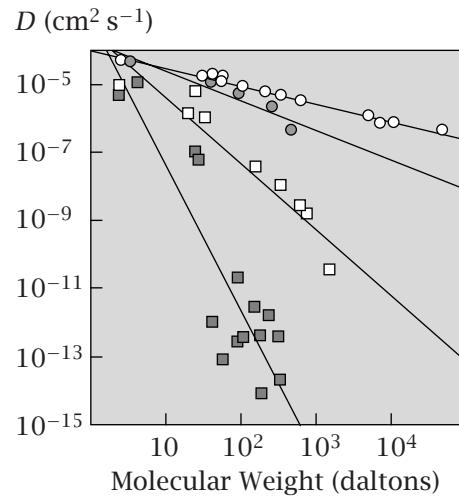


Figure 17.11 Friction coefficients ξ for objects in flow. η is the solvent viscosity, a and b are the dimensions shown, and the straight arrow indicates the flow direction. The objects include a disk moving face-on, edge-on, or averaged over all possible orientations, and an ellipsoid moving lengthways, sideways, or averaged over all orientations. Source: HC Berg, *Random Walks in Biology*, Princeton University Press, Princeton, 1993.

Figure 17.12 Diffusion coefficient as a function of solute molecular weight in water (\circ) and in three polymeric solvents: silicone rubber (\bullet), natural rubber (\square), and polystyrene (\blacksquare). The regression lines through the measurements have slopes of -0.51 (water), -0.86 (silicone rubber), -1.90 (natural rubber), and -4.20 (polystyrene). Source: adapted from RW Baker, *Controlled Release of Biologically Active Agents*, Wiley, New York, 1987.



the **Stokes-Einstein law** of diffusion for spherical particles,

$$D = \frac{kT}{6\pi\eta a}, \quad (17.44)$$

which predicts that larger spheres have greater friction and lower diffusion rates than smaller spheres. Viscosity has units of *poise*, $P = \text{dyn s cm}^{-2} = \text{g s}^{-1} \text{cm}^{-1}$. The viscosity of water at 20°C is 1.000 centipoise (cP).

Because the mass of a sphere is proportional to its volume, $m \sim a^3$, the Stokes-Einstein law implies $D \sim m^{-1/3}$. Figure 17.12 confirms that diffusion coefficients decrease with particle size, but it also shows that Equation (17.44) is not quantitatively accurate for particles in polymeric solvents, where the one-third-power law does not always hold.

Now we switch from diffusion to chemical and physical transformations.

Physical and Chemical States Evolve Toward Equilibrium

Consider the interconversion between two different states or chemical structures, as in Equation (13.1):



where k_f and k_r indicate the forward and reverse rate coefficients. How do the amounts of A and B change with time t , given the initial amounts at time $t = 0$?

The rate of increase of A is $d[A]/dt$, and the rate of increase of B is $d[B]/dt$, where $[]$ indicates the amount of each species. The forward **rate coefficient** k_f gives the probability that a molecule of A will convert to a B , per unit time. The number of molecules of A that convert to B per unit time is the product of (the number of molecules of A) \times (the probability that a molecule of A converts to a B per unit time). A rate coefficient k_r describes the probability per unit time of a reverse conversion, from B to A . You can use whatever units are most convenient: $[A(t)]$ can be given in terms of the numbers of molecules or in

terms of the concentration, such as the molarity or the mole fraction. Once you choose the units for the amounts, the units for the rates are determined. For the present problem, we'll use the numbers of molecules.

The overall rates of change of $[A(t)]$ and $[B(t)]$ are the creation rates minus the disappearance rates:

$$\frac{d[A(t)]}{dt} = -k_f[A(t)] + k_r[B(t)] \quad (17.46)$$

and

$$\frac{d[B(t)]}{dt} = k_f[A(t)] - k_r[B(t)] \quad (17.47)$$

Equations (17.46) and (17.47) are *coupled*: both quantities $[A(t)]$ and $[B(t)]$ appear in both equations. Such equations can be solved by standard matrix methods. Texts on chemical kinetics give the details [6, ?].

If the backward rate is much smaller than the forward rate, $k_r \ll k_f$, Equation (17.46) simplifies to

$$\frac{d[A(t)]}{dt} = -k_f[A(t)]. \quad (17.48)$$

To express the time dependence explicitly, rearrange and integrate Equation (17.48) from time $t = 0$ to t to get

$$\begin{aligned} \int \frac{dA'}{A'} &= - \int_0^t k_f dt' \implies \ln \frac{[A(t)]}{[A(0)]} = -k_f t \\ \implies [A(t)] &= [A(0)] e^{-k_f t}, \end{aligned} \quad (17.49)$$

where $[A(0)]$ is the concentration of A at time 0, the start of the reaction. Because $k_f > 0$, the amount of A diminishes exponentially with time. If $[A(t)] + [B(t)] = \text{constant}$, then $[B(t)]$ increases with time: $[B(t)] = \text{constant} - [A(0)] \exp(-k_f t)$.

Sources and Sinks Also Contribute to Flows: Examples from Population Biology

Some dynamical processes involve multiple *sources*, *inputs*, or *birth* processes or *sinks*, *outputs*, or *death* processes. To model such processes, add the fluxes. Here, we consider a problem of population dynamics, starting with the simplest kinetic model and adding other flux components.

Consider a population $c(t)$ of some biological species as a function of time. In 1798, TR Malthus first modeled situations in which a change in population, dc/dt , is proportional to the population itself:

$$\frac{dc}{dt} = ac, \quad (17.50)$$

because the number of offspring is proportional to the number of parents. The constant a equals the number of births minus the number of deaths per

member of the population. Rearranging to $dc/c = a dt$ and integrating both sides gives

$$c(t) = c_0 e^{at}, \quad (17.51)$$

where c_0 , the population at $t = 0$, is a constant. Equation (17.51) predicts that populations grow exponentially with time if $a > 0$, and that populations die out exponentially if $a < 0$.

But real populations can't sustain exponential growth indefinitely. In 1836, PF Verhulst introduced a term $-bc^2(t)$ to account for how high populations are reduced by competition or crowding. For high populations, c^2 becomes large in relation to c , and the minus sign says that high populations reduce the growth rate. This model for population growth gives

$$\frac{dc}{dt} = ac(t) - bc^2(t). \quad (17.52)$$

Integrating Equation (17.52) gives the solution

$$c(t) = \frac{c_0 e^{at}}{1 + (c_0 b/a)(e^{at} - 1)}. \quad (17.53)$$

In this case, $c(t)$ saturates with time, as Figure 17.13 shows.

In population biology and molecular science, plants, and animals or molecules evolve not only in time, but also in space [8]. For example, to account for geographic migrations of species, RA Fisher in 1937 included a spatial dependence by introducing the diffusion equation

$$\frac{\partial c(x, y, t)}{\partial t} = D \nabla^2 c + ac - bc^2. \quad (17.54)$$

Figure 17.14 shows a solution of this model. The population spreads by diffusion, but it reaches a limit to growth.

These examples illustrate that fluxes and diffusion are not limited to atoms and molecules, and they show how sources and sinks can be added to the diffusion equation. Example 17.7 applies the same principles to molecules that diffuse and react in time and space.

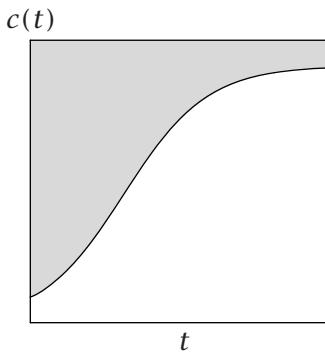


Figure 17.13 The Verhulst model in which the population $c(t)$ grows until it saturates, according to Equations (17.52) and (17.53). Source: JD Murray, *Mathematical Biology*, 2nd edition, Springer-Verlag, Berlin, 1993.

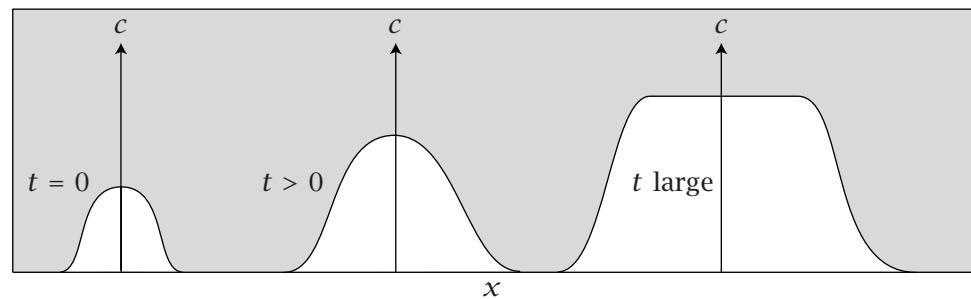


Figure 17.14 Diffusion and growth of a population in one dimension, subject to limits on growth due to crowding (see Equation (17.54)). Source: JD Murray, *Mathematical Biology*, 2nd edition, Springer-Verlag, Berlin, 1993.

EXAMPLE 17.7 Diffusion coupled with a chemical reaction. Suppose that drug molecules diffuse out of a tablet (which we model as a planar wall) into a solution (see Figure 17.15). The drug undergoes a chemical reaction in the solution. The chemical reaction rate is k_{rx} and the diffusion constant of the drug in the solution is D . Let's find the concentration profile for the drug as a function of distance x away from the tablet wall in the solution. If the reaction causes the drug to be depleted in proportion to its concentration c , then

$$\frac{dc}{dt} = -k_{rx}c. \quad (17.55)$$

Including this reaction process (sink) in the general diffusion equation (17.54) gives

$$\left(\frac{\partial c}{\partial t}\right) = D\left(\frac{\partial^2 c}{\partial x^2}\right) - k_{rx}c. \quad (17.56)$$

To simplify the problem, consider steady-state conditions, $(\partial c / \partial t) = 0$. c now depends only on x . You can get $c(x)$ by solving

$$\frac{d^2c}{dx^2} - \frac{k_{rx}}{D}c = 0. \quad (17.57)$$

The general solution is

$$c(x) = A_1 e^{-ax} + A_2 e^{ax}, \quad (17.58)$$

where $a = \sqrt{k_{rx}/D}$, and A_1 and A_2 are constants that must satisfy the boundary conditions. You can check this solution by substituting it into Equation (17.57). Because the concentration at $x \rightarrow \infty$ in a large container must be $c(\infty) = 0$, you have $A_2 = 0$ and

$$c(x) = c(0) \exp\left(-x\sqrt{\frac{k_{rx}}{D}}\right), \quad (17.59)$$

where $c(0)$ is the concentration at the wall, $x = 0$. The drug concentration diminishes exponentially with distance from the tablet. When $x = \sqrt{D/k_{rx}}$, $c(x)/c(0) = 1/e$. For comparison, if you had no chemical reaction ($k_{rx} = 0$ in Equation (17.57)) with the same boundary conditions, the drug concentration would diminish as a linear function of distance (Equation (17.16)).

At the surface ($x = 0$), the steady-state flux of the drug out of the tablet can be computed from Equation (17.59):

$$J = -D\left(\frac{dc}{dx}\right)_{x=0} = c(0)aDe^{-ax}|_{x=0} = c(0)\sqrt{Dk_{rx}}. \quad (17.60)$$

The drug is drawn out of the tablet rapidly if it has a high diffusion constant or has a high reaction rate in solution.

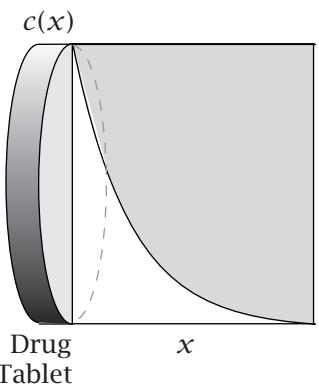


Figure 17.15
Concentration profile of a drug from a tablet, $c(x)$, at steady state. Drug is depleted in solution with rate constant k_{rx} .

So far, we have considered a single type of force f and a single type of flow J , related by $J = Lf$ (see Equation (17.4)). Now, we consider situations involving the coupling of multiple forces and flows.

Onsager Reciprocal Relations Describe Coupled Flows

In some materials, a temperature gradient drives a heat flow while a voltage difference also drives an electrical current. If such processes were independent of each other, you would have

$$J_1 = L_1 f_1 \quad \text{and} \quad J_2 = L_2 f_2, \quad (17.61)$$

where the subscript 1 denotes the heat flux and 2 denotes the electrical flux.

But, in general, multiple flow processes in a system are not independent; they are coupled. For example, the flow of charged particles leads to both an electrical current and to a change in the gradient of particle concentration. Applying a voltage drives the charges to move (an electrical current), at the same time it drives particle flow. The simplest (linear) relationships between two forces and two flows are

$$J_1 = L_{11} f_1 + L_{12} f_2 \quad \text{and} \quad J_2 = L_{21} f_1 + L_{22} f_2, \quad (17.62)$$

where L_{12} and L_{21} are the *coupling coefficients* that describe how the temperature gradient affects the electrical current and vice versa. A most remarkable experimental observation is that the coupling coefficients obey a very simple reciprocal relationship: $L_{21} = L_{12}$. Here is an example.

EXAMPLE 17.8 Thermoelectricity involves coupled flows: Seebeck and Peltier Effects. Thermoelectricity involves a coupling of two types of flows: an electrical current and a temperature difference. In 1821, TJ Seebeck discovered that temperature differences in metallic junctions can cause electrical currents to flow. Make a junction by twisting together two different wires made of two types of metal A and B. Now, hook together two such junctions and hold the junctions at different temperatures as shown in Figure 17.16.

If the two junctions are held at different temperatures, it will cause an electrical current to flow around the circuit shown. This is the basis for modern

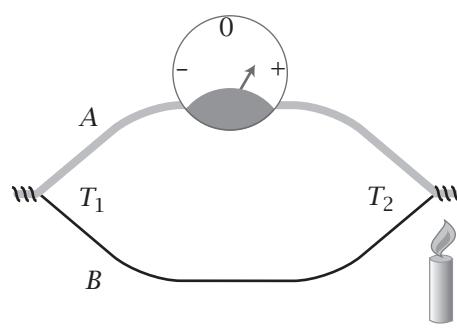


Figure 17.16 The Seebeck effect of coupled flows is the basis for thermocouples. Wires made of two different metals, A and B, are connected at two junctions (shown as twisted pairs of wires). If the two junctions are held at two different temperatures, an electrical current will flow. This principle is used to make thermocouples, which measure temperature. The principle, applied in reverse, with a current flow driving a temperature difference, is called the Peltier effect. It is used to make small refrigerators and ovens.

Table 17.2 Phenomenological flow coefficients L_{ij} for the system NaCl/KCl/water. The subscript 1 is NaCl and 2 is KCl.

Salt Concentrations (M)				
C_{NaCl}	0.25	0.5	0.25	0.5
C_{KCl}	0.25	0.25	0.5	0.5
$L_{ij} \times 10^9 RT$				
L_{11} ^b	2.61	4.76	2.79	5.15
L_{12} ^b	-0.750	-1.03	-0.99	-1.52
L_{21} ^a	-0.729	-1.02	-0.97	-1.45
L_{22} ^b	3.50	3.83	6.36	7.02
L_{12}/L_{21}	1.03	1.01	1.02	1.05

Sources: ^aH Fujita and LJ Gostling, *J Phys Chem* **64**, 1256–1263 (1960); ^bPJ Dunlop and LJ Gostling, *J Phys Chem* **63**, 86–93 (1959).

thermocouples, which are twisted-wire pairs that measure temperatures. In 1834, JCA Peltier discovered the reverse: if you apply a voltage to the circuit in Figure 17.16 (replacing the voltmeter), it will cause the two junctions to have different temperatures. This is the basis for modern Peltier refrigerators, which plug into wall sockets and convert that electricity into the heating or cooling of a small volume (enough to cool a few sodas.) The Onsager relationship says that the temperature difference needed to drive electrical current is related to the voltage difference that is needed to drive a temperature difference.

Here is another example of coupled flows. The diffusion of one salt, say NaCl, affects the diffusion of another, such as KCl. Table 17.2 shows careful measurements of such coupled diffusion processes, and shows that—at least in this case—the reciprocal relations hold to within experimental error.

Such couplings are called Onsager reciprocal relations, after L Onsager (1903–1976), a Norwegian physical chemist who explained this symmetry as arising from *microscopic reversibility* [9, 10]. (Although the Norges Tekniske Høgskole judged Onsager's work unacceptable and did not award him a PhD degree, they later awarded him an honorary doctorate. This work was awarded the Nobel Prize in Chemistry in 1968.)

Summary

Flows of particles or energy can be driven by external forces. Flows are often linearly proportional to the applied forces. Also, particles flow where there are concentration gradients. Energy flows where there are temperature gradients. These are described by Fick's law and Fourier's law and the diffusion equation. The proportionality constant between the flow rate and the gradient is a transport property. For particle flow, the ratio is the diffusion constant. For energy flow, the ratio is the thermal conductivity. And, dynamical changes in systems also result from chemical and physical transformations. Multiple forces and flows are coupled; Onsager's relations describe symmetries among the coupled flows.

Problems

1. Diffusion-controlled dimerization reactions.

- (a) Using the Stokes-Einstein relation, show that the rate of diffusion of a particle having radius a to a sphere of radius a can be expressed in terms that depend only on temperature and solvent viscosity.
- (b) Using the expression from (a), compute the diffusion-controlled dimerization rate k_a of two identical spheres in water ($\eta = 0.01 \text{ P}$) at $T = 300 \text{ K}$.
- (c) For the diffusion of two particles of radii a_A and a_B and diffusion constants D_A and D_B , the generalization of Equation (17.31) for the dimerization rate is

$$k_a = 4\pi(D_A + D_B)(a_A + a_B). \quad (17.63)$$

Show that Equation (17.63) reduces to $k_a = 8kT/3\eta$ when $a_A = a_B$ and $D_A = D_B$.

2. Diffusion of drug from a delivery tablet.

Suppose a drug is encapsulated between two planes at $x = 0$ and $x = h$. The drug diffuses out of both planes at a constant rate R , so the diffusion equation is

$$D \frac{d^2c}{dx^2} = -R.$$

- (a) Solve for $c(x)$ inside the tablet, subject to boundary conditions $c(0) = c(h) = 0$, that is, the drug is used up the instant it is released.
- (b) Compute the flux of the drug out of the tablet.

3. Diffusion of long rods is faster than diffusion of small spheres.

If ligand molecules have diffusion coefficient D and concentration c_∞ in solution, the current $I(L, a)$ of their diffusional encounters with a rod of length L and radius a is

$$I(L, a) \approx -\frac{4\pi DLC_\infty}{\ln(2L/a)}.$$

For a long rod, such as DNA, for which $L/a = 10^6$, compute the ratio of $I(L, a)$ to $I(a)$, the current to a sphere of radius a .

4. HIV growth kinetics.

According to [11], the following model is sufficient to describe the time-dependent amount $V(t)$ of virus in the body of an HIV-infected patient. Two factors contribute to the rate dV/dt : (1) a constant rate P of viral production, and (2) a clearance rate cV at which virus is removed from the body.

- (a) What differential equation gives $V(t)$?
- (b) Solve that equation to show the function $V(t)$.

5. Permeation through bilayers.

Acetamide has an oil/water partition coefficient of $K_w^o = 5 \times 10^5$. It has a permeability $P = 5 \times 10^4 \text{ cm s}^{-1}$ through oil-like bilayers, which are 30 \AA thick. What is the diffusion constant D for acetamide across the bilayer?

6. Computing the radius of hemoglobin.

Assume that hemoglobin is a sphere. Compute its radius from its diffusion constant in Table 17.1.

7. Rotational diffusion.

Consider particles that are oriented at an angle θ with respect to the x axis. The distribution of particles at different angles is $c(\theta)$. The flux $J(\theta)$ of particles through different angles depends on the gradient as in Fick's law:

$$J(\theta) = -\Theta \left(\frac{\partial c}{\partial \theta} \right),$$

where Θ is the orientational diffusion coefficient.

- (a) Write an expression for the rotational diffusion equation.
- (b) If the orientational friction coefficient for a sphere of radius r is $f_{\text{or}} = 8\pi\eta r^3$, write an expression for $\Theta(r)$.

8. Friction of particles.

- (a) A sphere and disk have the same radius. For a given applied force, which moves faster in flow?
- (b) Does a long thin ellipsoid move faster lengthwise or sideways?

9. Diffusion/reaction length from a drug tablet.

A drug that has diffusion constant $D = 10^{-6} \text{ cm}^2 \text{ s}^{-1}$ reacts at a rate $k_{\text{rx}} = 10^2 \text{ s}^{-1}$ at the aqueous surface of a tablet. At what 'decay length' is the concentration of drug $1/e$ of its value inside the tablet?

10. Drug delivery.

You have successfully developed a new drug with a molecular mass of 199.1 g mol^{-1} . Owing to its chemical properties and the type of disease to be treated, you decide to develop a transdermal patch for the drug delivery. You engineer a $51 \text{ mm} \times 51 \text{ mm}$ square transdermal patch that is just 2 mm thick. You are able to incorporate 700 mg of your drug into each patch. The region of the skin where the patch is to be applied has an average thickness of just 650 nm . There is a strong capillary network in this region, so any drug that crosses the skin is immediately removed by the circulatory system and transported away from the drug delivery site. Assume the rate of flux does not change as the concentration of drug in the patch decreases.

- (a) After a variety of tests, you determine that the partition coefficient through the skin is 0.56 . The diffusion constant is calculated as $6.9 \times 10^{-9} \text{ cm}^2 \text{ s}^{-1}$. Calculate the permeability P (in units of cm s^{-1}).
- (b) What is the flux in units of $\text{mol cm}^{-2} \text{ s}^{-1}$?
- (c) The recommended dosage time is $12 \text{ h}/\text{patch}$. Calculate, in grams, the amount of drug delivered over this time.
- (d) At the end of the treatment, what percentage of drug remains in the patch? If a user applied the patch on Monday at 5 pm and the rate of release of the drug remains linear until the drug is completely gone, at what time would the patch be empty?

11. Chemotaxis of a cell in a microfluidics tube. Some cells have the ability to undergo chemotaxis. That is, the cell can sense a gradient of a chemical attractant from its front to its back, and move in the up-gradient direction. Figure 17.17 shows a cell that is initially stationary in the middle of a microfluidics tube. The tube is 100 μm long and the cell can be approximated as sphere 10 μm in diameter. Independent flows can be used to control the concentration c of an attractant molecule A at each end of the tube. The concentration at the front end of the tube is maintained at 1 μM , and the concentration at the back end is maintained at 0 μM . Assume that a steady-state concentration profile has been reached inside the microfluidics

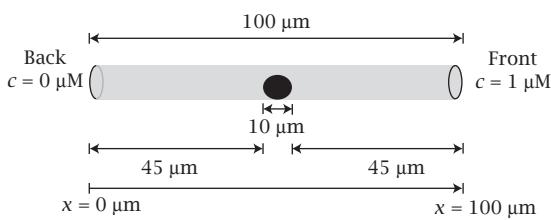


Figure 17.17

tube. All experiments are at $T = 300\text{ K}$. The viscosity of water is $1\text{ cP} = 0.001\text{ N s m}^{-2}$.

- The attractant molecule A is a sphere of radius 1 nm. What is its diffusion constant in water?
- Solve the diffusion equation and compute the concentration of attractant, $c(x)$, across the tube. What is the difference in concentration felt by the front of the cell versus the back of the cell?
- Upon binding to a receptor on the cell surface, attractant A triggers phosphorylation of another protein, B, at the front of the cell. Assume that B can freely diffuse in the cytoplasm, and does so with a diffusion constant $D = 10^{-10}\text{ m}^2\text{ s}^{-1}$. How long does it take for the phosphorylated protein B to diffuse to the back end of the cell?
- Now, instead assume that B is a transmembrane protein, and therefore can only diffuse in the membrane (the protein is limited to the surface of the cell). It does so with a diffusion constant $D = 10^{-11}\text{ m}^2\text{ s}^{-1}$. How long does it take now for B to diffuse to the back end of the cell? Hint: Think about distances on a globe, rather than a straight-line distance.

References

- J Crank, *The Mathematics of Diffusion*, 2nd edition, Clarendon Press, Oxford, 1975.
- HS Carslaw and JC Jaeger, *Conduction of Heat in Solids*, 2nd edition, Clarendon Press, Oxford, 1959.
- RD Astumian, *Am J Phys* **74**, 683–688 (2006).
- JW Beams and FB Haynes, *Phys Rev* **50**, 491–492 (1936).
- GK Batchelor, *An Introduction to Fluid Dynamics*, Cambridge University Press, Cambridge, 1967.
- WC Gardiner Jr, *Rates and Mechanisms of Chemical Reactions*, WA Benjamin, New York, 1969.
- JI Steinfeld, JS Francisco, and WL Hase. *Chemical Kinetics and Dynamics*, 2nd edition, Prentice-Hall, Upper Saddle River, NJ, 1998.
- JD Murray. *Mathematical Biology*, Volumes I and II, 3rd edition, Springer-Verlag, New York, 2002 and 2003.
- L Onsager, *Phys Rev* **37**, 405–426 (1931).
- L Onsager, *Phys Rev* **38**, 2265–2279 (1931).
- DD Ho, AU Neumann, AS Perelson, et al., *Nature* **373**, 123–126 (1995).

Suggested Reading

HC Berg, *Random Walks in Biology*, Princeton University Press, Princeton, NJ, 1993. Elementary and concise overview of diffusion, flow, and random walks.

TF Weiss, *Cellular Biophysics*, Volume 1, MIT Press, Cambridge, 1996. Detailed and extensive treatment of forces and flows through biological membranes.

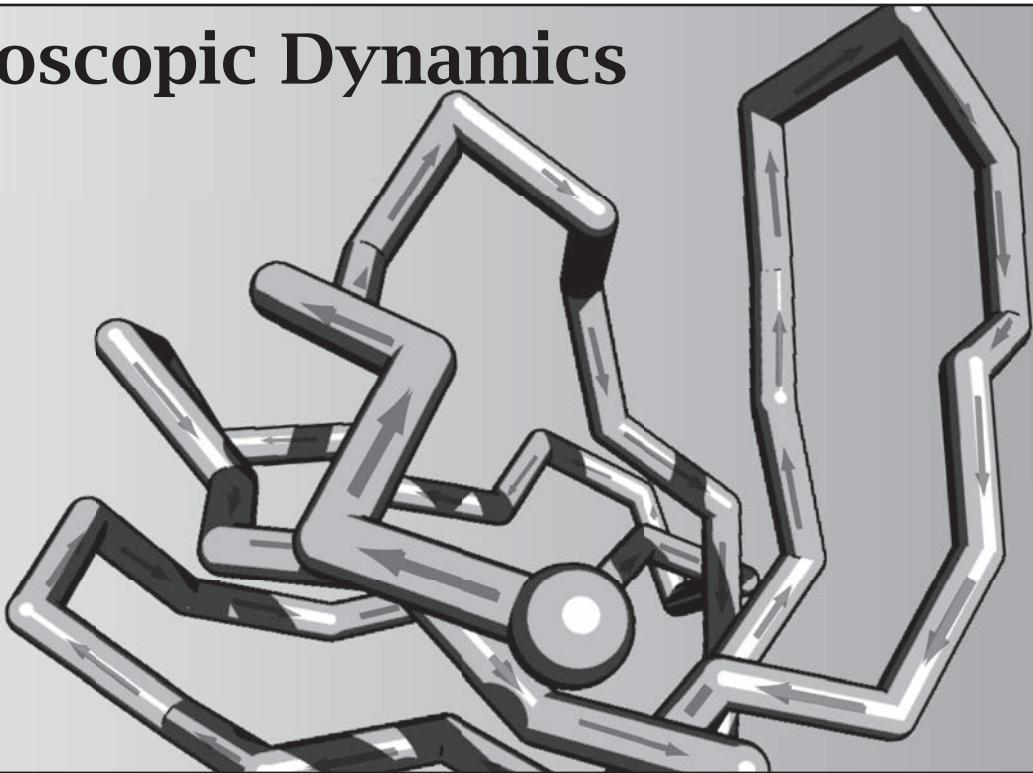
Excellent treatments of Onsager relations and coupling include:

R Haase, *Thermodynamics of Irreversible Processes*, Addison-Wesley, Reading, MA, 1969 (reprinted by Dover Publications, New York, 1990).

A Katchalsky and PF Curran, *Nonequilibrium Thermodynamics in Biophysics*, Harvard University Press, Cambridge, 1965.

This page is intentionally left blank.

18 Microscopic Dynamics



The Dynamical Laws Arise from Stochastic Processes

We now explore *nonequilibrium statistical mechanics*, a more microscopic perspective on physical kinetics. We use coin-flip statistics as the basis for the *random-flight model* of the possible trajectories of a single particle as it diffuses in time and space; for Fick's law of diffusion; and for the Poisson and waiting-time distributions for the inter-arrival times of stochastic events. We also explore the individual stochastic trajectories of particles in $A \rightarrow B$ two-state dynamical processes. Bulk experiments measure averages over such trajectories, while single-molecule experiments can give you additional information about the individual trajectories themselves and the *fluctuations*, which are the dynamical deviations away from the averages.

Diffusing Particles Can be Modeled as Random Flights

A diffusing particle undergoes *Brownian motion*, the jerky random battering of the particle by atoms or molecules in its surrounding medium. You can model the diffusion of particles, the conduction of heat, or the conformations of polymer chains using the *random-walk model* (in two dimensions), also sometimes called the *random-flight model* (in three dimensions). The random-walker problem was originally posed by the English mathematician K Pearson (1857–1936)

in 1905. He was interested in how mosquitoes enter a cleared jungle. The solution was first provided for a different problem by the English physicist Lord Rayleigh (1849–1919), who won the Nobel prize in physics in 1904 for the discovery of argon.

Consider first a random walk in one dimension. Imagine a ‘walker’ (a particle) that takes one step during each unit time interval. Each step is one unit long, but the step direction is randomly chosen to be either ‘forward’ in the $+x$ or ‘backward’ in the $-x$ direction. Steps are independent of each other. We want to compute the probability $P(m, N)$ that if a walk has N total steps, exactly m of those steps will be in the $+x$ direction and $(N - m)$ steps will be in the $-x$ direction. Because the (+) and (−) step directions are chosen like coin flips, $P(m, N)$ is a binomial probability distribution, Equation (1.31):

$$P(m, N) = \frac{N!}{m!(N-m)!} f^m b^{N-m}, \quad (18.1)$$

where f is the *a priori* probability of a forward step and $b = 1 - f$ is the probability of a backward step. (Think of a coin that has a probability f of heads and b of tails.) The function $P(m, N)$ has a maximum. We aim to approximate the distribution function P near the point of the maximum, which we call m^* , since that region of the function describes the most probable random walks. (Figure 2.6 shows that for $f = b = 1/2$, the maximum occurs in the middle, at $m^* = N/2$.) Owing to the mathematical intractability of the factorials, it is easiest to work with the logarithm, $\ln P$, which can be computed as a Taylor series expansion around the most probable ending point of the walk m^* for fixed N :

$$\begin{aligned} \ln P(m) &= \ln P^* + \left(\frac{d \ln P}{dm} \right)_{m^*} (m - m^*) \\ &\quad + \frac{1}{2} \left(\frac{d^2 \ln P}{dm^2} \right)_{m^*} (m - m^*)^2 + \dots, \end{aligned} \quad (18.2)$$

where $P^* = P(m^*)$ is the function P evaluated at $m = m^*$. Taking the logarithm of Equation (18.1) and using Stirling’s approximation gives

$$\ln P = N \ln N - m \ln m - (N - m) \ln(N - m) + m \ln f + (N - m) \ln b. \quad (18.3)$$

The most probable number of forward steps, $m = m^*$, is the point at which $P(m)$ or $\ln P(m)$ is at a maximum, or where $(d \ln P / dm)_{m^*} = 0$. Taking the derivative of Equation (18.3), with N constant, gives

$$\begin{aligned} \frac{d \ln P}{dm} \Big|_{m^*} &= -1 - \ln m^* + \ln(N - m^*) + 1 + \ln f - \ln b = 0 \\ \Rightarrow \ln\left(\frac{m^*}{N - m^*}\right) &= \ln\left(\frac{f}{b}\right) \end{aligned} \quad (18.4)$$

$$\Rightarrow m^* = N \left(\frac{f}{f + b} \right) = Nf, \quad (18.5)$$

where we substituted $f + b = 1$ at the end. (In problems of this type, you keep variables such as f and b explicit until after you have taken all the derivatives you want, then you substitute relations such as $f + b = 1$ at the end.) At the maximum of P , which occurs at $m = m^* = Nf$, the first-derivative term in

Equation (18.2) is zero. The second derivative, also evaluated at $m = m^*$, is

$$\frac{d^2 \ln P}{dm^2} \Big|_{m^*} = \left(-\frac{1}{m} - \frac{1}{N-m} \right)_{m^*=Nf} = -\frac{1}{Nfb}, \quad (18.6)$$

where again we substituted $f+b=1$ at the end. Substituting Equation (18.6) into Equation (18.2) and exponentiating gives

$$P(m, N) = P^* e^{-(m-m^*)^2/(2Nfb)}. \quad (18.7)$$

$P(m, N)$ is the distribution of ending locations m of one-dimensional random walks that begin at the origin and have N total steps. According to Equation (18.7), $P(m, N)$ is a Gaussian distribution function. This is a good approximation as $N \rightarrow \infty$. Let's now convert this equation to a more useful form. The simplest way to find P^* is to compare Equation (18.7) with the properly normalized Gaussian probability distribution, Equation (1.50), to get the variance

$$\sigma^2 = Nfb, \quad (18.8)$$

and to get the distribution

$$P(m, N) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-x^2/2\sigma^2} = \frac{1}{\sqrt{2\pi Nfb}} e^{-(m-Nf)^2/(2Nfb)}. \quad (18.9)$$

Equation (18.9) says that the most probable ending point (where P is maximal) is $m^* = Nf$. If the random walk is symmetrical, $f = b = 1/2$. And if you take the number of forward minus backward steps, $(m - (N - m)) = 2m - N$, as a measure of the *net forward progress*, you find that the particle will most likely terminate its walk approximately where it started ($2m^* - N = 0$). This is because the distribution is symmetrical in forward and backward steps. Only a few rare trajectories will have substantially different numbers of forward versus backward steps.

Is there a single number that best characterizes a random walk? The average distance traveled is not a very informative quantity; it equals zero. A more informative quantity is the *root-mean-square (RMS)* distance traveled, which is a measure of the width, or the variance, of the distribution function $P(m, N)$. The *mean-square* ending position in x is $\langle (m - Nf)^2 \rangle$, and its square root, $\langle (m - Nf)^2 \rangle^{1/2}$, is the RMS distance. The mean-square distance traveled by the one-dimensional walker is

$$\langle (m - Nf)^2 \rangle = \sigma^2 = Nfb = 2Dt, \quad (18.10)$$

where the last equality comes from comparing with the diffusion equation (17.33). There is a correspondence between the random-flight model and the diffusion equation: the number of steps N in the random-flight model corresponds to t in the diffusion equation, the time of diffusion.

Equation (18.10) shows a key feature of random walks. The RMS displacement of the walker from the origin, $\langle x^2 \rangle^{1/2} = \langle (m - Nf)^2 \rangle^{1/2}$, scales not linearly with the number of steps, as for a directed walker, but rather as the square root of the number of steps because of the random meandering. That

is, random walking leads away from a starting point much more slowly than directed motion does.

Generalizing to a d -dimensional walk, you have

$$\langle r^2 \rangle = \langle x^2 \rangle + \langle y^2 \rangle + \langle z^2 \rangle + \dots = N = 2dDt. \quad (18.11)$$

For example in $d = 3$ dimensions, $\langle r^2 \rangle = 6Dt$.

Equation (18.11) tells you how to convert D , a measurable quantity, into microscopic information about the rms distance $\langle x^2 \rangle^{1/2}$ that the particle moves in time t . For example, suppose you perform computer simulations using an atomic model of a particle undergoing Brownian motion. From that model, you simulate the particle distribution and compute the mean-square particle displacement $\langle x^2 \rangle$. Equation (18.11) provides a way to relate your microscopic model to experimental measurements of D .

Another Way to Calculate the Properties of $P(m, N)$

Here's another useful way to compute dynamical averages and fluctuations. Define a dynamical partition-function-like quantity, q_d , as the sum over the probabilities $P(m, N)$ (from Equation (18.1)) over all the possible trajectories, over all values of m :

$$\begin{aligned} q_d &= \sum_{m=0}^N P(m, N) \\ &= b^N + Nb^{N-1}f + \frac{N(N-1)}{2}b^{N-2}f^2 + \dots + Nb^{N-1}f + f^N \\ &= (b+f)^N. \end{aligned} \quad (18.12)$$

From left to right in Equation (18.12), the first term, b^N , describes the one random walk in which all N steps are backward. The second term, $Nb^{N-1}f$, describes the N different random-walk trajectories that have one step forward and $N-1$ steps backward (there are N of them because the one step forward could have been at any one of the N time steps). The last term, f^N , describes the one trajectory in which all N steps are forward. (Note that since $b+f = 1$, you have $q_d = 1$. When you perform mathematical operations involving q_d , such as taking derivatives, you keep the form $q_d = (b+f)^N$, and only substitute $q_d = 1$ at the last step.) The average is

$$\begin{aligned} \langle m \rangle &= \sum_{m=0}^N mP(m, N) \\ &= (0) \left(\frac{b^N}{q_d} \right) + \dots + (N-1) \left(\frac{bf^{N-1}}{q_d} \right) + (N) \left(\frac{f^N}{q_d} \right). \end{aligned} \quad (18.13)$$

Here is a way to simplify this expression. Look at the terms in Equation (18.13). Notice that you can express each term by taking a derivative with respect to f and then multiplying by f (to bring down the factor of $N-j$ in the exponent) (see also Equation (C.5) in Appendix C):

$$\langle m \rangle = \frac{f}{q_d} \frac{dq_d}{df} = \frac{d \ln q_d}{d \ln f}. \quad (18.14)$$

Equation (18.14) gives a useful way to compute the average. To see this, use the expression $q_d = (b+f)^N$ from Equation (18.12), and take the derivative according to Equation (18.14), to get

$$\langle m \rangle = \frac{f N (b+f)^{N-1}}{(b+f)^N} = Nf, \quad (18.15)$$

where we used $f+b=1$ in the last step.

Often, it is easy to construct an expression for q_d for a given dynamical problem, and then to find the average by taking the derivative in Equation (18.14). For example, in this case, $q_d = (b+f)^N$ can be constructed using the laws of probability. You add b to f inside the parentheses because the particle has two mutually exclusive options, namely to step forward or backward on each step. You multiply the factors $(b+f)$ together to get $(b+f)^N$ because each step decision is independent of each other.

You can also take other derivatives of q_d to compute variances or higher moments of $P(m, N)$. For example, by looking at the individual terms as we did above, you can show (see Equations (C.7) and (C.8)) that the variance is

$$\sigma^2 = \langle (m - m^*)^2 \rangle = f \frac{d}{df} \left(\frac{f}{q_d} \frac{dq_d}{df} \right) = \frac{d^2 \ln q_d}{d \ln f^2}. \quad (18.16)$$

Since you already found $d \ln q_d / d \ln f = \langle m \rangle = Nf / (f+b)$, you just need to take one more derivative to get the variance for the random-flight model:

$$\sigma^2 = f \frac{d}{df} \left(\frac{Nf}{f+b} \right) = \frac{Nfb}{(f+b)^2} = Nfb, \quad (18.17)$$

where we substituted $f+b=1$ at the last step. This partition-function-based approach is often an easy way to get averages, variances, and other properties. The random-flight model describes the diffusion of a single particle over time. We now consider the related problem of the diffusion of many particles at a given time.

What is the Basis for Fick's Law?

Why do particles flow from higher to lower concentrations? How does a particle ‘know’ which direction is downhill? After all, each particle moves only independently and randomly. This is sometimes called the *problem of reversibility*: according to Newton’s laws, the motions of one particle are fully time-reversible and random, and yet when particles are taken together collectively, they move in a particular direction. There is an arrow of time. Put a drop of dye in water. If you watched a movie of just one single dye molecule jiggling around by Brownian motion, you could not tell if the movie was running forwards or backwards in time. But if you watch the whole drop of dye, it is simple to tell whether the movie is running forward in time (the dye spreads out in water) or backwards (the dye contracts into a single point).

Recall Fick’s law of particle diffusion, Equation (17.5):

$$\langle J \rangle = -D \frac{dc}{dx}. \quad (18.18)$$

Here, we use the brackets $\langle J \rangle$ to indicate the average flux. We did not need the brackets in Chapter 17, because all the flux quantities were averages.

However, in this chapter, we retain the more explicit notation (...) for averages, to distinguish them from the other moments of rate distributions. What is the microscopic basis for Fick's law? Let's first reduce diffusion to its essence, the stochastic hopping of particles between two adjacent spatial planes. Figure 18.1 shows *the dog-flea model* of diffusion [1]. Instead of a continuum of values of x , you have just two positions: x and $x + \Delta x$. In our dog-flea metaphor, position 1, x , is called dog 1 and position 2, $x + \Delta x$, is called dog 2. At time t , dog 1 has $M_1(t)$ fleas on it (representing particles) and dog 2 has $M_2(t)$ fleas on it. During the next time interval, any flea (particle) can jump from the dog (plane) it is currently on to the other dog, or the flea can stay on its current dog. Here is the question of principle of Fick's law: If there are more fleas on dog 1 at time t (i.e., $M_1 > M_2$), why will the average net flux of fleas during Δt be proportional to $M_1 - M_2$?

First focus on a single dog. Suppose there are $M(t)$ fleas on a dog at time t . We are interested in the probability $P(j, M)$ that j of those fleas will jump to

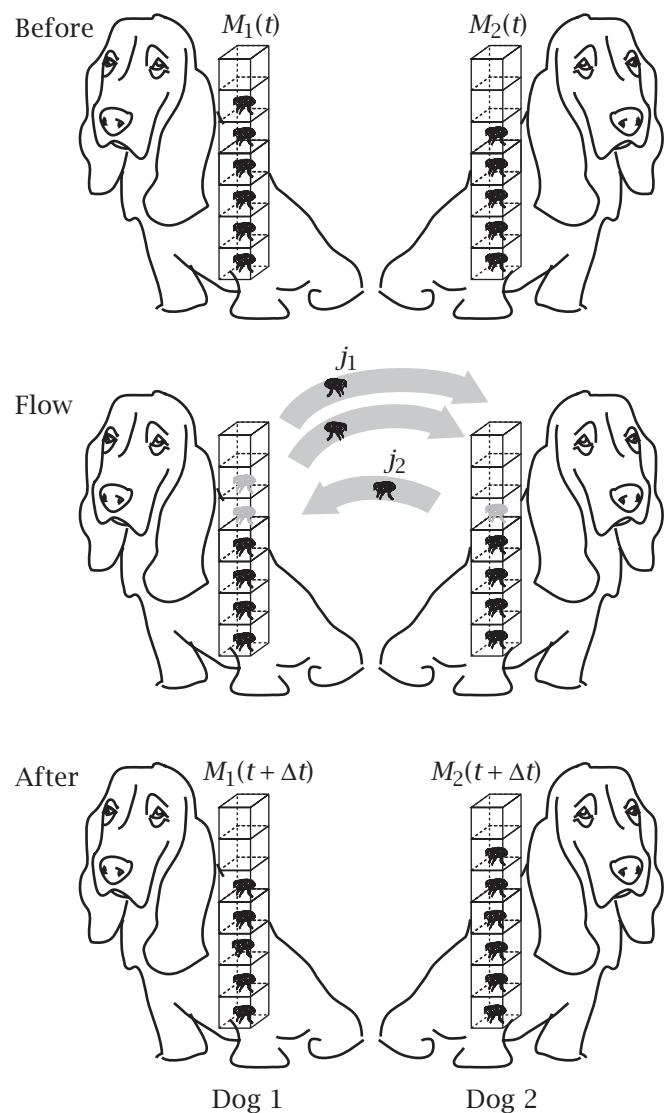


Figure 18.1 Dog-flea model of diffusion. Why do particles (fleas) flow from regions of high to low concentration? There are combinatorially more ways for particles to flow that way.

the other dog between time t and $t + \Delta t$ while $M - j$ of the fleas do not jump. It is a simple matter of binomial combinatorics to count all the different ways that j of the M different fleas can jump:

$$P(j, M) = \frac{M!}{j!(M-j)!} \ell^j s^{M-j}, \quad (18.19)$$

where ℓ is the probability that a flea leaps and s is the probability that a flea stays on its dog during the time interval Δt . Notice that Equation (18.19) has exactly the same binomial form as Equation (18.1). The difference is that the random-flight equation (18.19) describes the N time steps of one particle while Equation (18.1) describes one time step of M different particles. However, the math is the same. For example, for a dog with M fleas, construct the dynamical partition function (see Equation (18.12)),

$$q_d = (\ell + s)^M, \quad (18.20)$$

using the addition and multiplication rules of probability: leap (ℓ) or stay (s) are mutually exclusive options available to each flea, so you add those statistical weights together inside the parentheses. And there are M independent fleas, so you multiply those factors together. Upon expansion of Equation (18.20), the term ℓ^M describes the one way that all M fleas can jump off the dog in the time interval Δt , $M\ell^{M-1}s$ describes the M different ways that one flea (among the M possible fleas) could stay on the dog while the other $M - 1$ fleas jump, and so on. To get the average number of fleas that jump during the time interval, follow the logic of Equation (18.14), and take the derivative:

$$\langle j \rangle = \frac{d \ln q_d}{d \ln \ell} = \frac{\ell M (\ell + s)^{M-1}}{(\ell + s)^M} = M\ell, \quad (18.21)$$

where we substituted $\ell + s = 1$ at the end. Equation (18.21) says that the average number of fleas that jump is proportional to the total number of fleas on a single dog.

Now, to find the basis for Fick's law, we are interested in the difference in the flow of fleas between two dogs. Assume the rate quantities ℓ and s are the same for both dogs. The average net number of fleas jumping from dog 1 to dog 2 is

$$\langle j_1 \rangle - \langle j_2 \rangle = \ell(M_1 - M_2) = \ell \Delta M. \quad (18.22)$$

Convert this number to a flux (i.e., the number of particles that jump per unit time per unit area) and replace numbers with concentrations using $\Delta M = -A \Delta c \Delta x$, where $A \Delta x$ is the volume containing the particles. The minus sign is used because we are considering the forward direction to be from 1 to 2: the positive flux direction is $\langle j_1 \rangle - \langle j_2 \rangle$, while the gradient $\Delta c = c_2 - c_1$ is defined in the opposite direction. This gives

$$\langle J \rangle = \frac{\langle j_1 \rangle - \langle j_2 \rangle}{A \Delta t} = \frac{\ell \Delta M}{A \Delta t} = -\frac{\ell \Delta c A \Delta x}{A \Delta t} \left(\frac{\Delta x}{\Delta x} \right) = -\frac{\ell \Delta x^2}{\Delta t} \left(\frac{\Delta c}{\Delta x} \right). \quad (18.23)$$

Equation (18.23) is Fick's law as it pertains to just two adjacent planes of particles, rather than to many planar slices of space. It shows that the average flow of particles is proportional to the difference in concentrations.

Fick's law is explained by a simple counting argument. The reason that particles flow down their concentration gradients is not because of some property of the individual particles, but simply because there are more *routes of flow*, i.e., more *ways that fleas can hop* for larger numbers than for smaller numbers of particles. This logic resembles that of coin flips. The reason that a long series of coin flips is usually half heads and half tails is just that there are more sequences that have that composition than any other. No one coin 'knows' what the other coins are doing. Like coin flips, diffusion becomes highly predictable when you have large numbers of particles.

What Is the Basis for Fourier's Law of Heat Flow?

What about heat? Why does heat flow down temperature gradients? We borrow from the results above. Imagine now that each metaphoric flea jumping from dog 1 to dog 2 carries with it an average energy $\varepsilon_1 = \langle mv^2 \rangle / 2$ (where m is the particle's mass and v its velocity), which we express as $kT_1 / 2$. And imagine that every flea that jumps from dog 2 to dog 1 carries with it an energy $\varepsilon_2 = kT_2 / 2$. This represents a temperature gradient; T_1 and T_2 are different temperatures at different locations. For our purposes here, we are envisioning energy transport in discrete packets of averaged energies. Assume that the only driving force is the temperature gradient; the number of particles is fixed ($M_1 = M_2 = M$). Then, the fleas represent packets of energy flowing (heat) in a net direction. Similar to Equation (18.22), the heat flux will be

$$\langle J_q \rangle = \frac{1}{A \Delta t} \left(\langle j_1 \rangle \frac{kT_1}{2} - \langle j_2 \rangle \frac{kT_2}{2} \right) = -\frac{\ell M k \Delta x^2}{2V \Delta t} \left(\frac{\Delta T}{\Delta x} \right), \quad (18.24)$$

where $V = A \Delta x$ is the volume of each slab of particles and the *thermal conductivity* is $\kappa = (\ell M k \Delta x^2) / (2V \Delta t)$. This is Fourier's law of heat flow, applied or two planes. It says that heat flow is proportional to the temperature gradient. (This is only one possible heat-transfer mechanism. Sometimes heat flows even when particles don't flow.)

Fluctuations of Diffusion Rates

Fick's and Fourier's laws describe *average* fluxes in large systems, such as in beakers. But for small systems, such as biological cells, you may want to know the variations, or fluctuations, around that average. Look at Figure 18.2. It shows small numbers of particles diffusing near a plane. If you divide the figure into small planar slices taken perpendicular to the x axis, you will see that more particles jump left or right across some slices per unit time than predicted by Fick's law, while across other slices, fewer particles jump. We are interested in the distribution of these jump numbers. To keep the math simple, let's look at the situation where $M_1(t) = 4$ and $M_2(t) = 2$ and $\ell = 1/2$. For that case, shown with fleas in Figure 18.3, Table 18.1 shows the multiplicities of routes of flea flow between the two dogs.

First, notice that the maximum multiplicity of trajectories (12 trajectories out of 64) occurs when $j_1 = 2$ fleas jump from the 4-flea dog and $j_2 = 1$ flea jumps from the 2-flea dog. This is the Fick's law result: particles flow down their concentration gradients in proportion to their concentrations. All the other

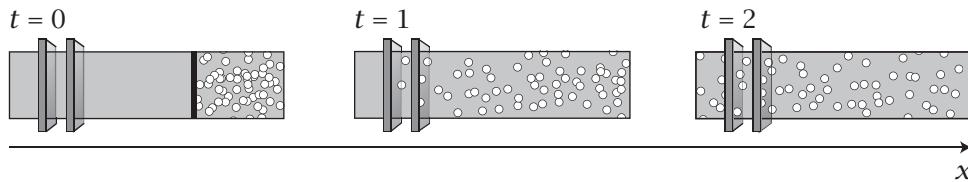


Figure 18.2 A small number of colloidal particles diffusing in a plane.

entries in this table represent the fluctuations around this average value. For example, look at the top right corner of this table. This entry says that 1/64 of all trajectories involve a ‘backwards’ or ‘upstream’ flux in which both fleas jump from dog 2 to dog 1, and no fleas jump from dog 1 to dog 2. These are instances in which particles flow *up* their concentration gradients, like salmon swimming upstream, rather than down their gradients. Now, look at the bottom left corner of the table. This describes a ‘superflux’: all 4 fleas jump from dog 1 and no fleas jump from the dog 2. This flux ‘overdoes it,’ and, in one step, overshoots equilibrium and creates a gradient in the opposite direction.

Fick’s law describes only the average flow, in the same way that predicting half heads and half tails describes only the average outcome of coin flips. If you flip a coin only a few times, sometimes you get more than half heads or more than half tails; you will get a whole distribution of probabilities. In the same way, dynamical processes such as diffusion involve whole distributions of rates. Let’s look beyond just the average values and explore the rest of the dynamical distribution function for diffusion.

The Distribution of Diffusion Rates is Gaussian

The diffusion-rate distribution for the system shown in Figure 18.2 is approximately Gaussian. To see this, start again from the binomial distribution and follow the steps to the random-flight equation (18.9), to get

$$P(j, M) = \frac{1}{\sqrt{2\pi M \ell s}} \exp\left[-\frac{(j-M\ell)^2}{2M\ell s}\right], \quad (18.25)$$

where M is the total number of fleas (particles), j is the number of them that jump during time interval Δt , and ℓ and s are the probabilities that a particle leaps or stays during that time interval. The average number of fleas that jump (or particles that diffuse from one plane or the next) per time interval is $\langle j \rangle = M\ell$. The variance is $\sigma^2 = M\ell s$. But Equation (18.25) is not quite the distribution function that we want. We want the distribution of *net jumps*, $P(j_1 - j_2)$, not the distribution $P(j_1)$ of right jumps j_1 or $P(j_2)$ of left jumps j_2 alone.

Table 18.1 The distribution of flea jumps. j_1 and j_2 are the different numbers of fleas that jump from dog 1 or 2, respectively, during a time interval. Each entry shows how many different ways these numbers of fleas can jump from each dog. It is most probable that 2 fleas jump from dog 1 and 1 flea jumps from dog 2; there are 12 ways the various fleas can do this.

j_1	j_2		
	0	1	2
0	1	2	1
1	4	8	4
2	6	12	6
3	4	8	4
4	1	2	1

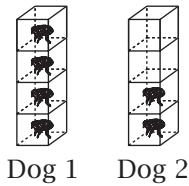


Figure 18.3 Dog-flea model with $M_1(t) = 4$ and $M_2(t) = 2$.

However, owing to a remarkable property, it is simple to compute the quantity we want. If you have two Gaussian functions, one with mean $\langle x_1 \rangle$ and variance σ_1^2 , and the other with mean $\langle x_2 \rangle$ and variance σ_2^2 , then the distribution of the difference, $P(x_1 - x_2)$, will also be a Gaussian function, having mean $\langle x_1 \rangle - \langle x_2 \rangle$ and variance $\sigma^2 = \sigma_1^2 + \sigma_2^2$.

For our binomial distributions, the mean values are $\langle j_1 \rangle = \ell M_1$ and $\langle j_2 \rangle = \ell M_2$ (since we assumed ℓ and s are the same for fleas jumping from either dog). The variances are $\sigma_1^2 = M_1 \ell s$ and $\sigma_2^2 = M_2 \ell s$. So the distribution of the net numbers of particles jumping in a time interval, $J = j_1 - j_2$, is

$$P(J) = \frac{1}{\sqrt{2\pi M \ell s}} \exp\left\{-\frac{[J - \ell(M_1 - M_2)]^2}{2M\ell s}\right\}, \quad (18.26)$$

where $M = M_1 + M_2$ is the total number of particles.

Now, with this distribution, you can readily show that if the number M of particles is large, the fluctuations will be small, and you will almost always observe the average behavior. To see this, take the ratio of the standard deviation σ to the mean:

$$\frac{\sqrt{\sigma^2}}{\langle J \rangle} = \frac{\sqrt{M \ell s}}{(M_1 - M_2) \ell} \sim M_1^{-1/2} \quad (18.27)$$

in the limit, $M_1 \gg M_2$. The fluctuations become negligible as $M_1 \rightarrow \infty$. On the other hand, in situations involving small numbers of particles, such as inside biological cells, the fluctuations may sometimes be important.

Figure 18.4 shows experimental confirmation of the Gaussian distribution predicted in Equation (18.26).

The Direction of Time's Arrow: A Fluctuation Theorem

An historic matter of fundamental principle is whether the Second Law of Thermodynamics is ever violated. Ever since 1865, when R Clausius first defined the word 'entropy,' there has been interest in whether systems might ever tend

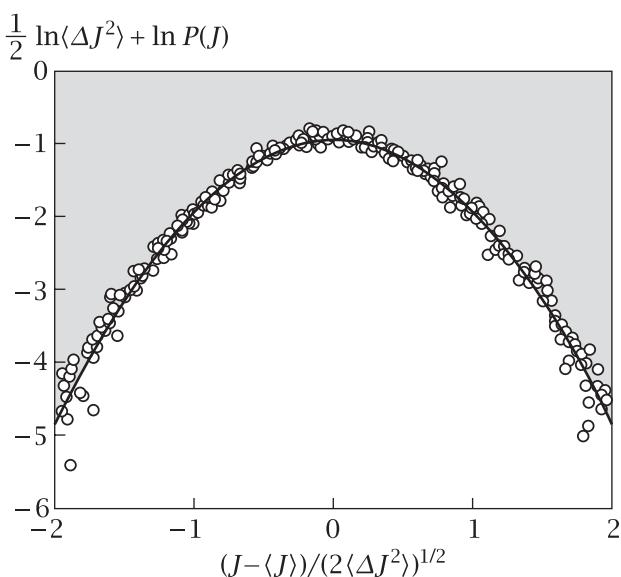


Figure 18.4 The distribution of diffusional fluxes is found to be Gaussian.
Source: S Seitaridou, MM Inamdar, R Phillips, et al., *J Phys Chem B* **111**, 2288–2292 (2007) and K Ghosh, KA Dill, MM Inamdar, et al., *Am J Phys* **74**, 123–133 (2006).

toward lower entropy, rather than toward higher entropy, as the Second Law predicts. In 1867, JC Maxwell imagined a thought experiment. Maxwell envisioned a fictitious microscopic being—now called *Maxwell's demon*—who is situated at a miniature doorway that separates a hot container of gas from a cold container of gas. The demon watches and opens a gate only when it can allow a cold molecule to pass from the hot side to the colder chamber or a hot molecule to pass to the hotter chamber. The demon's actions would cause the cold side to get colder and the hot side hotter, in apparent violation of the Second Law. Is it possible to construct a real Maxwell's demon device that could violate the Second Law? Subsequent arguments have shown that demon-like violations of the Second Law are impossible because the physical act of separating hot from cold particles requires some way to distinguish hot from cold particles, and that, in turn, would require energy or information.

So, as far as is known, the Second Law, which describes only *the average tendencies* of systems toward states of maximum entropy, is never violated. There is no violation of any law when systems that are small enough or close enough to equilibrium to fluctuate, causing particles to flow ‘upstream’ against a concentration gradient or heat to flow upstream against a temperature gradient.

A popular way to make this more quantitative is in the form of a *fluctuation theorem*. Let's focus on fluctuations of the flux J . Using Equation (18.26), take the ratio of the probability $P(J)$ that the system will have a flux J in the positive direction and divide it by the probability $P(-J)$ that the system will have a flux of the same magnitude, but in the backwards direction:

$$\frac{P(J)}{P(-J)} = \frac{\exp[-(J-a)^2/b]}{\exp[-(-J-a)^2/b]} = \exp\left(\frac{4Ja}{b}\right) = \exp\left(\frac{2J\Delta M}{q_0 M}\right), \quad (18.28)$$

where $a = \ell \Delta M$, $b = 2\ell sM$, $\Delta M = M_1 - M_2$, and $M = M_1 + M_2$. If your system is far from equilibrium, $J \gg 0$, then Equation (18.28) predicts that backward fluxes become exponentially unlikely. For systems that have small particle numbers or are near equilibrium, backward fluctuations should occur occasionally. Figure 18.5 shows confirmation of the flux fluctuation equation (18.28) from experiments on small numbers of colloidal particles diffusing in a microfluidic cell under a microscope.

Let's now look at a different type of dynamical process: the inter-arrival times between stochastic events.

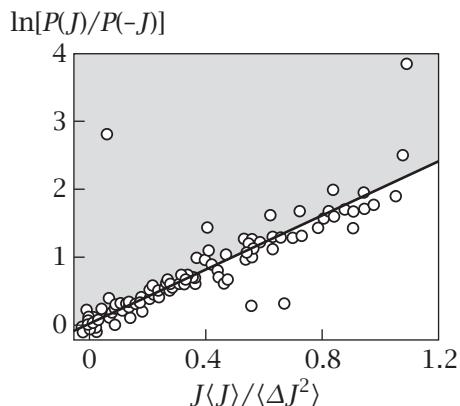


Figure 18.5 Experimental validation of the flux fluctuation theorem. It is found that $P(J)/P(-J)$ increases exponentially with J , as predicted by Equation (18.28). Source: S Seitaridou, MM Inamdar, R Phillips, et al., *J Phys Chem B* **111**, 2288–2292 (2007).

Inter-Arrival Rates Are Described by Poisson and Waiting-Time Distributions

Imagine standing on the side of a road watching cars pass by. When each car passes, you start a timer. You stop the clock when the next car passes you. You plot a histogram of the number of instances in which the *inter-arrival time* was 10 s, 20 s, 30 s, and so on. Waiting-time distributions describe the times between earthquakes, wars, deaths (insurance companies care about this), your telephone calls, the failure rates of materials and commercial products, radioactive decays of radium atoms, the stepping of molecular motors along tracks, binding and unbinding events, chemical reactions, and electrons emitted from hot metal surfaces. For simple systems, waiting times follow an exponential distribution: you are likely to observe many intervals that are relatively short and few intervals that are extremely long.

Closely related is the *Poisson distribution*. While the waiting time distribution gives the *time per event*, the Poisson distribution gives the *number of events per unit time* (for independent events). For waiting times, you run your stopwatch from one event until the next, and find the distribution of those times. For the Poisson distribution, you run your stopwatch for a fixed time t and find the distribution of how many events you see. Below, we derive these two distributions.

The Number of Stochastic Events Per Unit Time Is Often Described by a Poisson Distribution.

Imagine watching for each event (such as the passing of a car or the folding of a single protein molecule) during an interval from time 0 to time t . Divide the total time t into N discrete time steps that are small enough that you see only zero or one event during any one step. If each step takes a time Δt , then the total time is $t = N \Delta t$. Our problem resembles coin-flipping. What is the probability of seeing exactly k heads in N coin flips? The probability that you will see exactly k events during the N time steps is

$$P(k, N) = \frac{N!}{k!(N-k)!} p^k q^{N-k}, \quad (18.29)$$

where p is the probability of seeing an event during one time step and $q = 1 - p$ is the probability that no event happens during a time step. The average number of events that will occur over the time t will be $\langle k \rangle = Np$; see Equation (18.15).

Let's focus on situations involving relatively rare events, small p and large N ; in most of the time intervals, events do not occur. This is just another way of saying that the average time between events is much longer than the time interval Δt . Let's define a rate $\lambda = \langle k \rangle / t$, the average number of events that occur per unit time. Substitute $p = \lambda t / N$ into Equation (18.29) and factorize the term q^{N-k} to get

$$P(k, N) = \frac{N!}{k!(N-k)!} \left(\frac{\lambda t}{N} \right)^k \left(1 - \frac{\lambda t}{N} \right)^N \left(1 - \frac{\lambda t}{N} \right)^{-k}. \quad (18.30)$$

You can simplify Equation (18.30). First, note that in the limit of small p and large N , the last term on the right approximately equals 1. Second, because

$k \ll N$, note that the collection of factors $N!/[(N-k)!N^k]$ also approximately equals 1. (Try it. If $N = 1000$ and $k = 3$, you have $(1000)(999)(998)/(1000^3) \approx 1$.) Third, for large N , use the approximation

$$\left(1 - \frac{\lambda t}{N}\right)^N \approx e^{-\lambda t}. \quad (18.31)$$

Making these three substitutions into Equation (18.30) gives the **Poisson distribution**,

$$P(k, t) = \frac{(\lambda t)^k e^{-\lambda t}}{k!}. \quad (18.32)$$

Equation (18.32) gives the probability that you will observe k events during a time interval from 0 to t , if the events are independent.

Now, let's ask a slightly different question about the same process. What is the distribution $P_w(t) dt$ of *waiting times* or *dwell times*, i.e., the probability that one event happens between time t and $t+dt$?

The Waiting Time Between Independent Random Events Is Described by an Exponential Distribution.

To derive the waiting-time distribution, first use Equation (18.32) to find the probability $P(0, t)$ that you would observe $k = 0$ events during a time t :

$$P(0, t) = e^{-\lambda t}. \quad (18.33)$$

$P(0, t)$ is the probability that you observe no events before time t , i.e., that you will not observe your first event until sometime after time t . Your waiting time for the first event will be longer than t . Said differently, the probability $P(0, t)$ that you observe no events before time t is exactly the same as the probability that you must wait some amount of time from t to ∞ to find your first event; the latter quantity is the following integral over the waiting time distribution:

$$\int_t^\infty P_w(t') dt' = P(0, t) = e^{-\lambda t}. \quad (18.34)$$

Now, to get the waiting-time distribution $P_w(t)$, take the derivative of Equation (18.34):

$$P_w(t) = \lambda e^{-\lambda t}. \quad (18.35)$$

Equation (18.35) shows that the distribution of waiting times is exponential. Once you start your stopwatch, it is more likely that you will wait a shorter time than a longer time to see the first event. You can use Equation (18.35) to show that the average waiting time is $\langle t \rangle = 1/\lambda$ and the variance is $\langle t^2 \rangle - \langle t \rangle^2 = (1/\lambda)^2$.

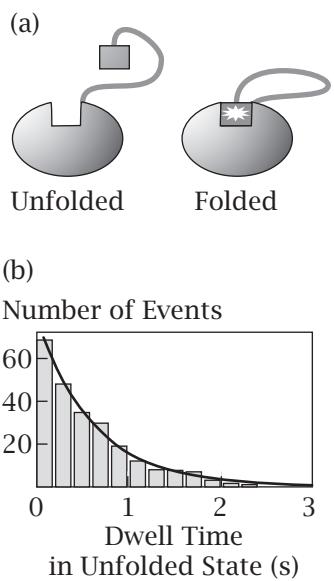
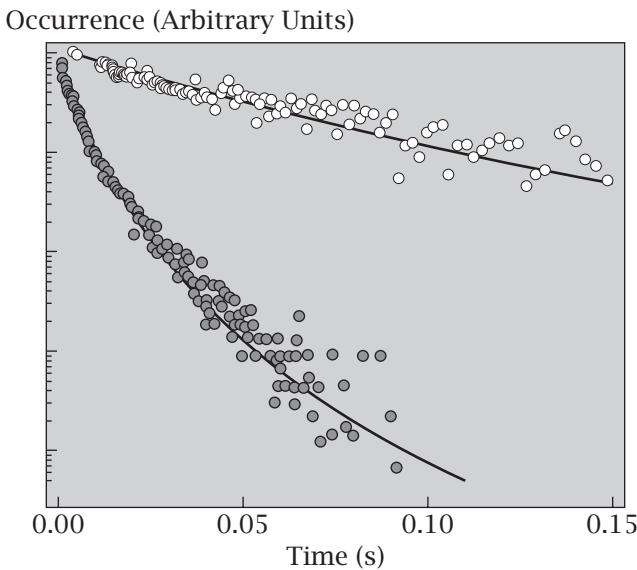


Figure 18.6
Single-molecule RNA folding and unfolding (a) has an exponential waiting-time distribution (b), as expected from Equation (18.35).
Source: X Zhuang, LE Bartley, HP Babcock, et al., *Science* **288**, 2048–2051 (2000).

EXAMPLE 18.1 Waiting-time distributions for RNA folding. RNA molecules undergo folding and unfolding transitions. Single-molecule experiments can ‘see’ the folded and unfolded states and can follow the trajectories of the transitions back and forth. From a trajectory, you measure the waiting-time distribution by marking each transition to the unfolded state and measuring the time until the trajectory next switches back to the folded state, for example. Figure 18.6 shows an RNA folding waiting-time distribution. It is exponential, as predicted by Equation (18.35), indicating independent random events.

EXAMPLE 18.2 Single-molecule waiting-time experiments reveal multiple timescales for enzyme catalysis. Min et al. [2] performed single-molecule experiments in which they watched an enzyme called β -galactosidase catalyze its biochemical reaction. A Michaelis-Menten enzyme reaction would be described by the kinetic process, $E + S \leftrightarrow ES \rightarrow E + P$. Min et al. were able to observe two states, the enzyme unbound as E and the enzyme bound as ES . Figure 18.7 shows the waiting-time distribution they measured for this process. Two curves are shown. When the substrate S is at low concentration, they measured a single-exponential waiting-time distribution, as expected from Equation (18.35). However, when the substrate S is at high concentrations, the waiting time is a sum of two exponentials. At low concentrations, the single slow event corresponds to the enzyme catalytic event and its natural slow recovery. At high concentrations, the additional fast event arises because a substrate molecule hops onto a free enzyme E before the enzyme has had enough time to relax from its previous catalytic event. This example shows the types of insights single-molecule waiting-time distributions can give into chemical mechanisms.

Figure 18.7 Waiting-time distributions for the binding and unbinding of an enzyme E to its substrate S . At low substrate concentrations (\circ), waiting times follow a single exponential distribution. At high substrate concentrations (\bullet), the waiting-time distribution involves two exponentials, indicating two mechanistic steps.
Source: BP English, W Min, AM van Oijen, et al. *Nat Chem Biol* **2**, 87–94 (2006).



Now let's consider another type of dynamical process. In Chapter 17, we explored the kinetics of two-state processes:



Equations (17.49) and (17.51) showed that the concentrations of A and B change exponentially with time in *bulk* systems (i.e., systems having large numbers of particles). Now, we look at the two-state dynamics of only a single particle, for example the folding and unfolding of a single RNA molecule, the binding or unbinding of one ligand molecule to a protein, the opening or closing of one ion channel molecule, the blinking on or off of a quantum-dot particle, or the catalysis of a biochemical reaction by an individual enzyme molecule.

Single-Particle Kinetics is Described by Trajectories

A single-particle state does not have a ‘concentration.’ To characterize the dynamics of a single particle, look at its possible *trajectories*. A trajectory is one sequence of events that the particle can take as it hops back and forth between its two states, A and B , as a function of time. Let's divide time into discrete intervals of Δt each. Consider a particle that starts at time $t = 0$ in state A and proceeds for only $t = 4$ units of time. Figure 18.8 enumerates all of its eight possible trajectories over time.

First, break each trajectory into two-step pieces; see Figure 18.9. This is called a *Markov model*. In a Markov model, you assume that the probability of jumping to a particular state at a time t depends only on that particular state and what state it jumps from, and does not depend on probabilities of earlier states in the trajectory. The Markov assumption is often a very good approximation. We represent the four possible transition rates as *dynamical statistical weight* constants: d is the probability that a system leaps *down* from

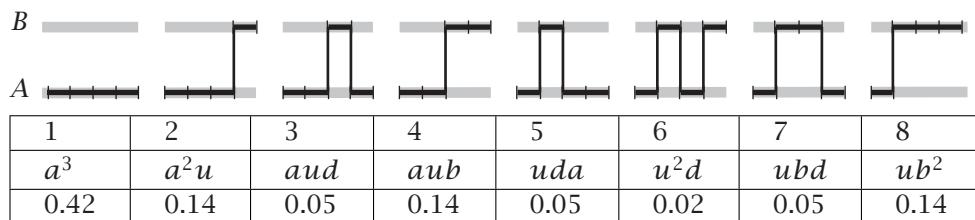


Figure 18.8 All the possible trajectories for the two-state process $A \rightarrow B$ for 4 time steps, starting in state A at time $t = 1$. The middle row gives the statistical weight for each trajectory. The bottom row gives the particular values used in Example 18.4.

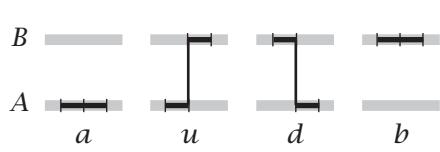


Figure 18.9 The four dynamical statistical weight quantities for modeling two-state trajectories: a , the weight of a transition that starts in A and stays in A ; u , for starting in A and jumping up to B ; d , for starting in B and jumping down to A ; b , for starting in B and staying in B .

B to A during a time step; u is the probability of leaping *up* from A to B ; a is the probability the system *stays in A* if it was previously in A ; and b is the probability the system *stays in B* during one time step.

To illustrate the method, let's miniaturize the problem. Which dynamical trajectories are you most likely to observe? If you know the constants a , u , d , and b , you can compute the probability p_j that the system will follow any particular trajectory j . Summing the statistical weights (unnormalized probabilities) over all the trajectories will give you the *dynamical partition function* q_d [3, 4]:

$$q_d = a^3 + a^2 u + 2a^2 d + a u b + a u^2 d + u b d + u b^2. \quad (18.37)$$

Each term in Equation (18.37) describes the weight of one particular trajectory. For example, a^3 is the statistical weight for the trajectory in which the particle starts in state A and stays in A at each of the three time transition points. The probability of a trajectory is its statistical weight divided by q_d . So, the probability of trajectory 1 in Figure 18.8 is $p_1 = a^3/q_d$. The probability of trajectory 3 is $p_3 = a u d/q_d$: the particle stays in state A at the first transition point, jumps up to B at the second transition point, and jumps back down to A at the third transition point.

Of the four statistical weight quantities, only two are independent of each other, because you have

$$a + u = 1 \quad (18.38)$$

$$b + d = 1. \quad (18.39)$$

The first constraint, $a + u = 1$, says that a particle in state A has only two options in the next time step: it can either stay in state A (weight a) or it can jump to B (weight u). The second constraint, $b + d = 1$, applies the same rule to a particle in state B .

In one experiment, you will observe one particular trajectory. Does it resemble an average trajectory? To know the average dynamical behavior of a single particle requires observing many different trajectories of the same system. You can use the dynamical partition function q_d to compute average and variance properties of the dynamical trajectories.

EXAMPLE 18.3 Computing average properties of trajectories. You can take derivatives of q_d to get various average quantities. For example, to compute $\langle N_{AA} \rangle$, the average number of $A \rightarrow A$ transitions, take the derivative with respect to the transition statistical weight a :

$$\langle N_{AA} \rangle = \sum_{j=1}^8 p_j N_{AAj} = \frac{d \ln q_d}{d \ln a} = (3a^3 + 2a^2 u + 2a u d + a u b) q_d^{-1}, \quad (18.40)$$

where N_{AAj} is the number of $A \rightarrow A$ transitions that occur in trajectory j . The factor $3a^3/q_d$ appears because a^3/q_d is the probability of trajectory number 1 and the front factor of 3 appears because there are three transitions of type AA that occur in that trajectory. The factor of $2a^2 u$ appears because the probability for trajectory number 2 is $a^2 u/q_d$ and that trajectory has two $A \rightarrow A$ transitions, and so on. Or, to get the average number $\langle N_{AB} \rangle$ of $A \rightarrow B$

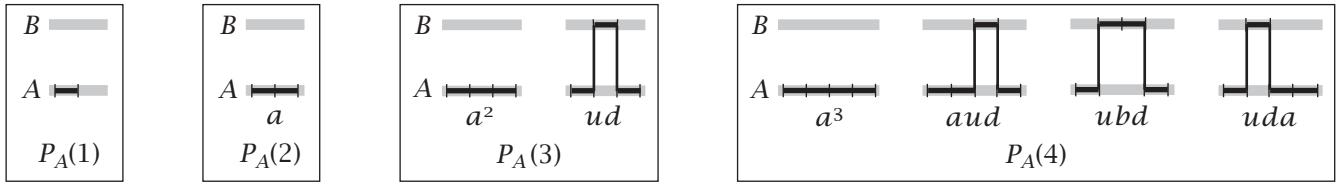


Figure 18.10 To compute $P_A(t)$, for time $t = 1, 2, 3, 4$, collect together all the trajectories for which the system is in state A at time t .

transitions, take the derivative with respect to the statistical weight u :

$$\langle N_{AB} \rangle = \frac{d \ln q_d}{d \ln u} = (a^2 u + 2aud + aub + 2u^2 d + ubd + ub^2) q_d^{-1} \quad (18.41)$$

If you want to compute a variance, take the second derivative. For example,

$$\sigma_{AA}^2 = \langle N_{AA}^2 \rangle - \langle N_{AA} \rangle^2 = \frac{d^2 \ln q_d}{d \ln a^2}; \quad (18.42)$$

see Equations (C.7) and (C.8). Or, if you want to know the average number of time steps the system is in state B , add two quantities together: ‘leaps from A to B ’ plus ‘stays in B ,’ $\langle N_B \rangle = \langle N_{AB} \rangle + \langle N_{BB} \rangle$. By taking derivatives of q_d , you can compute many different properties of the dynamical rate distribution.

So, here’s how to develop a dynamical model from your experimental data. Before your experiments, you won’t know the values of a , u , d , and b . Now, observe several trajectories of your system. Measure two average quantities over those trajectories, say $\langle N_{AA} \rangle$ and $\langle N_{AB} \rangle$. Then, solve Equations (18.38)-(18.41) to get the four statistical-weight constants a , u , b , and d . Finally, substitute those four statistical weights into equations such as Equation (18.42) to compute variances, trajectory probabilities, and other dynamical properties of interest. By comparing those additional properties with your experiments, you can learn whether this model is sufficient or whether you need a better one. Because you get more information from a full dynamical distribution function than you get from just an average rate, single-particle experiments can provide more insights for model making than bulk experiments can.

Now, let’s compute a different probability. Instead of the probability p_j that the system takes trajectory j , let’s compute the probability $P_A(t)$ that the particle, averaged over all its trajectories, is specifically in state A at time t . We call this a *state probability*. Figure 18.10 shows the trajectories that contribute to $P_A(t)$.

Master Equations Are Differential Equations for Markov Dynamics

Here is a simple way to compute the state probabilities $P_A(t)$ and $P_B(t)$. First, express the statistical weights as a matrix:

$$\mathbf{G} = \begin{bmatrix} a & d \\ u & b \end{bmatrix}. \quad (18.43)$$

Next, if you knew $P_A(t-1)$ and $P_B(t-1)$, the probabilities of being in states A or B at time $t-1$, then you could get the populations $P_A(t)$ and $P_B(t)$ at time t using

$$\begin{bmatrix} P_A(t) \\ P_B(t) \end{bmatrix} = \mathbf{G} \begin{bmatrix} P_A(t-1) \\ P_B(t-1) \end{bmatrix} = \begin{bmatrix} aP_A(t-1) + dP_B(t-1) \\ uP_A(t-1) + bP_B(t-1) \end{bmatrix}. \quad (18.44)$$

This just says that $P_A(t)$ is made up of two types of trajectories. In one type, the particle was already in A and then had an a -type transition to stay in A . In the other type, the particle was already in B and had a d -type transition from B to A . You can compute $P_B(t)$ the same way.

Now, we want to convert Equation (18.44) into a differential equation. To do this, subtract $P_A(t-1)$ from $P_A(t)$, to get

$$\begin{bmatrix} P_A(t) - P_A(t-1) \\ P_B(t) - P_B(t-1) \end{bmatrix} = \begin{bmatrix} a-1 & d \\ u & b-1 \end{bmatrix} \begin{bmatrix} P_A(t-1) \\ P_B(t-1) \end{bmatrix}. \quad (18.45)$$

Taking small time intervals $\Delta t \rightarrow dt$ and substituting the constraint equations (18.38) and (18.39) gives

$$\frac{d}{dt} \begin{bmatrix} P_A \\ P_B \end{bmatrix} = \begin{bmatrix} dP_A/dt \\ dP_B/dt \end{bmatrix} = \begin{bmatrix} -k_f & k_r \\ k_f & -k_r \end{bmatrix} \begin{bmatrix} P_A \\ P_B \end{bmatrix}, \quad (18.46)$$

that is,

$$\frac{dP_A}{dt} = -k_f P_A + k_r P_B,$$

$$\frac{dP_B}{dt} = k_f P_A - k_r P_B,$$

Refresher on Matrix and Vector Multiplication

(1) When a vector $[a, b]$ multiplies a matrix $\begin{bmatrix} x & y \\ z & w \end{bmatrix}$, the result is

$$[a, b] \begin{bmatrix} x & y \\ z & w \end{bmatrix} = [ax + bz, ay + bw].$$

(2) When a matrix $\begin{bmatrix} x & y \\ z & w \end{bmatrix}$ multiplies a vector $[a, b]$, the result is

$$\begin{bmatrix} x & y \\ z & w \end{bmatrix} \begin{bmatrix} a \\ b \end{bmatrix} = \begin{bmatrix} xa + yb \\ za + wb \end{bmatrix}.$$

(3) When two matrices are multiplied, the result is

$$\begin{bmatrix} a & b \\ c & d \end{bmatrix} \begin{bmatrix} x & y \\ z & w \end{bmatrix} = \begin{bmatrix} ax + bz & ay + bw \\ cx + dz & cy + dw \end{bmatrix}.$$

Note that the order of matrix multiplication matters. In general, for two matrices \mathbf{A} and \mathbf{B} , $\mathbf{AB} \neq \mathbf{BA}$.

where $k_f = u/\Delta t$ and $k_r = d/\Delta t$ are rate coefficients. Equation (18.46) is called the *master equation* for this one-particle system. It resembles the differential equations for bulk concentrations of the two-state system, except that particle concentrations are replaced by $P_A(t)$ and $P_B(t)$.

Now, notice that Equation (18.44) gives a way to propagate $P_A(t-1)$ forward in time. If you multiply by the matrix t times, you have

$$\begin{bmatrix} P_A(t) \\ P_B(t) \end{bmatrix} = \mathbf{G}^t \begin{bmatrix} P_A(0) \\ P_B(0) \end{bmatrix}. \quad (18.47)$$

The partition function is the sum

$$q_d = P_A(t) + P_B(t) = [1, 1] \begin{bmatrix} P_A(t) \\ P_B(t) \end{bmatrix} = [1, 1] \mathbf{G}^t \begin{bmatrix} 1 \\ 0 \end{bmatrix} \quad (18.48)$$

for a particle that starts in state A. The multiplication by the column vector $[1, 0]$ on the right selects out only the trajectories that started in state A. And, the row vector $[1, 1]$ on the left sums over those trajectories to give the dynamical partition function. You can check that when $t = 4$, Equation (18.48) reduces to Equation (18.37). Master equations relate the microscopic dynamical quantities, a , u , b , and d , which characterize single-molecule trajectories, to the state populations $P_A(t)$ and $P_B(t)$. And, master equations give a microscopic basis for explaining dynamics you see in bulk experiments.

EXAMPLE 18.4 Computing two-state exponential relaxation from trajectories. Now we show that the single-molecule quantity $P_A(t)$, the probability that the system is in state A at time t , decays exponentially in time, just as bulk concentrations in a two-state system would do. Suppose the particle starts in state A. It becomes increasingly likely that the particle sooner or later will jump to state B (and then back and forth after that). Let's see how the trajectory approach shows this. To keep it simple, consider only four time points, $t = 1, 2, 3, 4$. Suppose you already know that the statistical weights are $a = b = 3/4$ and $u = d = 1/4$. To compute $P_A(t)$, find all the trajectories that happen to be in state A at the particular time t ; Figure 18.10 shows all the trajectories that contribute to $P_A(1), P_A(2), P_A(3)$, and $P_A(4)$. For example, $P_A(4)$ is the sum of weights of trajectories 1, 2, 5, and 6 in Figure 18.8, because those are all possible ways the system can arrive in state A at time $t = 4$. The statistical weight of trajectory 1 is 0.42, that of trajectory 2 is 0.14, that of trajectory 5 is 0.05, and that of trajectory 6 is 0.02, so $P_A(3) = 0.42 + 0.14 + 0.05 + 0.02 = 0.63$. Now, do this for each time point to get the full time dependence: $P_A(1) = 1$, $P_A(2) = 0.75$, $P_A(3) = 0.63$, and $P_A(4) = 0.57$. Figure 18.11 shows this exponential decay with time.

This example shows how trajectory-based modeling works. It predicts the time dependence of $P_A(t)$, even though the statistical weights a, u, b , and d are constants, independent of time. Each trajectory does not itself change with time; each trajectory describes a full possible time course. For example, one trajectory may be mostly in A at early times and mostly in B at later times. Dynamical observable quantities are weighted averages over different trajectories. Below we describe some other dynamical properties of interest.

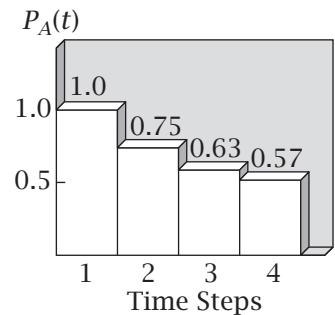


Figure 18.11 The probability $P_A(t)$ that a particle is in state A decays exponentially with time t if it starts in state A. Example 18.4 shows how the quantity $P_A(t)$ is obtained as a sum over the microscopic trajectory probabilities.

EXAMPLE 18.5 Waiting times from q_d for $A \rightarrow B$ trajectories. Suppose a trajectory begins in state A at time $t = 0$. What is the probability $P_w(k)$ that the system waits, or *dwells*, in state A for exactly $k - 1$ time steps and only first enters state B at time step k ? From the toy-model trajectories given above, $P_w(k)$ is given by

$$P_w(k) = a^{k-1} u / q_d, \quad (18.49)$$

because the only relevant trajectories are those that have $k - 1$ $A \rightarrow A$ transitions (weight = a) and finally one $A \rightarrow B$ transition (weight = u). Equation (18.49) shows that $P_w(k)$ is an exponentially decreasing function of k , since $a < 1$ (see Equation (18.35)).

EXAMPLE 18.6 M particles at a time. Above, we treated the dynamics of a single particle. How can you generalize this to M independent particles undergoing two-state $A \rightarrow B$ trajectories at the same time? The dynamical partition function for M independent particles is

$$Q_M = [P_A(t) + P_B(t)]^M. \quad (18.50)$$

You can take derivatives of Q_M to compute averages and variances of the M -particle system in the same way that you used q_d for the single-particle system.

Now we explore a model of how molecular machines such as proteins can combine Brownian motion with binding and release processes to create directed motion.

Brownian Ratchets Convert Binding and Release Events into Directed Motion

Some protein molecules are molecular motors. Examples include kinesin, which walks along microtubules; myosin, which walks along actin in muscle; helicases, which unwind DNA; translocases, which pull proteins across membranes; and other protein machines that convert chemical energy into motions in some particular direction. How do molecular machines produce directed motion? Brownian motion, by itself, cannot be the source of directed motion, because it is random. Figure 18.12 illustrates the **Brownian ratchet** model [5, 6] of how random diffusion, coupled with energy-driven but nondirectional binding and release events, can lead to directed motion.

Consider a ligand molecule L that moves along some ‘molecular rope,’ a partner molecule P having a chain of binding sites. Suppose there is an asymmetric binding free energy as a function of spatial coordinate x , which is aligned with the rope axis. Assume this function is shaped like sawteeth, with a steeper dependence of free energy on x in one direction than the other (see Figure 18.12). Before time $t = 0$, the system is stable, and L is bound at a location where the binding free energy $F(x)$ is a minimum. At time $t = 0$, energy is put into the system in some way to release L from its binding site on P . The ligand, no longer in contact with P , then diffuses freely along the x axis with no bias favoring either the $+x$ or $-x$ direction. The ligand remains unbound to P and

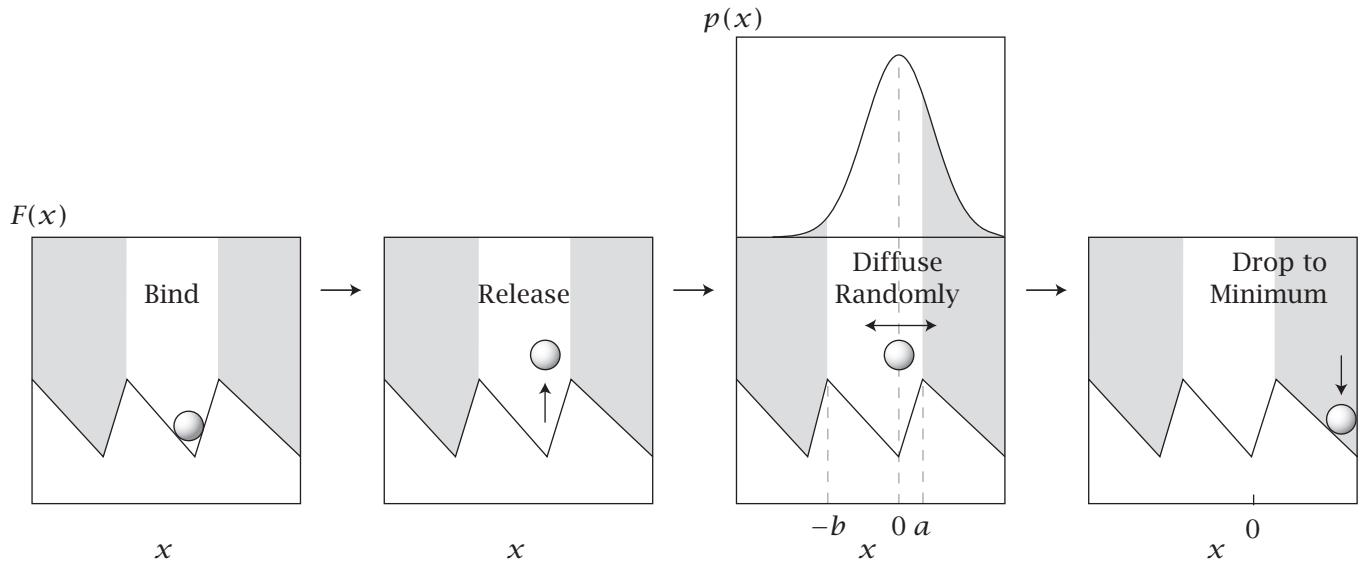


Figure 18.12 A simple Brownian ratchet. $F(x)$ is the free energy versus x , the one-dimensional coordinate of motion. There are four steps, as follows: (1) The ligand L binds to a low-energy well. (2) L is released. (3) L diffuses randomly in the $+x$ and $-x$ directions for a time τ_{off} . Because $F(x)$ is asymmetric, the particle is more likely to move to the right of $x = a$ than to the left of $x = -b$, as indicated by the shading in the diagram of probability $p(x)$. (4) At the final diffusional value of x , L binds again, and rolls downhill to the nearest energy minimum. The cycle repeats.

diffuses for a time τ_{off} . Diffusion leads to a Gaussian distribution along x . During that time, some of the ligand molecules will diffuse to $x \geq a$, where $x = a$ is the location of the next maximum to the right. At that time, those ligand molecules will rebind and slide energetically downhill to the next energy well to the right of the original binding site. A smaller number of molecules will diffuse to the left to $x \leq -b$, where they can fall into a well to the left of the original binding site.

Even though diffusion is symmetrical in x , the ligand binding potential $F(x)$ is not. At time τ_{off} , more particles fall into the energy well on the right than fall into the well on the left. Repeated cycles of release, diffusion, and rebinding lead to a net hopping from one well to the next toward the right. If the time interval τ_{off} is too short, most of the ligands will return to the same well from which they started, so there will be little net motion. If τ_{off} is too long, the particles will have time to spread so broadly in both directions that again there will be no net directed motion. So a key parameter that determines the average velocity is the off-rate of the ligand from P , $k_{\text{off}} = 1/\tau_{\text{off}}$.

Let's make the model more quantitative. L diffuses along the x axis for a time τ_{off} . p_{right} is the probability that the ligand diffuses to beyond $x = a$ in that time. p_{right} is given by integrating Equation (17.33):

$$p_{\text{right}} = (4\pi D\tau_{\text{off}})^{-1/2} \int_a^{\infty} e^{-x^2/4D\tau_{\text{off}}} dx. \quad (18.51)$$

Equation (18.51) can be expressed more compactly in terms of a mathematical function erfc (called the complementary error function):

$$\text{erfc}(z) = \frac{2}{\sqrt{\pi}} \int_z^{\infty} e^{-u^2} du. \quad (18.52)$$

Comparison of the exponents in Equations (18.51) and (18.52) shows that u is a dimensionless quantity given by

$$u^2 = \frac{x^2}{4D\tau_{\text{off}}} = \left(\frac{x}{\ell_0}\right)^2 \quad (18.53)$$

where $\ell_0 = 2\sqrt{D\tau_{\text{off}}}$ has units of length. ℓ_0 is the root-mean-square distance traveled in time τ_{off} by a particle having diffusion constant D (see Equation (18.11); for one dimension, $d = 1$). z is the lower limit of the integral in these reduced units. That is, when $x = a$, Equation (18.53) gives $u = a/\ell_0$, so the lower limit of the integral in Equation (18.52) is $z = a/\ell_0$. Substitute $z = a/\ell_0$ and Equation (18.53) into Equation (18.52). Then express Equation (18.51) as

$$p_{\text{right}} = \frac{1}{2} \operatorname{erfc}\left(\frac{a}{2\sqrt{D\tau_{\text{off}}}}\right). \quad (18.54)$$

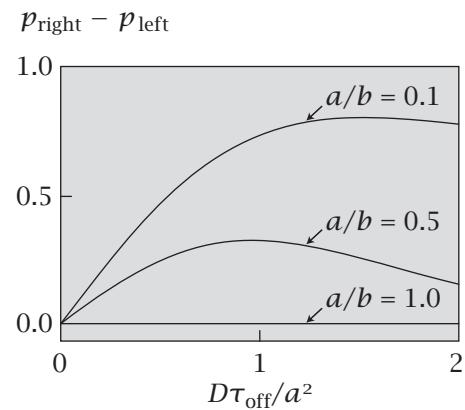
Similarly, the probability p_{left} that the particle diffuses to the left and reaches the next well at a distance $|x| = b$ is $p_{\text{left}} = (1/2) \operatorname{erfc}(b/(2\sqrt{D\tau_{\text{off}}}))$. Because the ligand travels a unit distance $a+b$ in a unit time $\tau_{\text{off}} + \tau_{\text{on}}$, the net velocity to the right, v_{right} , is

$$v_{\text{right}} = \left(\frac{a+b}{\tau_{\text{off}} + \tau_{\text{on}}}\right) (p_{\text{right}} - p_{\text{left}}). \quad (18.55)$$

Equation (18.55) is an approximation based on the assumption that the root-mean-square diffusion distance $2\sqrt{D\tau_{\text{off}}}$ is not large compared with a , because otherwise some particles might jump two or more units in the time $\tau_{\text{off}} + \tau_{\text{on}}$.

Figure 18.13 is a plot of $p_{\text{right}} - p_{\text{left}}$ versus $(D\tau_{\text{off}})/a^2$ for various ratios a/b . It shows that the net velocity toward the right increases as D increases, as τ_{off} increases, or as the asymmetry, a/b , of the potential increases. If the potential is symmetric, $a = b$, this model shows that there is no net motion. This model shows how Brownian motion, which is nondirectional, and energy-driven binding and release events, which are also nondirectional, can combine to create directed motion if there is an asymmetric binding potential.

Figure 18.13 Brownian ratchet particle velocity to the right, v_{right} , is proportional to $p_{\text{right}} - p_{\text{left}}$. This velocity increases as D increases, as τ_{off} increases, or as a decreases.



The Fluctuation-Dissipation Theorem Relates Equilibrium Fluctuations to the Rate of Approach to Equilibrium

An important theorem of statistical mechanics relates a property of kinetics to a property of equilibrium. At equilibrium, a system undergoes thermal fluctuations. Remarkably, the magnitudes of these equilibrium fluctuations are related to how fast the system approaches equilibrium. This theorem is quite general, and applies to many different physical and chemical processes. It allows you to determine the diffusion constant, viscosity, and other transport properties from knowledge of the equilibrium fluctuations.

To illustrate the idea, we develop the *Langevin model* of a particle moving in one dimension, subject to Brownian motion. The particle—whether a protein molecule, a colloidal particle, or a biological cell—is large relative to the solvent molecules. The solvent molecules bombard the particle rapidly from all directions. According to Newton's laws, a particle's mass m multiplied by its acceleration $d\nu/dt$, where ν is the particle velocity, equals the sum of all the forces acting on the particle. In the Langevin model, two forces act on the particle: friction and a random force $f(t)$ that fluctuates rapidly as a function of time t , representing Brownian motion. Every time a Brownian bombardment causes the particle to move in the $+x$ direction, frictional drag acts in the $-x$ direction to slow it down. For this situation, Newton's law gives the **Langevin equation**:

$$m \frac{d\nu}{dt} = f(t) - \xi \nu, \quad (18.56)$$

where ξ is the friction coefficient. Because the fluctuating force acts just as often in the $+x$ direction as in the $-x$ direction, the *average force* is zero:

$$\langle f(t) \rangle = 0. \quad (18.57)$$

We use this model to illustrate the idea of a *time correlation function*.

The Time Correlation Function Describes How Quickly Brownian Motion Erases the Memory of the Initial Particle Velocity

Here's how to construct a time-correlation function. Take the velocity $\nu(0)$ of the particle at time $t = 0$. Multiply by the velocity of the particle $\nu(t)$ at a later time t . Take the equilibrium ensemble average of this product over many different collisions to get $\langle \nu(0)\nu(t) \rangle$. This is the *velocity autocorrelation function*. It is an example of a more general quantity called a time correlation function. For example, you can do this with positions $\langle x(0)x(t) \rangle$ or forces $\langle f(0)f(t) \rangle$.

The velocity autocorrelation function tells you how fast a particle ‘forgets’ its initial velocity, owing to Brownian randomization. When the time t is short relative to the correlation time of the physical process, the particle velocity will be nearly unchanged from time 0 to t , and $\nu(t)$ will nearly equal $\nu(0)$, so $\langle \nu(0)\nu(t) \rangle \approx \langle \nu^2(0) \rangle$. But when t is much greater than the system's correlation time, Brownian motion will have had time to randomize the particle velocity

relative to its initial velocity, so $v(t)$ will be *uncorrelated* with $v(0)$. This means that $v(0)v(t)$ will be negative just as often as it is positive, so the ensemble average will be zero: $\langle v(0)v(t) \rangle = 0$.

Let's compute the velocity autocorrelation function from the Langevin model. Multiply both sides of the Langevin equation (18.56) by $v(0)$, which is a constant independent of t . Then take the ensemble average to get

$$m \left\langle v(0) \frac{dv(t)}{dt} \right\rangle = \langle v(0)f(t) \rangle - \xi \langle v(0)v(t) \rangle. \quad (18.58)$$

Because the fluctuating force is uncorrelated with the initial velocity of the particle, you have $\langle v(0)f(t) \rangle = \langle v(0) \rangle \langle f(t) \rangle$. This product is zero because $\langle f(t) \rangle = 0$, so

$$\begin{aligned} \frac{d}{dt} \langle v(0)v(t) \rangle + \frac{\xi}{m} \langle v(0)v(t) \rangle &= 0 \\ \Rightarrow \quad \langle v(0)v(t) \rangle &= \langle v^2(0) \rangle e^{-\xi t/m}. \end{aligned} \quad (18.59)$$

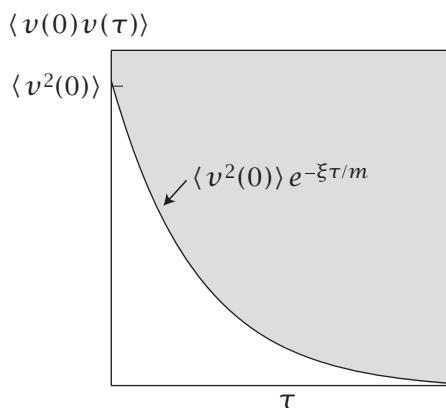
Now because $\langle mv^2(0)/2 \rangle = kT/2$ (see the equipartition equation (10.17)), Equation (18.59) can be expressed as

$$\langle v(0)v(t) \rangle = \frac{kT}{m} e^{-\xi t/m}. \quad (18.60)$$

This function is shown in Figure 18.14. This figure shows that the longer the delay between two collisions, the less correlation there is between the velocities $v(0)$ and $v(t)$.

In the Langevin model, there are two timescales. First, each bombardment of the large particle by a solvent molecule is very fast. There is a second, much slower, timescale over which the particle's velocity at one time becomes uncorrelated with its velocity at another time. Because the exponent in Equation (18.60) is dimensionless, the *time constant* for this slower process is m/ξ .

Figure 18.14 Autocorrelation function, showing the ensemble average of the product of the velocity at two different times, $t = 0$ and $t = \tau$. If τ is large enough, the velocity $v(\tau)$ becomes uncorrelated with $v(0)$.



EXAMPLE 18.7 Autocorrelation times. The *correlation time* m/ξ is the time required for the velocity autocorrelation function to reach $1/e$ of its initial value. The correlation time is short when the mass is small or when the friction coefficient is large. The correlation time for the diffusional motion of a small protein of mass $m = 10,000 \text{ g mol}^{-1}$ is in the picosecond time range:

$$\begin{aligned}\frac{m}{\xi} &= \frac{(10,000 \text{ g mol}^{-1})(1 \text{ kg}/1000 \text{ g})}{(6.023 \times 10^{23} \text{ molecules mol}^{-1})(3.77 \times 10^{-11} \text{ kg s}^{-1})} \\ &= 4.4 \times 10^{-13} \text{ s.}\end{aligned}$$

A particle's velocity is correlated with its earlier velocity over times shorter than m/ξ , but is uncorrelated over times much longer than this.

Now let's see how $\langle v(0)v(t) \rangle$, a property of the fluctuations in equilibrium, is related to a kinetic property, in this case the diffusion coefficient D . Integrate the time correlation function, Equation (18.60), over all the possible time lags t :

$$\int_0^\infty \langle v(0)v(t) \rangle dt = \frac{kT}{m} \int_0^\infty e^{-\xi t/m} dt = \frac{kT}{\xi} = D, \quad (18.61)$$

because $\int_0^\infty e^{-\xi t/m} dt = m/\xi$. In three dimensions, the velocities are vectors $\mathbf{v}(0)$ and $\mathbf{v}(t)$, and Equation (18.61) becomes

$$\int_0^\infty \langle \mathbf{v}(0)\mathbf{v}(t) \rangle dt = 3D. \quad (18.62)$$

Equations (18.61) and (18.62) are examples of **Green-Kubo relations** [7, 8] between an equilibrium correlation function and a *transport coefficient*, the diffusion constant D in this case. There are several such relationships. Without going into the details, we just note that another such relationship gives the friction coefficient in terms of the **force correlation function**:

$$\xi = \frac{1}{2kT} \int_{-\infty}^{\infty} \langle f(0)f(t) \rangle dt. \quad (18.63)$$

In addition, the viscosity is the autocorrelation function of momentum transfer in liquid flows. Green-Kubo relations equate a kinetic property (such as the diffusion coefficient or the friction coefficient) with an ensemble average of an equilibrium property, such as $\langle v(0)v(t) \rangle$.

Summary

We have used the binomial distribution for describing different dynamical processes, including the random-flight model of diffusion of a single particle, the dog-flea model of Fick's law diffusion, and the Poisson and waiting-time distributions for inter-arrival times of stochastic events. Particles diffuse down concentration gradients because there are more microscopic routes of flow in that direction than in the opposite direction. Often, computing dynamical properties can be simplified by using the dynamical partition function. Brownian ratchets are microscopic devices in which diffusion becomes directed by harnessing an energy-utilizing step. The Langevin model gives a useful description of fluctuation-dissipation relationships.

Problems

1. Diffusion of light from the center of the Sun. Neutrinos fly from the Sun's center to its surface in about 2.3 s because they travel at the speed of light ($c \approx 3 \times 10^{10} \text{ cm s}^{-1}$) and they undergo little interaction with matter. But a photon of light takes much longer to travel from the Sun's center to its surface because it collides with protons and free electrons and undergoes a random walk to reach the Sun's surface. A photon's step length is approximately 1 cm per step [9]. The Sun's radius is $7 \times 10^{10} \text{ cm}$. How long does it take a photon to travel from the center to the surface of the Sun?

2. How far does a protein diffuse? Consider a protein that has a diffusion constant in water of $D = 10^{-6} \text{ cm}^2 \text{ s}^{-1}$.

- What is the time that it takes the protein to diffuse a root-mean-square distance of $10 \mu\text{m}$ in a three-dimensional space? This distance is about equal to the radius of a typical cell.
- If a cell is spherical with radius $r = 10 \mu\text{m}$, and if the protein concentration outside the cell is $1 \mu\text{M}$ (micromolar), what is the number of protein molecules per unit time that bombard the cell at the diffusion-limited rate?

3. Einstein's estimate of Brownian motion and Avogadro's number. Einstein assumed that a particle of radius $5 \times 10^{-5} \text{ cm}$ could be observed to undergo Brownian motion under a microscope [10]. He computed the root-mean-square distance $\langle x^2 \rangle^{1/2}$ that the particle would move in one minute in a one-dimensional walk in water, $\eta \approx 1 \text{ cP}$ at about $T = 300 \text{ K}$.

- Compute $\langle x^2 \rangle^{1/2}$.
- Show how the argument can be turned around to give Avogadro's number.

4. $\langle x^2 \rangle$ from the Langevin model.

- Derive an expression for $\langle x^2(t) \rangle$ from the Langevin model.
- Find limiting expressions for $\langle x^2(t) \rangle$ as $t \rightarrow 0$ and as $t \rightarrow \infty$.

5. Molecular movements inside cells. Consider a biological cell of radius $r = 10^{-6} \text{ m}$ contained within a bilayer membrane of thickness 30 \AA .

- You have a water-soluble small-molecule drug that has a diffusion coefficient $D = 10^{-5} \text{ cm}^2 \text{ s}^{-1}$ and it partitions from water into oil with partition coefficient $K = 10^{-3}$. The drug has concentration $1 \mu\text{M}$ outside the cell, and zero inside the cell. How many molecules per second of the drug flow into the cell?
- How long does it take the drug to diffuse to the center of the cell, after it passes through the membrane?
- How long would it take a large protein, having diffusion coefficient $D = 10^{-7} \text{ cm}^2 \text{ s}^{-1}$, to diffuse the same distance?

6. How does *Escherichia coli* sense a chemical gradient?

A cell that is surrounded by chemoattractant molecules will sense or 'measure' the number of molecules, N , in roughly its own volume a^3 , where a is the radius of the cell. Suppose the chemoattractant aspartate is in concentration $c = 1 \mu\text{M}$ and has diffusion constant $D = 10^{-5} \text{ cm}^2 \text{ s}^{-1}$. For *E. coli*, $a \approx 10^{-6} \text{ m}$.

- Compute N .
- The error or noise in the measurement of a quantity like N will be approximately \sqrt{N} . Compute the precision P of *E. coli*'s measurement, which is the noise/signal ratio $P = \sqrt{N}/N$.
- Compute the 'clearance time' τ_c that it takes for a chemoattractant molecule to diffuse roughly the distance a . This is the time over which *E. coli* makes its 'measurement.'
- If the cell stays in place for m units of time $\tau = m\tau_c$, to make essentially m independent measurements, the cell can improve its precision. Write an expression for the precision P_m of the cell's measurement after m time steps, as a function of a, D, c , and τ .
- Given $\tau = 1 \text{ s}$, compute P_m .

References

- [1] K Ghosh, KA Dill, MM Inamdar, et al., *Am J Phys* **74**, 123–133 (2006).
- [2] W Min, BP English, G Luo, et al., *Acc Chem Res* **38**, 923 (2005).
- [3] G Stock, K Ghosh, and KA Dill, *J Chem Phys* **128**, 194102 (2008).
- [4] D Wu, K Ghosh, M Inamdar, et al., *Phys Rev Lett* **103**, 050603 (2009).
- [5] Y Okada and N Hirokawa, *Science* **283**, 1152–1157 (1999).
- [6] RD Astumian, *Science* **276**, 917–922 (1997).
- [7] D Chandler, *Introduction to Modern Statistical Mechanics*, Oxford University Press, Oxford, 1987.
- [8] DA McQuarrie, *Statistical Mechanics*, 2nd edition, University Science Books, Sausalito, CA, 2000.
- [9] M Harwit, *Astrophysical Concepts*, 4th edition, Springer-Verlag, New York, 2006.
- [10] A Einstein, *Investigations on the Theory of Brownian Movement*. Dover Publications, New York, 1956.

Suggested Reading

HC Berg, *Random Walks in Biology*, Princeton University Press, Princeton, NJ 1993. Elementary and concise overview of diffusion, flow, and random walks.

RD Smiley and GG Hammes, *Chem Rev* **106**, 3080-3094 (2006).

Excellent treatments of the Langevin equation and time correlation functions:

RS Berry, SA Rice, and J Ross, *Physical Chemistry*, 2nd edition, Oxford University Press, New York, 2000, Chapter 29

D Chandler, *Introduction to Modern Statistical Mechanics*, Oxford University Press, Oxford, 1987.

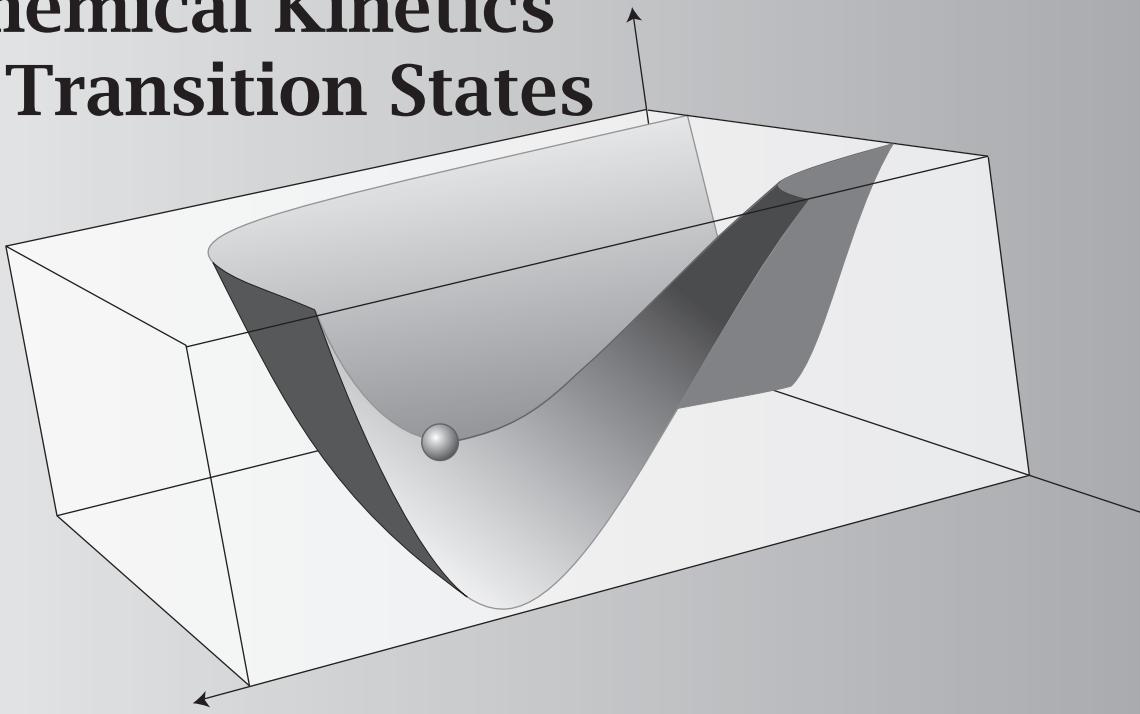
N Wax, ed., *Selected Papers on Noise and Stochastic Processes*, Dover Publications, New York, 1954.

An outstanding textbook on cellular biophysics that treats many different microscopic dynamical processes in biology:

R Phillips, J Kondev, and J Theriot, *Physical Biology of the Cell*, Garland Science, New York, 2008.

This page is intentionally left blank.

19 Chemical Kinetics & Transition States



Chemical Reaction Rates Depend on Temperature

Now, we focus on the kinetics of chemical reactions. To predict how the rate of a chemical reaction depends on its molecular structures, you can use the same statistical thermodynamics approach that we used in Chapter 13 to model equilibria. Chemical reactions typically speed up more strongly with temperature than physical processes do. To understand this, you need one additional concept: the *transition state* or *activation barrier*.

The Mass Action Laws Describe Mechanisms in Chemical Kinetics

Consider a chemical reaction in which a product P is produced from reactants A , B , and C , with stoichiometric coefficients a , b , and c :



In general, the reaction rate depends on the concentrations of the reactants, the temperature and pressure, and the coefficients a , b , and c . Experiments often measure how reaction rates depend on the concentrations of the reactants. Such experiments can provide valuable information about the mechanism of the reaction.

The kinetic *law of mass action*, first developed by CM Guldberg and P Waage in 1864, says that reaction rates should depend on stoichiometry in the same way that equilibrium constants do. According to this law, the initial rate of product formation, $d[P]/dt$ for the reaction in Equation (19.1), depends on the reactant concentrations:

$$\frac{d[P]}{dt} = k_f[A]^a[B]^b[C]^c, \quad (19.2)$$

where k_f is the rate coefficient for the forward reaction. However, kinetic mechanisms often do not follow the thermodynamic stoichiometries. If a reaction is a single step, called an *elementary reaction*, then such expressions apply. The main types of elementary reactions are *unimolecular* decay of a molecule ($a = 1, b = c = 0$) or a *bimolecular* reaction when two molecules come together ($a = b = 1, c = 0$). However, chemical reactions typically involve multiple steps and kinetic intermediate states. For such non-elementary reactions, you cannot express the rate equation in such simple stoichiometric terms. At present, rate laws can only be determined from experiments. We do not address reaction mechanisms here; they are described in chemical kinetics textbooks such as [1, 2].

Chemical reactions obey a principle called *detailed balance*, described below.

At Equilibrium, Rates Obey Detailed Balance

Consider a two-state reaction (Equation (17.45)):



where k_f and k_r are the forward and reverse rate coefficients. The *principle of detailed balance* says that the forward and reverse rates must be identical for an elementary reaction at equilibrium:

$$k_f[A]_{\text{eq}} = k_r[B]_{\text{eq}}, \quad (19.4)$$

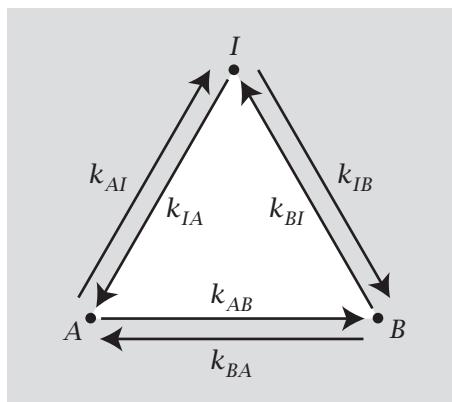
where $[]_{\text{eq}}$ is the equilibrium concentration. To see that this is a condition of equilibrium, substitute Equation (19.4) into Equations (17.46) and (17.47) to get $d[A]/dt = 0$ and $d[B]/dt = 0$. Note that it is the rates, not the rate coefficients, that are equal at equilibrium.

Equation (19.4) relates the rate coefficients k_f and k_r to the equilibrium constant K :

$$K = \frac{[B]_{\text{eq}}}{[A]_{\text{eq}}} = \frac{k_f}{k_r}. \quad (19.5)$$

For more complex systems, the principle of detailed balance goes beyond just the statement that the system is in a steady state. Figure 19.1(a) shows a three-state reaction between states A , B , and I . The steady-state condition says only that $d[A]/dt = d[B]/dt = d[I]/dt = 0$: the flux *in* equals the flux *out* for each state. The scheme in Figure 19.1(b) can satisfy that condition, for example if $k_{AI}[A]_{\text{eq}} = k_{IB}[I]_{\text{eq}} = k_{BA}[B]_{\text{eq}}$. A metaphor is 10 trucks per hour going from city A to I and 10 from I to B and 10 from B to A . There is no net

(a) Possible Mechanism



(b) Impossible Mechanism

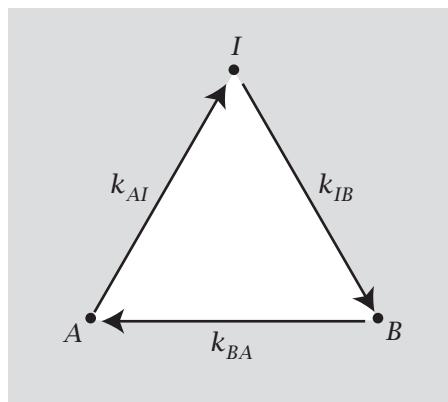


Figure 19.1 The principle of detailed balance is satisfied by mechanism (a) between three states A , I , and B , but violated by mechanism (b). For a system in equilibrium, forward rates must equal reverse rates for each pair of states.

accumulation of trucks in one city, so the system is in a steady state, but there is a net circulation around the cycle, so this system is not in equilibrium. Being in equilibrium is a more stringent requirement than being in a steady state. Detailed balance says that to achieve equilibrium, each individual forward rate must equal its corresponding individual backward rate:

$$\frac{[I]_{\text{eq}}}{[A]_{\text{eq}}} = \frac{k_{AI}}{k_{IA}}, \quad \frac{[B]_{\text{eq}}}{[I]_{\text{eq}}} = \frac{k_{IB}}{k_{BI}}, \quad \text{and} \quad \frac{[A]_{\text{eq}}}{[B]_{\text{eq}}} = \frac{k_{BA}}{k_{AB}}. \quad (19.6)$$

The principle of detailed balance is useful for ruling out reaction mechanisms. The detailed balance equations (19.6) say that the mechanism shown in Figure 19.1(b) is impossible. In the forward direction, A converts to B through an intermediate state I , while in the reverse direction, B converts directly to A , without an intermediate state. Multiplying Equations (19.6) gives the detailed balance condition that must be satisfied:

$$1 = \frac{k_{AI}k_{IB}k_{BA}}{k_{IA}k_{BI}k_{AB}}. \quad (19.7)$$

Equation (19.7) cannot be satisfied by mechanism (b) because the denominator would be zero if the reverse rates were zero. The principle of detailed balance says that forward and backward reactions at equilibrium cannot have different intermediate states. That is, if the forward reaction is $A \rightarrow I \rightarrow B$, the backward reaction cannot be $B \rightarrow A$. Detailed balance is helpful in understanding the *mechanisms* of chemical reactions, the molecular steps from reactants to products.

You can also get insights into chemical reactions from their temperature dependences, described below.

Chemical Reactions Depend Strongly on Temperature

Consider a bimolecular reaction in the gas phase:



The initial rate of appearance of P can be expressed as

$$\frac{d[P]}{dt} = k_2[A][B], \quad (19.9)$$

where $k_2[A][B]$ is the rate of the reaction, and the subscript 2 on the rate coefficient indicates that two reactants are involved. By definition, the rate coefficient k_2 is independent of the concentrations of A and B . However, k_2 can depend strongly on temperature.

Figure 19.2 shows a typical temperature dependence of a rate coefficient. This observed dependence of the reaction rate coefficient on temperature is much greater than you would find from just the enhanced thermal motions of the molecules. Increasing the temperature from 700 K to 800 K would increase the average thermal energy RT by only 16%, but the rate coefficient increases 40-fold (see Figure 19.2).

In 1889, S Arrhenius (1859–1927), a Swedish chemist and physicist, proposed a simple explanation for the strong temperature dependence of reaction rates. Based on the van't Hoff equation (13.35) for the strong dependence of the equilibrium constant K on temperature,

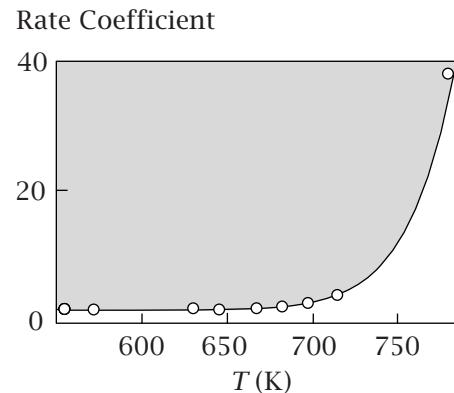
$$\frac{d \ln K}{dT} = \frac{\Delta h^\circ}{RT^2},$$

Arrhenius proposed that the forward and reverse rate coefficients k_f and k_r might also have the van't Hoff form:

$$\frac{d \ln k_f}{dT} = \frac{E_a}{RT^2} \quad \text{and} \quad \frac{d \ln k_r}{dT} = \frac{E'_a}{RT^2}, \quad (19.10)$$

where E_a and E'_a have units of energy that are chosen to fit the experimental data. E_a and E'_a are called *activation energies*.

Figure 19.2 Rate coefficient k_f for the reaction $2\text{HI} \rightarrow \text{H}_2 + \text{I}_2$ in the gas phase as a function of temperature T . These data show the exponential dependence described by Equation (19.11). Source: H Eyring and EM Eyring, *Modern Chemical Kinetics*, Reinhold, New York, 1963.



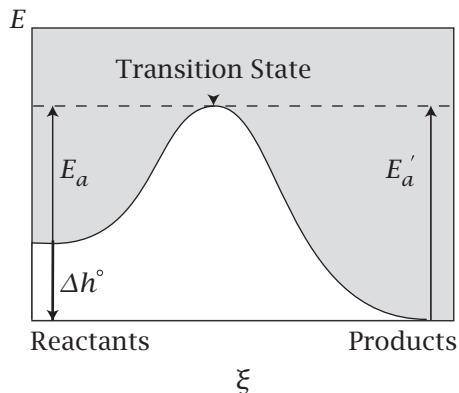


Figure 19.3 An activation energy diagram relates three quantities: the activation energy E_a for the forward reaction, the activation energy E'_a for the reverse reaction, and the enthalpy Δh° for the equilibrium. ξ is the reaction coordinate. A more microscopic interpretation of the reaction coordinate is given on page 363.

According to Arrhenius, it is not the average energy of the reactants that determines the reaction rates but only the high energies of the ‘activated’ molecules. Figure 19.3 shows this idea in a diagram. There are two plateaus in energy, representing reactants and products. The diagram shows a maximum, called the *transition state* or *activation barrier*, which is the energy that activated molecules must have to proceed from reactants to products. All three of the energy differences between minima or maxima shown on this diagram are commonly obtainable from experiments. Measuring the forward rate k_f as a function of temperature and using Equation (19.10) gives E_a . Measuring the reverse rate gives E'_a . Measuring the equilibrium constant versus temperature and using Equation (13.35) gives Δh° . Because the equilibrium constant is related to the rates by $K = k_f/k_r$, the three energy quantities are related by $\Delta h^\circ = E_a - E'_a$.

Integrating Equation (19.10) and exponentiating gives the temperature dependence of the forward reaction:

$$k_f = Ae^{-E_a/RT}, \quad (19.11)$$

where A is a constant.

Just as equilibria can be represented by a van't Hoff plot of $\ln K$ versus $1/T$, kinetics can be represented by an *Arrhenius plot* of $\ln k$ versus $1/T$. Example 19.1 shows how Equations (19.10) can account for the temperature dependence of reaction rates, taking E_a and E'_a to be constants.

EXAMPLE 19.1 Calculate E_a for $H_2 + I_2 \rightarrow 2HI$. Use Figure 19.4 and the Arrhenius equation (19.11) to get

$$\begin{aligned} \ln\left(\frac{k_{f2}}{k_{f1}}\right) &= -\frac{E_a}{R}\left(\frac{1}{T_2} - \frac{1}{T_1}\right) \\ \Rightarrow E_a &= \frac{-R(\ln k_{f2} - \ln k_{f1})}{(1/T_2) - (1/T_1)} \approx (-2 \text{ cal K}^{-1} \text{ mol}^{-1}) \left(\frac{\ln 10^4 - \ln 10^{-1}}{0.0012 - 0.0018} \right) \\ &\approx 38 \text{ kcal mol}^{-1}. \end{aligned}$$

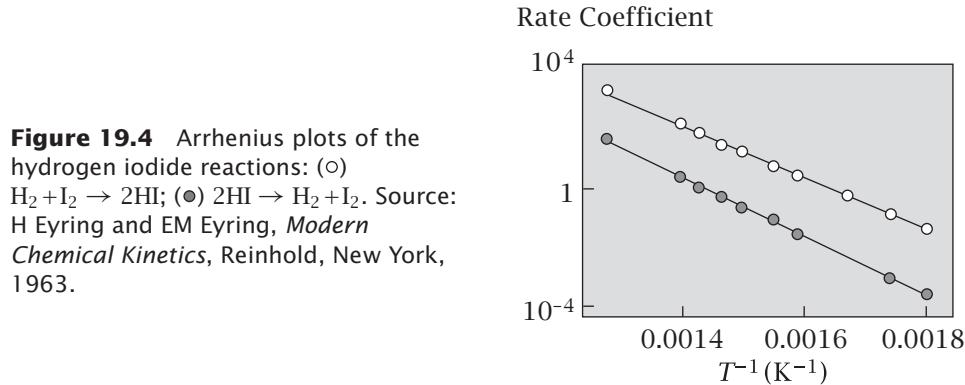


Figure 19.4 Arrhenius plots of the hydrogen iodide reactions: (○) $H_2 + I_2 \rightarrow 2HI$; (●) $2HI \rightarrow H_2 + I_2$. Source: H Eyring and EM Eyring, *Modern Chemical Kinetics*, Reinhold, New York, 1963.

Arrhenius kinetics applies to many physical processes, not just to chemical reactions. One example is the diffusion of atoms in solids. Figure 19.5 shows that the diffusion rates of carbon atoms through solid iron metal follow Arrhenius behavior over a remarkable 14 orders of magnitude. This evidence supports the *interstitial model*, which says that the carbon atoms occupy the interstices in the iron lattice, and ‘jump’ over energy barriers to travel from one interstitial site to another. Many other rate processes also follow the Arrhenius temperature law: the chirping of crickets (see Figure 19.6), the flashing of fireflies, the heartbeats of turtles, and the metabolic rates of birds and animals, for example. The temperature dependence of each of these processes may reflect some underlying chemical origins.

When should you treat a process as activated? If a small increase in temperature gives a large increase in rate, a good first step is to try the Arrhenius model. The making or breaking of bonds is often, but not always, well modeled as an activated process. Highly reactive radicals are counterexamples. Reactions such as $H_3^+ + HCN \rightarrow H_2 + H_2CN^+$ can be much faster than typical activated processes, and they slow down with increasing temperature, although only slightly. Such ‘negative activation energies’ are usually small, 0 to -2 kcal mol $^{-1}$ [3, 4].

We now describe a more microscopic approach to reaction rates, called transition state theory.

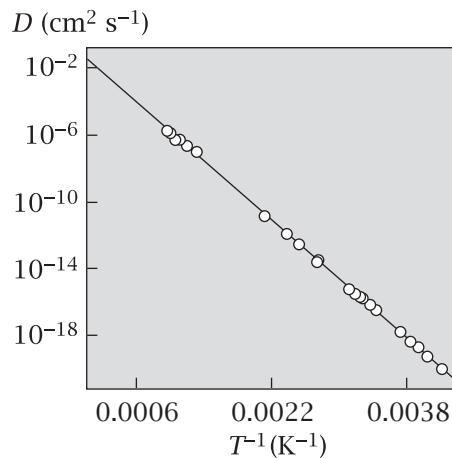


Figure 19.5 The diffusion coefficient D of carbon in iron as a function of temperature: $\log D$ versus $1/T$ is linear, indicating a constant activation energy. Source: RJ Borg and GJ Dienes, *The Physical Chemistry of Solids*, Academic Press, Boston, 1992.

Logarithm of the Frequency
of Chirping of Tree Crickets

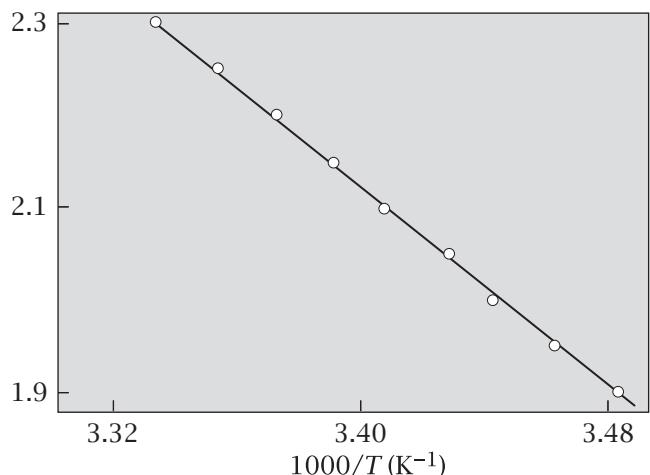


Figure 19.6 Crickets chirp faster on warm summer evenings. The logarithm of the chirping rate decreases linearly with the inverse temperature, consistent with the Arrhenius model. Source: redrawn from KJ Laidler, *J Chem Ed* **49**, 343–344 (1972).

Activated Processes Can Be Modeled by Transition-State Theory

The Energy Landscape of a Reaction

An *energy landscape* defines how the energy of a reacting system depends on its degrees of freedom—the positions and orientations of all the reactant and product atoms with respect to each other. The cover of this book shows a complex energy landscape. A simpler example is the reaction between atoms *A*, *B*, and *C*:



On the left, *B* is bound to *C*; on the right *B* is bound to *A*. This reaction can be described by two degrees of freedom: the distance R_{AB} between *A* and *B*, and the distance R_{BC} between *B* and *C*, if the attack and exit of the isolated atom are collinear with the bond. Otherwise, the angles of attack and exit are also relevant degrees of freedom. Figure 19.7 shows the energy landscape computed from quantum mechanics as a function of these two degrees of freedom for the reaction of deuterium *D* with a hydrogen molecule: $D + H_2 \rightarrow HD + H$.

When *D* is far away from H_2 , the H_2 bond length is constant. *D* approaches along an energy valley of constant depth. When *D* gets close to H_2 , it causes the $H-H$ bond to stretch and the energy increases to a maximum, which is indicated as a saddle point resembling a mountain pass in Figure 19.7. At the maximum, the central hydrogen is weakly bound to the incoming *D* and is also weakly bound to the outgoing *H*. After the $H-H$ bond is broken and the lone *H* exits, the $D-H$ bond length R_{HD} shortens and approaches constant length and energy, indicated by the outgoing valley. The width of each valley represents the vibrations of the $H-H$ and $D-H$ bonds.

The reaction trajectory for any one set of reacting atoms typically involves some excursions up the walls of the valleys, corresponding to the central *H* oscillating between the incoming and outgoing atoms. But when averaged over

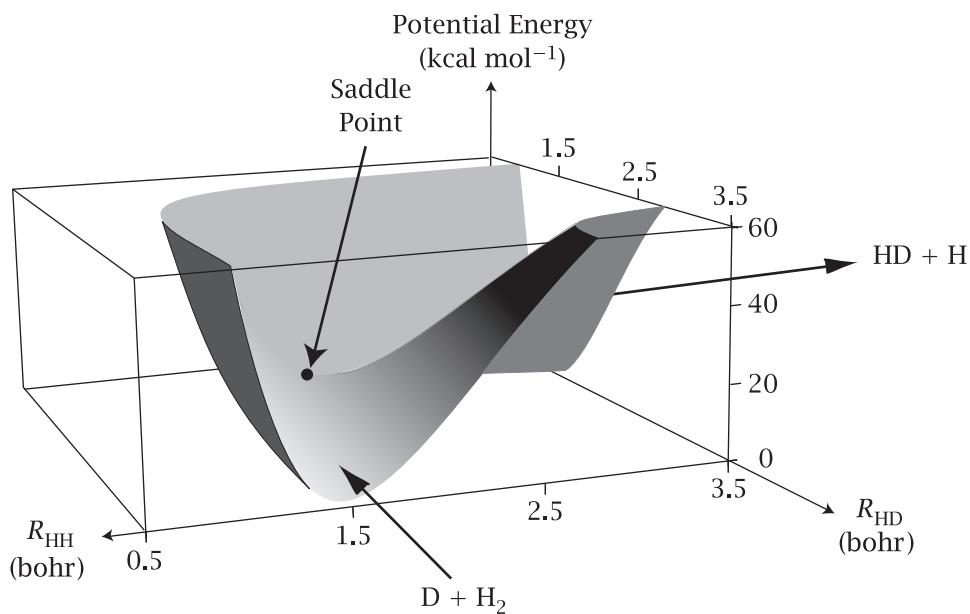


Figure 19.7 The energy surface for the reaction $D + H_2 \rightarrow HD + H$ as a function of the bond distances R_{HH} and R_{HD} (1 bohr = 0.53×10^{-10} m) when the three atoms are collinear. The arrows indicate the path of the reaction. The point of highest energy on the minimum-energy path is the saddle point. Source: S Borman, *Chem Eng News*, June 4 (1990). Data are from DG Truhlar, University of Minnesota.

multiple trajectories, the reaction process can be described as following the lowest-energy route, along the entrance valley over the saddle point and out of the exit valley, because the Boltzmann populations are highest along that average route.

The reaction coordinate ξ defines the position of the reaction along this average route. Even when energy landscapes involve many degrees of freedom, if there is a single lowest-energy route involving valleys and mountain passes, that route defines the reaction coordinate. The saddle point, which is the highest point along the reaction coordinate, is called the *transition state* and is denoted by the symbol \ddagger . The transition state is unstable: a ball placed at the saddle point will roll downhill along the reaction coordinate in one direction or the other. Because the transition state has high energy, it has low population.

Calculating Rate Coefficients from Transition-State Theory

Let's calculate the rate coefficient k_2 for the reaction $A + B \xrightarrow{k_2} P$ by means of *transition-state theory*, also called *absolute rate theory*. The approach is to divide the reaction process into two stages. The first stage is an equilibrium between the reactants and the transition state $(AB)^{\ddagger}$ with 'equilibrium constant' K^{\ddagger} . The second stage is a direct step downhill from the transition state to form the product with rate coefficient k^{\ddagger} :



A key assumption of transition-state theory is that the first stage can be expressed as an equilibrium between the reactants A and B and the transition state $(AB)^\ddagger$, with equilibrium constant K^\ddagger ,

$$K^\ddagger = \frac{[(AB)^\ddagger]}{[A][B]}, \quad (19.14)$$

even though the transition state $(AB)^\ddagger$ is not a true equilibrium state. It is an unstable state because it is at the top of an energy hill. The overall rate is expressed as the number of molecules in the transition state, $[(AB)^\ddagger]$, multiplied by the rate coefficient k^\ddagger for converting transition state molecules to product:

$$\frac{d[P]}{dt} = k^\ddagger[(AB)^\ddagger] = k^\ddagger K^\ddagger [A][B]. \quad (19.15)$$

The second equality in Equation (19.15) follows from Equation (19.14). Combining Equations (19.15) and (19.9) gives the measurable rate coefficient k_2 :

$$k_2 = k^\ddagger K^\ddagger. \quad (19.16)$$

Because the quantity K^\ddagger is regarded as an equilibrium constant, you can express it in terms of the molecular partition functions developed in Chapter 13 (see Equation (13.17)):

$$K^\ddagger = \left(\frac{q_{(AB)^\ddagger}}{q_A q_B} \right) e^{\Delta D^\ddagger / kT}, \quad (19.17)$$

where ΔD^\ddagger is the dissociation energy of the transition state minus the dissociation energy of the reactants.

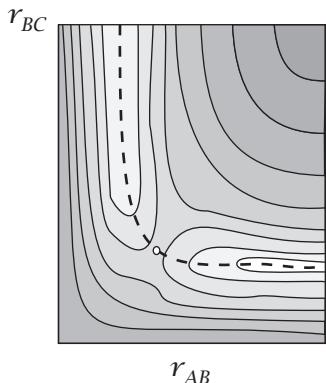
$q_{(AB)^\ddagger}$ is the partition function for the transition state. At the transition state, the AB bond is ready to form, but it is stretched, weak, and unstable. The system will move rapidly downhill in energy, either forward along the reaction coordinate, forming the bond, or backward, breaking it. Although the system is *unstable* (at a maximum of energy) *along* the reaction coordinate, it is *stable* (at a minimum) of energy in all other directions that are normal to the reaction coordinate (see Figure 19.8(b)). For the stable degrees of freedom, the system is presumed to act like any other equilibrium system. Those degrees of freedom are described by equilibrium partition functions.

Provided that the unstable degrees of freedom are independent of the stable degrees of freedom, $q_{(AB)^\ddagger}$ can be factorized into two components:

$$q_{(AB)^\ddagger} = \overline{q^\ddagger} q_\xi, \quad (19.18)$$

where $\overline{q^\ddagger}$ represents the partition function for all of the ordinary thermodynamic degrees of freedom of the transition-state structure and q_ξ represents the one nonequilibrium vibrational degree of freedom of the bond along the reaction coordinate. Thus, $\overline{q^\ddagger}$ is computed as an ordinary equilibrium partition function, but the ‘final state’ of this ‘equilibrium’ is the transition state. For example, to compute the translational partition function component q_t^\ddagger , use the sum of the masses of molecules A and B . To compute the rotational partition function q_r^\ddagger , use the moment of inertia given by the masses and bond separations of the transition state structure.

(a) Potential Energy Surface



(b) Three-Dimensional Transition State

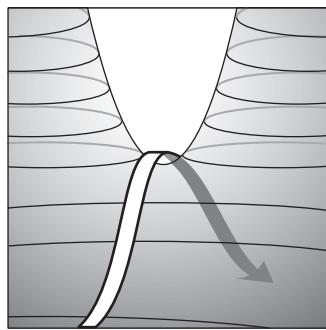


Figure 19.8 (a) Contour plot of a reaction pathway (— —) on an energy landscape for the reaction $A + BC \rightarrow AB + C$. The shading shows the energy levels, and the dashed line shows the lowest-energy path between reactants and products. (b) The transition state is an unstable point along the reaction pathway (indicated by the arrow), and a stable point perpendicular to the reaction pathway.
Source: A Pross, *Theoretical and Physical Principles of Organic Reactivity*, Wiley, New York, 1995.

Because the bond along ξ is only partly formed in the transition-state structure, we treat it as a weak vibration (see page 203). The discussion on page 202 shows that the frequency ν of vibration depends on the spring constant k_ξ and the mass. For the frequency ν_ξ of the reacting bond,

$$\nu_\xi = \frac{1}{2\pi} \left(\frac{k_\xi}{\mu} \right)^{1/2}, \quad (19.19)$$

where $\mu = m_A m_B / (m_A + m_B)$ is the reduced mass. The vibration is weak, so k_ξ and ν_ξ are small, and

$$\frac{\hbar\nu_\xi}{kT} \ll 1. \quad (19.20)$$

(In contrast, for stable covalent bonds, $\hbar\nu/kT \gg 1$.) The partition function $q_{\text{vibration}}$ for vibrations is given by Equation (11.26). When a vibration is weak, you can approximate the exponential in the partition function by $e^{-x} \approx 1 - x$ for small x (see Equation (J.1) in Appendix J), so

$$q_\xi = \frac{1}{1 - e^{-\hbar\nu_\xi/kT}} \approx \frac{kT}{\hbar\nu_\xi}. \quad (19.21)$$

Once the system has reached the transition state, it is assumed to proceed to product as rapidly as the system permits, namely at the frequency of the reaction coordinate vibration. The rate coefficient for the downhill step from transition state to product is

$$k^\ddagger = \nu_\xi. \quad (19.22)$$

A factor κ , called the *transmission coefficient*, is often introduced to account for observed deviations from the simple rate theory: $k^\ddagger = \kappa\nu_\xi$. For reactions in a condensed-phase medium, typically $\kappa < 1$ because the reactant/product can ‘bounce’ or *recross* the barrier owing to solvent interactions; for some other reactions, quantum mechanical tunneling through the barrier leads to $\kappa > 1$ [5].

Now substitute Equations (19.17), (19.18), (19.21), and (19.22) into Equation (19.16) to get the rate coefficient k_2 from the partition functions:

$$\begin{aligned} k_2 &= \left(\frac{kT}{h} \right) \left(\overline{\frac{q^\ddagger}{q_A q_B}} \right) e^{\Delta D^\ddagger/kT} \\ &= \left(\frac{kT}{h} \right) \overline{K^\ddagger}, \end{aligned} \quad (19.23)$$

where the overbars indicate that the unstable reaction coordinate degree of freedom has been factored out (into the term kT/h). Example 19.2 shows how you can predict chemical reaction rates if you know the transition state structure and its energy.

EXAMPLE 19.2 The rate of the reaction $\text{F} + \text{H}_2 \rightarrow \text{HF} + \text{H}$. Compute the rate of this reaction at $T = 300\text{ K}$. The masses are $m_{\text{H}_2} = 2$ atomic mass units (amu), and $m_{\text{F}} = 19$ amu. For the transition state $\text{F}\cdots\text{H}-\text{H}$, the mass is $m = 21$ amu. Quantum mechanical modeling shows that the transition state has an $\text{F}-\text{H}$ bond length of 1.602 \AA , an $\text{H}-\text{H}$ bond length of 0.756 \AA , a moment of inertia $I^\ddagger = 7.433\text{ amu \AA}^2$, and an electronic degeneracy $g_e = 4$. There are two bending vibrations, each with wavenumber 397.9 cm^{-1} . The reactant H_2 has bond length 0.7417 \AA , moment of inertia $I_{\text{H}_2} = 0.277\text{ amu \AA}^2$, symmetry factor $\sigma = 2$, and electronic degeneracy $g_e = 1$. It is in its vibrational ground state. The F atom has no rotational or vibrational partition function and its electronic degeneracy is $g_e = 4$. The difference in dissociation energies is $\Delta D^\ddagger = -657\text{ kJ mol}^{-1}$. Because experimental data are typically measured in terms of the pressures of the reactant and product gases (see page 241), and are measured per mole, rather than per molecule, we express Equation (19.23) as

$$\text{rate coefficient} = \left(\frac{RT}{h}\right) \frac{q^\ddagger/V}{(q_{\text{F}}/V)(q_{\text{H}_2}/V)} e^{\Delta D^\ddagger/RT}. \quad (19.24)$$

The contributions to the ratio of the partition functions are as follows:

$$\text{Electronic: } \frac{4}{4 \cdot 1} = 1,$$

$$\text{Rotational: } \frac{8\pi^2 I^\ddagger kT/h^2}{8\pi^2 I_{\text{H}_2} kT/\sigma_{\text{H}_2} h^2} = \frac{\sigma_{\text{H}_2} I^\ddagger}{I} = \frac{2(7.433)}{0.277} = 53.67,$$

and

Each vibration:

$$\frac{h\nu}{kT} = \frac{(6.626 \times 10^{-34}\text{ Js})(397.9\text{ cm}^{-1})(2.99 \times 10^{10}\text{ cm s}^{-1})}{(1.38 \times 10^{-23}\text{ JK}^{-1})(300\text{ K})} \\ = 1.904.$$

So the contribution of the two vibrational modes is

$$(1 - e^{-h\nu/kT})^{-2} = (1 - e^{-1.904})^{-2} = (1 - 0.149)^{-2} = 1.38.$$

Translations:

$$\begin{aligned} & \left(\frac{m_{\text{F}} + m_{\text{H}_2}}{m_{\text{F}} m_{\text{H}_2}} \right)^{3/2} \left(\frac{2\pi kT}{h^2} \right)^{-3/2} = \left[\left(\frac{1 + m_{\text{H}_2}/m_{\text{F}}}{m_{\text{H}_2}} \right) \left(\frac{h^2}{2\pi kT} \right) \right]^{3/2} \\ &= \left[\left(\frac{1 + 2/19}{(0.002\text{ kg mol}^{-1})/(6.02 \times 10^{23}\text{ mol}^{-1})} \right) \right. \\ &\quad \times \left. \left(\frac{(6.626 \times 10^{-34}\text{ Js})^2}{2\pi(1.38 \times 10^{-23}\text{ JK}^{-1})(300\text{ K})} \right) \right]^{3/2} \\ &= \left[(3.33 \times 10^{26})(1.688 \times 10^{-47}) \right]^{3/2} \\ &= 4.21 \times 10^{-31}\text{ m}^3. \end{aligned}$$

Substituting these terms into Equation (19.24) gives

$$\begin{aligned}\text{rate coefficient} &= \left(\frac{(8.314 \text{ JK}^{-1})(300 \text{ K})}{6.626 \times 10^{-34} \text{ Js}} \right) (53.67)(1.38) \\ &\quad \times (4.21 \times 10^{-31} \text{ m}^3) e^{-6570/(8.314T)} \\ &= 1.17 \times 10^8 \times e^{-790/T} \text{ m}^3 \text{ s}^{-1} \text{ mol}^{-1}\end{aligned}$$

The experimental value is $2.0 \times 10^8 \times e^{-800/T} \text{ m}^3 \text{ s}^{-1} \text{ mol}^{-1}$ [2]. The 1.6-fold discrepancy at $T = 300 \text{ K}$ is probably due to the limitations of the transition-state theory: the assumed separability of the reaction coordinate from other degrees of freedom, and the assumption that each trajectory crosses the transition state only once, $\kappa = 1$.

Transition-state theory gives a microscopic basis for the Arrhenius model. To see this, let's convert to a thermodynamic notation.

The Thermodynamics of the Activated State

We are treating $\overline{K^\ddagger}$ in Equation (19.23) as an equilibrium constant for the stable degrees of freedom (the overbar indicates that the unstable reaction coordinate degree of freedom ξ has been factored out). So you use $\overline{K^\ddagger}$ as you would any other equilibrium constant, and express it in terms of thermodynamic quantities, which are called the **activation free energy** ΔG^\ddagger , **activation enthalpy** ΔH^\ddagger , and **activation entropy** ΔS^\ddagger :

$$-kT \ln \overline{K^\ddagger} = \Delta G^\ddagger = \Delta H^\ddagger - T\Delta S^\ddagger. \quad (19.25)$$

Substituting Equation (19.25) into Equation (19.23) gives

$$k_2 = \left(\frac{kT}{h} \right) e^{-\Delta G^\ddagger/kT} = \left(\frac{kT}{h} \right) e^{-\Delta H^\ddagger/kT} e^{\Delta S^\ddagger/k}. \quad (19.26)$$

Measurements of the temperature dependence of a chemical reaction rate can be fitted to Equation (19.26) to give the two parameters ΔH^\ddagger and ΔS^\ddagger . The quantity ΔH^\ddagger is related to the activation energy E_a of Arrhenius. The factor $(kT/h) \exp(\Delta S^\ddagger/k)$ gives the front factor A in the Arrhenius expression. For any reaction that is activated (i.e., has a barrier $\Delta G^\ddagger > 0$), Equation (19.26) says that the fastest possible rate, which is achieved as $\Delta G^\ddagger \rightarrow 0$, is $kT/h = (1.38 \times 10^{-23} \text{ J K}^{-1})(300 \text{ K}) / (6.626 \times 10^{-34} \text{ Js}) = 6.24 \times 10^{12} \text{ s}^{-1}$ or about one conversion per 0.16 ps.

Here's a way to think about activated processes. The rate k_2 is a product of two factors. k^\ddagger defines the 'speed limit,' the maximum speed the reaction can go. However, the reaction will go more slowly than this because of a bottleneck that restricts how many molecules are poised or 'activated' to proceed at that rate. Only a small fraction of reactant molecules, $K^\ddagger = [(AB)^\ddagger]/[A][B] \ll 1$, reach the bottleneck at any given time. The logarithm of this fraction can be expressed as the free energy barrier, $-RT \ln K^\ddagger = \Delta G^\ddagger > 1$. When the neck of an upturned bottle is small, the flow of water is restricted by how fast the water reaches the bottleneck. For chemical reactions, an enthalpic bottleneck can arise when a system must first break a bond before making a bond, for

example. An entropic bottleneck can arise when a molecule cannot reach the product state without first achieving a restricted orientation, configuration, or location in space, for example.

We have focused here on reactions involving multiple reactants, such as atoms *A* and *B*. Unimolecular reactions involving only a single reactant atom are somewhat different because there is no vibration or bond formation that defines the reaction coordinate. Such reactions are discussed in [6].

Figure 19.9 shows the rates for two processes: (1) the migration of individual carbon monoxide (CO) molecules on the surface of metallic copper and (2) the migration of CO dimers on the same surface. Figure 19.10 shows a model of this activated kinetics as a landscape of periodic energy barriers that must be surmounted by the CO molecules as they move across the surface. ΔH^\ddagger can be determined from the slope and ΔS^\ddagger can also be found from Figure 19.9. Figure 19.9 shows that the activation enthalpy barrier is about the same for monomers and dimers. The dimers migrate faster for entropic reasons. The dimers gain orientational freedom at the point where they just begin to detach from the surface, at the transition state. The monomers don't, so the migration rate is greater for the dimers.

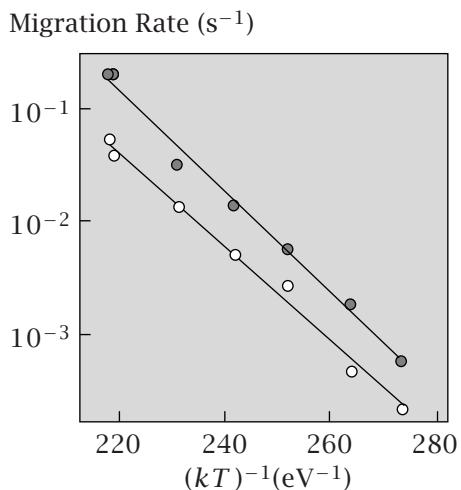


Figure 19.9 The migration rate of CO monomers (○) and dimers (●) on copper surfaces shows an Arrhenius temperature dependence. Source: BG Briner, MD Doering, HP Rust, and AM Bradshaw, *Science* **278**, 257–260 (1997).

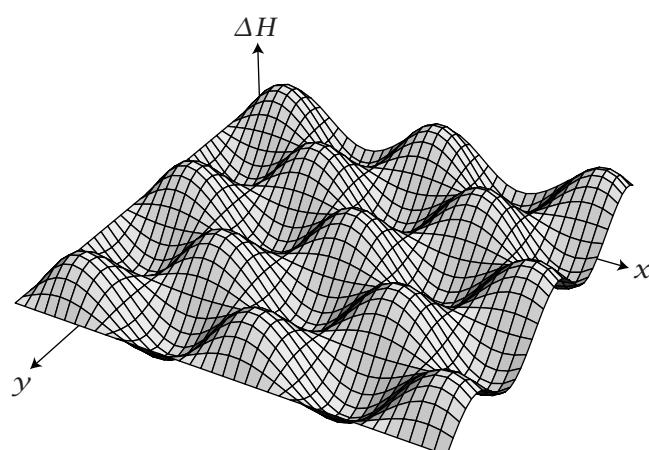


Figure 19.10 Atoms and molecules can hop over energy or enthalpy barriers to migrate laterally on solid surfaces. Figure 19.9 shows data that can be interpreted in this way.

In the next section, we show how the transition-state theory explains *isotope effects*, which are used experimentally for determining reaction mechanisms.

The Primary Kinetic Isotope Effect

Isotope substitution is useful for studying chemical reaction mechanisms. For our purposes, an isotope is an atom that is identical to another except for its mass. In general, chemical reaction rates do not change much if one isotope is substituted for another. However, if the isotope substitution occurs at a reacting position in the molecule, it can change the reaction kinetics. For example, at room temperature, a C–H bond cleaves about 8 times faster than a C–D bond, and about 60 times faster than a carbon–tritium bond. A ^{12}C –H bond cleaves about 1.5 times faster than a ^{14}C –H bond. Isotope effects are called *primary* when the isotope is substituted at a reacting position, and *secondary* when it is substituted elsewhere in the molecule.

Here is the basis for kinetic isotope effects. Compare the rate k_{H} for breaking a carbon–hydrogen bond,



with the rate k_{D} for breaking a carbon–deuterium bond,



From Equation (19.23), the ratio of the rates for breaking these two types of bond is

$$\frac{k_{\text{H}}}{k_{\text{D}}} = \frac{\left(\frac{q_{(\text{CH})^{\ddagger}}}{q_{\text{CH}}} \right) e^{\Delta D_{\text{CH}}^{\ddagger}/kT}}{\left(\frac{q_{(\text{CD})^{\ddagger}}}{q_{\text{CD}}} \right) e^{\Delta D_{\text{CD}}^{\ddagger}/kT}} \approx e^{(\Delta D_{\text{CH}}^{\ddagger} - \Delta D_{\text{CD}}^{\ddagger})/kT}, \quad (19.29)$$

where q_{CH} and q_{CD} are the partition functions for the respective molecules before bond breakage, and $q_{(\text{CH})^{\ddagger}}$ and $q_{(\text{CD})^{\ddagger}}$ are the partition functions for the respective transition states. Changing the mass mainly affects the ground-state vibrational energy of the bond, and thus the energy difference $\Delta D_{\text{CH}}^{\ddagger} - \Delta D_{\text{CD}}^{\ddagger}$, giving the expression on the right in Equation (19.29).

The transition state for this reaction is the point at which the bond breaks and the atoms dissociate, so its energy D^{\ddagger} is assumed to be the same for both isotopes. Only the zero-point vibrational energy of the *reactant* bond is affected by the isotope substitution (see Figure 19.11), so

$$\Delta D_{\text{CH}}^{\ddagger} - \Delta D_{\text{CD}}^{\ddagger} = -\frac{1}{2}h(\nu_{\text{CD}} - \nu_{\text{CH}}). \quad (19.30)$$

The vibrational frequencies ν_{CD} and ν_{CH} depend on the reduced masses μ_{CH} and μ_{CD} :

$$\begin{aligned} \mu_{\text{CH}} &= \frac{m_{\text{C}}m_{\text{H}}}{m_{\text{C}}+m_{\text{H}}} \approx m_{\text{H}}, \\ \mu_{\text{CD}} &= \frac{m_{\text{C}}m_{\text{D}}}{m_{\text{C}}+m_{\text{D}}} \approx m_{\text{D}} = 2m_{\text{H}} = 2\mu_{\text{CH}}, \end{aligned} \quad (19.31)$$

where m_{C} , m_{D} , and m_{H} are the masses of carbon, deuterium, and hydrogen, respectively. If the spring constant is k_s , the vibrational frequencies are (see

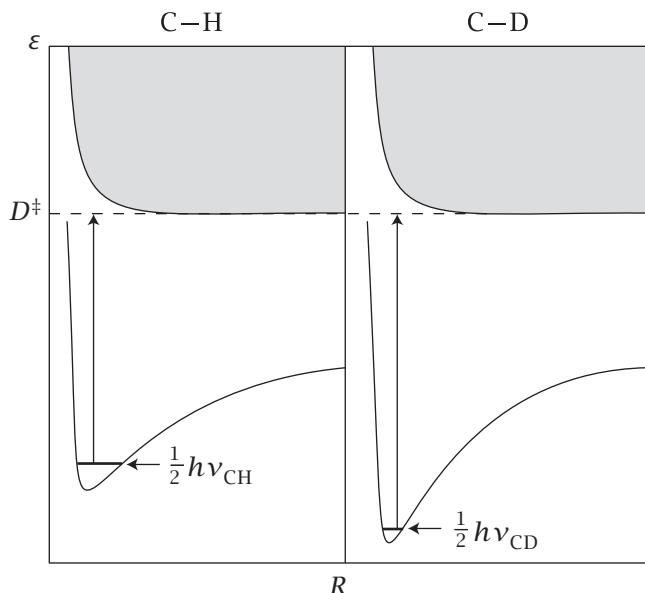


Figure 19.11 Interpretation of primary kinetic isotope effects. The bond energy ε is shown as a function of R , the bond length. Most molecules are in their lowest vibrational energy state at ordinary temperatures. This state is lower by about 4.8 kJ mol^{-1} in a C–D bond than in a C–H bond. So the C–D bond requires about 4.8 kJ mol^{-1} more energy to reach the same transition state. Source: A Cornish-Bowden, *Fundamentals of Enzyme Kinetics*, Portland Press, London, 1995.

Equation (11.23)):

$$\nu_{\text{CD}} = \frac{1}{2\pi} \left(\frac{k_s}{\mu_{\text{CD}}} \right)^{1/2} = \frac{1}{2\pi} \left(\frac{k_s}{2\mu_{\text{CH}}} \right)^{1/2} = \frac{\nu_{\text{CH}}}{\sqrt{2}}$$

$$\Rightarrow \nu_{\text{CD}} - \nu_{\text{CH}} = \left(\frac{1}{\sqrt{2}} - 1 \right) \nu_{\text{CH}}. \quad (19.32)$$

Combine Equations (19.29), (19.30), and (19.32) to get

$$\frac{k_{\text{H}}}{k_{\text{D}}} = \exp \left[\frac{-h\nu_{\text{CH}}}{2kT} \left(\frac{1}{\sqrt{2}} - 1 \right) \right]. \quad (19.33)$$

EXAMPLE 19.3 C–H bonds break about 8 times faster than C–D bonds. Compute the kinetic isotope rate effect at $T = 300 \text{ K}$ if the C–H stretch mode observed in the infrared is 2900 cm^{-1} . Use Equation (19.33):

$$\begin{aligned} \frac{k_{\text{H}}}{k_{\text{D}}} &= \exp \left[\frac{(-6.626 \times 10^{-34} \text{ Js})(2900 \text{ cm}^{-1})(3.0 \times 10^{10} \text{ cm s}^{-1})}{2(1.38 \times 10^{-23} \text{ J K}^{-1})(300 \text{ K})} \left(\frac{1}{\sqrt{2}} - 1 \right) \right] \\ &= 7.68 \end{aligned}$$

at $T = 300 \text{ K}$. In general, bonds of lighter isotopes cleave more rapidly than those of heavier isotopes, because the lighter isotopes have higher zero-point energies.

In the following section, we consider how *catalysts* can speed up chemical reactions by reducing the transition-state free energy.

Catalysts Speed Up Chemical Reactions

Catalysts speed up the rates of chemical reactions. For example, biochemical reactions are accelerated by enzyme catalysts. Enzymes can achieve remarkably large accelerations—by a factor of 2×10^{23} for orotidine 5'-phosphate decarboxylase, for example [7]. Metals are also frequently used for catalysis. Petroleum is converted to gasoline and natural gas by using platinum–nickel catalysts. An important industrial method for manufacturing ammonia from hydrogen and nitrogen gases ($3\text{H}_2 + \text{N}_2 \rightarrow 2\text{NH}_3$) is the Haber–Bosch process, for which Nobel Prizes were awarded to Fritz Haber in 1918 and Carl Bosch in 1931. Ammonia was difficult to produce and was needed to make fertilizer to sustain a large fraction of the Earth's population. Even though nitrogen is abundant in the air, it doesn't readily combine with hydrogen. If you put nitrogen and hydrogen gases together in a container, they won't react, even though ammonia is the more stable state. But add a little iron dust, and you'll get ammonia. Iron oxide or chromium oxide are catalysts that speed up the reaction. A catalyst affects only the rate of a reaction, and not the equilibrium ratio of products to reactants.

American chemist L Pauling proposed in 1946 that catalysts work by *stabilizing the transition state* [8, 9] (see Figure 19.12). Suppose that an uncatalyzed reaction has a forward rate coefficient k_0 :



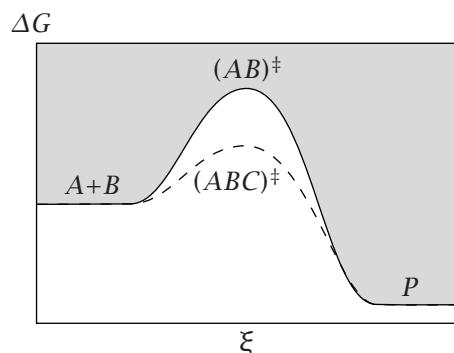
k_0 is given by transition-state theory, Equation (19.23), as the product of a rate factor kT/h and an equilibrium factor \overline{K}_0^\ddagger :

$$k_0 = \left(\frac{kT}{h}\right) \overline{K}_0^\ddagger = \left(\frac{kT}{h}\right) \frac{[\overline{AB}^\ddagger]}{[A][B]}. \quad (19.35)$$

\overline{K}_0^\ddagger is the equilibrium ratio of the concentration of the transition state $[\overline{AB}^\ddagger]$ to the product of the concentrations of the reactants (see Equation (19.14)). The overbar indicates that the unstable degree of freedom has been factored out into the term kT/h . The corresponding process for the same reaction when a catalyst C is involved is shown by the bottom arrows in Figure 19.13. For the catalyzed reaction, the rate is $k_c = (kT/h)\overline{K}_c^\ddagger$, where

$$\overline{K}_c^\ddagger = \frac{[\overline{ABC}^\ddagger]}{[A][B][C]}. \quad (19.36)$$

Figure 19.12 The free energy barrier ΔG^\ddagger is reduced by a catalyst. A and B are reactants, P is product, and C is the catalyst. $(AB)^\ddagger$ is the transition state for the uncatalyzed reaction, and $(ABC)^\ddagger$ is the transition state for the catalyzed reaction.



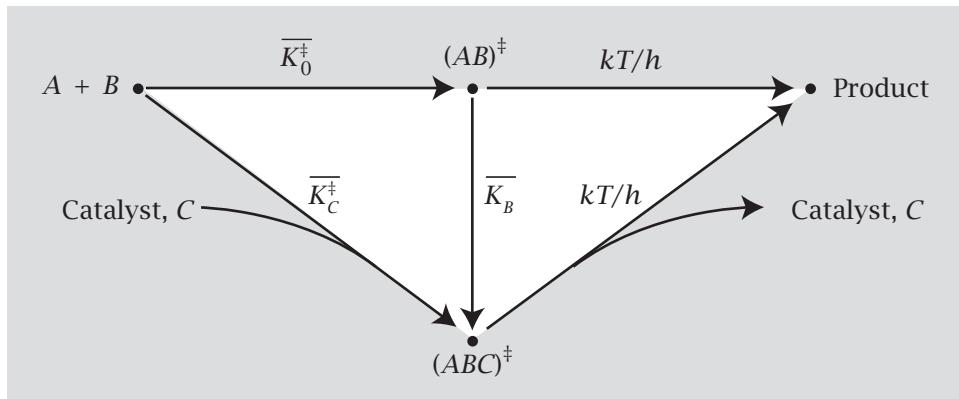


Figure 19.13 Across the top, the uncatalyzed reaction goes from reactant to transition state (\cdot) ‡ , with equilibrium constant $\overline{K}_0^{\ddagger}$, then to product, with rate kT/h . The bottom arrows show the catalyzed reaction. Reactant and catalyst form a transition state with equilibrium constant $\overline{K}_c^{\ddagger}$, and go on to product with rate kT/h . \overline{K}_B is the “binding constant” of the catalyst to the transition state.

The rate enhancement due to the catalyst is given by the ratio

$$\frac{k_c}{k_0} = \frac{(kT/h)\overline{K}_c^{\ddagger}}{(kT/h)\overline{K}_0^{\ddagger}} = \frac{\overline{K}_c^{\ddagger}}{\overline{K}_0^{\ddagger}} = \overline{K}_B = \frac{[(\overline{ABC})^{\ddagger}]}{[(\overline{AB})^{\ddagger}][C]}, \quad (19.37)$$

where \overline{K}_B represents the ‘binding constant’ of the catalyst to the transition state (see Figure 19.13). Pauling’s principle, expressed in Equation (19.37), is that the rate enhancement by the catalyst is proportional to the binding affinity of the catalyst for the transition state. This has two important implications. First, to accelerate a reaction, Pauling’s principle says you should design a catalyst that binds tightly to the transition state (and not to the reactants or products). The second is that a catalyst that reduces the transition-state free energy for the forward reaction is also a catalyst for the backward reaction (see Figure 19.12). Reducing the transition-state barrier is not the only mechanism by which catalysts accelerate reactions. In other cases, catalysts change the reaction path or raise the energy well of the reactants, called *ground-state destabilization*.

Speeding Up Reactions by Intramolecular Localization or Solvent Preorganization

What are the physical mechanisms of catalysis? One possibility is that the slow step in the reaction of A with B is their diffusion together through the solvent into the appropriate mutual orientation to react. In that case, an appropriate catalyst would be a molecule or surface that provides a common binding site for bringing reactants A and B together in an appropriate orientation. To test the importance of this kind of spatial and orientational localization, experiments have been performed in which reactants A and B are contained within a single molecule, sometimes in a particular orientation. The reaction rates of A with B are then compared with reference reactions in which A and B are not covalently linked. Table 19.1 compares six reaction rates of a given pair of reactants, A and

Table 19.1 The relative reaction rates k_{rel} for two reactants; A is COO^- and B is $\text{COOC}_6\text{H}_4\text{Br}$, either as isolated molecules ($k_{\text{rel}} = 1$) or covalently linked together in various configurations in different molecules.

Reactants	k_{rel}	Reactants	k_{rel}
$\begin{array}{c} + \\ \text{CH}_3\text{COOC}_6\text{H}_4\text{Br} \\ \text{CH}_3\text{COO}^- \end{array}$	1.0	$\begin{array}{c} \square \quad \text{COOC}_6\text{H}_4\text{Br} \\ \square \quad \text{COO}^- \end{array}$	$2.3 \times 10^5 \text{ M}$
$\begin{array}{c} \text{COOC}_6\text{H}_4\text{Br} \\ \text{Cyclohexane ring} \\ \text{COO}^- \end{array}$	$1 \times 10^3 \text{ M}$	$\begin{array}{c} \text{Ph} \diagup \quad \text{COOC}_6\text{H}_4\text{Br} \\ \text{Ph} \diagdown \quad \text{COO}^- \end{array}$	$2.7 \times 10^5 \text{ M}$
$\begin{array}{c} \text{COOC}_6\text{H}_4\text{Br} \\ \text{Bicyclic ring} \\ \text{COO}^- \end{array}$	$3.6 \times 10^3 \text{ M}$	$\begin{array}{c} \text{i-Pr} \diagup \quad \text{COOC}_6\text{H}_4\text{Br} \\ \text{i-Pr} \diagdown \quad \text{COO}^- \end{array}$	$1.3 \times 10^6 \text{ M}$

Source: FC Lightstone and TC Bruice, *J Am Chem Soc* **118**, 2595–2605 (1996).

B: separated in solution versus tethered at two ends of various linker chains, some longer and some shorter, some with more steric restriction and some with less. It shows that intramolecular localization of the reactants A and B, and their proper relative spatial positioning, can lead to large rate enhancements. Such localization can involve both entropic and enthalpic factors.

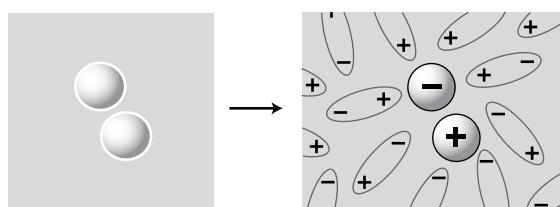
Another mechanism has been proposed for reducing activation barriers in enzyme-catalyzed reactions [10–12]. The enzyme may act as a *preorganized ‘solvent’ cavity*. Table 19.2 shows that some reactions can be extremely slow, particularly in polar solvents. In such cases, the reactants A and B may be neutral but the transition state may involve charge separation, A^+B^- . A big energy

Table 19.2 Reaction rate coefficients for various solvents (top, least polar; bottom, most polar).

$\text{Cl}^- + \text{CH}_3\text{Br} \rightarrow \text{ClCH}_3 + \text{Br}^-$	
Solvent	Relative Rate ($\text{cm}^3 \text{molecule}^{-1} \text{s}^{-1}$)
None (gas phase)	1
$(\text{CH}_3)_2\text{CO}$	10^{-10}
Dimethylformamide	10^{-11}
CH_3OH	10^{-15}
H_2O	10^{-16}

Source: WR Cannon and SJ Benkovic, *J Biol Chem* **273**, 26257–27260 (1998).

(a) Reactants Polarize, so Water Reorganizes



(b) Enzyme Binds Ions in Preorganized Pocket

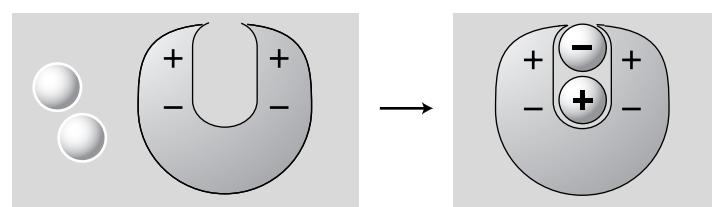


Figure 19.14 (a) Two neutral reactants become charged in the transition state. Creating this charge separation costs free energy because it orients the solvent dipoles. (b) Enzymes can reduce the activation barrier by having a site with preorganized dipoles [10, 11].

cost in creating this charge separation involves orienting the surrounding polar solvent molecules. That is, water molecules near the transition state molecule reorient so that the partial negative charges on their oxygen atoms will associate with A^+ and the partial positive charges on their hydrogen atoms will associate with B^- . Figure 19.14 shows the idea. A way to reduce the activation energy is to put the reactants into a cavity that is appropriately polarized in advance. An enzyme may be able to provide an environment of this type to stabilize the charge distribution on the transition state.

The Brønsted Law of Acid and Base Catalysis: The Stronger the Acid, the Faster the Reaction it Catalyzes

Some chemical reactions are catalyzed by acids or bases. For example, a reactant R converts to a product P , assisted by a coupling to an acid or base equilibrium, according to the reaction



AH is an acid catalyst having an acid dissociation constant

$$K_a = \frac{[H^+][A^-]}{[AH]}$$

Under some conditions, the rate of disappearance of the reactant is

$$\frac{d[R]}{dt} = -k_a[AH][R],$$

where k_a is the rate coefficient for the reaction.

The **Brønsted law** is the observation that the stronger the acid (K_a large), the faster the reaction $R \rightarrow P$ that it catalyzes. That is,

$$\log k_a = \alpha \log K_a + c_a, \quad (19.39)$$

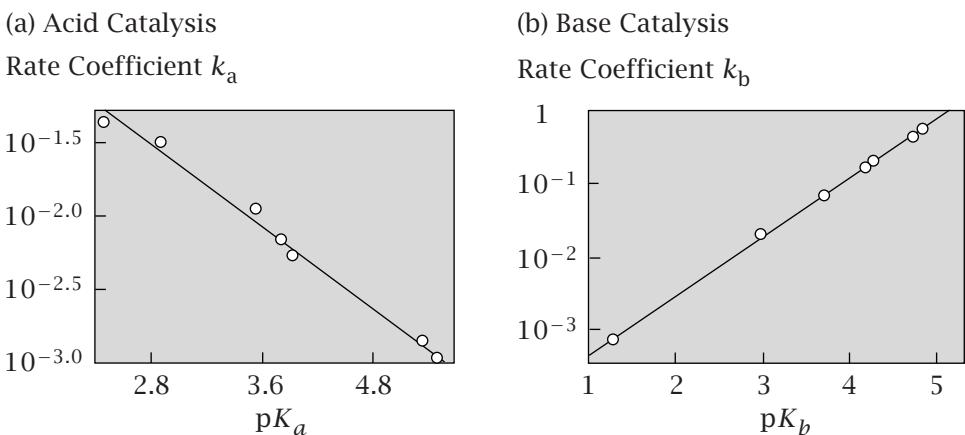


Figure 19.15 Brønsted plots. (a) Acid catalyzed hydrolysis of ethyl vinyl ether in H_2O at 25°C . (b) Base catalysis of the decomposition of nitramide in H_2O at 15°C . Data are from AJ Kresge, HL Chen, Y Chiang, et al., *J Am Chem Soc* **93**, 413 (1971).

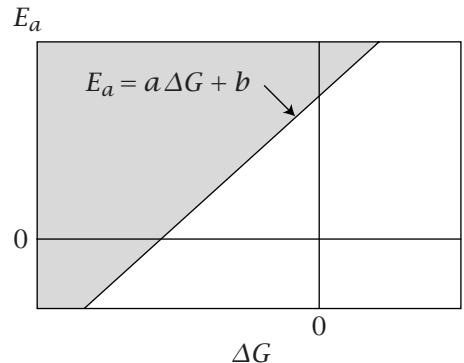
where c_a is a constant and $\alpha > 0$. Using the notation $\text{p}K_a = -\log K_a$, the Brønsted law is expressed as $\log k_a = -\alpha \text{p}K_a + c_a$. The acid dissociation equilibrium matters because the rate-limiting step in the reaction $R \rightarrow P$ is often the removal of a proton from the catalyst AH . Base catalysis often follows a similar linear relationship, but with a slope of opposite sign: $\log k_b = \beta \text{p}K_b + c_b$, where K_b is the base dissociation constant, $\beta > 0$, and c_b is a constant. Examples of acid and base catalysis are given in Figure 19.15. Such observations imply a linear relationship between the activation energy E_a for the reaction and the free energy ΔG of acid dissociation:

$$E_a = a \Delta G + b, \quad (19.40)$$

where a and b can be positive or negative constants. Equation (19.40) is the **Brønsted relationship**. It is an example of a broader class of *linear free energy relationships* in which the logarithm of a rate coefficient is proportional to the logarithm of an equilibrium constant, as in Equation (19.39) (see Figure 19.16).

Equation (19.38) describes two coupled processes in which the equilibrium of one ($\text{AH} \rightarrow \text{H}^+ + \text{A}^-$) is coupled to the kinetics of the other ($R \rightarrow P$). In other

Figure 19.16 According to the Brønsted relation, the activation energy E_a of a process is a linear function of ΔG , the stability of the product relative to the reactant. As ΔG becomes more negative, E_a is reduced (see Figure 19.17).



Energy

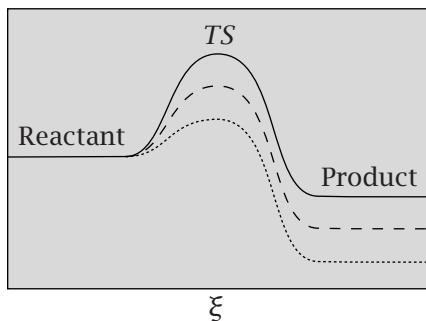


Figure 19.17 The Brønsted relationship shown schematically. The more stable is the product, the lower is the barrier, so the forward reaction is faster.

cases, the coupling of equilibria and kinetics can occur within the same reaction, such as



Increasing the strength of the acid in this case (increasing K_a) increases the rate of the reaction.

What is the basis for the Brønsted law? Figure 19.17 shows the reaction coordinate diagrams for a series of related reactions, indicating that the more stable the product, the faster the reaction. For acid catalysis, the figure implies that the more weakly bonded the proton is to the catalyst, the faster it comes off, and the faster is the catalyzed reaction. Here is a model.

A First Approximation: The Evans–Polanyi Model of the Brønsted Law and Other Linear Free Energy Relationships

In 1936, MG Evans and M Polanyi developed a simple model to explain Equation (19.40) [1]. Consider the reaction



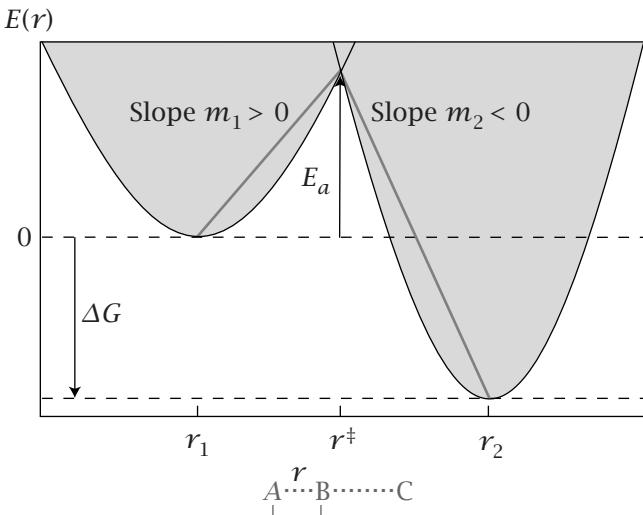
This reaction can be described in terms of two degrees of freedom: r_{AB} , the bond length between A and B , and r_{BC} , the bond length between B and C . To simplify, Evans and Polanyi considered the distance between A and C to be fixed, so r_{BC} decreases as r_{AB} increases. Now the problem can be formulated in terms of a single reaction coordinate $r = r_{AB} = \text{constant} - r_{BC}$. This is usually a good approximation for the range of bond lengths in which one bond is being made while the other is being broken.

Figure 19.18 shows how the bond energy $E(r)$ depends on bond length r . The left parabola in the figure gives the energy $E_{AB}(r)$ of the $A-B$ bond as a function of its length r . The right parabola gives the energy $E_{BC}(r)$ of the $B-C$ bond. The right parabola as shown on this diagram is ‘backwards’ in the sense that increasing r (from r^\ddagger to r_2) represents a decreasing $B-C$ bond length.

As the $A-B$ bond stretches, the energy increases up to the transition state, where the bond breaks; then the energy decreases as the $B-C$ bond forms. Evans and Polanyi approximated the two energy functions between reactants, transition state, and products by two straight lines that intersect at the transition state. For the AB molecule, the energy is

$$E_{AB}(r) = m_1(r - r_1), \quad (19.43)$$

Figure 19.18 The Evans–Polanyi model for linear free-energy relationships. ΔG is the stability of the product relative to the reactant, E_a is the activation energy, r_1 and r_2 are the stable bond lengths of AB and BC , and r^\ddagger is the A – B bond length in the activated state.



where m_1 is the slope of the straight line from reactants to transition state and r_1 is the equilibrium length of the A – B bond. At the transition state, $E_a = m_1(r^\ddagger - r_1)$, so

$$r^\ddagger = \frac{E_a}{m_1} + r_1. \quad (19.44)$$

Similarly, for the BC molecule,

$$E_{BC}(r) = m_2(r - r^\ddagger) + E_a \implies \Delta G = m_2(r_2 - r^\ddagger) + E_a, \quad (19.45)$$

where $\Delta G = G_{\text{products}} - G_{\text{reactants}}$ is the equilibrium difference in free energy (see Figure 19.18), and the slope of the straight line from transition state to product is negative, $m_2 < 0$. To express the activation energy E_a in terms of the free energy difference ΔG , substitute Equation (19.44) into Equation (19.45) to get

$$\Delta G = m_2 \left[r_2 - \left(\frac{E_a}{m_1} + r_1 \right) \right] + E_a. \quad (19.46)$$

Rearranging Equation (19.46) gives a linear relationship between E_a and ΔG :

$$E_a = \frac{m_1}{m_1 - m_2} [\Delta G - m_2(r_2 - r_1)]. \quad (19.47)$$

This simple model shows how linear free energy relationships can arise from reactant and product energy surfaces with a transition state between them that is defined by the *curve-crossings*. Figure 19.19 shows how shifting the equilibrium to stabilize the products can speed up the reaction. It also illustrates that such stabilization can shift the transition state to the left along the reaction coordinate, to earlier in the reaction. If $|m_1| \gg |m_2|$, the transition state will be closer to the reactants than to the products, and if $|m_1| \ll |m_2|$, the reverse will hold. The Evans–Polanyi model rationalizes why stabilities should correlate linearly with rates.

Activation energies are not always linearly proportional to product stabilities. Some experiments show an *inverted region*, where further increasing the product stability can *slow down* the reaction. This was explained by R Marcus, who was awarded the 1992 Nobel Prize in Chemistry. Marcus theory replaces Equation (19.43) with a parabola, $E_{AB}(r) = m_1(r - r_1)^2 + b$, and a

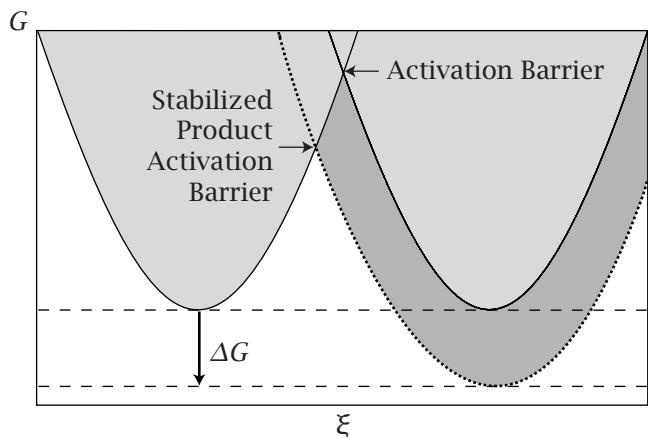


Figure 19.19 When the product is stabilized, the right-hand parabola shifts from (—) to (····). The activation barrier is also lowered and shifted to the left along the reaction coordinate ξ .

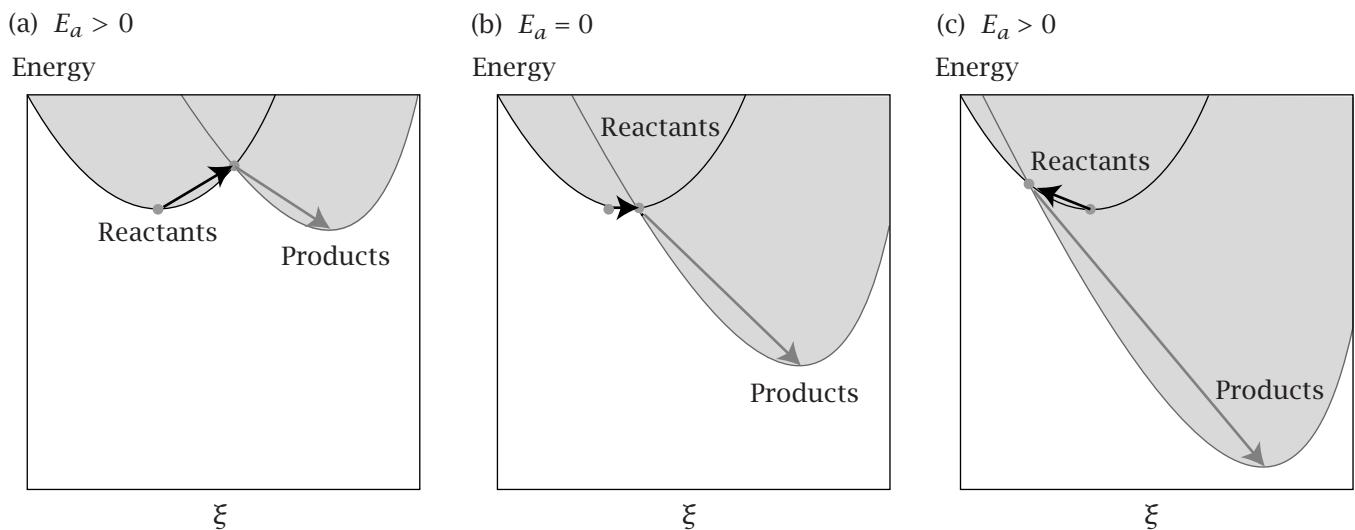


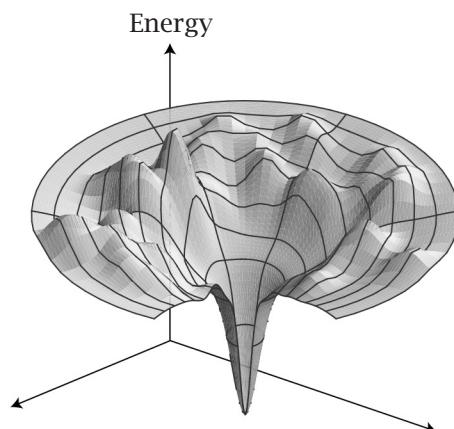
Figure 19.20 Reactant and product curves against the reaction coordinate ξ , as in Figure 19.19. Increasing stability of products, from (a) to (c), leads first to a reduction of the activation barrier to $E_a = 0$ in (b), then to an increase.

similar parabola replaces Equation (19.45). The result is a square-law dependence $E_a \sim (\text{constant} + \Delta G)^2$ between activation energy and product stability instead of Equation (19.47) (see [13]). Figure 19.20 indicates how this leads to the inverted behavior.

Funnel Landscapes Describe Diffusion and Polymer Folding Processes

All the processes described in this chapter so far involve well-defined reactants and products, and a well-defined reaction coordinate. But diffusional processes and polymer conformational changes often cannot be described in this way. Consider a solution of mobile molecules diffusing toward a sphere, or an ensemble of polymer conformations put into folding conditions (conditions that strongly favor some compact polymer conformations). Now the starting

Figure 19.21 A bumpy energy landscape, such as occurs in some diffusion processes, polymer conformational changes, and biomolecule folding. A single minimum in the center may represent “product,” but there can be many different “reactants,” such as the many open configurations of a denatured protein. Source: KA Dill and HS Chan, *Nat Struct Biol* **4**, 10–19 (1997).



state is not a single point on an energy landscape. It is a broad distribution—the many different spatial locations of the starting positions of the mobile molecules, or the many different open conformations of the polymer chain.

Figure 19.21 shows an example of such an energy landscape. The observed kinetics may often still be describable in terms of an activation barrier. But in these cases, there may be no single structure or single mountain-pass enthalpy barrier that defines the kinetic bottleneck. Instead, the activation barrier may be entropic. The rate limit is just the diffusional meanderings from the different starting points as each molecule progresses toward the stable state.

Summary

Chemical reactions can speed up substantially with temperature. This can be explained in terms of a transition state or activation barrier and an equilibrium between reactants and a transient and unstable species, called the transition state. For chemical reactions, the transition state involves an unstable weak vibration along the reaction coordinate, and an equilibrium between all other degrees of freedom. You can predict kinetics by using microscopic structures and the equilibrium partition functions of Chapter 13 if the transition state structure is known. Catalysts can act by binding to the transition-state structure. They can speed reactions by forcing the reactants into transition-state-like configurations.

Problems

1. Isotope substitution experiments can detect hydrogen tunneling. Isotope substitution experiments can sometimes be used to determine whether hydrogens are cleared from molecules through mechanisms that involve tunneling. To test this, two isotopes are substituted for hydrogen: (1) deuterium (D, mass = 2) is substituted and the ratio of rate coefficients k_H/k_D is measured, and (2) tritium (T, mass = 3) is substituted and k_H/k_T is measured.

- (a) Using the isotope substitution model in this chapter, show that

$$\frac{k_H}{k_T} = \left(\frac{k_H}{k_D} \right)^\alpha.$$

- (b) Compute the numerical value of α .

2. Relating stability to activation barriers. Using the Evans-Polanyi model, with $r_1 = 5$, $r_2 = 15$, $m_1 = 1$, and $m_2 = -2$:

- (a) Compute the activation barriers E_a for three systems having product stabilities $\Delta G = -2 \text{ kcal mol}^{-1}$, -5 kcal mol^{-1} , and -8 kcal mol^{-1} .
 (b) Plot the three points E_a versus ΔG , to show how the activation barrier is related to stability.

3. Reduced masses. Equation (19.31) gives the reduced masses for C–H and C–D bonds as $\mu_{\text{CH}} \approx m_{\text{H}}$ and $\mu_{\text{CD}} \approx 2m_{\text{H}}$. The approximation is based on assuming the mass of carbon is much greater than of H or D. Give the more correct values of these reduced masses if you don't make this assumption.

4. Classical collision theory. According to the kinetic theory of gases, the reaction rate k_2 of a sphere of radius r_A with another sphere of radius r_B is

$$k_2 = \pi R^2 \left(\frac{8kT}{\pi\mu_{AB}} \right)^{1/2} e^{-\Delta\varepsilon_0^\ddagger/kT},$$

where $R = r_A + r_B$ is the distance of closest approach, μ_{AB} is the reduced mass of the two spheres, and $\Delta\varepsilon_0^\ddagger$ is the activation energy. Derive this from transition state theory.

5. The pressure dependence of rate constants.

- (a) Show that the pressure dependence of the rate constant k for the reaction



is proportional to an *activation volume* v^\ddagger :

$$\left(\frac{\partial \ln k_f}{\partial p} \right) = -\frac{(v^\ddagger - v_A)}{kT}.$$

- (b) Show that the expression in (a) is consistent with Equation (13.46), $K = k_f/k_r$, where k_r is the rate of the reverse reaction.

6. Relating the Arrhenius and activated-state parameters. Derive the relationship of the activation parameter ΔH^\ddagger in Equation (19.26) to the Arrhenius activation energy E_a in Equation (19.11) for a gas-phase reaction.

7. Enzymes accelerate chemical reactions. Figure 19.22 shows an Arrhenius plot for the uncatalyzed reaction of 1-methylorotic acid (OMP) [7].

- (a) Estimate ΔH^\ddagger from the figure.
 (b) Estimate ΔS^\ddagger at $T = 300 \text{ K}$.
 (c) At $T = 25^\circ\text{C}$, the enzyme OMP decarboxylase accelerates this reaction 1.4×10^{17} -fold. How fast is the catalyzed reaction at 25°C ?
 (d) What is the binding constant of the enzyme to the transition state of the reaction at $T = 300 \text{ K}$?

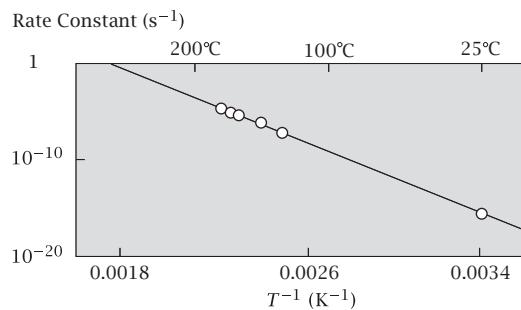


Figure 19.22 Source: A Radzicka and R Wolfenden, *Science* **267**, 90–93 (1995).

8. Rate increase with temperature. A rule of thumb used to be that chemical reaction rates would roughly double for a 10-degree increase in temperature, say from $T_1 = 300 \text{ K}$ to $T_2 = 310 \text{ K}$. For what activation energy E_a would this be exactly correct?

9. Catalytic rate enhancement. The reaction rate of an uncatalyzed reaction is $k_0 = 10^3 \text{ M}^{-1} \text{ s}^{-1}$ at $T = 300 \text{ K}$. A catalyst C binds to the transition state with free energy $\Delta G = -5 \text{ kcal mol}^{-1}$. What is the rate k_0 of the catalyzed reaction at $T = 300 \text{ K}$?

10. Negative activation energies. Activation energy barriers often indicate a process that is rate-limited by the need to break a critical bond, or achieve some particular strained state. In contrast, some processes involve *negative* activation energies—their rates decrease with increasing temperature (see Figures 19.23 and 19.24). Table 19.3 shows some Arrhenius parameters for the formation of duplex nucleic acids from single strands of various nucleotide sequences. Figure 19.24 shows an example of the kind of data from which the parameters in the table were derived. Which molecules have negative activation energies?

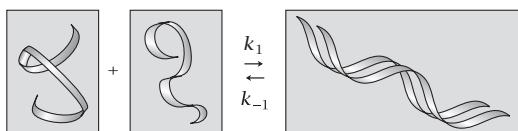


Figure 19.23

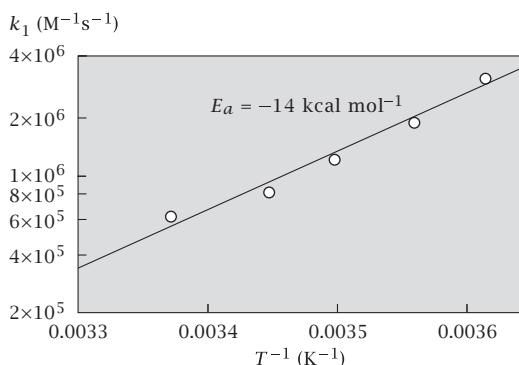


Figure 19.24 Source: CR Cantor and PR Schimmel, *Biophysical Chemistry*, Part III, WH Freeman, San Francisco, 1980.

Table 19.3 Relaxation kinetics of oligonucleotides (21°C to 23°C).

Ordered Form	k_1 ($M^{-1}s^{-1}$)	E_a ($kcal\ mol^{-1}$)
A ₉ · U ₉	5.3×10^5	-8
A ₁₀ · U ₁₀	6.2×10^5	-14
A ₁₁ · U ₁₁	5.0×10^5	-12
A ₁₄ · U ₁₄	7.2×10^5	-17.5
A ₄ U ₄ · A ₄ U ₄	1.0×10^6	-6
A ₅ U ₅ · A ₅ U ₅	1.8×10^6	-4
A ₆ U ₆ · A ₆ U ₆	1.5×10^6	-3
A ₇ U ₇ · A ₇ U ₇	8.0×10^5	+5
A ₂ GCU ₂ · A ₂ GCU ₂	1.6×10^6	+3
A ₃ GCU ₃ · A ₃ GCU ₃	7.5×10^5	+7
A ₄ GCU ₄ · A ₄ GCU ₄	1.3×10^5	+8

Source: CR Cantor and PR Schimmel, *Biophysical Chemistry*, Part III, WH Freeman, San Francisco, 1980.

11. Drug dissociation from DNA. Drug A is an uncharged molecule containing a three-ring aromatic chromophore linked to three amino acids: valine, N-methylvaline and proline. This drug binds to double-stranded DNA, with the planar ring system intercalating between base pairs and the peptide portion lying in the minor groove.

- (a) The aqueous solubility of drug A decreases with increasing temperature. What is the sign of ΔH° for the dissolution process?
- (b) This drug is generally poorly soluble in water, i.e., $\Delta G^\circ > 0$ for dissolution. What is the sign of ΔS° for this process?
- (c) Give a brief explanation for your answer to (b) in terms of solvent properties.
- (d) Figure 19.25 shows data for the temperature dependence of the rate constant for dissociation of drug A from duplex DNA. Use this plot to estimate the activation energy E_a and the entropy of activation ΔS^\ddagger . Comment on the role of solvent in determining the value of ΔS^\ddagger .

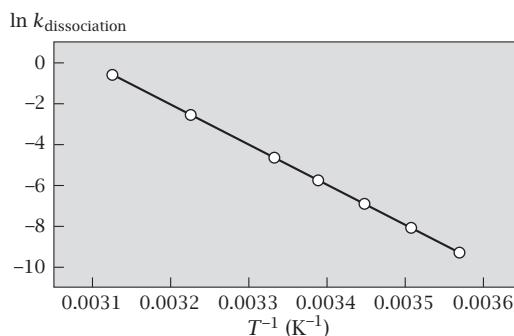
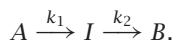


Figure 19.25

12. Series versus parallel reactions.

- (a) You have a reaction in which A converts to B through two steps in series:



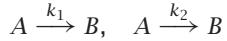
where the temperature dependences of the steps are given by the Arrhenius law:

$$k_1 = a e^{-E_1/RT},$$

$$k_2 = b e^{-E_2/RT}.$$

Derive an expression for the total rate k_{tot} , from A to B as a function of a, E_1 , E_2 , and T. (Hint: the time from A to B is the sum of times, from A to I and from I to B.)

- (b) Now consider instead two steps in parallel:



Using the Arrhenius equations from (a) for the individual steps, derive the temperature dependence for k in this case. (Hint: now the rates add.)

- 13. A cricket thermometer.** The frequency that crickets chirp increases with temperature. You can calculate the

temperature in degrees Celsius by adding 4 to the number of chirps you hear from a cricket in 8 s. In this way, you can use a cricket as a thermometer.

- What four temperatures would you have measured if the number of chirps in 8 seconds were 16, 21, 26, and 31?
- From this data, make an Arrhenius plot and calculate the activation energy in kcal mol^{-1} . This will tell you about a rate-limiting biochemical reaction that underlies cricket chirping.
- You observe that crickets outside a fast-food restaurant chirp faster than their cousins in the wild. The fast-food crickets chirp 30 times in 8 s at 25°C and 40 times in 8 s at 35°C . By how much did the fast-food additive reduce the activation barrier for cricket chirping?
- If the equilibrium enthalpy for the chirping reaction, Δh° , is 10 kcal mol^{-1} , what is the activation energy for the reverse chirp reaction for the normal crickets?

14. Onset of Huntington's disease. Huntington's disease is an example of a polyglutamine disease. The severity of the disease correlates with the degree of polyglutamine fibril formation. The age of onset decreases exponentially with increasing length of the polyglutamine chains (see the curve in Figure 19.26). Fibrils form quickly after an initial nucleation event, the binding of two polyglutamine chains. Assume dimerization is the rate-limiting step in the onset of the disease. Our aim is to interpret this data with a simple dimerization model.

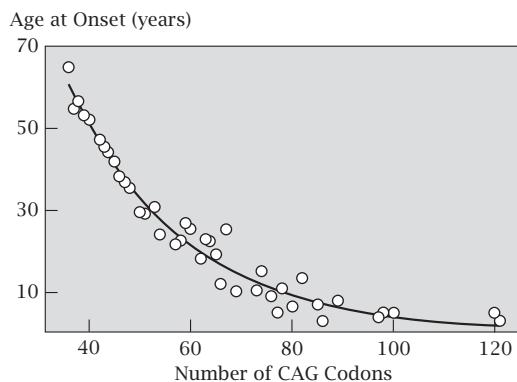


Figure 19.26 Source: JF Gusella and ME MacDonald, *Nat Rev Neurosci* **1**, 109–115 (2000).

- Use the lattice model shown in Figure 19.27. Define w_{ss} , w_{sg} , and w_{gg} to be the solvent-solvent, solvent-glutamine, and glutamine-glutamine interaction energies. If each glutamine chain is of length L , fill in Table 19.4 for the pairwise interactions that are broken and formed when two chains dimerize.

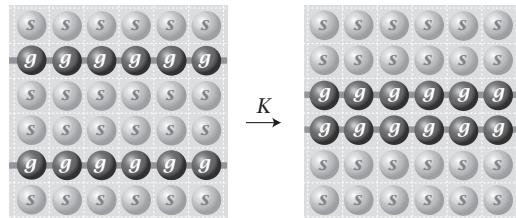


Figure 19.27 Polyglutamine chains associate with equilibrium constant K .

Table 19.4

	w_{ss}	w_{sg}	w_{gg}
Broken			
Formed			

- Write an expression for the dimerization equilibrium constant in terms of the interaction energies.
- Use the expression from answering (b) to explain the exponential form of Figure 19.26.
- Estimate the polymer-solvent interaction energy χ_{sg} from Figure 19.26.

References

- [1] WC Gardiner Jr, *Rates and Mechanisms of Chemical Reactions*. WA Benjamin, New York, 1969.
- [2] JI Steinfeld, JS Francisco, and WL Hase, *Chemical Kinetics and Dynamics*, 2nd edition, Prentice-Hall, Upper Saddle River, NJ, 1999.
- [3] SW Benson and O Dobis, *J Phys Chem A* **102**, 5175–5181 (1998).
- [4] DC Clary, *Annu Rev Phys Chem* **41**, 61–90 (1990).
- [5] RD Levine, *Molecular Reaction Dynamics*, Cambridge University Press, Cambridge, 2005.
- [6] T Baer and WL Hase. *Unimolecular Reaction Dynamics: Theory and Experiments*. Oxford University Press, New York, 1996.
- [7] A Radzicka and R Wolfenden, *Science* **267**, 90–93 (1995).
- [8] J Kraut, *Science* **242**, 533–540 (1988).
- [9] WR Cannon, SF Singleton, and SJ Benkovic. *Nat Struct Biol*, **3**, 821–833 (1996).
- [10] A Warshel, *J Biol Chem* **273**, 27035–27038 (1998).
- [11] A Warshel, *Computer Modeling of Chemical Reactions in Enzymes and Solutions*. Wiley, New York, 1991.
- [12] WR Cannon and SJ Benkovic, *J Biol Chem* **273**, 26257–26260 (1998).
- [13] RI Masel, *Principles of Adsorption and Reaction on Solid Surfaces*. Wiley, New York, 1996.

Suggested Reading

Excellent texts on isotope effects, the Brønsted law, and other aspects of physical organic chemistry:

ML Bender, RJ Bergeron, and M Komiyama, *The Bioorganic Chemistry of Enzymatic Catalysis*, Wiley, New York, 1984.

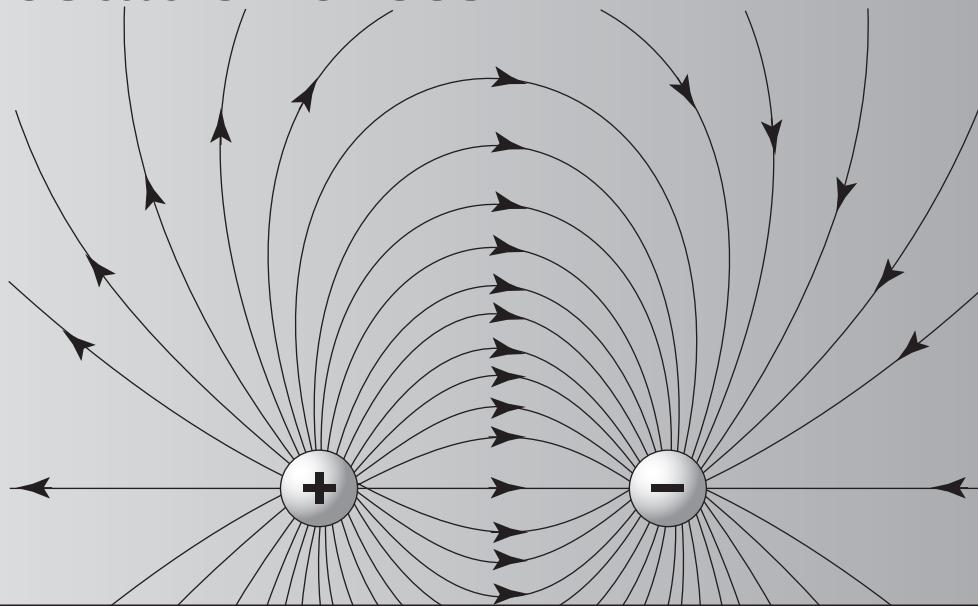
H Eyring and EM Eyring, *Modern Chemical Kinetics*, Reinhold, New York, 1963.

JE Leffler and E Grunwald, *Rates and Equilibria of Organic Reactions*, Wiley, New York, 1963 (reprinted by Dover Publications, New York, 1989).

H Maskill, *The Physical Basis of Organic Chemistry*, Oxford University Press, New York, 1985.

A Gross, *Theoretical and Physical Principles of Organic Reactivity*, Wiley, New York, 1995.

20 Coulomb's Law of Electrostatic Forces



Charge Interactions Are Described by Coulomb's Law

Electrical interactions between charges govern much of physics, chemistry, and biology. They are the basis for chemical bonding, weak and strong. Salts dissolve in water to form the solutions of charged ions that cover three-quarters of the Earth's surface. Salt water forms the working fluid of living cells. pH and salts regulate the associations of proteins, DNA, cells, and colloids, and the conformations of biopolymers. Nervous systems would not function without ion fluxes. Electrostatic interactions are also important in batteries, corrosion, and electroplating.

Charge interactions obey Coulomb's law. When more than two charged particles interact, the energies are sums of Coulombic interactions. To calculate such sums, we introduce the concepts of the *electric field*, Gauss's law, and the *electrostatic potential*. With these tools, you can determine the electrostatic force exerted by one charged object on another, as when an ion interacts with a protein, DNA molecule, or membrane, or when a charged polymer changes conformation.

Coulomb's law was discovered in careful experiments by H Cavendish (1731–1810), J Priestley (1733–1804), and CA Coulomb (1736–1806) on macroscopic objects such as magnets, glass rods, charged spheres, and silk cloths. Coulomb's law applies to a wide range of size scales, including atoms, molecules, and biological cells. It states that the interaction energy $u(r)$

between two charges in a vacuum is

$$u(r) = \frac{\mathcal{C}q_1q_2}{r}, \quad (20.1)$$

where q_1 and q_2 are the magnitudes of the two charges, r is the distance separating them, and $\mathcal{C} = 1/4\pi\epsilon_0$ is a proportionality constant (see box below).

Units

The proportionality constant \mathcal{C} in Equation (20.1) depends on the units used to measure the charge and the distance. In the older c.g.s. system, the units were defined so that $\mathcal{C} = 1$. We use the SI system, in which the unit of charge is the coulomb C, the unit of energy is the joule J, and the unit of distance is the meter m. The corresponding constant \mathcal{C} equals $(4\pi\epsilon_0)^{-1}$. The factor ϵ_0 is called the permittivity of the vacuum. In SI units, $\epsilon_0 = 8.85 \times 10^{-12} \text{ F m}^{-1}$. The farad F is the SI unit of capacitance, equal to 1 CV^{-1} . The volt V is the SI unit of electrical potential, which is equal to 1 JC^{-1} . So $\epsilon_0 = 8.85 \times 10^{-12} \text{ C}^2 \text{ J}^{-1} \text{ m}^{-1}$. In c.g.s. units, the unit of charge is 1 statcoulomb or 1 esu = $3.00 \times 10^9 \text{ C}$, the unit of potential is 1 statvolt = $(1/300) \text{ V}$, and the unit of capacitance is 1 statfarad = $9 \times 10^{11} \text{ F}$.

In SI units, the charge on a proton is $e = 1.60 \times 10^{-19} \text{ C}$ (with $\mathcal{N}e = 9.647 \times 10^4 \text{ C mol}^{-1}$, where \mathcal{N} is Avogadro's number), and the charge on an electron is $-e = -1.60 \times 10^{-19} \text{ C}$. A useful quantity for calculations is $\mathcal{C}e^2\mathcal{N} = e^2\mathcal{N}/4\pi\epsilon_0 = 1.386 \times 10^{-4} \text{ J m mol}^{-1}$.

EXAMPLE 20.1 The Coulombic attraction between Na⁺ and Cl⁻. Compute the pair interaction $u(r)$ between a sodium ion and a chloride ion at $r = 2.8 \text{ \AA}$, the bond length of NaCl, in a vacuum. Equation (20.1) gives

$$\begin{aligned} u(r) &= \frac{-\mathcal{C}e^2\mathcal{N}}{r} = \frac{-1.39 \times 10^{-4} \text{ J m mol}^{-1}}{(2.8 \times 10^{-10} \text{ m})} \\ &= -496 \text{ kJ mol}^{-1} \approx -119 \text{ kcal mol}^{-1}. \end{aligned}$$

EXAMPLE 20.2 Coulombic interactions hold materials together. Coulomb's law explains the attraction of positive for negative charges. Coulomb's law also explains how even when charges have the same sign, they can cluster together into stable structures. Positive charges will cluster together if there is enough negative charge 'glue' holding them together. There are many examples: salt crystals have alternating positive and negative charges, molecules are constellations of positively charged atomic nuclei held together by negatively charged electron density in covalent bonds, and charge clustering is common among colloids and biological complexes. Electrostatic bonding can be understood using the following simple model.

Figure 20.1 shows an in-line arrangement of two $+1$ charges, each separated by a distance r from a charge of $-q$ in the middle. The total Coulombic energy is the sum of the three pairwise interactions:

$$u = C e^2 \left(-\frac{q}{r} - \frac{q}{r} + \frac{1}{2r} \right). \quad (20.2)$$

This arrangement of charges is stable when $u < 0$, which happens when there is sufficient negative charge in the middle, $-q < -0.25$. This model illustrates how charged objects can bond together. However, this model is not quantitatively accurate for capturing covalent chemical bonding in molecules, where positive nuclei are held together by negative electron density. For that, you need quantum mechanics, which gives different mathematical forms for $u(r)$, including steric repulsions as $r \rightarrow 0$.

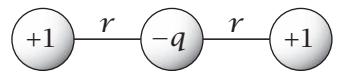


Figure 20.1 Two charges of $+1$ are held together by electrostatic attraction if there is a sufficient negative charge $-q$ in the middle.

Charge Interactions Are Long-Ranged

In previous chapters, we considered only the interactions between uncharged atoms. Uncharged atoms interact with each other through *short-ranged* interactions. That is, attractions between uncharged atoms are felt only when the two atoms are so close to each other that they nearly contact. Short-ranged interactions typically diminish with distance as r^{-6} . We will explore this in more detail in Chapter 24. For now, it suffices to note that charge interactions are very different: they are long-ranged. Coulombic interactions diminish with distance as r^{-1} . Counting lattice contacts is no longer sufficient.

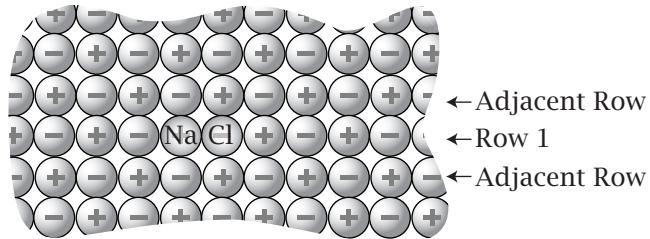
The difference between short-ranged and long-ranged interactions is much more profound than it may seem. The mathematics and physics of long-ranged and short-ranged interactions are very different. A particle that interacts through short-ranged interactions feels only its nearest neighbors, so system energies can be computed by counting the nearest-neighbor contacts, as we did in previous chapters. But when interactions involve charges, more distant neighbors contribute to the energies, so we need more sophisticated methods of summing energies.

Example 20.3 shows how long-ranged electrostatic interactions hold a salt crystal together.

EXAMPLE 20.3 Why are sodium chloride crystals stable? Let's assume that a crystal of sodium chloride is composed of hard spherical ions that interact solely through electrostatics. To compute the total energy, we sum all the pairwise Coulombic interaction energies of one ion with every other ion in the crystal. Crystalline NaCl is arranged in a simple cubic lattice, with interionic spacings $a = 2.81 \text{ \AA}$ at 25°C .

First sum the interactions of a sodium ion in row 1 shown in Figure 20.2 with all the other ions in the same row. Because there are two negative ions at distance a with unit charge $-e$, two positive ions at distance $2a$ with unit charge e , and two more negative ions at distance $3a$, etc., the total Coulombic

Figure 20.2 Crystalline sodium chloride packs in a cubic array of alternating sodium and chloride ions.



interaction along the row $u_{\text{row}1}$ is given by the series

$$\begin{aligned} u_{\text{row}1} &= \frac{C e^2}{a} \left(-\frac{2}{1} + \frac{2}{2} - \frac{2}{3} + \frac{2}{4} - \dots \right) \\ &= \frac{2 C e^2}{a} \left(-1 + \frac{1}{2} - \frac{1}{3} + \frac{1}{4} - \dots \right). \end{aligned} \quad (20.3)$$

This series is slow to converge, implying that two charges can have substantial interaction even when they are far apart. The series converges to $-\ln 2$ per row (see Appendix J, Equation (J.4) with $a = x = 1$), so the energy of interaction of one sodium ion with all other ions in the same row is

$$u_{\text{row}1} = -\frac{2 C e^2}{a} \ln 2 = -\frac{1.386 C e^2}{a}. \quad (20.4)$$

Next, sum the interactions of the same sodium ion with all the ions in the four adjacent rows above, below, in front of, and behind row 1. In these four rows, the closest ions are the four negative charges at distance a , the eight positive charges at distance $a\sqrt{2}$, and the eight negative charges at distance $a\sqrt{5}$. The pattern of pairwise distances gives a j th term involving eight charges at distance $a(1+j^2)^{1/2}$. These four adjacent rows contribute energies $u_{\text{four adjacent rows}}$:

$$u_{\text{four adjacent rows}} = \frac{4 C e^2}{a} \left(-1 + \frac{2}{\sqrt{2}} - \frac{2}{\sqrt{5}} + \frac{2}{\sqrt{10}} - \dots \right). \quad (20.5)$$

When the pairwise energies are summed over all the rows of charges, the total Coulombic energy of the interaction between one ion and all the others in the crystal is found [1] to be

$$\begin{aligned} U &= -1.747 \frac{C e^2 N}{a} \\ &= \frac{(-1.747)(1.386 \times 10^{-4} \text{ J m mol}^{-1})}{2.81 \times 10^{-10} \text{ m}} \\ &= -862 \text{ kJ mol}^{-1} = -206 \text{ kcal mol}^{-1}. \end{aligned} \quad (20.6)$$

The energy holding each positive charge in the crystal is $U/2$, where the factor of 2 corrects for the double-counting of interactions. To remove a NaCl molecule, both a Na^+ ion and a Cl^- ion, the energy is $U = 206 \text{ kcal mol}^{-1}$, according to this calculation.

The experimentally determined energy of vaporization (complete ionization) of crystalline NaCl is $183 \text{ kcal mol}^{-1}$. The 23 kcal mol^{-1} discrepancy between the value we calculated and the experimental value is due mainly to the assumption that the ions are hard incompressible particles constrained to a separation of exactly $a = 2.81 \text{ \AA}$. In reality, ions are somewhat compressible. Nevertheless, we can conclude that the Coulombic interactions are the dominant attractions that hold ionic crystals together.

The quantity -1.747 in Equation (20.6) is called the *Madelung constant*. The magnitude of the Madelung constant depends on the symmetry of the crystal lattice.

Charges that interact with each other over long spatial distances contribute substantially to the energy of an electrostatic system. To see this, suppose that we had summed only the contributions of the six nearest neighbors, as we have done for systems dominated by short-range interactions. All six nearest neighbors are negative charges, so our estimate for the Madelung constant would have been -3 (correcting for double counting), rather than -1.747 , leading to a huge error, about $148 \text{ kcal mol}^{-1}$!

Electrostatic interactions are strong as well as long-ranged. The electrostatic repulsion is about 10^{36} times stronger than the gravitational attraction between two protons. Strong, short-ranged nuclear forces are required to hold the protons together against their electrostatic repulsions in atomic nuclei. When charge builds up on a macroscopic scale, violent events may result, such as lightning and thunderstorms, and explosions in oil tankers. Because charge interactions are so strong, matter does not sustain charge separations on macroscopic scales: macroscopic matter is neutral. The transfer of less than a thousand charges is enough to cause the static electricity that you observe when you rub your shoes on a carpet. In solution, too, imbalances are usually less than thousands of charges, but because this is one part in 10^{21} , it means that you can assume that solutions have charge neutrality—no net charge of one sign—to a very high degree of approximation. Although macroscopic charge separations aren't common in ordinary solutions, charge separations on microscopic scales are important, and will be described in Chapters 22 and 23.

Charge Interactions Are Weaker in Media

While charges may interact strongly in a vacuum, charges interact more weakly with each other when they are in liquids. To varying degrees, liquids can be *polarized*. This means that if two fixed charges, say a negative charge A and a positive charge B , are separated by a distance r in a liquid, the intervening molecules of the liquid tend to reorient their charges. The positive charges in the medium orient toward A , and the negative charges toward B , to shield and weaken the interaction between A and B (see Figure 20.3). This weakening of the Coulombic interactions between A and B is described by the *dielectric constant* of the medium. Media that can be readily polarized shield charge interactions strongly: they have high dielectric constants.

In this and the next three chapters, we model the medium as a *polarizable isotropic continuum*. In treating a medium as a continuum, we neglect its atomic structure and focus on its larger-scale properties. In treating a medium

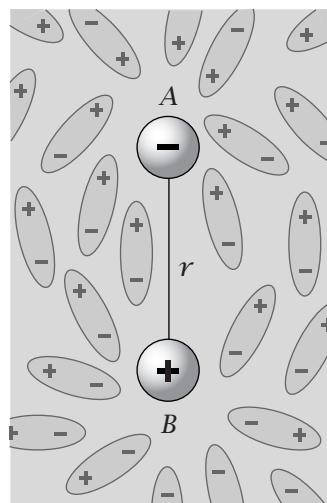


Figure 20.3 Fixed charges A and B cause dipoles in the surrounding liquid to orient partly, which causes some *shielding*, a reduction of the interaction between A and B . The dielectric constant D describes the degree of shielding.

Table 20.1 The dielectric constants D of several organic liquids.

Liquid	$T(^{\circ}\text{C})$	D
Heptane	0	1.958
Heptane	30	1.916
Methanol	25	33
Formamide	20	109
Formic acid	16	58
Nitrobenzene	25	35
Hydrogen cyanide	0	158
Hydrogen cyanide	20	114
Glycol	25	37
Water	0	88.00
Water	25	78.54

Source: RC Weast, ed., *Chemical Rubber Company Handbook of Chemistry and Physics*, 53rd edition, CRC, Cleveland, 1972.

as *isotropic*, we assume that its polarizability is the same in all directions. By treating a medium as *polarizable*, we assume that the charge redistributes in response to an electric field, even if the medium is neutral overall.

The more polarizable the medium, the greater is the reduction of an electrostatic field across it. The reduction factor is the *dielectric constant* D , a scalar dimensionless quantity. When charges q_1 and q_2 are separated by a distance r in a medium of dielectric constant D , their Coulombic interaction energy is

$$u(r) = \frac{C q_1 q_2}{Dr}. \quad (20.7)$$

For air at 0°C , $D = 1.00059$. Table 20.1 gives the dielectric constants of some liquids. Some polar liquids have dielectric constants larger than $D = 100$, implying that an electrostatic interaction in such a liquid has less than 1% of its strength in a vacuum.

Coulombic interactions are described as an energy $u(r)$, but when dielectric constants are involved, $u(r)$ is a free energy, because the dielectric constant depends on temperature. Figure 20.3 illustrates that polarization often involves orientations, and therefore entropies.

The polarizability of a medium arises from several factors. First, if the medium is composed of molecules that have a *permanent dipole moment*, a positive charge at one end and a negative charge at the other, then applying an electric field tends to orient the dipoles of the medium in a direction that opposes the field.

Second, the polarizability of a medium can arise from the polarizabilities of the atoms or molecules composing it, even if they have no permanent dipole moment. Atoms or molecules that lack permanent dipoles have electronic *polarizability*, a tendency of nuclear or electronic charge distributions to shift slightly within the atom in response to an electric field (see Chapter 24). The electronic polarizabilities of hydrocarbons and other nonpolar substances are the main contributors to their dielectric constants ($D \approx 2$).

Third, polarizabilities can also arise from networks of hydrogen bonds in liquids such as water. If there are many alternative hydrogen-bond donors and

acceptors, a shift in the pattern of hydrogen bonding is not energetically costly, so the networks of hydrogen bonds in hydrogen-bonded liquids are labile when electric fields are applied.

The Bjerrum Length

To see the effect of the polarizability of the medium, let's first compute a useful quantity called the *Bjerrum length* ℓ_B , defined as the charge separation at which the Coulomb energy $u(r)$ between a mole of ion pairs just equals the thermal energy RT (see Figure 20.4). Substitute ℓ_B for r and RT for u in $u(r) = \mathcal{C}q_1q_2/Dr$ (Equation (20.7)), and solve for ℓ_B for single charges $q_1 = q_2 = e$:

$$\ell_B = \frac{\mathcal{C}e^2\mathcal{N}}{DRT}. \quad (20.8)$$

For a temperature $T = 298.15\text{ K}$ in vacuum or air ($D = 1$),

$$\ell_B = \frac{1.386 \times 10^{-4} \text{ J m mol}^{-1}}{(8.314 \text{ JK}^{-1} \text{ mol}^{-1} \times 298.15 \text{ K})} = 560 \text{ \AA}.$$

560 \AA is a relatively long distance, many times the diameter of an atom. When two charges are much closer together than this distance, they feel a strong repulsive or attractive interaction, greater than the thermal energy kT . But when two charges are much further apart than the Bjerrum length, their interactions are weaker than kT , and then two charges are more strongly directed by Brownian motion than by their Coulombic interactions.

Dividing $u(r) = \mathcal{C}q_1q_2/Dr$ by $RT = \mathcal{C}q_1q_2/D\ell_B$ gives

$$\frac{u}{RT} = \frac{\ell_B}{r}. \quad (20.9)$$

The Bjerrum length simplifies calculations. What is the interaction energy between K^+ and Cl^- separated by $r = 20\text{ \AA}$ in a vacuum? Equation (20.9) gives $u = (560/20)RT = 28(0.6) = 16.8\text{ kcal mol}^{-1}$.

Since we will often be interested in charge interactions in water, let's also define another Bjerrum length, ℓ_{Bw} , for systems in water. Substitute the dielectric constant of water, $D = 78.54$ at 25°C , into Equation (20.8) to get $\ell_{Bw} = 7.13\text{ \AA}$.

Example 20.4 shows how the polarizability of a medium affects charge interactions. Although Na^+ is strongly attracted to Cl^- in a vacuum (by

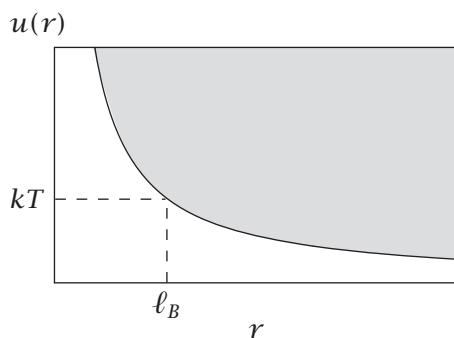


Figure 20.4 The Bjerrum length ℓ_B is defined as the distance r at which two unit charges interact with an energy $u(r)$ equal to kT .

$119 \text{ kcal mol}^{-1}$ as we found in Example 20.1), the attraction is much weaker in water, where $D = 78.54$ at 25°C .

EXAMPLE 20.4 Why do salts ionize in water? In vacuum or in oil, an ionic bond is strong.

But NaCl readily ionizes in water, that is, it falls apart into individual ions, indicating that its Coulombic bond is weak in water. At a distance of 2.81 \AA , Equation (20.9) shows that the ionic interaction in water is $u(r, D = 78.54) = -(7.13/2.81) \times RT$ per mole, or $-1.5 \text{ kcal mol}^{-1}$ at $T = 300 \text{ K}$, only a little more than twice RT .

So, although two ions in air or oil interact strongly, two ions in water interact only very weakly. This is why salts in water dissociate into ions.

So far, our calculations for ionic interactions have involved simple arrangements of charges. When several charges are involved, or when charges have complex spatial arrangements, more sophisticated methods are needed to treat them. We now develop some useful tools that lead to Gauss's law and Poisson's equation.

Electrostatic Forces Add Like Vectors

Vectors and Coulomb's Law

Coulomb's law may be expressed either in terms of energy $u(r) = Cq_1q_2/Dr$ or in terms of force $\mathbf{f} = -(\partial u/\partial r) = Cq_1q_2/Dr^2$. The difference is that energies are scalar quantities, which add simply, while forces are vectors that add in component form. Either way, an important fundamental law is the superposition principle: electrostatic energies and forces are both additive.

The total electrostatic force on a particle is the vector sum of the electrostatic forces from all other particles. You can express the force on particle B from particle A in Figure 20.5 as

$$\mathbf{f} = \frac{Cq_A q_B}{Dr^2} \frac{\mathbf{r}}{r}, \quad (20.10)$$

where \mathbf{r}/r is a vector pointing in the direction of the vector \mathbf{r} , along the line connecting the particles, but with unit length. \mathbf{f} is the force acting in the same direction. The component of the force \mathbf{f} in the x direction is described by using the component of the vector \mathbf{r} between the charges, $r_x = r \cos \alpha$, where α is

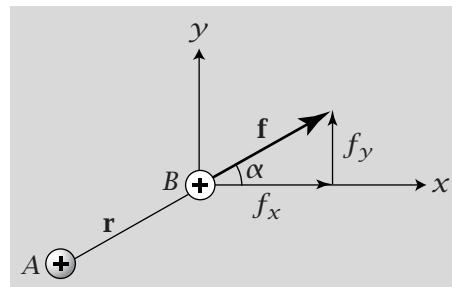


Figure 20.5 The electrostatic force \mathbf{f} of fixed charge A on test charge B is a vector with components f_x and f_y along the x and y axes, respectively.

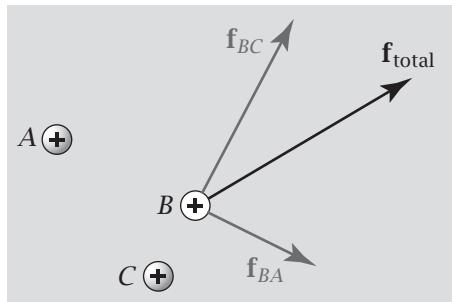


Figure 20.6 The force on the test charge B is the vector sum of the forces from the two fixed charges A and C .

the angle between \mathbf{f} and the x axis. The x component of the force is

$$f_x = \left(\frac{q_A q_B}{4\pi \epsilon_0 D r^2} \right) \cos \alpha. \quad (20.11)$$

The y component is

$$f_y = \left(\frac{q_A q_B}{4\pi \epsilon_0 D r^2} \right) \sin \alpha. \quad (20.12)$$

Figure 20.6 shows how electrostatic forces add vectorially.

What Is an Electrostatic Field?

Now we want a general way to describe the force that would act on a charged particle if it were placed at *any position* in space containing other fixed charges. This task is simplified by the concept of the *electrostatic field*. An electrostatic field is a vector field (see Appendix G). Suppose that charges A and B in Figure 20.7 are in fixed positions.

Figure 20.7 diagrams the force vector that would act on a charge C at two different positions. The force $\mathbf{f}(x, y, z)$ that acts on a particle C varies with the spatial position (x, y, z) of the particle C . When C is at position 1, the force from A and B on C is large and in the $+x$ direction. When C is at position 2, the force on C is in a different direction, and it is weaker because of the r^{-2} dependence of force on distance.

To describe how the force acting on C depends on spatial position, you compute the electrostatic field. Consider C to be a *test particle*, having a charge

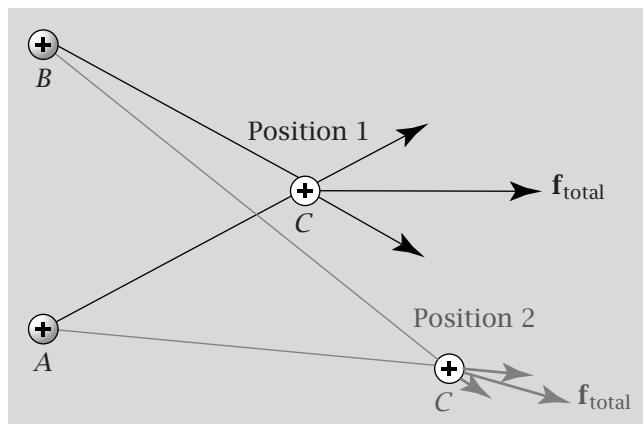


Figure 20.7 A and B are charges at fixed positions in space. C is a *test particle*. When C is placed at position 1, the net force on C is large and along the $+x$ direction. When C is placed at position 2, the force is weaker and in a different direction. The *electrostatic field* is the set of vectors that describes the force per unit charge that the test particle would feel at any position in space.

q_{test} . By test particle, we mean a positive charge that you are free to place at any position to probe the force there. The test particle is acted upon by the fixed charges. You need not be concerned with how the fixed charges act on each other, because that isn't a degree of freedom. The electric (or electrostatic) field \mathbf{E} is the vector located at position $\mathbf{r} = (x, y, z)$ in space that indicates the direction and magnitude of the electrostatic force that acts on a charged test particle C at that point, due to all the fixed charges. Wherever you place C in space, you will find a vector there that gives the direction and magnitude of the force that will act on the particle at that position. Suppose the vector force that would act on a test particle C in Figure 20.7 at position \mathbf{r} is $\mathbf{f}(\mathbf{r})$. Dividing the force in Equation (20.10) by the test charge q_{test} defines the field:

$$\mathbf{E}(\mathbf{r}) = \frac{\mathbf{f}(\mathbf{r})}{q_{\text{test}}} = \frac{q_{\text{fixed}}}{4\pi\epsilon_0 D r^2} \frac{\mathbf{r}}{r}, \quad (20.13)$$

the force per unit charge on the test particle. The SI units of electrostatic field are NC^{-1} or Vm^{-1} .

EXAMPLE 20.5 The field from a point charge. A point charge is radially symmetric, so you can dispense with vector notation to compute its field and just treat the dependence on radial distance r . The field from a fixed point charge acting on a test particle is

$$E(r) = \frac{C q_{\text{fixed}}}{D r^2}. \quad (20.14)$$

Figure 20.8 shows the electrostatic fields around a positive charge and around a negative charge. It shows discrete vectors at specific positions in space, because that's the best you can do in a diagram. The actual fields are continuous functions of spatial position. The figure shows that a positive test charge is pulled toward a fixed negative charge and repelled from a fixed positive charge. The relative lengths of the vector arrows indicate that the force is greatest near the charge, and diminishes with distance from the fixed charge.

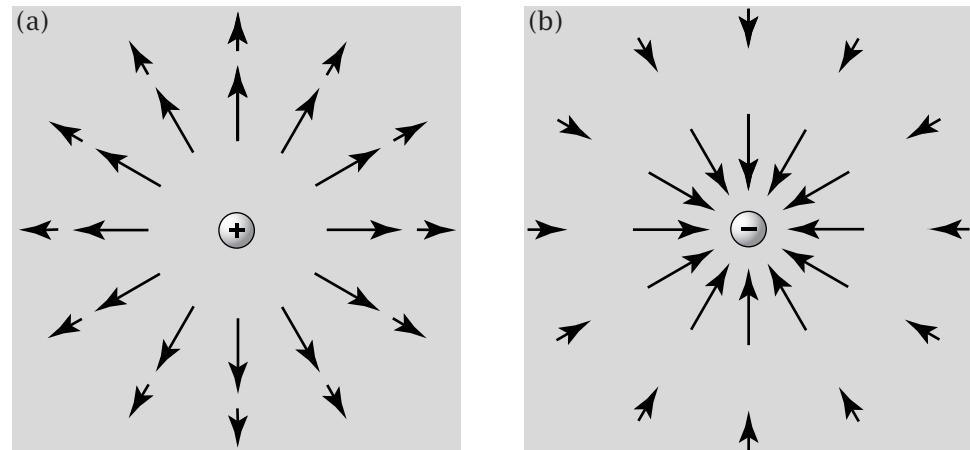


Figure 20.8 The electric field is indicated by vectors pointing (a) away from a positive charge and (b) toward a negative charge. The field represents the force per unit charge that acts on a positive test charge put at each particular position in space.

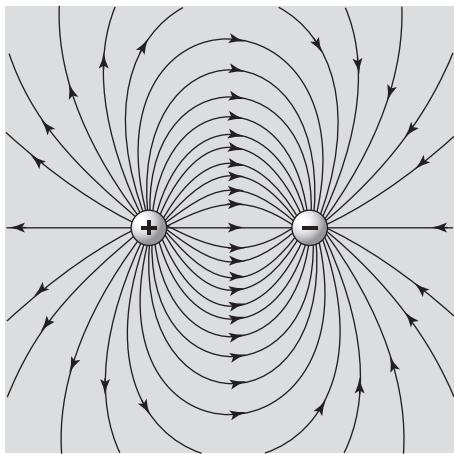


Figure 20.9 The electric field near a dipole is represented by field lines pointing away from the positive charge and toward the negative charge. This diagram shows the direction that a positive test charge would be moved if it were placed at a given position.

Figure 20.9 uses field lines instead of field vectors to show the force that a positive test charge would experience near a dipolar arrangement of one positive and one negative charge. In this type of diagram, each line represents only the direction of the force at any position in space, and not its magnitude.

Now suppose you are interested in the electrical force acting on a charged particle, from a line of charges like the phosphate groups in DNA or other charged groups on a polyelectrolyte molecule, or from a plane of charge like a membrane or surface. You will need to sum many Coulomb force vectors. If the charges on the line or plane are close together, you can approximate this sum as an integral over force vectors. But there is a much easier way to do all this, called Gauss's law. Gauss's law is nothing but Coulomb's law in a different form.

Electric Fields Have Fluxes

To derive Gauss's law you need a quantity called the *electric field flux*. This concept is useful for simplifying various problems that are otherwise hard to solve. Positive and negative charges are the sources and sinks of electrical fields. Flowing fluids also involve fluxes and sources and sinks (see Appendix G). Consider some constellation of positive and negative charges fixed at certain positions in space. We invent an imaginary balloon that we place around the charges. We want to determine the flux of the electric field through the imaginary balloon surface. The electric field flux Φ is defined as the integral of D times the electric field \mathbf{E} over all the surface area elements $d\mathbf{s}$ of the balloon:

$$\Phi = \int_{\text{surface}} D \mathbf{E} \cdot d\mathbf{s}, \quad (20.15)$$

where D is the dielectric constant. Because \mathbf{E} is proportional to $1/D$ in a uniform medium (see Equation (20.13)), the flux Φ is independent of D .

Equation (20.15) is general and applies to any constellation of fixed charges, and to any shape and position of an imaginary balloon around them. Many problems of interest have a special simplicity or symmetry that allows you to simplify Equation (20.15). For example, suppose that the charge constellation is just a single point charge q and that the imaginary balloon is just a sphere of

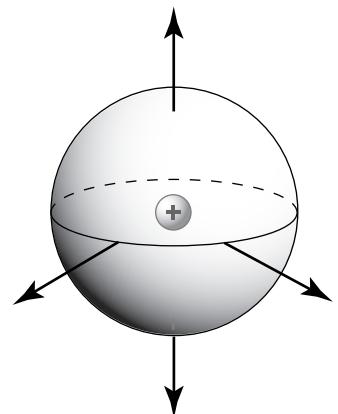


Figure 20.10 The vector electric field \mathbf{E} is normal to a sphere surrounding a point charge located at the center of the sphere.

radius r centered on the point charge (see Figure 20.10). We are interested in computing the electric field flux from the point charge through the imaginary spherical surface.

The electrostatic field of a point charge is spherically symmetrical—the magnitude of the field E is the same through the north pole as it is through the south pole or through the equator, or through any other patch of surface of a given area. So you can drop the vector notation and express the field as $E(r)$, which depends only on r . Because E doesn't differ from one surface patch to another, the integral in Equation (20.15) becomes a simple product. That is, E is parallel to $d\mathbf{s}$ everywhere, so $\mathbf{E} \cdot d\mathbf{s} = E ds$, and the flux integral becomes $\Phi = DE \int ds = DE(r) \times (\text{total area of sphere})$. Use $E(r) = q/(4\pi\epsilon_0 Dr^2)$ from Equation (20.14), and $4\pi r^2$ for the surface area of a sphere, to get

$$\begin{aligned}\Phi &= DE(r)(4\pi r^2) \\ &= \left(\frac{q}{4\pi\epsilon_0 r^2}\right)(4\pi r^2) \\ &= \frac{q}{\epsilon_0},\end{aligned}\tag{20.16}$$

where q is the charge at the center of the sphere.

Two profound implications of Equation (20.16) lead to a general method for calculating electrostatic forces and fields. First, the right-hand side of Equation (20.16) does not depend on r . For the electric flux through a spherical container, the r^{-2} dependence of E cancels with the r^2 dependence of the spherical surface area. Because of this cancelation, the electric field flux out of a sphere is independent of its radius. It doesn't matter how large you choose your imaginary balloon to be, the flux is the same through spherical balloons of any radius. Second, we show now that the *shape* of the balloon doesn't matter either.

The Electric Field Flux Is Independent of Balloon Shape

Now construct another imaginary balloon (the outer balloon) at radius $R > r$, fully containing the original spherical balloon (the inner balloon) that encloses the point charge q (see Figure 20.11). The outer balloon may have any arbitrary shape: it need not be spherical. Consider a radial cone emanating from the point charge. The flux through a small area element $s(r)$ of the inner balloon is $\Phi_{\text{inner}} = DE(r)s(r)$. The same cone cuts through a patch of area $s(R)$ on the outer balloon that is oriented at an angle θ with respect to the radial vector. Because of this angle, the flux through the outer balloon element is reduced by $\cos \theta$, so

$$\Phi_{\text{outer}} = DE(R)s(R)\cos \theta.$$

To relate Φ_{outer} to Φ_{inner} , compute the field through the outer element,

$$E(R) = \frac{\mathcal{C}q}{DR^2} = E(r)\left(\frac{r}{R}\right)^2,$$

and notice that the area of the outer balloon element is greater than the area of the inner balloon element owing to two factors: the area is larger by a factor

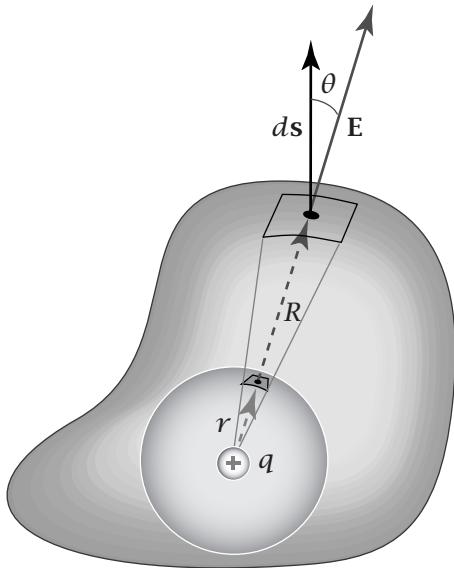


Figure 20.11 The flux through *any* closed surface around a charge q is the same as the flux through a sphere. R is the distance from the charge to the outer surface element. ds indicates the outer surface area element. E indicates the field at distance R from the charge.

$(R/r)^2$ owing to the greater distance from the origin; the area is larger by a factor of $1/\cos \theta$ owing to its tilt. Hence $s(R) = s(r)(R/r)^2/\cos \theta$. Combining these factors gives

$$\Phi_{\text{outer}} = DE(r) \left(\frac{r}{R}\right)^2 \left(\frac{s(r)}{\cos \theta}\right) \left(\frac{R}{r}\right)^2 \cos \theta = \Phi_{\text{inner}}.$$

The flux is the same through the outer balloon area element as through the inner balloon element.

Because this argument applies to all surface elements at any angle θ or distance R , it means that the electric field flux does not depend on the size or shape of the bounding surface. Therefore, because we know that the flux through the sphere of radius r is $\Phi = q/\epsilon_0$, it must be true that for *any closed surface* of any radius or shape that encloses the point charge q ,

$$\Phi = \int_{\text{surface}} D\mathbf{E} \cdot d\mathbf{s} = \frac{q}{\epsilon_0}. \quad (20.17)$$

Before we show how Equation (20.17) is useful, consider one last generalization. Because this argument applies to *any* bounding surface, it surely must apply to a spherical balloon that is shifted so that its center does not coincide with the charge point.

These generalizations now lead to a remarkable result called Gauss's law. Because the electric field $\mathbf{E}(\mathbf{r})$ from any constellation of fixed charges is always the vector sum of the fields from the component charges, $\mathbf{E} = \mathbf{E}_1 + \mathbf{E}_2 + \mathbf{E}_3 + \dots + \mathbf{E}_n$, the flux through any closed surface around any constellation of charges is

$$\Phi = \int_{\text{surface}} D\mathbf{E} \cdot d\mathbf{s} = \int_{\text{surface}} D(\mathbf{E}_1 + \mathbf{E}_2 + \mathbf{E}_3 + \dots) \cdot d\mathbf{s} = \frac{1}{\epsilon_0} \sum_{i=1}^n q_i, \quad (20.18)$$

where $\sum_{i=1}^n q_i$ is just the net charge contained inside the boundary surface. The vector quantity $D\mathbf{E}$ is called the electric displacement.

Equation (20.18) is **Gauss's law**. It says that a very complex quantity can be computed by a very simple recipe. If you want to find the flux of the electrostatic field through any bounding balloon, no matter how complex its shape, you don't need to compute the electrostatic field vector at each point in space, find its dot product with all the surface elements, and integrate. Instead, you can compute the flux simply by counting up the total net charge contained within the bounding surface, and divide by ϵ_0 , a constant.

We have derived Gauss's law for a uniform medium. But Equation (20.18) is also valid in a medium with spatial variations of the dielectric constant D . In that case the local value of D is used in the surface integrals of Equation (20.18). This will be useful for treating dielectric boundaries in Chapter 21.

If the charges are many or densely distributed, it is sometimes more convenient to represent the charge distribution in terms of a continuous spatial distribution function $\rho(x, y, z)$, the charge per unit volume. The rightmost term in Equation (20.18) is a sum over charges. If you express that sum as an integral instead, then Gauss's law (20.18) can be expressed in a form we will find useful later:

$$\int D\mathbf{E} \cdot d\mathbf{s} = \frac{1}{\epsilon_0} \int_V \rho \, dV. \quad (20.19)$$

The real power of Gauss's law is in the third equality in Equation (20.18), not the first. It prescribes a very simple way to compute the electrostatic field of force from a constellation of fixed charges. Here are some examples.

EXAMPLE 20.6 The field from a line charge. Let's model a charged wire or a polyelectrolyte molecule such as DNA as a long charged line of length L , a cylinder of zero radius. Suppose that this line has λ charges per unit length and is so long that you can neglect end effects. Then the only electrostatic field is *radial*, perpendicular to the line of charge, because the cylindrical symmetry of the line causes the nonradial components to cancel each other. What is the radial force on a charged test particle near the line charge? One way to compute the field is to sum the vector forces on the test charge from each small piece of the line charge. A much simpler way is to use Gauss's law. Use a cylindrical bounding surface with its axis along the line of charges to enclose a length L of the line of charges (see Figure 20.12). The flux that exits perpendicular to the cylinder surface at a distance r is $D \times E(r) \times (\text{area}) = DE(2\pi r L)$. Because the total charge is (number of charges per unit length) \times (length) $= \lambda L$, Gauss's law for this system is given by $D \times E(r) \times (\text{area}) = (\text{total charge})/\epsilon_0 = \lambda L/\epsilon_0$, so

$$DE(r)(2\pi r L) = \frac{\lambda L}{\epsilon_0}. \quad (20.20)$$

Rearranging Equation (20.20) shows that the field $E(r)$ at a radial distance r from a line of charge is

$$E(r) = \frac{\lambda}{2\pi\epsilon_0 D r}. \quad (20.21)$$

Equation (20.21) gives the force per unit charge that acts on a test charge at any distance r from the line charge.

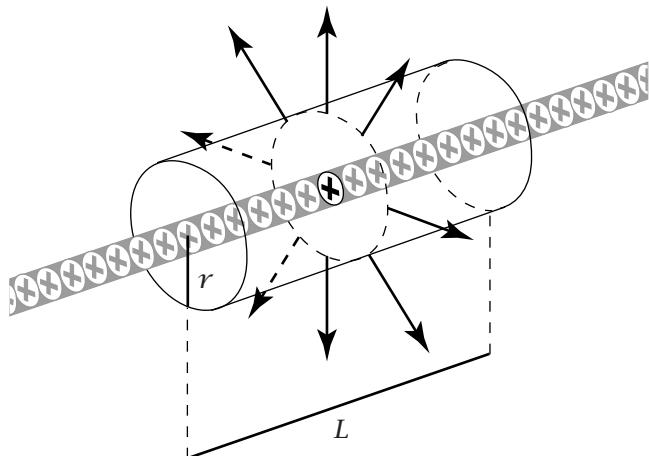


Figure 20.12 To compute the electric field at a distance r from a charged line, construct an imaginary cylinder of radius r and length L and use Gauss's law.

EXAMPLE 20.7 The field from a charged planar surface. Let's compute the field due to a charged planar surface, such as the plate of a capacitor, an electrode, or a cell membrane in an isotropic dielectric medium. Suppose that the charge is distributed uniformly on the planar surface. Assume that the plane is so large that you can neglect edge effects, and that it has a uniform surface charge σ per unit surface area.

Consider two different situations. When you are interested in the fields that act over *large* distances relative to the thickness of the plane, a charged plane is considered *thin*. For example, for computing the field E_x between electrodes in batteries or capacitors, the electrode is usually thin relative to the separation between electrodes. In contrast, when you are interested in the fields that act over *small* distances relative to the thickness of the plane, the charged plane is considered *thick*. For example, salt ions form ‘shielding layers’ around proteins or DNA molecules or membranes. Such shielding layers usually extend over only a few tens of molecular diameters (see Chapter 23). In those cases, the charged plane is thick relative to the spatial scale over which the field acts.

First, consider a thin plane. To use Gauss's law, construct a cylindrical bounding surface sandwiching the plane, as shown in Figure 20.13. By symmetry, the magnitude of the electric field must be equal on the two opposite sides of the sheet, so the vectors E point in opposite directions. Because the only nonzero component of the field is in the x direction, we can dispense with vector notation. If $E(x)$ is the magnitude of the field on one side of the plane, then the total flux out (both ends of) the bounding surface is $2DE(x)$ multiplied by the area A of each end of the cylinder. Because the total charge enclosed by the surface is $A\sigma$, Gauss's law gives

$$2DE(x)A = \frac{\text{total charge}}{\epsilon_0} = \frac{A\sigma}{\epsilon_0}. \quad (20.22)$$

Rearrange this to find E for a plane of charges:

$$E_{\text{thin plane}} = \frac{\sigma}{2\epsilon_0 D}. \quad (20.23)$$

(a) Thin Plane

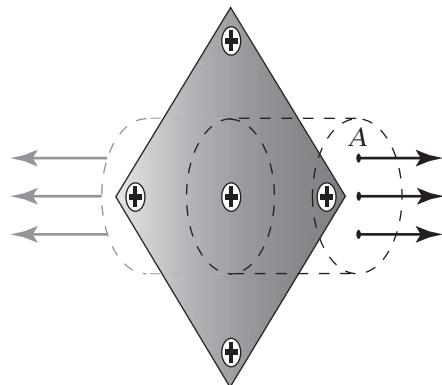
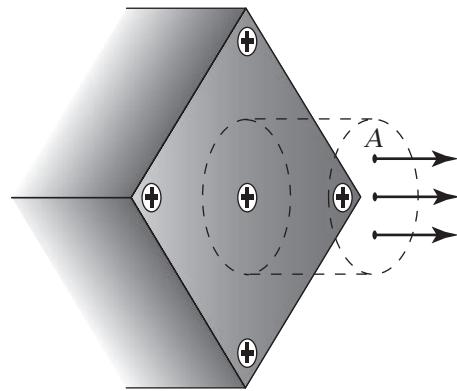


Figure 20.13 (a) A field emanates in both directions perpendicular to a charged *thin* plane. (b) A field emanates in only one direction perpendicular to a charged *thick* plane. To compute the field from either type of plane, construct a cylindrical box having end area A and containing the plane, and use Gauss's law.

(b) Thick Plane



The height or volume of the cylinder you choose doesn't matter because the field strength is independent of the distance x from the plane.

E changes sign from one side of the charged surface to the other. The total difference in electrostatic field from one side of the plane to the other, from E to $-E$, equals $\sigma/\epsilon_0 D$. In general (the normal component of) the electric field changes discontinuously upon passing through a charged surface or upon passing through a dielectric boundary (see Chapter 21).

For a thick plane, the electrostatic flux exits only from one end of the cylinder, so

$$E_{\text{thick plane}} = \frac{\sigma}{\epsilon_0 D}. \quad (20.24)$$

Examples of both thin and thick planes appear in the next three chapters.

We have found that the electrostatic field from a point charge is shorter-ranged ($1/r^2$) than the field from a line charge ($1/r$), and that the field from a plane charge is independent of distance. These relations are a matter of scale. If you are interested in a field on a small scale very close to a large charged sphere, the charged surface acts like a plane, and the field is independent of r .

Very far away from a charged plane, much farther than the size of the plane, the plane appears to be a point charge, and the field decays as $1/r^2$.

Summary

Electrostatic interactions are governed by Coulomb's law. They are long-ranged: for two point charges, $u(r)$ diminishes with charge separation as $1/r$. While charge interactions in a vacuum can be very strong, charges interact more weakly in polarizable media, because the medium arranges itself to counteract the field set up by the charges. The degree to which a medium can polarize and shield charge interactions is described by the dielectric constant of the medium. Electrostatic forces sum like vectors, and are described by electrostatic fields. Gauss's law, defined in terms of the electric field flux through any bounding surface, provides a useful relationship between the electrostatic field and the net charge for any constellation of fixed charges. In the next chapter, we define the electrostatic potential and Poisson's equation for more complex problems than we can treat with Gauss's law.

Problems

1. NaCl in a gas or crystal. In Example 20.1, we calculated the electrostatic interaction energy between Na^+ and Cl^- ions in a vacuum at 2.81 \AA distance. How is this energy related to the crystal energy calculated in Example 20.3?

2. Divalent ion attraction. A Mg^{2+} ion and a Cl^- ion are 6 \AA apart in water at 25°C . What is their electrostatic interaction energy in units of kT ?

3. Charges in alcohol. The dielectric constant of methanol is $D = 33$. Two charges are separated by a distance ℓ in methanol. For what value of ℓ is the electrostatic interaction equal to kT at $T = 25^\circ\text{C}$?

4. Gauss's law by summing Coulombic interactions. Derive Equations (20.21) and (20.23) from Coulomb's law by integrating the appropriate field component.

5. The field around DNA. Consider a line charge with the linear charge density of the DNA double helix, $\lambda = 2e$ per 3.4 \AA . What is the electric field in V m^{-1} at a radial distance of 30 \AA in water (dielectric constant $D = 80$)?

6. The field inside a spherical shell. What is the electric field inside a uniformly charged spherical shell?

7. The field of a sphere. Consider a uniformly charged spherical volume with radius R and space charge density ρ . What is the electric field for $r < R$ and for $r > R$?

8. The field of a cylinder. A cylinder with radius R is uniformly filled with a charge of space density ρ . What is the electric field inside the cylinder? Neglect end effects.

9. Charges affect chemical isomers. 1,2-Dibromoethane is a molecule that can be in either the trans form or the gauche form, as shown in Figure 20.14.

other $-\sigma$. Extend Example 20.7 to show that the field outside a parallel-plate capacitor vanishes, in accordance with Gauss's law.

Table 20.2 Equilibrium constants $N_{\text{gauche}}/N_{\text{trans}}$ for the trans/gauche isomerism of 1,2-dibromoethane at 25°C .

Medium	D	$N_{\text{gauche}}/N_{\text{trans}}$
Gas	1	0.164
<i>n</i> -Hexane	1.9	0.251
Cyclohexane	2.0	0.260
CCl_4 , dilute	2.2	0.307
CS_2	2.7	0.453
1,2-Dibromoethane, pure liquid	4.8	0.554
Methanol	33	0.892

Source: JE Leffler and E Grunwald, *Rates and Equilibria of Organic Reactions*, Wiley, New York, 1963 (reprinted by Dover Publications, New York, 1989).

11. Small charge cluster. Based on the NaCl crystal lattice of Example 20.3, is a single Na^+ surrounded by its six nearest neighboring Cl^- ions a stable structure?

12. Bjerrum length. Compute the Bjerrum length ℓ_B , in \AA , for alcohol ($D = 33$) at $T = 300\text{ K}$.

References

- [1] RP Feynman, *The Feynman Lectures on Physics*, Volume II: *Mainly Electromagnetism and Matter*, Addison-Wesley, Reading, MA, 1964.

Suggested Reading

Excellent introductory texts on electrostatics:

- D Halliday, R Resnick, and J Walker, *Fundamentals of Physics Extended*, 9th edition, Wiley, New York, 2010.
 EM Purcell, *Electricity and Magnetism* (Berkeley Physics Course, Volume 2), 2nd edition, McGraw-Hill, New York, 1984.

More advanced treatments:

- R Becker, *Electromagnetic Fields and Interactions*, Blaisdell, New York, 1964 (reprinted by Dover Publications, New York, 1982).
 IS Grant and WR Phillips, *Electromagnetism*, 2nd edition, Wiley, New York, 1990.
 P Lorrain, DR Corson, and F Lorrain, *Electromagnetic Fields and Waves*, 3rd edition, WH Freeman, New York, 1988.

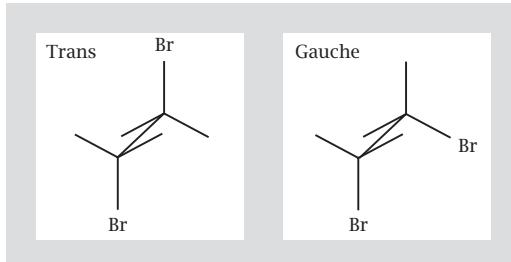
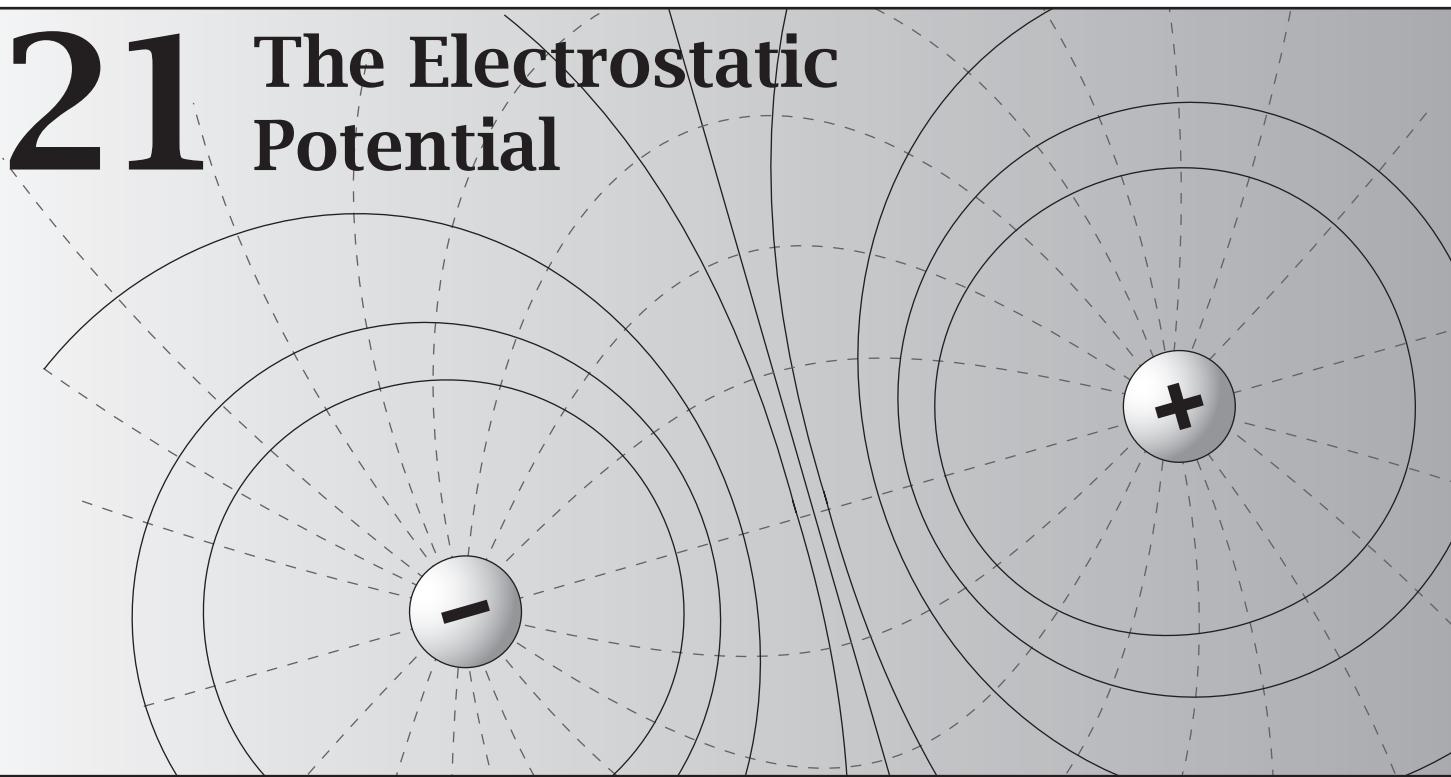


Figure 20.14

Using Table 20.2, which describes the relative populations of gauche/trans conformations in solvents having different dielectric constants D , determine whether bromines attract or repel each other.

10. The field of a parallel-plate capacitor. A parallel-plate capacitor is an arrangement of two parallel conducting plates that carry opposite surface charges, one σ , the

21 The Electrostatic Potential



We Switch from Fields (Forces) to Electrostatic Potentials (Energies)

The *electrostatic field*, introduced in Chapter 20, describes forces, which are vector quantities. In this chapter, we switch to the *electrostatic potential*, which describes energies, which are scalar quantities. We switch for two reasons. First, scalars are easier to use than vectors. Second, our aim in Chapters 22 and 23 is to predict equilibria in systems in which charges play a role. To predict equilibria, we need energies.

We also develop here Poisson's equation. Poisson's equation simplifies electrostatics calculations in many problems. We will use it in Chapter 22 to compute the energy for moving a charged particle, such as a salt ion or protein, in the vicinity of other charged objects such as proteins, colloids, DNA, polyelectrolytes, membranes, electrodes, or dielectric boundaries.

What Is the Electrostatic Potential?

The work $d\omega$ that you must perform to move a charge q through a small distance $d\ell$ in the presence of a fixed electrostatic field \mathbf{E} is the negative of the dot product of the force $\mathbf{f} = q\mathbf{E}$ and the displacement $d\ell$ (see Equation (3.8)):

$$\delta\omega = -\mathbf{f} \cdot d\ell = -q\mathbf{E} \cdot d\ell. \quad (21.1)$$

There is a minus sign in Equation (21.1) because this work is performed *against* the field, not *by* the field as in Equation (3.8). To move a charge from point *A* to *B*, the total work w_{AB} is given by the *path integral* (see Appendix G),

$$w_{AB} = -q \int_A^B \mathbf{E} \cdot d\boldsymbol{\ell}. \quad (21.2)$$

By equating w_{AB} with the electrostatic energy, we are assuming, according to the First Law of Thermodynamics, processes that involve no heat exchange. In such cases, w_{AB} is the maximum possible work and is called the *reversible work*.

We define the difference in the **electrostatic potentials** ψ_B and ψ_A as the work w_{AB} of moving a unit test charge q_{test} from point *A* to point *B*, divided by the test charge q_{test} :

$$\psi_B - \psi_A = \frac{w_{AB}}{q_{\text{test}}} = - \int_A^B \mathbf{E} \cdot d\boldsymbol{\ell}. \quad (21.3)$$

If you know the field \mathbf{E} due to some constellation of charges, Equation (21.3) tells you how to compute the electrostatic potential difference for moving a test charge from point *A* to *B* within that field.

The Electric Field Is the Gradient of the Electrostatic Potential

To find the electric field, given the potential, put the integral, Equation (21.3), into the form of a differential equation. To do this, first express Equation (21.3) in terms of its vector components:

$$\begin{aligned} \psi_B - \psi_A &= - \int_A^B \mathbf{E} \cdot d\boldsymbol{\ell} \\ &= - \int_{x_A}^{x_B} E_x dx - \int_{y_A}^{y_B} E_y dy - \int_{z_A}^{z_B} E_z dz. \end{aligned} \quad (21.4)$$

Now convert from an integral to a differential equation. Suppose points *A* and *B* are very close together: *A* is at (x, y, z) and *B* is at $(x + \Delta x, y, z)$, so $dy = dz = 0$. Then Equation (21.4) gives

$$\psi_B - \psi_A = \Delta\psi = - \int_{x_A}^{x_B} E_x dx = -E_x \Delta x. \quad (21.5)$$

At the same time, the Taylor series for $\Delta\psi$ gives

$$\Delta\psi = \left(\frac{\partial\psi}{\partial x} \right) \Delta x. \quad (21.6)$$

Comparing Equations (21.5) and (21.6) shows that

$$E_x = - \left(\frac{\partial\psi}{\partial x} \right). \quad (21.7)$$

By the same reasoning, taking B at $(x, y + \Delta y, z)$ and at $(x, y, z + \Delta z)$, you will find that $E_y = -(\partial\psi/\partial y)$ and $E_z = -(\partial\psi/\partial z)$. In more compact notation (see Appendix G, Equation (G.3)),

$$\mathbf{E} = -\nabla\psi. \quad (21.8)$$

The electric field is the negative of the gradient of the electrostatic potential ψ .

To compute the potential, it is customary to put the starting point A at an infinite distance from the end point B and to set $\psi_\infty = 0$, so the potential difference is $\psi_B - \psi_A = \psi_B$. The *electrostatic potential* is defined as the work *per unit charge*, so it is *not an energy*, and does not have units of energy. The electrostatic potential *multiplied by charge* is an energy. Like electrostatic forces and fields, electrostatic potentials are additive. They are the sums of the potentials from the individual fixed charges, no matter how the charges are distributed.

The Electrostatic Potential Around a Point Charge

Consider the simplest possible charge constellation, namely a single fixed point charge q_{fixed} . How much work is required to bring a test charge from far away to a distance r from the fixed point charge? The potential only varies along a single coordinate, the radial direction, so you can dispense with vector notation for this problem. That is, $\mathbf{E} \cdot d\ell = E dr$ because the field \mathbf{E} and the vector $d\ell$ are parallel to each other (both point in the radial direction). You need only multiply the magnitudes of E and dr , then integrate. The field that acts radially from the fixed charge has strength $E(r) = Cq_{\text{fixed}}/Dr^2$, according to Equation (20.14), where D is the dielectric constant. The change in electrostatic potential $\Delta\psi$ upon moving a test charge radially inward toward the fixed charge from $r' = \infty$ to $r' = r$ is (see Figure 21.1)

$$\psi(r) = - \int_{\infty}^r E dr' = - \frac{Cq_{\text{fixed}}}{D} \int_{\infty}^r \frac{1}{(r')^2} dr' = \frac{Cq_{\text{fixed}}}{Dr} = \frac{q_{\text{fixed}}}{4\pi\epsilon_0 Dr}. \quad (21.9)$$

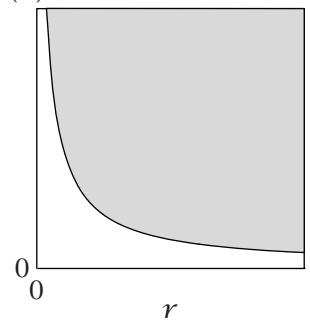
where $\Delta\psi = \psi(r) - \psi(\infty) = \psi(r)$ because the electrostatic potential is $\psi(\infty) = 0$ at $r' \rightarrow \infty$. The prime ' is used to indicate that r' is the variable of integration that ranges from r to ∞ . If the test particle has charge q_{mobile} , then the work of bringing it to a distance r from the fixed particle is $w = q_{\text{mobile}}\psi$ (see Equation (21.3)).

EXAMPLE 21.1 Compute the work of bringing two charges together. Bring a magnesium ion of valence $z = +2$ from $r = \infty$ to a distance $r = 10\text{\AA}$ away from a sodium ion of valence $z = +1$. How much work w must you do? Use Equations (21.3) and (21.9):

$$\begin{aligned} \frac{w}{RT} &= \frac{q_{\text{mobile}}}{RT} [\psi(10\text{\AA}) - \psi(\infty)] \\ &= \frac{q_{\text{mobile}}}{RT} \left(\frac{Cq_{\text{fixed}}}{Dr} \right) = \frac{2\ell_B w}{r} = \frac{2 \times 7.13\text{\AA}}{10\text{\AA}} = 1.43. \end{aligned} \quad (21.10)$$

(a) Positive Charge

$$\psi(r)$$



(b) Negative Charge

$$\psi(r)$$

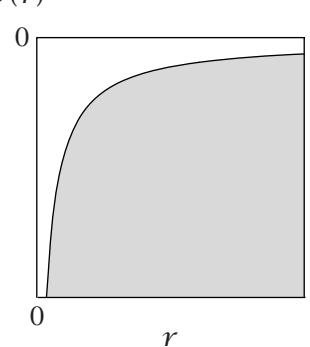


Figure 21.1 The electrostatic potential $\psi(r)$ as a function of the distance r from (a) a positive point charge and (b) a negative point charge.

(In this and other cases below, we simplify the units by using $\ell_{Bw} = \mathcal{C}e^2 \mathcal{N} / (D_w RT)$ from Equation (20.8).) So, the work is $w = 1.43 \times 0.6 \text{ kcal mol}^{-1} = 0.86 \text{ kcal mol}^{-1}$.

If instead of a single charge q_{fixed} , you have a fixed charge density $\rho_{\text{fixed}} = \sum_i q_i / V$ in a volume V around a point 1, the electrostatic potential ψ_2 at point 2 is

$$\psi_{\text{test}} = \frac{1}{4\pi\epsilon_0 D} \int_V \frac{\rho_{\text{fixed}}}{r_{12}} dV, \quad (21.11)$$

where r_{12} is the distance between all the pairs of fixed and test charges.

What Are Electrostatic Potential Surfaces?

To illustrate the idea of electrostatic potential, consider a distribution of charges in a plane. Positive charges tend to move toward regions of negative electrostatic potential. Negative charges spontaneously move toward regions of positive electrostatic potential.

Think of the contours of an electrostatic potential surface in the same way you think of the contours of altitude on a geological map. A geological map plots lines of constant altitude $z(x, y)$ as a function of north-south and east-west coordinates (x, y) . Such maps describe the gravitational potential energy field. If you walk from coordinate (x_1, y_1) to coordinate (x_2, y_2) , your altitude increases from z_1 to z_2 . In this analogy, the positions of the mountains are *fixed* and the mass that you carry up the mountain is called the *test particle*. Multiplying that mass by the altitude difference gives the gravitational energy change: mass \times gravitational acceleration constant \times height. We do the same calculation here with charges and electrostatic potentials, rather than with masses and altitudes. No work is required to move charges along contours of constant electrostatic potential, just as no gravitational work is done in walking along level paths on a hillside. For a point charge, a surface of constant electrostatic potential is any sphere of radius r centered on the point. It requires no work to move a test charge anywhere on that sphere (see Figure 21.2). The electric field \mathbf{E} is always perpendicular to an equipotential surface.

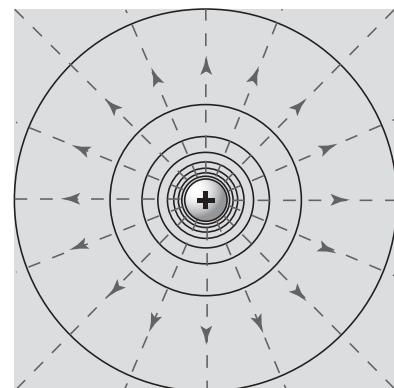


Figure 21.2 The electrostatic potential is constant at distance r from a point charge. Lines of equipotential are circles in two dimensions. Broken lines indicate the direction of force and the field \mathbf{E} .

The Equipotential Surfaces Around Two Positive Point Charges

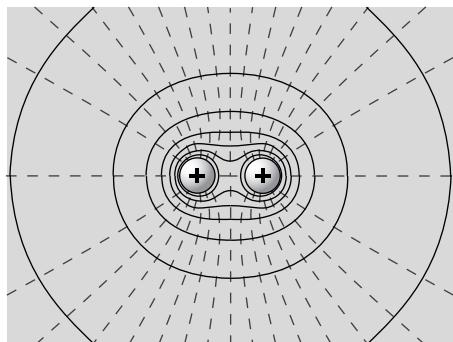


Figure 21.3 Lines of equal electrostatic potential around two positive point charges.

The dashed lines in Figure 21.3 show the electric field \mathbf{E} in the (x, y) plane due to two charges of the same sign: q at $x = -\ell/2$, and q at $x = +\ell/2$. The x axis is an axis of symmetry of this electric field and charge constellation. The solid curves indicate the intersection of the plane of the page and the equipotential surfaces. If you rotate the solid curves about the x axis, out of the plane of the page, you generate the full three-dimensional equipotential surfaces. The equipotential surfaces become more nearly perfect spheres as their distance from the charges increases. Therefore, at a distance far away from the point charges, you can view the electrostatic potential as though it were due to a single point charge equal to $2q$ at $x = 0$.

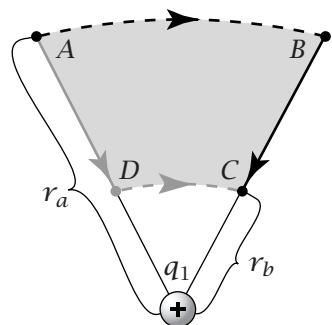


Figure 21.4 The same amount of work is required to move a charge from point A to D to C as from point A to B to C , because work is done only in the radial steps.

Electrostatic Interactions Are Conservative Forces

The reversible electrostatic work is a path-independent quantity that sums to zero around a cycle. Figure 21.4 shows two different pathways for moving a charge q_2 from point A to point C in the field of a single point charge q_1 . Path 1 is from A to B to C . Path 2 is from A to D to C . Moving a particle along circumference segments AB and DC involves no work because there is no change in radius. Both radial segments, AD and BC , involve the same change in radius, so the reversible work in moving a charge along either pathway ABC or ADC is the same:

$$w = -q_2 \frac{C}{D} \int_{r_a}^{r_b} \frac{q_1}{r^2} dr = \frac{Cq_1q_2}{D} \left(\frac{1}{r_b} - \frac{1}{r_a} \right). \quad (21.12)$$

To compute the reversible work between any two points A and C , you can use any sequence of radial and equipotential segments. The work along the equipotential curves will be zero, and the work along the radial segments will be given by Equation (21.12). Figure 21.5 illustrates that the same ideas apply to more complex systems, in this case the protein superoxide dismutase. The reversible work of moving a charge around a cycle from point A to B to C and back to A is zero.

Now we calculate the electrostatic potential that a charge experiences if it is located between two charged planes.

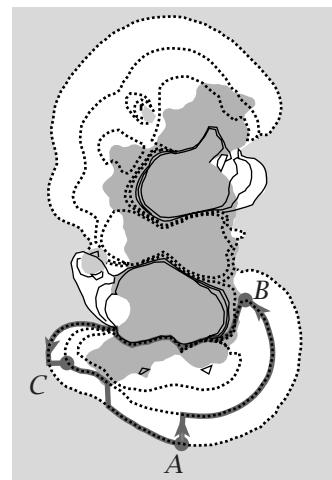
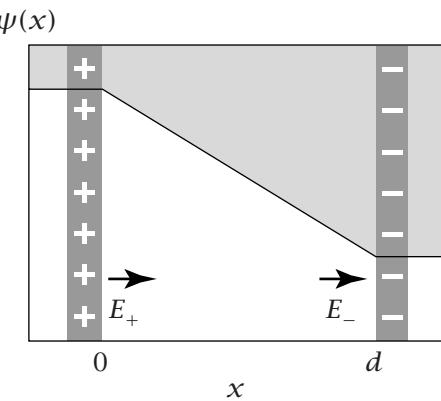


Figure 21.5 Even with an array as complex as the charges in the protein superoxide dismutase (gray), no work is done in moving a charge in a cycle. One such cycle is shown by the heavy line. Dotted lines are equipotential contours. Source: I Klapper, R Hagstrom, R Fine, et al., *Proteins: Structure, Function and Genetics* 1, 47–59 (1986).

EXAMPLE 21.2 The electrostatic potential in a parallel-plate capacitor. A capacitor is a device made of two conductors separated by an insulator. It stores positive charge on one conductor and negative charge on another. Often the conductors are arranged as two parallel charged planes.

We want to compute the work of moving a charge from one thin plate to the other in the capacitor shown in Figure 21.6. This calculation is also useful for treating the energies of ions that approach surfaces and membranes. Suppose that the positive plate has $+\sigma$ charges per unit area, giving an electrostatic field E_+ , and the negative plate has $-\sigma$ charges per unit area and electrostatic field E_- . The plates are separated by a distance d . Each square plate has area A . If d is much smaller than \sqrt{A} , you can neglect edge effects. To find the work of moving a charge, first determine the total electric field E_{inside} between the plates.

Figure 21.6 The electrostatic potential $\psi(x)$ as a function of the distance x between two charged planes at $x = 0$ and $x = d$. The negative slope indicates that a positive charge moves to the right to go energetically downhill. Because a positive charge between the plates feels a field E_+ pushing it to the right and a field E_- pulling to the right, the total field is the sum from both planes, $E_{\text{inside}} = E_+ + E_-$.



In Equation (20.23), we used Gauss's law to find that the field due to a thin charged plate is equal to $\sigma/2\varepsilon_0 D$, so we have $E_+ = \sigma/2\varepsilon_0 D$ acting in the $+x$ direction (the positive plate pushing a positive charge) and $E_- = \sigma/2\varepsilon_0 D$ also acting in the $+x$ direction (the negative plate pulling a positive charge). The total field inside the capacitor in Figure 21.6 acting to drive a positive charge to the right is the vector sum of the fields from each plate: $E_{\text{inside}} = E_+ + E_- = \sigma/\varepsilon_0 D$. The potential is (see Equation (21.3))

$$\Delta\psi = - \int E_{\text{inside}} dx' = - \int_0^x \left(\frac{\sigma}{\varepsilon_0 D} \right) dx' = - \frac{\sigma x}{\varepsilon_0 D}. \quad (21.13)$$

For moving a positive charge from $x = 0$ to $x = d$, $\Delta\psi = -\sigma d/D\varepsilon_0$. Equation (21.13) implies that it takes work to move a positive charge toward the positive plane in a parallel-plate capacitor, and the work required increases with charge density and plate separation.

The field outside the plates is zero because a positive charge that is located to the left of the positive plate, for example, is pulled to the right by E_- but pushed to the left by E_+ . The capacitance C_0 is defined as the total amount of charge separation, $A\sigma$, per unit of potential difference:

$$C_0 = \frac{A\sigma}{|\Delta\psi|} = \frac{A\varepsilon_0 D}{d}. \quad (21.14)$$

Capacitors can be used to measure the dielectric constants D of materials. To do this, you first measure the capacitance of a capacitor of given area A and separation d having only air or vacuum between the plates ($D = 1$). Then fill the space between the plates with the material of interest and measure the capacitance again. Equation (21.14) says that the ratio of the measured capacitances gives the dielectric constant.

Biological membranes are capacitors. Let's calculate the capacitance of the membrane of a nerve cell.

EXAMPLE 21.3 The capacitance of a nerve cell membrane. Assume that a nerve cell is a long cylinder enclosed by a thin planar bilayer membrane of lipids (see Figure 21.7). You can treat the lipid bilayer as a parallel-plate capacitor. Lipid bilayers have oil-like interiors, for which we will assume $D = 2$, and thickness approximately $d \approx 20\text{ \AA}$. Equation (21.14) gives the capacitance per unit area as

$$\frac{C_0}{A} = \frac{\epsilon_0 D}{d} = \frac{(8.85 \times 10^{-12} \text{ F m}^{-1})(2)}{20\text{ \AA}} \left(\frac{1\text{ m}}{10^2\text{ cm}} \right) \left(\frac{10^8\text{ \AA}}{\text{cm}} \right)$$

$$= 8.85 \times 10^{-7} \text{ F cm}^{-2} \approx 0.9 \mu\text{F cm}^{-2}. \quad (21.15)$$

This gives a reasonable model for the capacitance of nonmyelinated nerve cells, which is about $1 \mu\text{F cm}^{-2}$ (for the units, see the box on page 386).

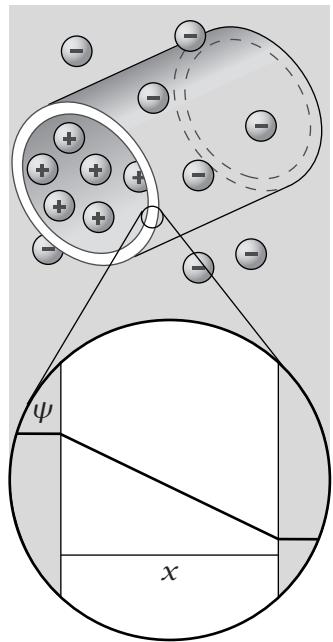


Figure 21.7 The cell membrane of a neuron acts as a capacitor. The membrane is composed of lipids that, like oil, have a low dielectric constant. If the inside and outside solutions have different electrostatic potentials, there is a gradient of electrostatic potential across the membrane.

Dipoles Are Equal and Opposite Charges Separated by a Distance

A dipole is an arrangement of charges $+q$ and $-q$ separated by a distance ℓ (see Figure 21.8). (If you had a more complex system, say a charge of -3.6 at one end and $+5.2$ at the other end, you would take it to be a dipole of -3.6 at one end and $+3.6$ at the other end, plus an additional charge of 1.6 at one end; see Chapter 24. Here we just focus on the dipole component.) Dipoles are oriented in space, indicated by their vector ℓ pointing from $-q$ to $+q$. The **dipole moment** is a vector

$$\boldsymbol{\mu} = q\ell. \quad (21.16)$$

Now we calculate the work of orienting an electric dipole.

EXAMPLE 21.4 The energy of a dipole in an electric field. Figures 21.9 and 21.10 show a dipole with its center fixed in space. The dipole is subject to an orienting force from an electric field E . Compute the work w that the field performs in rotating the dipole from angle 0 (parallel to the field) to an angle θ :

$$w = \int \mathbf{f} \cdot d\ell. \quad (21.17)$$

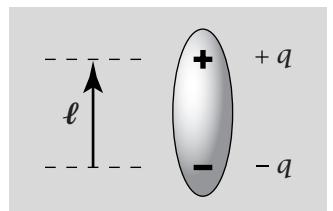


Figure 21.8 A dipole.

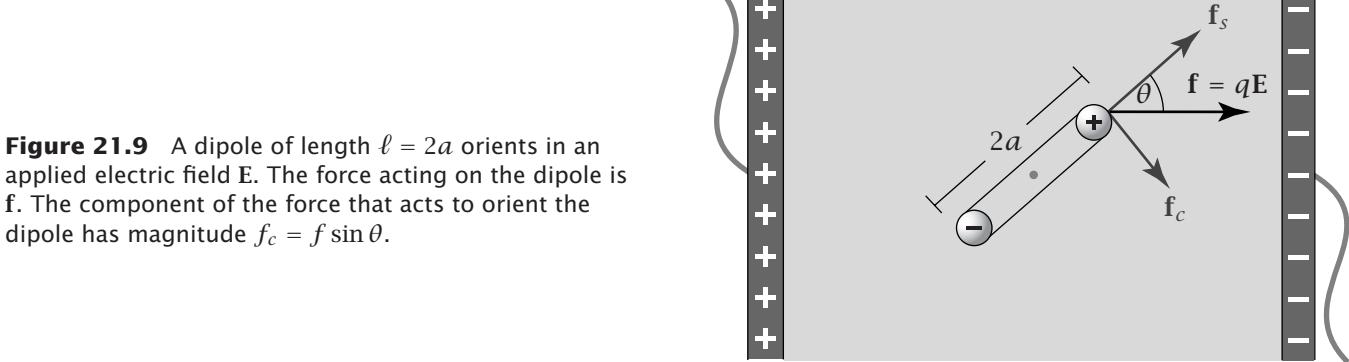
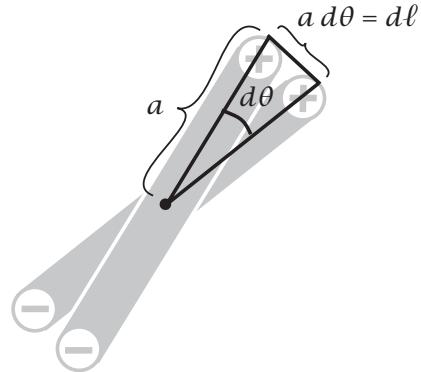


Figure 21.9 A dipole of length $\ell = 2a$ orients in an applied electric field E . The force acting on the dipole is f . The component of the force that acts to orient the dipole has magnitude $f_c = f \sin \theta$.

Figure 21.10 Displacement due to dipole orientation. The dipole of length $2a$ changes angle by an amount $d\theta$, so $d\ell = a d\theta$.



The magnitude of the force on a charge q acting in the direction of the field is qE . The force can be decomposed into two vector components: \mathbf{f}_c acting in the circumferential direction to rotate the dipole and \mathbf{f}_s acting to stretch the dipole.

We are interested only in \mathbf{f}_c . As Figure 21.9 indicates, this has magnitude $f_c = f \sin \theta = Eq \sin \theta$. What is the charge displacement $d\ell$ in the circumferential direction? If a point at radial distance a from a pivot point in the middle of the dipole is rotated through an angle $d\theta$, the circumferential displacement is $a d\theta$ (see Figure 21.10). Because the force acts in the opposite direction of increasing θ , it introduces a minus sign and we have

$$\mathbf{f} \cdot d\ell = -(f_c)(a d\theta) = -Eq a \sin \theta d\theta. \quad (21.18)$$

To get the work of rotating the dipole from angle 0 to θ , integrate Equation (21.18), and multiply by 2 because there is an equal torque acting on each end of the dipole:

$$w = -2Eq a \int_0^\theta \sin \theta' d\theta' = Eq \ell (\cos \theta - 1) = E\mu (\cos \theta - 1). \quad (21.19)$$

Equation (21.19) is useful for modeling polar molecules orienting in electric fields and gating in ion channels, described below.

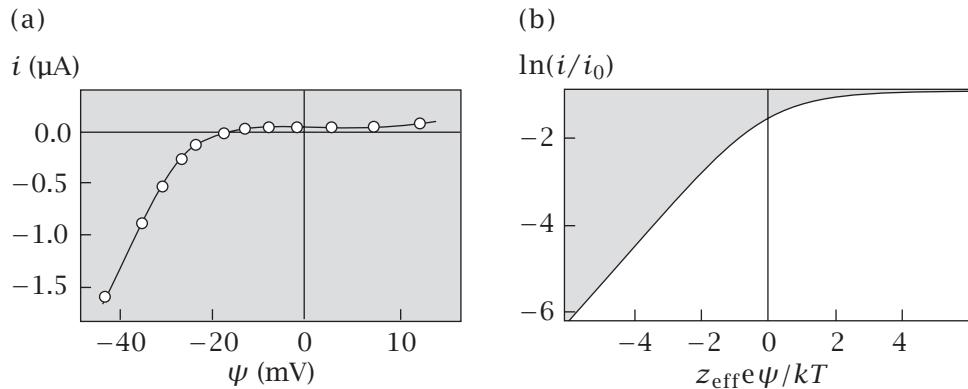


Figure 21.11 (a) Current–voltage measurements for a potassium channel in the egg cell membrane of a starfish. (b) Relationship between the current i and the applied voltage ψ from Equation (21.23). Source: (a) B Hille, *Ionic Channels of Excitable Membranes*, Sinauer, Sunderland, MA, 1984. Data are from S Hagiwara, S Miyazaki, and NP Rosenthal, *J Gen Physiol* **67**, 621–638 (1976). (b) TW Weiss, *Cellular Biophysics*, Volume 2: *Electrical Properties*, MIT Press, Cambridge, MA, 1996.

Voltage-Gated Ion Channels May Be Modeled as Orienting Dipoles

Ions sometimes cross a biological membrane in one direction but not the other. Figure 21.11(a) shows experimental evidence that applying an external voltage opens a protein channel to allow ions to pass through. Here is a simple model for voltage-gated ion conductance, based on the assumption that the applied field orients a dipole within the protein channel (see Figure 21.12).

The equilibrium constant for channel opening can be expressed in terms of the ratio of the probabilities that the channel is in the open or closed states:

$$\frac{p_{\text{open}}}{p_{\text{closed}}} = e^{-(\Delta G_0 - w)/RT}, \quad (21.20)$$

where ΔG_0 is the free energy required to open the channel in the absence of the applied field and w is the work performed by the applied electrostatic field in orienting the dipole.

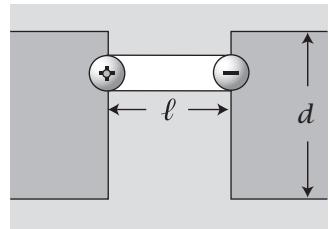
Applying an electrostatic field can shift the equilibrium toward the open state. According to Equation (21.19), the work performed by a field in aligning a dipole is

$$w(\theta) = Eq\ell(\cos \theta_{\text{open}} - \cos \theta_{\text{closed}}) = -\frac{\psi q\ell}{d} (\Delta \cos \theta) \quad (21.21)$$

(ψ is the potential of the lower solution relative to 0 in the upper solution). We used $E = -\psi/d$ to convert from the field E to the voltage difference across the membrane because the solutions on the two sides of the membrane act as parallel-plate electrodes across a membrane of thickness d . Because the quantities $\Delta \cos \theta = \cos \theta_{\text{open}} - \cos \theta_{\text{closed}}$, q , and ℓ may not be known independently, we can reduce the number of unknown parameters by collecting them into a quantity

$$z_{\text{eff}} e = \frac{-q\ell}{d} (\Delta \cos \theta). \quad (21.22)$$

(a) No Field:
Closed Channel



(b) Applied Field:
Open Channel

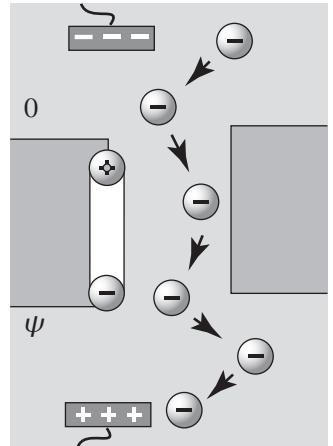


Figure 21.12 A voltage-gated channel. (a) Gate closed. (b) Applied potential opens gate.

z_{eff} has units of charge. It is a charge multiplied by a dimensionless fractional distance: how far the charge moved down the bilayer divided by the full membrane thickness. We introduced the minus sign in Equation (21.22) because we are free to define z_{eff} any way we want, and this simplifies the interpretation of the signs. In particular, any favorable arrangement of the dipole (negative end near the positive electrode, for example, as shown in Figure 21.12) leads to a negative value of $q\psi$ or a positive value of z_{eff} .

The ion flow through the channel is proportional to the fraction of time that the channel is open, which you can compute by substituting Equation (21.22) into Equation (21.21) and then into Equation (21.20), to get

$$\frac{p_{\text{open}}}{p_{\text{open}} + p_{\text{closed}}} = \frac{1}{1 + (p_{\text{closed}}/p_{\text{open}})} = \frac{1}{1 + e^{(\Delta G_0 - z_{\text{eff}}\psi)/RT}}. \quad (21.23)$$

Figure 21.11(b) shows the prediction of Equation (21.23) that the gate opens for positive applied voltages, and becomes increasingly closed for negative applied voltages, as is commonly observed in experiments.

EXAMPLE 21.5 The potential field around a dipole. Consider the electrical dipole shown in Figure 21.13. It is formed by charges $-q$ at $x = -\ell/2$ and $+q$ at $x = +\ell/2$, so the dipole moment vector points in the $+x$ direction. Here we will determine the electrostatic potential at any point in the (x, y) plane. The complete three-dimensional potential field is found by rotating the equipotential lines about the x axis.

Let's find the potential at any arbitrary point $P(x, y)$ at distance $r = (x^2 + y^2)^{1/2}$ from the origin in Figure 21.13(b). The vector \mathbf{r} is at an angle θ with respect to the x axis. Use Equation (21.9) to define the potential at P from each of the two fixed point charges. The total electrostatic potential ψ is the sum

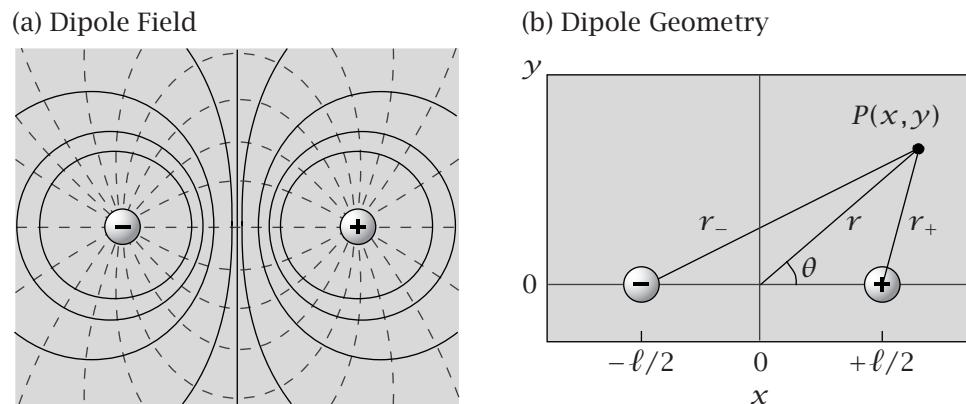


Figure 21.13 (a) Electric field (---) and equipotential lines (—) for a dipole in the (x, y) plane. (b) Definition of quantities r_+ and r_- in Equations (21.25) and (21.26).

of the potentials from each charge of the dipole:

$$\psi = -\frac{Cq}{Dr_-} + \frac{Cq}{Dr_+}. \quad (21.24)$$

The distances r_- from $-q$ and r_+ from $+q$ are found by considering the right triangles formed by drawing a perpendicular line from point P to the x axis. You can compute r_-^2 using Figure 21.13(b) and the expression $r^2 = x^2 + y^2$:

$$\begin{aligned} r_-^2 &= \left(x + \frac{\ell}{2}\right)^2 + y^2 = x^2 + 2\left(\frac{\ell x}{2}\right) + \left(\frac{\ell}{2}\right)^2 + y^2 \\ &= r^2 + \ell x + \left(\frac{\ell}{2}\right)^2 \approx r^2 \left(1 + \frac{\ell x}{r^2}\right), \end{aligned} \quad (21.25)$$

where the approximation arises from neglecting the second-order term $(\ell/2)^2$ in Equation (21.25). This applies when P is far away from the dipole, $r \gg \ell$. The same logic gives r_+^2 in terms of r :

$$r_+^2 = \left(x - \frac{\ell}{2}\right)^2 + y^2 \approx r^2 \left(1 - \frac{\ell x}{r^2}\right). \quad (21.26)$$

Substitute r_+ and r_- from Equations (21.25) and (21.26) into Equation (21.24):

$$\psi \approx \frac{Cq}{Dr} \left[-\left(1 + \frac{\ell x}{r^2}\right)^{-1/2} + \left(1 - \frac{\ell x}{r^2}\right)^{-1/2} \right]. \quad (21.27)$$

The terms in parentheses in Equation (21.27) can each be approximated with the first term in the series expansion for $(1+x)^{-1/2} \approx (1-(x/2)+\dots)$, given in Equation (J.6) of Appendix J:

$$\psi \approx \frac{Cq}{Dr} \left(-1 + \frac{\ell x}{2r^2} + 1 + \frac{\ell x}{2r^2} \right) = \frac{Cq}{Dr} \left(\frac{\ell x}{r^2} \right). \quad (21.28)$$

Collecting terms and using $x/r = \cos \theta$, you have

$$\psi \approx \frac{Cq\ell}{Dr^2} \cos \theta = \frac{C\mu \cos \theta}{Dr^2}, \quad (21.29)$$

where $\mu = q\ell$ is the magnitude of the dipole moment. When $0 < \theta < \pi/2$, $\cos \theta$ is positive, so ψ is positive, indicating that a positive charge at P will be repelled by the dipole, because it is closest to the $+$ end of the dipole. Because $\mu = q\ell$ is the dipole moment, you can write Equation (21.29) in terms of the dot product of the dipole moment vector μ and the radius vector \mathbf{r} :

$$\psi = \frac{C\mu \cdot (\mathbf{r}/r)}{Dr^2}. \quad (21.30)$$

Lines of constant ψ surrounding a dipole are shown in Figure 21.13(a) and those surrounding two charges of the same sign in Figure 21.3.

Now we use Equation (21.29) to determine the energy of interaction between a charged ion and a dipolar molecule.

Interactions Between a Charge and a Dipolar Molecule

An ion is located at point P and has a point charge Q . Another molecule has a dipole moment of magnitude $q\ell$. The interaction energy $u(r, \theta)$ is the work of bringing the two molecules from infinite separation to a separation r , where θ is the angle of the dipole relative to the line between the centers of the dipole and point charge (see Figure 21.13). From Equations (21.3) and (21.29), you get

$$u(r, \theta) = \psi Q = \frac{c\mu Q \cos \theta}{Dr^2}. \quad (21.31)$$

You would get the same result by summing with Coulomb's law for the charge Q and the two dipolar charges.

The most notable aspect of Equation (21.31) is the dependence on the inverse *square* of the distance, $u \propto 1/r^2$. According to Coulomb's law, Equation (20.1), two ions interact with a dependence $u \propto 1/r$. Replacing one ion by a dipole causes the interaction to become shorter-ranged. At close range, the single charge 'sees' both charges on the dipole, a sum of two $1/r$ terms. As the dipole moves further away, however, the single charge sees the dipole charges as neutralizing each other because they are close together relative to their distance from the ion. So the interaction weakens even more than the $1/r$ dependence would predict. A similar treatment would show that when two dipoles move apart, the interaction becomes even more short-ranged: $u \propto 1/r^3$ at large separations (see Chapter 24).

Poisson's Equation Is Used to Compute the Potential for Any Constellation of Charges

So far, we have computed the electrostatic potentials for very simple constellations of charges. In Chapter 20, we found that we could treat more complex problems of electrostatic forces by using a little vector calculus to find Gauss's law. Now, for energies rather than forces, we resort again to some vector calculus, which will lead us to Gauss's divergence theorem (distinct from Gauss's law) and to Poisson's equation.

Gauss's theorem (Equation (G.15) in Appendix G) says that for any vector field \mathbf{v} ,

$$\int_{\text{surface}} \mathbf{v} \cdot d\mathbf{s} = \int_{\text{volume}} \nabla \cdot \mathbf{v} dV. \quad (21.32)$$

Substituting $\mathbf{v} = D\mathbf{E}$ gives **Gauss's theorem** applied to electric fields:

$$\int_{\text{surface}} D\mathbf{E} \cdot d\mathbf{s} = \int_{\text{volume}} D\nabla \cdot \mathbf{E} dV. \quad (21.33)$$

Gauss's theorem equates the *flux* of the electrostatic field *through a closed surface* with the *divergence* of that same field *throughout its volume*.

Substituting Gauss's law, Equation (20.19),

$$\int_{\text{surface}} D\mathbf{E} \cdot d\mathbf{s} = \int_{\text{volume}} \frac{\rho}{\epsilon_0} dV, \quad (21.34)$$

into Equation (21.33) gives the **differential form of Gauss's law**:

$$D\nabla \cdot \mathbf{E} = \frac{\rho}{\epsilon_0}. \quad (21.35)$$

Equation (21.35) is one of Maxwell's four equations. Now substitute Equation (21.8), $\mathbf{E} = -\nabla\psi$, into Equation (21.35) to get **Poisson's equation**:

$$\nabla^2\psi = -\frac{\rho}{\epsilon_0 D}, \quad \text{where} \quad \nabla \cdot \mathbf{E} = -\nabla^2\psi \quad (21.36)$$

and ∇^2 is called the *Laplacian operator*; see Appendix G.

You now have two methods for deriving the electrostatic potential from a given distribution of fixed electrical charges. First, you can use Coulomb's law. But this becomes unwieldy for complex constellations of charges or for systems with two or more different dielectric media. Second, you can solve Poisson's equation. The Poisson method is more general and usually simpler. It can also be used to treat heterogeneous dielectric media (see page 420).

Examples 21.6 and 21.7 show how to use Poisson's equation to compute the electrostatic potential around a charged sphere and cylinder. For more complex geometries, you usually need to solve Poisson's equation by computer.

EXAMPLE 21.6 A charged spherical shell: Poisson's equation. Consider a thin uniformly charged spherical shell located at radius $r = a$. The shell has a total charge q . The interior of the sphere is uncharged, with dielectric constant D . The shell is immersed in an external medium of the same dielectric constant. This system might model the charged membrane of a spherical vesicle or a biological cell.

Place a test charge inside or outside the sphere and find the electrostatic potential at that point due to the fixed charges on the spherical shell. Divide space into three regions: inside the charged sphere ($r < a$, where there is no net fixed charge), on the spherical shell ($r = a$, where all the charge is located), and outside ($r > a$, where there is also no net fixed charge). Compute the electrostatic potential everywhere by using Poisson's equation (21.36).

The problem has spherical symmetry, so the first step is to look up the spherical form of the vector operators (see Table G.1 in Appendix G). The potential ψ depends only on r and not on the angles θ and ϕ , because the charge is distributed with spherical symmetry. Therefore, $(\partial\psi/\partial\theta) = 0$ and $(\partial^2\psi/\partial\phi^2) = 0$, and $\nabla^2\psi$ in Equation (21.36) becomes

$$\nabla^2\psi = \frac{1}{r^2} \frac{d}{dr} \left(r^2 \frac{d\psi}{dr} \right) = \frac{1}{r} \frac{d^2(r\psi)}{dr^2}. \quad (21.37)$$

Our strategy is to solve Poisson's equation inside and outside the sphere. There is no charge in either region, so $\nabla^2\psi = 0$ for both. Then we'll use Gauss's law to establish a boundary condition for the charged shell at $r = a$, which will allow us to complete the solution.

For both the inside and outside regions, because $\rho = 0$, Poisson's equation is

$$\frac{1}{r} \frac{d^2(r\psi)}{dr^2} = 0 \quad \text{for } r > a \text{ and for } r < a. \quad (21.38)$$

To solve Equation (21.38), multiply by r and integrate once to get

$$\frac{d(r\psi)}{dr} = A_1, \quad (21.39)$$

where A_1 is the first integration constant. Integrating again yields another integration constant and the general solution,

$$r\psi = A_1 r + A_2 \implies \psi(r) = A_1 + \frac{A_2}{r}. \quad (21.40)$$

You can use boundary conditions to get the constants A_1 and A_2 . But here's an easier way, based in this case on the principle that ψ must be continuous at the boundaries. First, outside the sphere ($r > a$), $E_{\text{out}}(r) = Cq/Dr^2$ (see Equation (20.14)). This equation was derived for a point charge q , but it applies to any spherical charge distribution because a Gaussian balloon of radius $r > a$ fully contains the charge (see Figure 21.14(a)). Use Equation (21.3) to get

$$\psi_{\text{out}}(r) = - \int_{\infty}^r E_{\text{out}}(r') dr' = \frac{Cq}{Dr} \quad \text{for } r > a. \quad (21.41)$$

Now to get $\psi_{\text{in}}(r)$, put a spherical Gaussian balloon inside $r < a$ (see Figure 21.14(b)). There is no charge inside this balloon because all the charge is on the shell at $r = a$, so $E_{\text{in}}(r) = 0$, and Equation (21.41) gives

$$\psi_{\text{in}}(r) = - \int_{\infty}^r E_{\text{in}}(r') dr' = \text{constant}. \quad (21.42)$$

(a) All the Charge is Contained in the Gaussian Balloon

(b) No Charge is Contained in the Gaussian Balloon

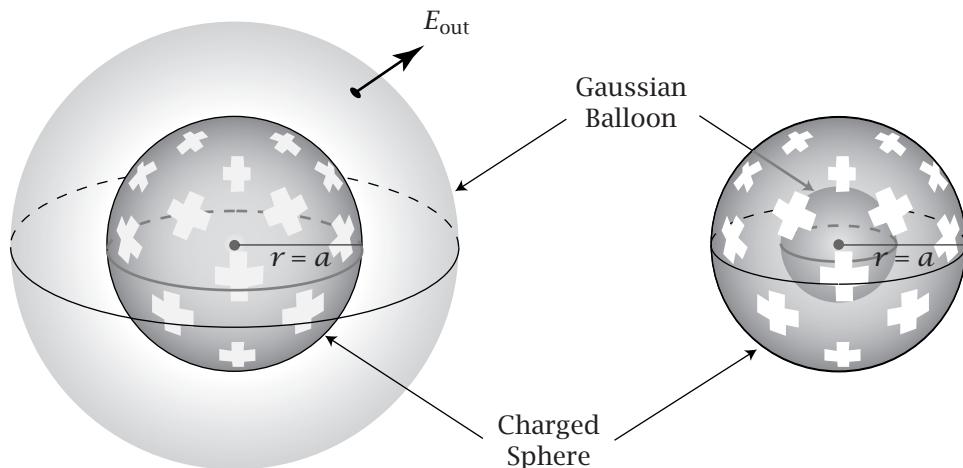


Figure 21.14 (a) To find the potential $\psi(r)$ outside a charged sphere, construct a Gaussian balloon that contains the sphere. (b) To find the potential $\psi(r)$ inside, construct the balloon inside the charged sphere.

Finally, a condition of *continuity* requires that $\Delta\psi = \psi_{\text{out}} - \psi_{\text{in}} = 0$ for an infinitesimal step across the boundary, so comparing Equations (21.41) and (21.42) at $r = a$ gives

$$\psi_{\text{in}} = \frac{Cq}{Da} \quad \text{for } r \leq a. \quad (21.43)$$

This is the electrostatic potential everywhere inside a charged spherical shell.

What is continuity? Use Equation (21.4) to find $\Delta\psi = \psi_{\text{out}} - \psi_{\text{in}} = \psi(a + \delta) - \psi(a - \delta)$, the difference in electrostatic potential across an infinitesimal change δ in radius from the inside to the outside of a sphere (see Figure 21.15):

$$\Delta\psi = \psi_{\text{out}} - \psi_{\text{in}} = - \left[\int_a^{a+\delta} E_{\text{out}}(r) dr - \int_{a-\delta}^a E_{\text{in}}(r) dr \right]. \quad (21.44)$$

Because $E_{\text{in}} = 0$, and $E_{\text{out}} = Cq/Dr^2$, integrating Equation (21.44) gives

$$\Delta\psi = \frac{Cq}{Dr} \Big|_a^{a+\delta} = \frac{Cq}{D} \left(\frac{1}{a+\delta} - \frac{1}{a} \right). \quad (21.45)$$

Equation (21.45) proves continuity because as $\delta \rightarrow 0$, it shows that $\Delta\psi \rightarrow 0$. Continuity holds across an interface of any shape.

The electrostatic potential is constant everywhere inside the sphere, so the field is zero. This is true inside any closed conducting surface. This is why it is a good idea to be inside a car or other conducting container in a lightning storm—there is no electrostatic force inside. This is also why your car radio loses its signal when you are driving inside a metal bridge or tunnel.

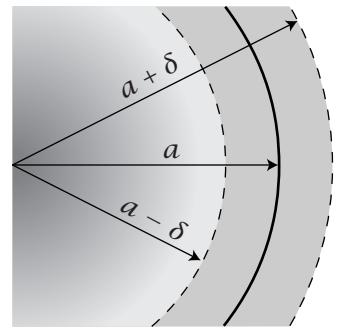


Figure 21.15 The definition of δ for the proof of the continuity principle for a sphere of radius a .

EXAMPLE 21.7 The electrostatic potential around a line charge. Let's compute the electrostatic field around a long charged cylinder of radius a . You can do this in different ways. One way is to use Poisson's equation as in Example 21.6. This problem has cylindrical symmetry, so use the cylindrical form of the vector calculus expressions. For the charge-free region outside the cylinder, Poisson's equation (using Equation (G.21) for a cylinder) is

$$\nabla^2 \psi = \frac{1}{r} \frac{d}{dr} \left(r \frac{d\psi}{dr} \right) = 0. \quad (21.46)$$

Multiply by r and integrate Equation (21.46) once over r to get

$$r \frac{d\psi}{dr} = A_1 \implies \frac{d\psi}{dr} = \frac{A_1}{r}, \quad (21.47)$$

where A_1 is a constant. Integrate $d\psi = (A_1/r) dr$ to get

$$\psi = A_1 \ln r + A_2. \quad (21.48)$$

There's an alternative way to derive Equation (21.48). Equation (20.21) shows that the electrostatic field around a charged cylinder is $E(r) = \lambda/(2\pi\epsilon_0 Dr)$. Equation (21.3) says that you can integrate $E(r)$ over r from the surface at

$r = a$ to get the potential:

$$\Delta\psi(r) = - \int_a^r E(r') dr' = -\frac{\lambda}{2\pi\epsilon_0 D} \ln\left(\frac{r}{a}\right). \quad (21.49)$$

The electrostatic potential around an infinitely long line charge is different from the potential around a point charge in one important respect: $\psi \rightarrow -\infty$ at large distances r away from a line charge. This means that you cannot define a boundary condition $\psi = 0$ at infinite distance.

You Can Model Ions Near Interfaces Using Image Charges

Sometimes you can simplify electrostatics calculations using a mathematical trick called the *method of image charges*. It stems from a property, called *uniqueness*, of differential equations such as those in this chapter [1, 2]. You are given some particular arrangement of charges and boundaries. You want to compute the electrostatic potential. Rather than trying to solve Poisson's equation for the given charge distribution, look for some other imaginary distribution of charges that satisfies the same boundary conditions. According to the uniqueness theorem, if you can find *any* solution to Poisson's equation that satisfies the boundary conditions for the problem of interest, you have found the *only solution*, even if you found it by wild guesses or clever tricks.

Here's an example. Ions are attracted to metal surfaces, *even to surfaces that are uncharged*. What is the force of attraction? The attraction arises from *induction*. The ion induces an electrostatic field in the metal. That field, in turn, attracts the ion to the surface. You can compute this attractive force by combining the image-charge method with a solution to a problem we already solved earlier in this chapter.

Figure 21.16 shows the field lines from a charge Q that sits at a distance a from a conducting surface. To compute the electrostatic potential and the force of attraction, you just need to recognize that all the field lines from the charge must be perpendicular to the plane where they intersect it. (Why is this? It is because if there were some component of the electrostatic field from the ion that were acting laterally within the conducting plane, it would imply lateral

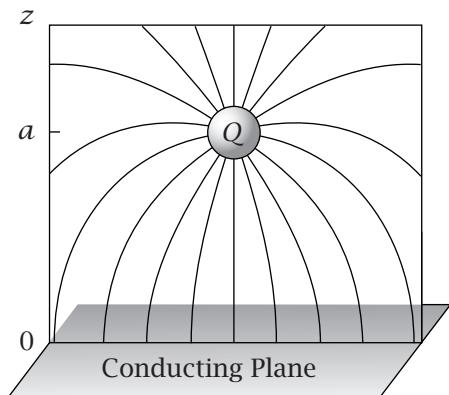


Figure 21.16 The field lines from a point charge Q must intersect a conducting plane at 90° .

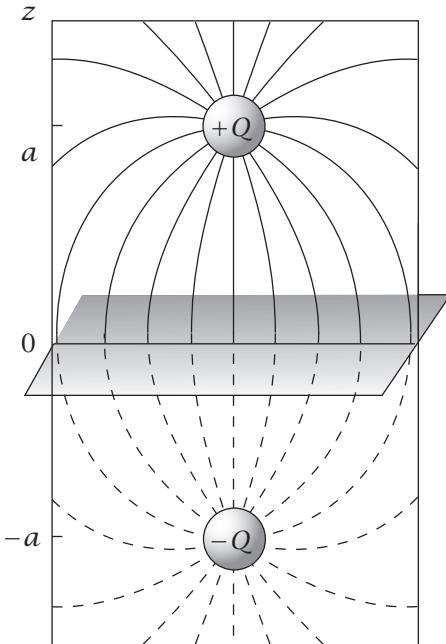


Figure 21.17 The attraction of a charge Q for a conducting plane separated by a distance a can be modeled as though the charge Q were attracted to an image charge, $-Q$ at $z = -a$.

forces, which would drive lateral electrical currents within the metal. But metals and conductors have the property that electrons flow so freely inside them that any unbalanced lateral forces quickly come into balance. So the only field lines intersecting a conductor must be perpendicular to it).

So, if you can find any other fictitious constellation of charges that would cause all the field lines to be parallel to the z axis at a distance a from the point charge, then it will give you exactly the electrostatic potential that you need to solve this problem of a charge near a conducting plane. But, we already solved such a problem in Example 21.5, the potential field around a dipole. Figure 21.13 shows that the plane that bisects a dipole has only perpendicular intersections with the field lines from the two dipolar charges, so this must be the solution we are looking for.

The method of image charges says that you can replace the conducting plane in the present problem by placing an imaginary charge of opposite sign ($-Q$ at $z = -a = -\ell/2$) as though it were the mirror image of the real charge Q , reflected over to the other side of the conducting plane (see Figure 21.17). This pair—the real charge and the imaginary charge—will give exactly the same electrostatic potential as the ion in the presence of the actual conducting plane everywhere on the $z > 0$ side of the conducting plane. With this trick, you now have a very graphic way to view the attraction of the charge Q for the plane. You can think of the charge Q as being attracted to its image $-Q$ on the other side of the plane. With this mirror image trick, you save the trouble of computing the electrostatic potential because you have already done it in Equation (21.24). This is shown in more quantitatively Example 21.8.

EXAMPLE 21.8 The attraction of an ion toward a planar conductor. To compute the attraction of an ion Q for a conducting plane, use Figure 21.13(a). Now the y axis coincides with the conducting plane. The ion is indicated by the

positive charge at $x = \ell/2$, and the image charge is indicated by the negative charge at $x = -\ell/2$. We are interested in the potential $\psi(\ell/2)$ arising from the image charge $-Q$ at $x = -\ell/2$ that acts on the charge at $x = \ell/2$. You can use Equation (21.24) with $r_- = \ell = 2a$, where a is the distance of the ion from the plane. The second term of Equation (21.24) is not appropriate here because it is not a potential that acts on $+Q$. Equation (21.24) gives

$$\psi = -\frac{CQ}{2Da} = -\frac{CQ}{D\ell}. \quad (21.50)$$

The field E is the force per unit charge acting on the particle along the z axis in Figure 21.16:

$$E = -\frac{d\psi}{d\ell} = -\frac{CQ}{D\ell^2} = -\frac{CQ}{4Da^2}. \quad (21.51)$$

The negative sign in Equation (21.51) indicates that a charge Q of either sign is attracted to the conducting plane owing to the induced field. As a gets smaller, the attraction gets stronger.

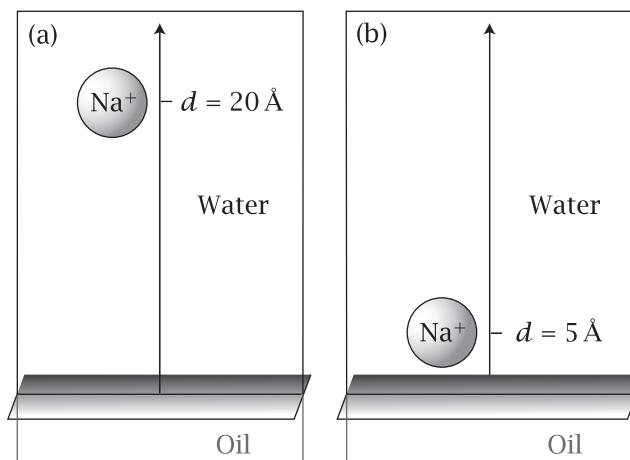
Now let's apply the same reasoning to a charge near a *dielectric* boundary.

A Charge Near a Dielectric Interface

A charged ion in water will be repelled as it approaches an oil/water interface. In the language of image charges, the charge in the water phase sees its own image across the interface in the oil phase (not a negative image this time) and is repelled by it. Why does it see a positive image? And how strong is the repulsion?

Suppose that you have a charge q in water at a distance d from an oil/water interface (see Figure 21.18). The dielectric constant of the medium in which the ion is located is D_{in} . In this case, $D_{in} = D_w \approx 80$. The dielectric constant of the medium across the interface is D_{other} ; in this case, $D_{other} = D_{oil} \approx 2$. If you work through the details (which we won't do here), the key result is that the image,

Figure 21.18 A sodium ion in water is repelled by an oil interface.



which will be at a distance d into the other phase, will have a charge q' :

$$q' = \frac{D_{\text{in}} - D_{\text{other}}}{D_{\text{in}} + D_{\text{other}}} q. \quad (21.52)$$

The electrostatic potential felt by the charge q from the image q' will be

$$\psi = \frac{\mathcal{C}q'}{D_{\text{in}}(2d)}. \quad (21.53)$$

For a charge in water near an oil interface, you have $D_w \gg D_{\text{oil}}$, so $q' \approx q$, and q' has the same sign as q . You can think of the ion as seeing its own reflection in a mirror, and being repelled by the oil interface, driving the ion to dive deeper into the water phase. The image is at a distance $2d$ from the ion. Interestingly, if the other medium across the interface has instead a *higher* dielectric constant than the medium the ion is currently in, Equation (21.52) shows that the ion will be *attracted* to the interface. In the limit $D_{\text{oil}} \rightarrow \infty$, $q' \rightarrow -q$, and the interface attracts as if it were a conducting plane. A charge is attracted to whichever medium has the higher dielectric constant.

EXAMPLE 21.9 A sodium ion in water is repelled by an oil surface. Let's compute the work of moving a sodium ion at room temperature from a distance $d_A = 20 \text{ \AA}$ to a distance $d_B = 5 \text{ \AA}$ from a planar oil interface (see Figure 21.18). Assume that $D_w \gg D_{\text{oil}}$. Then Equation (21.52) gives an image charge $q' \approx q = e$. The work, w , of moving the ion from $d_1 = 20 \text{ \AA}$ to $d_2 = 5 \text{ \AA}$ is (see Equation (21.53))

$$w = e\psi(d_2) - e\psi(d_1) = \frac{\mathcal{C}e^2}{D_w} \left(\frac{1}{2d_2} - \frac{1}{2d_1} \right). \quad (21.54)$$

In terms of the dimensionless Bjerrum length ℓ_{Bw} defined on page 391,

$$\begin{aligned} \frac{w}{RT} &= \ell_{Bw} \left(\frac{1}{2d_2} - \frac{1}{2d_1} \right) = (7.13 \text{ \AA}) \left(\frac{1}{10} - \frac{1}{40} \right) = 0.535, \\ \Rightarrow w &= (0.535)(1.987)(300 \text{ K}) = 318 \text{ cal mol}^{-1}. \end{aligned} \quad (21.55)$$

Summary

A constellation of charges fixed in space generates an electrostatic potential. The potential, due to the fixed charges, describes the energy (per unit charge) for moving a test charge between any two points in space. The electrostatic potential can be computed as a path integral of the force on the test charge as it moves from one point to another. The electrostatic potential can also be computed using Poisson's equation and the appropriate boundary conditions. For problems of mixed symmetry, such as charged spheres or lines that are near planar interfaces, an important tool is the method of image charges. By replacing the actual constellation of charges with a fictitious set of image charges, it is sometimes possible to simplify the calculation of electrostatic fields. In the next chapter, we will combine electrostatics with thermodynamics to consider the equilibria of charged systems.

Problems

1. Compute $\Delta\psi$ for a capacitor. You have a parallel-plate capacitor filled with an oil of dielectric constant $D = 2$, with a distance of 0.20 cm between two $20\text{ cm} \times 20\text{ cm}$ plates. The upper plate has a total charge of 10^{-10} C relative to the lower plate. Neglecting edge effects, what is the potential difference between the upper and lower plates?

2. The potential around an ion. What is the potential at 25°C at a distance of 10 \AA from a monovalent ion in air? And in pure water?

3. A dipole inside a sphere. What is the average potential over a spherical surface that has a dipole at its center?

4. A dipole moment. What is the dipole moment of two charges $-e$ and $+e$ that are 1 \AA apart? (The unit of dipole moment is 1 debye = $1\text{ D} = 3.33 \times 10^{-30}\text{ C m}$.)

5. The charge-dipole interaction of Na^+ with water. What is the maximum attraction in units of RT :

- (a) between a bare Na^+ ion and a water molecule in air?
- (b) between a hydrated Na^+ ion in water and a water molecule?

Take a sphere with radius 0.95 \AA for the bare Na^+ ion and a sphere with radius 2.3 \AA for the hydrated Na^+ ion. Model a water molecule as a sphere with radius 1.4 \AA and a point dipole of moment 1.85 D at its center.

6. Water is a dipole. A water molecule in vacuum has a dipole moment of $0.62 \times 10^{-29}\text{ C m}$.

- (a) Assuming that this dipole moment is due to charges $+e$ and $-e$ at distance ℓ , calculate ℓ in \AA .
- (b) Consider the water molecule as a sphere with radius 1.5 \AA . If the dipole moment in (a) is due to charges $+q$ and $-q$ at the north and south poles of the sphere, how large is q ?

7. The potential around a charged sphere. You have an evenly charged spherical surface with radius a and a total charge q in a medium with dielectric constant D . Derive the potential inside and outside the surface by using Coulomb's law and integrating over the charged surface.

8. The potential around a cylindrical shell. You have a hollow inner cylinder with radius a_1 surrounded by a concentric outer cylinder with radius $a_2 > a_1$. Charges $-\lambda$ and $+\lambda$ per unit length are distributed uniformly over the inner and outer cylinders, respectively. The shells are in a medium with dielectric constant D .

- (a) What is the potential $\psi(r)$ as a function of the axial distance r ?
- (b) What is the capacitance per unit length of the arrangement of cylinders?

9. The work of moving a micelle to an oil/water interface. A spherical micelle in water has a charge $q = -60\text{ e}$.

What is the work required to bring it from deep in a water phase to 100 \AA away from an oil/water interface at $T = 300\text{ K}$?

10. The work of bringing an ion near a protein. An ion has negative charge $Q = -2$. A protein has negative charge $q = -10$. What is the work of bringing the ion from infinite distance to within 10 \AA of the protein at $T = 300\text{ K}$?

11. Water bridging in protein crystals. Protein molecules can crystallize. A puzzling observation is that two protein molecules in the crystal are sometimes separated by a spacer water molecule. However, if proteins are attracted by a dipolar electrostatic interaction, the system should be most stable when the proteins are as close together as possible.

To model this, consider each protein molecule to be spherical with a radius R_o , as shown in Figure 21.19. Each protein has a net dipole moment due to two charges, $+q$ and $-q$ respectively, each a distance d from the center of the protein, and collinear.

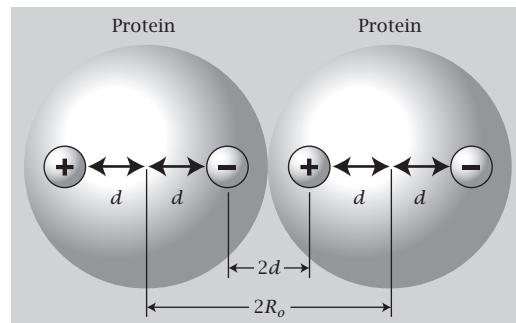


Figure 21.19

(a) Calculate the electrostatic potential for the protein-protein contact pair without the water, with the charges distributed as shown in Figure 21.19. Assume the system is in a vacuum.

(b) Now, calculate the electrostatic potential for the water-bridged protein-protein pair shown in Figure 21.20. To do this, first recognize that the water molecule also has a dipole moment. Assume that water has a charge of $-q_w$ at the center of a sphere of radius R_w , and a charge of $+q_w$ at

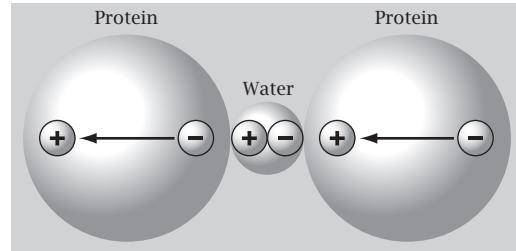


Figure 21.20

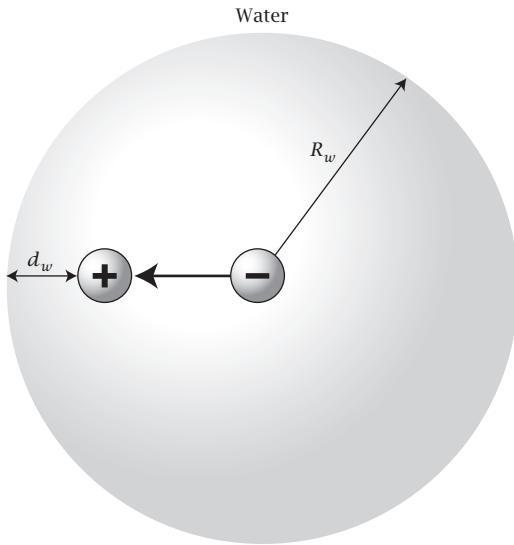


Figure 21.21

a distance $|R_w - d_w|$ from the center, as shown in Figure 21.21. Suppose that in the crystal, the water dipole is collinear with the dipoles of the proteins shown in Figure 21.20.

- (c) Assume $q = 2e$ in both systems in (a) and (b). Assume that $q_w = 1e$, $d = 5\text{ \AA}$, $d_w = 1.2\text{ \AA}$, $R_o = 10\text{ \AA}$ and $R_w = 5\text{ \AA}$. Which system, protein contacting protein or proteins contacting a central water, is the most stable?
- (d) Replace the vacuum surroundings of the protein-water-protein system by water surroundings, and assume a standard Coulombic treatment of the medium. What is the difference in total interaction energy in this case?

References

- [1] EM Purcell. *Electricity and Magnetism* (Berkeley Physics Course, Volume 2), 2nd edition, McGraw-Hill, New York, 1984.
- [2] IS Grant and WR Phillips. *Electromagnetism*, 2nd edition, Wiley, New York, 1990.

Suggested Reading

Clearly written elementary treatments of electrostatics include:

RP Feynman, RB Leighton, and M Sands, *The Feynman Lectures on Physics*, Volume II: *Mainly Electromagnetism and Matter*, Addison-Wesley Publishing Company, Reading, MA, 1964.

EM Purcell, *Electricity and Magnetism* (Berkeley Physics Course, Volume 2), 2nd edition, McGraw-Hill, New York, 1984.

More advanced treatments include:

R Becker, *Electromagnetic Fields and Interactions*, Blaisdell, New York, 1964 (reprinted by Dover Publications, New York, 1982).

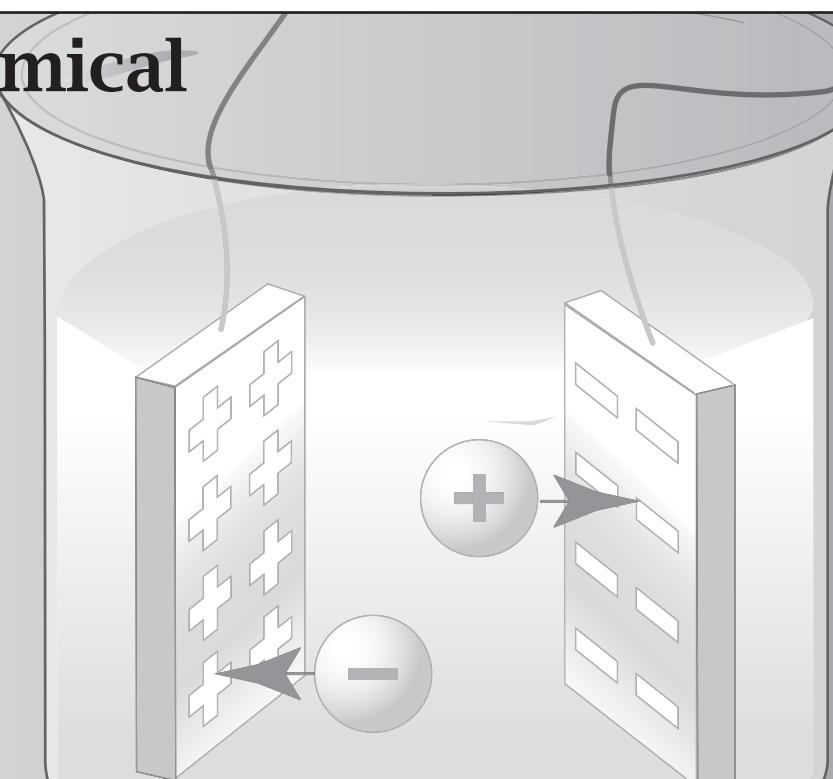
DK Cheng, *Field and Wave Electromagnetics*, 2nd edition, Addison-Wesley, Reading, MA, 1989.

L Egyes, *The Classical Electromagnetic Field*, Addison-Wesley, Reading, MA, 1972 (reprinted by Dover Publications, New York, 1980).

IS Grant and WR Phillips, *Electromagnetism*, 2nd edition, Wiley, New York, 1990.

This page is intentionally left blank.

22 Electrochemical Equilibria



Electrochemical Potentials Describe Equilibria in Solutions of Charged Species

Electrostatic forces and charges on atoms and molecules can affect chemical equilibria. In acid-base equilibria, molecules gain or lose protons. In oxidation-reduction equilibria, molecules gain or lose electrons. These processes occur in batteries, electroplating, respiration, photosynthesis, combustion, and corrosion. And, electric fields can gate the flows of ions through membrane channels, powering brains, nerves, and muscles. To explore these processes, we combine the laws of electrostatics, through Coulomb's law and Poisson's equation, with the laws of thermodynamic equilibrium, through the Boltzmann distribution law. We begin with the partitioning of charged particles, which is governed by a generalization of the chemical potential called the *electrochemical potential*.

What Drives the Partitioning of Ions?

If a molecule is charged, it will be driven to partition from one environment to another by chemical forces—described by chemical potentials—and by applied electrical fields. To see how these forces combine, we first generalize the thermodynamics of Chapters 6–8 to treat electrostatics.

In Chapter 6, we defined *simple systems* for which the internal energy is a function of the extensive variables S , V , and N . If electrical changes occur, the

charges $\mathbf{q} = q_1, q_2, \dots, q_M$ on the species $1, 2, \dots, M$ must also be taken into account as extensive variables in the energy function $U = U(S, V, \mathbf{N}, \mathbf{q})$. The electrostatic energy is $q\psi$, where ψ is the electrostatic potential felt by an ion due to the presence of electrodes, nearby charged surfaces, or any other constellations of charges. The fundamental equation (see Chapter 8), augmented to include charge effects, is

$$dU = T dS - p dV + \sum_{j=1}^t \mu_j dN_j + \sum_{i=1}^M \psi dq_i, \quad (22.1)$$

where $i = 1, 2, \dots, M$ are the charged species and the electrostatic potential $\psi = \psi(x, y, z)$ can depend on spatial position in general.

Because $G = U + pV - TS$, the differential equation for the Gibbs free energy including charge interactions is

$$dG = -S dT + V dp + \sum_{j=1}^t \mu_j dN_j + \sum_{i=1}^M \psi dq_i. \quad (22.2)$$

The total charge on species i is $q_i = z_i e N_i$, where z_i is the *valence* (the number of charges per ion), e is the unit charge on a proton, and N_i is the number of ions of type i . When an exchange process involves only charged particles, the indices j and i coincide and the free energy is

$$dG = -S dT + V dp + \sum_{i=1}^M (\mu_i + z_i e \psi) dN_i. \quad (22.3)$$

The quantity $\mu_i + z_i e \psi$ defines the *electrochemical potential* μ'_i :

$$\mu'_i = \mu_i + z_i e \psi. \quad (22.4)$$

For uncharged species at constant T and p , equilibrium occurs when the chemical potentials are equal. For charged species, equilibrium occurs when the *electrochemical* potentials are equal. Partitioning, solvation, and the transport of charged particles are governed by both chemical and electrostatic forces.

The Nernst Equation: Electrostatic Fields Drive Ion Distributions

Here's how you use Equation (22.4). Suppose that you want to compute how ions distribute spatially. The ions are driven by an externally applied electrostatic potential. Consider two different locations in space, \mathbf{r}_1 and \mathbf{r}_2 , within a single solution. To eliminate some vector complexity, let's look at a one-dimensional problem. Suppose that $\mathbf{r}_1 = x_1$ and $\mathbf{r}_2 = x_2$. At location x_1 , some constellation of fixed charges or electrodes creates the electrostatic potential $\psi(x_1)$. At x_2 , you have $\psi(x_2)$. Now consider a single species of mobile ions (so that you can drop the subscript i) that is free to distribute between locations x_1 and x_2 . The mobile ions will be at equilibrium when the electrochemical

potentials are equal:

$$\mu'(x_1) = \mu'(x_2). \quad (22.5)$$

Substituting the expression for the chemical potential, $\mu(x) = \mu^\circ + kT \ln c(x)$, into Equation (22.4) gives

$$\mu'(x) = \mu^\circ + kT \ln c(x) + z\epsilon\psi(x). \quad (22.6)$$

By regarding μ° as independent of x in this case, we are restricting our attention to situations in which x_1 and x_2 are in the same phase of matter.

In earlier chapters, we used mole fraction concentrations. Other concentration units are often useful, and just require a conversion factor that can be absorbed into μ° . Here we express concentrations in a general way as $c(x)$, in whatever units are most useful for the problem at hand. Substituting Equation (22.6) into Equation (22.5) gives the **Nernst equation**, named after WH Nernst (1864–1941), a German chemist who was awarded the 1920 Nobel Prize in Chemistry for his work in chemical thermodynamics:

$$\ln \left[\frac{c(x_2)}{c(x_1)} \right] = \frac{-ze[\psi(x_2) - \psi(x_1)]}{kT}, \quad (22.7)$$

or

$$c(x_2) = c(x_1) \exp \left\{ \frac{-ze[\psi(x_2) - \psi(x_1)]}{kT} \right\}. \quad (22.8)$$

Equation (22.8) is the Boltzmann distribution law for systems involving charges.

According to the Nernst equations (22.7) and (22.8), positive ions tend to move away from regions of positive potential $\psi(x) > 0$ and toward regions of negative potential. Negative charges move the opposite way (see Figure 22.1). These tendencies increase with the charge on the ion or decrease with the temperature. Figure 22.2 shows an example of a potential that is positive at a plane at $x = 0$ and diminishes with x . While $\mu = \mu^\circ + kT \ln c(x)$ describes a balance between translational entropy and chemical affinity for nonelectrolytes, the Nernst equation (22.8) describes a balance between translational entropy and electrostatic forces for electrolytes. Beginning on page 428, we explore some applications of the Nernst equation.

EXAMPLE 22.1 Useful units conversion. The quantity $kT/(ze)$ in Equation (22.7) appears often in electrostatics problems. Put it into more convenient units by multiplying the numerator and denominator by Avogadro's number \mathcal{N} :

$$\frac{kT}{ze} = \frac{\mathcal{N}kT}{z\mathcal{N}e} = \frac{RT}{zF} \quad (22.9)$$

where $F = \mathcal{N}e = 23,060 \text{ cal mol}^{-1} \text{ V}^{-1} = 96,500 \text{ C mol}^{-1}$ is the Faraday constant, the unit charge per mole. At $T = 300 \text{ K}$, you have $RT = 596 \text{ cal mol}^{-1}$, so for $z = +1$,

$$\frac{RT}{F} = \frac{596 \text{ cal mol}^{-1}}{23,060 \text{ cal mol}^{-1} \text{ V}^{-1}} = 0.0258 \text{ V} \quad (22.10)$$

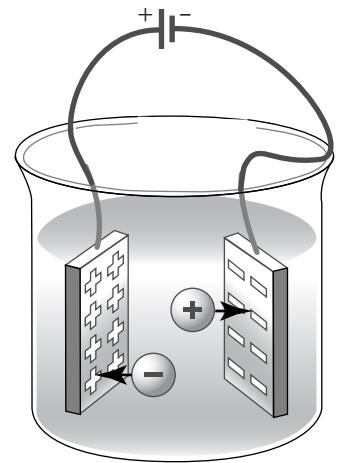
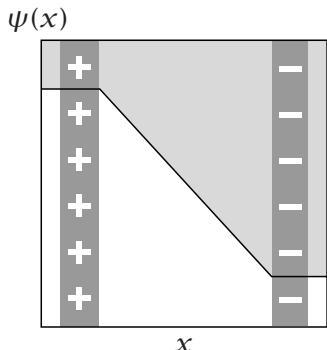


Figure 22.1 A voltage applied to electrodes produces an electrostatic potential $\psi(x)$ that acts on particle i having charge q_i . A negative ion will move toward a positive potential, while a positive ion will move toward a negative potential.

(a)



(b)

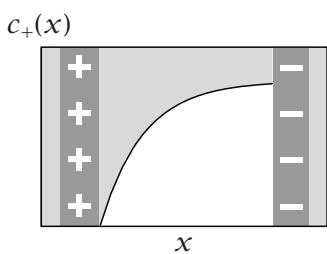


Figure 22.2 If the electrostatic potential $\psi(x)$ (a) is a linear function of distance x , then (b) the concentration $c(x)$ of ions depends exponentially on x , according to Equation (22.8). Positive ions are depleted near the positively charged surface.

You should be aware of one important difference between the electrochemical potential and the chemical potential. The chemical potential describes the free energy of inserting one particle into a particular place or phase, subject to any appropriate constraint. Constraints are introduced explicitly. In contrast, the electrochemical potential always carries an *implicit* constraint with it: overall electroneutrality must be obeyed. This is a very strong constraint. You can never insert a single ion in a volume of macroscopic dimensions, because that would violate electroneutrality. You can insert only an electroneutral combination of ions. In macroscopic systems such as beakers, there will be equal numbers of positive and negative charges. On those scales, you can work with the *chemical* potentials of neutral molecules such as NaCl. In contrast, when you are interested in microscopic nonuniformities, such as the counterion distribution near a DNA molecule, or the electrostatic potential microscopically close to an electrode surface, then you can use the electrochemical potential. Charge neutrality need not hold over microscopic distances.

The Chemical Potential of Neutral Salts

What is the chemical potential of a neutral salt that ionizes in solution? To be specific, consider NaCl. Add the electrochemical potentials of the separate ions to get the chemical potential of the neutral salt:

$$\mu_{\text{NaCl}} = \mu'_{\text{Na}^+} + \mu'_{\text{Cl}^-}$$

or, with Equation (22.6),

$$\mu_{\text{NaCl}} = \mu_{\text{Na}^+}^\circ + kT \ln c_{\text{Na}^+} + e\psi_{\text{Na}^+} + \mu_{\text{Cl}^-}^\circ + kT \ln c_{\text{Cl}^-} - e\psi_{\text{Cl}^-},$$

where ψ_{Na^+} and ψ_{Cl^-} are the electrostatic potentials felt by a Na^+ ion and a Cl^- ion, respectively, due to whatever field is present from external electrodes or nearby charges, etc. To get the standard chemical potential of the salt, μ_{NaCl}° , add the constants $\mu_{\text{Na}^+}^\circ$ and $\mu_{\text{Cl}^-}^\circ$:

$$\mu_{\text{NaCl}} = \mu_{\text{NaCl}}^\circ + kT \ln(c_{\text{Na}^+} c_{\text{Cl}^-}) + e(\psi_{\text{Na}^+} - \psi_{\text{Cl}^-}). \quad (22.11)$$

In solutions where the electrostatic potential is spatially uniform, $\psi_{\text{solution}} = \psi_{\text{Na}^+} = \psi_{\text{Cl}^-}$, and Equation (22.11) becomes

$$\mu_{\text{NaCl}} = \mu_{\text{NaCl}}^\circ + kT \ln(c_{\text{Na}^+} c_{\text{Cl}^-}). \quad (22.12)$$

In terms of the salt concentration $c_{\text{NaCl}} = c_{\text{Na}^+} = c_{\text{Cl}^-}$,

$$\mu_{\text{NaCl}} = \mu_{\text{NaCl}}^\circ + 2kT \ln c_{\text{NaCl}}. \quad (22.13)$$

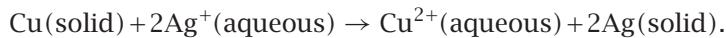
The factor of 2 in Equation (22.13) comes from assuming the complete ionization of NaCl.

This description is too simple in two situations. First, sometimes salts do not ionize completely. Second, near large charged surfaces such as colloids, large electrostatic potentials can lead to variations in ion concentrations: $c_{\text{Na}^+} \neq c_{\text{Cl}^-}$. Incomplete dissociation and microscopic nonuniformities are treated in Chapter 23 using the Debye-Hückel and Poisson-Boltzmann theories.

In the following sections, we explore two important types of electrochemical reactions. In *acid-base* reactions, materials change their charge states when *protons* hop on or off of them. In *oxidation-reduction* reactions (also called *redox* reactions), materials change their charge states when *electrons* hop on or off of them. These principles are the basis for batteries, fuel cells, electrolysis (situations where electric fields cause chemical reactions, sometimes used for separating materials), electroplating, corrosion, and two of the most important chemical reactions on Earth: respiration and photosynthesis. We begin with redox processes.

In Redox Reactions, Electrons Hop from One Molecule to Another

Put a piece of copper metal, Cu(solid), into a liquid solution of silver nitrate, AgNO₃, in water. What happens? First, because ions dissolve in water, the silver nitrate will dissolve into ions of silver, Ag⁺(aqueous), and nitrate, NO₃⁻(aqueous). Second, within minutes, the metal will turn black because the silver ions from the solution will react with the copper metal:



According to this equation, three things occur (see Figure 22.3): (1) A copper atom from the electrode sheds two electrons, ionizing to Cu²⁺(aqueous). The Cu²⁺(aqueous) ion jumps off the metal surface into the liquid. (2) Like a swimmer reaching shore, a silver ion Ag⁺(aqueous) comes out of solution and *plates onto* the copper surface, becoming Ag(solid). This turns the surface black. (3) This process is driven by the electrons left behind that spontaneously transfer from copper to silver.

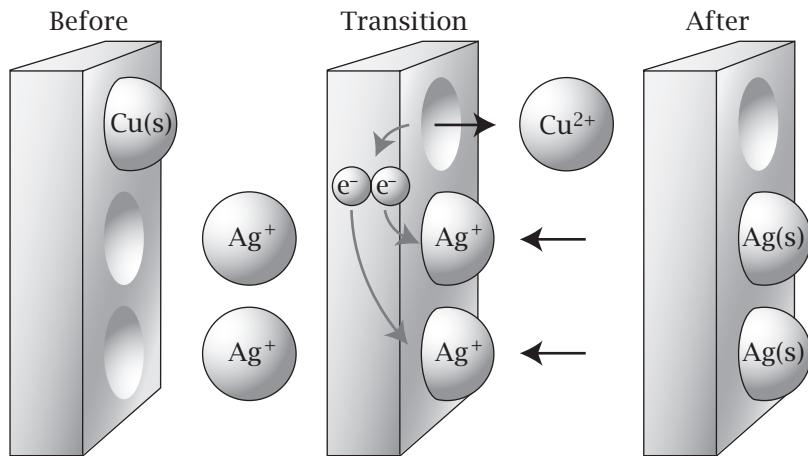
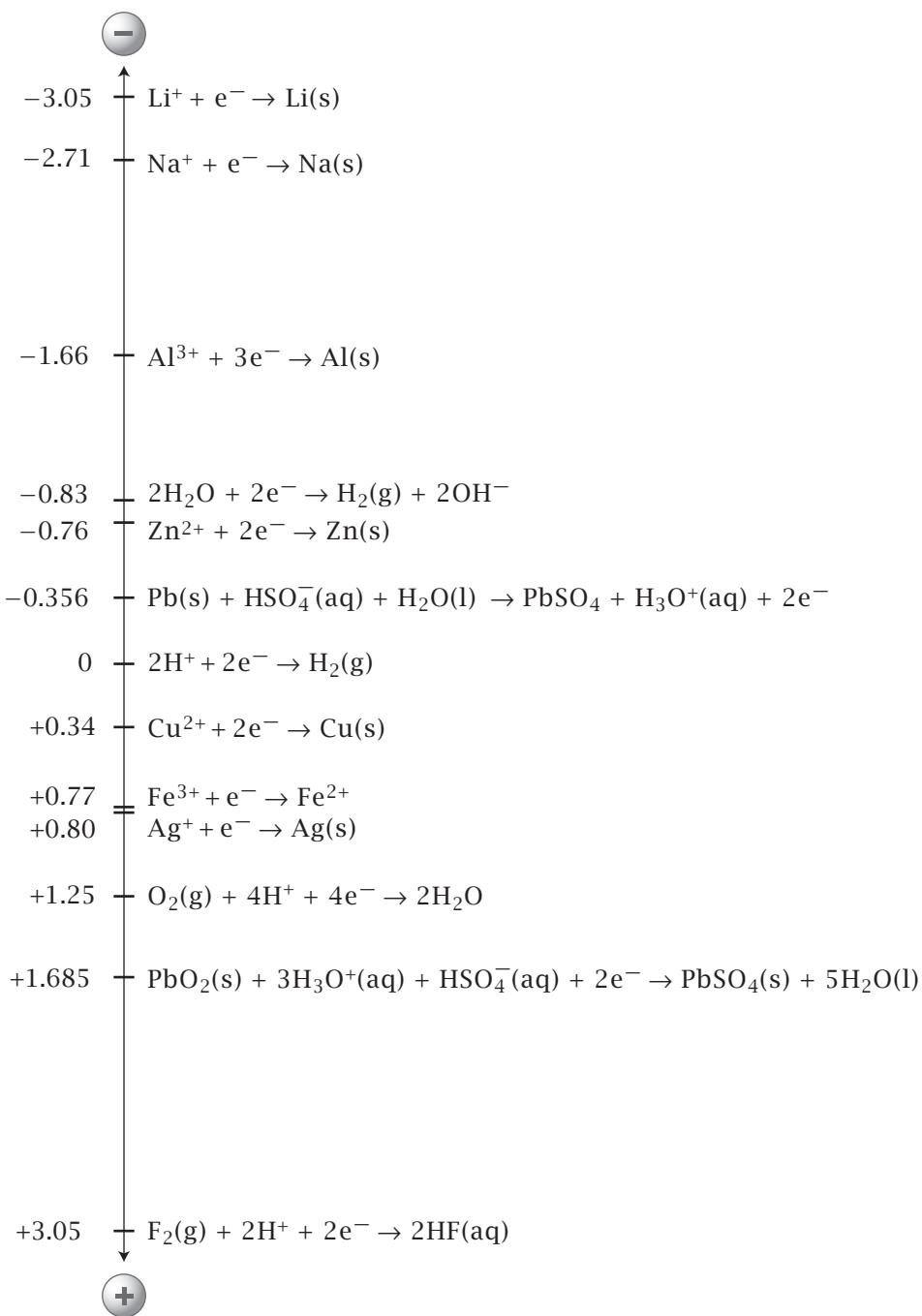


Figure 22.3 Copper metal in a silver nitrate solution. (Left, before) A neutral copper atom on the metal, and two silver ions in solution. (Middle, the process) The copper ionizes, losing two electrons, and dissolving in solution. At the same time, two silver ions from solution grab electrons, become neutral, and plate onto the metal. (Right, after) The copper metal has lost a copper atom and gained two silver atoms.

Reduction Potential ψ_0 (V)**Figure 22.4** Standard reduction potentials for various redox reactions.

What happens if you swap the materials? Now, put a piece of silver metal into a solution of copper nitrate, $\text{Cu}(\text{NO}_3)_2$, in water. Nothing happens. The copper ions in solution do not plate onto the silver metal, because electrons

do not spontaneously flow from silver to copper. To understand electrochemical reactions, we need a way to predict the direction of spontaneous flow of electrons from one material to another. We use a quantity called the *reduction potential* or *redox potential*. Think of the reduction potential as a ‘waterfall’ for electrons. Electrons flow down the waterfall of reduction potentials. The atoms and molecules with the most negative reduction potentials are at the top and the atoms and molecules with the most positive reduction potentials are at the bottom (see Figure 22.4). Electrons spontaneously flow from copper to silver, because copper is higher on the waterfall (copper’s reduction potential is more negative than silver’s). To make this more quantitative, you see that each step on the waterfall is written as a chemical reaction. To determine the electron flow, we describe below how to combine *half-cell reactions*.

If you put a silver metal electrode into a liquid solution that contains silver ions, the silver ions can either deposit onto the metal or remain in the solution. The ions exchange between the metal and the solution until they reach an equilibrium dictated by both the chemical potential and by the electrostatic interactions of the silver ions with the metal surface, where the metal surface has an electrostatic potential caused by an external circuit. For example, if you apply an electrostatic potential difference (a voltage) to such a solution/electrode interface, it can drive metal ions onto solid surfaces. This is the operating principle for electroplating. Conversely, if the solution is an acid that dissolves the metal electrode, it can cause the flow of ions and electrons, forming a battery.

Suppose that a silver electrode is in equilibrium with a solution of silver nitrate (see Figure 22.5). Ag^+ is called the *common ion* or the *the potential-determining ion*, because it is the exchangeable component common to both the liquid and the solid. Silver ions can plate onto the solid metal or dissolve to increase the AgNO_3 concentration in the liquid. Use the concentration c of Ag^+ in solution as the degree of freedom. The condition for equilibrium is that the electrochemical potential of the silver ions in the solid must equal the electrochemical potential of the silver ions in the liquid:

$$\mu'_{\text{Ag}^+}(\text{solid}) = \mu'_{\text{Ag}^+}(\text{liquid}). \quad (22.14)$$

According to Equation (22.6), the electrochemical potential for the liquid solution is

$$\mu'_{\text{Ag}^+}(\text{liquid}) = \mu^\circ_{\text{Ag}^+}(\text{liquid}) + RT \ln c + zF\psi_{\text{liquid}}, \quad (22.15)$$

where ψ_{liquid} is the electrostatic potential in the liquid. ψ_{liquid} depends on the voltages applied to the system and on chemical reactions in the solution. For the pure solid, the standard-state activity is defined as 1, and the electrochemical potential of the silver on the electrode is

$$\mu'_{\text{Ag}^+}(\text{solid}) = \mu^\circ_{\text{Ag}^+}(\text{solid}) + zF\psi_{\text{solid}}. \quad (22.16)$$

Substitute Equations (22.15) and (22.16) into Equation (22.14) to get the condition for equilibrium:

$$\mu^\circ_{\text{Ag}^+}(\text{solid}) + zF\psi_{\text{solid}} = \mu^\circ_{\text{Ag}^+}(\text{liquid}) + RT \ln c + zF\psi_{\text{liquid}}. \quad (22.17)$$

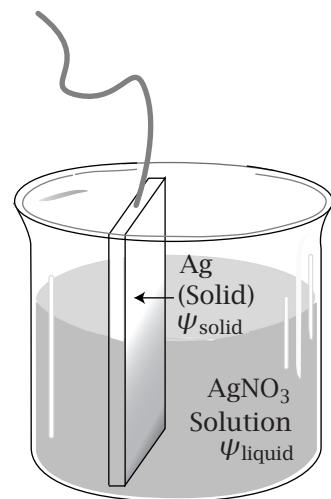


Figure 22.5 A silver electrode in a silver nitrate solution. The silver ions are free to plate onto the metal or dissolve in solution.

Collecting quantities together and rearranging Equation (22.17) gives

$$\psi = \psi_0 - \frac{RT}{zF} \ln c, \quad (22.18)$$

where we simplified to $\psi = \psi_{\text{liquid}}$. And, ψ_0 is the following collection of (individually unmeasurable) quantities in Equation (22.17):

$$zF\psi_0 = \mu^\circ(\text{solid}) - \mu^\circ(\text{liquid}) + zF\psi_{\text{solid}}. \quad (22.19)$$

Equation (22.17) is useful for understanding batteries and the electroplating of one metal onto another. And, it gives you a way to measure redox potentials. A system for measuring redox potentials, containing a metal and an electrolyte, is called a *half-cell*. ψ_0 is called the standard reduction potential or the *half-cell potential*. While the potential ψ depends on concentration, the standard-state reduction potential ψ_0 does not. ψ_0 describes the tendency for the spontaneous flow of electrons down their redox waterfalls. Note that the half-cell potential ψ_0 is a voltage, not a chemical potential. It takes two half-cells to make a battery. While you cannot readily measure the components of ψ_0 in Equation (22.19), you can easily measure the *difference* between the ψ_0 values of two different half-cells (see below).

Figure 22.6 shows how you can make a battery out of two half-cells. In the left cell L , a copper electrode is in equilibrium with a copper nitrate solution. In the right cell R , a silver electrode is in equilibrium with a silver nitrate solution. Figure 22.7 shows how this battery works. The battery requires a *salt bridge*, which is commonly a gel-like material (contained in a glass tube) soaked with another type of ion, say sodium nitrate, NaNO_3 , and bridging across the two half-cells (see Figure 22.6). The salt bridge is used for two reasons. First, it completes the electrical circuit. As electrons flow to the right through the external circuit, negative ions must flow to the left through the solution, to avoid causing

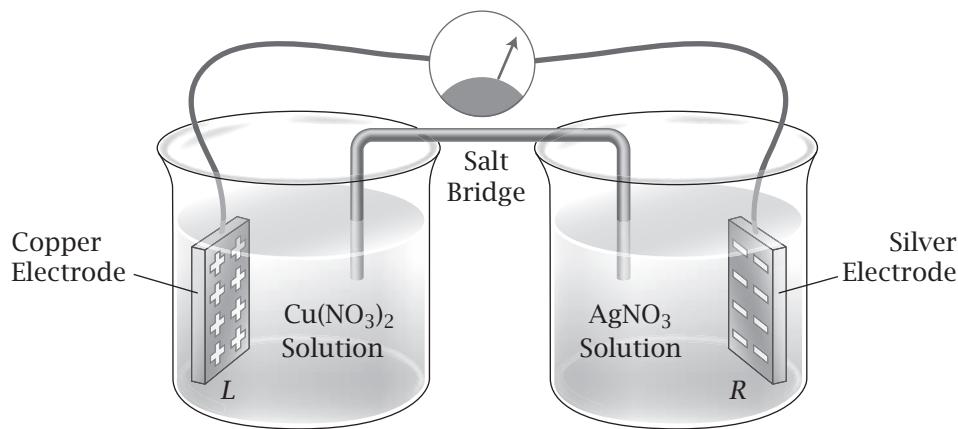
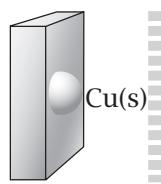
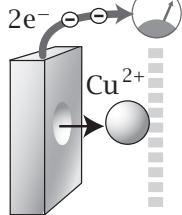


Figure 22.6 A simple battery. Left half-cell, L : Copper electrode in a copper nitrate solution. Right half-cell, R : Silver electrode in a silver nitrate solution. The electrodes are connected by a wire, through which current flows. The cells are connected by a salt bridge that allows nitrate ions to flow from R to L , to allow a full circuit of current flow. The principle is described in Figure 22.7.

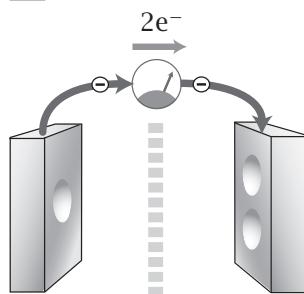
(a) Anode



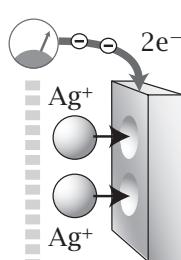
(b)



(c)



(d)



(e)

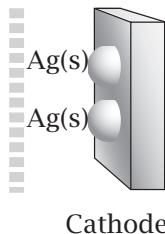


Figure 22.7 Principles of the battery shown in Figure 22.6. (a) A neutral copper atom starts on the copper electrode (anode). (b) the copper atom ionizes and jumps into the left solution, releasing two electrons. The electrons flow up the anode. (c) The two electrons flow through the external circuit, providing current to power an electrical device. (d) Two silver ions jump from the liquid onto the cathode, taking up the two electrons, and becoming (e) two neutral silver atoms that are now plated onto the silver electrode. At the same time, two nitrate ions are released from the right compartment, flow leftward through the salt bridge, and associate with the copper ions in the left compartment. This process is spontaneous because electrons are flowing down their redox waterfall, from copper to silver.

a build-up of charge that would halt the current flow. Second, if you had used a single beaker instead of two half-cells, then silver atoms could have drifted back to coat the copper electrode *L*, restricting further current flow. By clever choice of materials, some types of batteries use only a single cell, avoiding the need for salt bridges.

The *R* electrode, called the *cathode*, is positive (electrons flow towards it) and the *L* electrode, called the *anode*, is negative. For the battery in Figure 22.6, if the concentrations on the two sides are 1 M, the voltage difference is $\Delta\psi = \psi_{0,\text{cathode}} - \psi_{0,\text{anode}} = 1.14\text{ V}$. Different metals and different electrolyte solutions will give different voltages. This type of device gives only differences in ψ_0 , not absolute values of ψ_0 . So, in order to make up tables of redox potentials for different half-cell electrode/electrolyte combinations, you need to choose a reference state that defines $\psi_0 = 0$. The reference state is arbitrary, that is, chosen by convention. By standard convention, the reference half-cell is a platinum electrode in a 1 M solution of hydrochloric acid held under 1 atm pressure of H₂ gas. So the reaction H⁺ + e⁻ → (1/2)H₂(gas) is taken to define $\psi_0 = 0$. Figure 22.4 shows the potentials for various electrode/electrolyte combinations, relative to this reference. For example, for our combined copper and silver half-cells, the figure gives $\Delta\psi = \psi_{0,\text{cathode}} - \psi_{0,\text{anode}} = 0.340 - (-0.80) = +1.14\text{ V}$. Using devices of this type, you can measure redox potentials of materials.

How can you rationalize the redox potentials shown on Figure 22.4? For example, why is the reduction potential of lithium so negative and that of fluorine so positive? Roughly, redox potentials correlate with the electronegativities of atoms in the periodic table. Fluorine is highly electronegative, strongly attracting electrons, so it is at the bottom of the redox potential waterfall. Lithium is not very electronegative; it releases its electrons easily, so it is at the top of the waterfall. The most electronegative atoms, which are mostly at the bottom of the redox waterfall, tend to be in the upper right of the periodic table: F, O, Cl, and N. The least electronegative materials, which tend to be at the top of the redox waterfall, tend to be on the left side (Li⁺, Na⁺, K⁺, etc.) and at the bottom of the periodic table.

EXAMPLE 22.2 Some common batteries. Cars use *lead-acid batteries*. The anode is lead (Pb). The cathode is lead dioxide (PbO₂). The electrolyte is 6 M sulfuric acid (H₂SO₄) in water. Both electrodes are in the same cell; there is no salt bridge separating the half-cells. A typical car battery contains six cells, each with a voltage of about 2.1 V, totaling about 12.6 V. As lead-acid batteries wear out, both electrodes become lead sulfate (PbSO₄).

Flashlights typically use *dry-cell batteries*. The anode is a zinc can that surrounds and contains a low-moisture electrolyte paste of NH₄Cl and MnO₂. The paste encases a center cathode carbon rod, the end of which is a positive button at the top. Dry-cell batteries are typically 1.5 V.

Portable electronics and electric cars use *lithium-ion batteries*. The anode is often graphite and the cathode is one of various lithium materials. The electrolyte is typically a propylene carbonate solution containing a lithium salt, such as LiPF₆. Developed mostly in the 1990s, lithium-based batteries have become important because of their relatively high voltage (lithium has a very negative reduction potential) and light weight (lithium's atomic mass is only 7).

A *fuel cell* is a battery that creates electrical currents by catalytically converting fuels to electrons and byproducts, rather than by dissolving metal electrodes. Figure 22.8 shows an example in which hydrogen gas, H₂, is piped into the anode chamber and O₂ gas is piped into the cathode chamber.

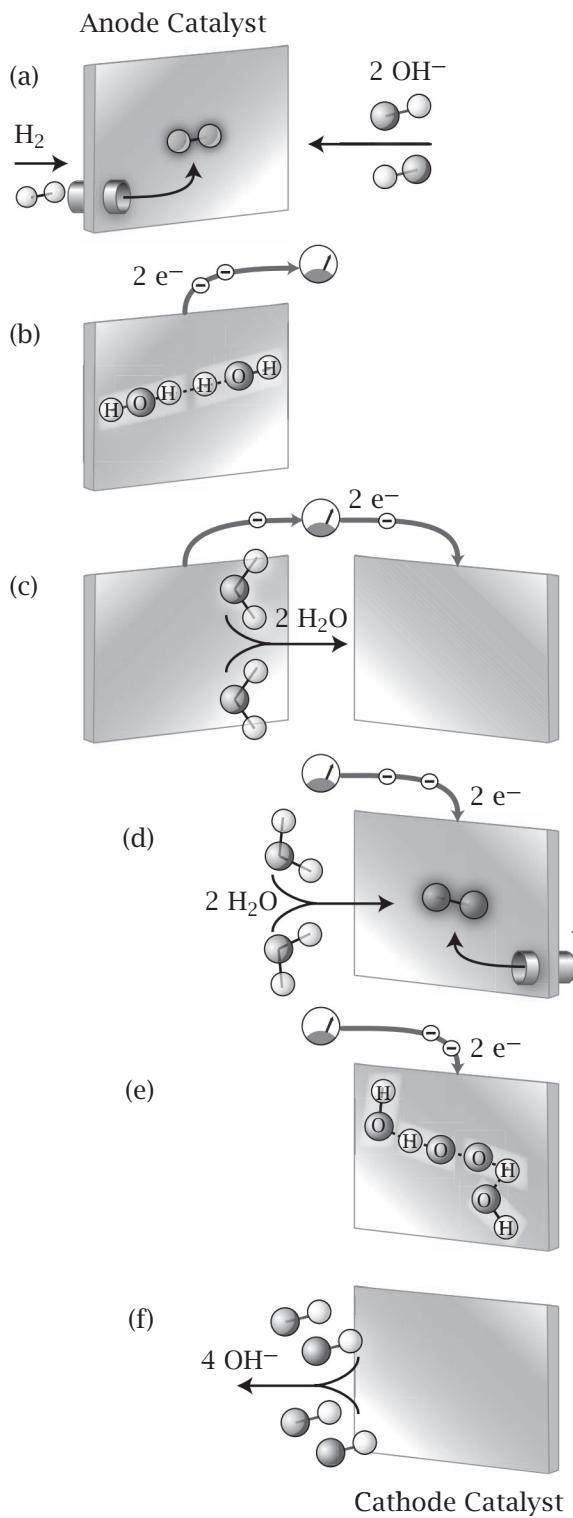
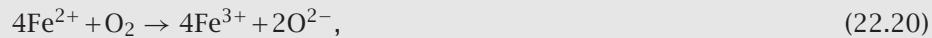


Figure 22.8 Principles of a fuel cell. (a) Hydrogen (H_2) fuel is piped onto the anode. (b) The anode catalyzes the conversion of the H_2 molecule plus two hydroxide ions (OH^-) from the fuel cell's water, to form two water molecules (H_2O). (c) Two electrons flow through the circuit at the top, powering some electrical device. Two H_2O molecules flow through the liquid from the anode on the left to the cathode on the right. (d) Oxygen (O_2) fuel is piped onto the cathode. Two H_2O molecules jump onto the cathode, combining with the two electrons and with one O_2 molecule. (e) The cathode catalyzes the conversion of the two H_2O molecules and the two O_2 molecules to form four OH^- ions. (f) The OH^- ions enter the water and return to the anode on the left. At the anode, the reaction stoichiometry is $2\text{H}_2(\text{gas}) + 4\text{OH}^-(\text{aqueous}) \rightarrow 4\text{H}_2\text{O}(\text{liquid}) + 4\text{e}^-$. $2\text{H}_2\text{O}(\text{gas})$ escape the fuel cell (not shown). At the cathode, the stoichiometry is $\text{O}_2 + 2\text{H}_2\text{O}(\text{liquid}) + 4\text{e}^- \rightarrow 4\text{OH}^-(\text{aqueous})$.

In the presence of catalysts, they react to form water, releasing electrons to flow in an external circuit. Other types of fuel cells use other fuels, such as alcohol.

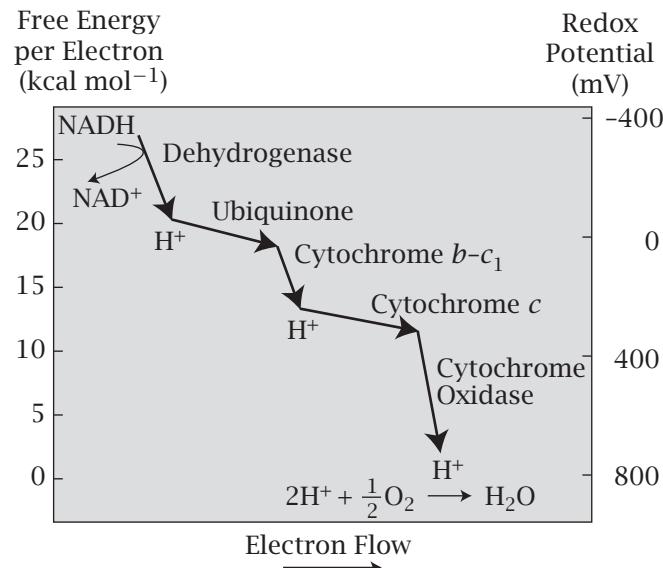
EXAMPLE 22.3 Corrosion, rusting, oxidation, and respiration. Iron rusts in the presence of strong oxidants, such as water, oxygen, or acids. One of the main reactions that produces rust is



in which oxygen takes up four electrons from iron, Fe^{2+} . Rusting is a spontaneous process (i.e., electrons fall down their redox waterfalls) because oxygen strongly attracts electrons away from materials such as iron. The term *oxidation* arose because oxygen is so ubiquitous and soaks up electrons in so many situations: fruits turn brown, copper pennies turn green, and hot gasoline in car engines combusts to form $\text{CO}_2 + \text{H}_2\text{O}$.

Figure 22.9 shows the electron transfer waterfall in mitochondrial respiration in biological cells: glucose converts to NADH, a molecule that is high up on the reduction-potential waterfall, which then converts, through a further series of downhill redox energy utilization steps, to $\text{CO}_2 + \text{H}_2\text{O}$.

Figure 22.9 Biological cells extract energy from glucose through NADH, then through a series of steps in which electrons flow down an oxidation potential waterfall. Source: B Alberts, A Johnson, J Lewis, M Raff, K Roberts, and P Walter, *Molecular Biology of the Cell*, 4th edition, Garland, New York, 2002.



EXAMPLE 22.4 Making a battery out of salt solutions. Another way to make a battery is apparent from the Nernst equation. Instead of using different electrode materials in the two half-cells, use identical electrodes, but different electrolyte concentrations. Suppose the Ag^+ concentration in one half-cell is c_1 and in the other it is c_2 . Because the solid is identical in the two cases, you can

set the right-hand side of Equation (22.17) equal for the two different solutions, liquid₁ and liquid₂, at equilibrium with the solid:

$$\begin{aligned}\mu_{\text{Ag}^+}^\circ(\text{liquid}) + RT \ln c_1 + z\epsilon\psi_{\text{liquid}_1} \\ = \mu_{\text{Ag}^+}^\circ(\text{liquid}) + RT \ln c_2 + z\epsilon\psi_{\text{liquid}_2},\end{aligned}\quad (22.21)$$

which rearranges to the Nernst equation:

$$\Delta\psi = \psi_{\text{liquid}_1} - \psi_{\text{liquid}_2} = \left(\frac{RT}{zF} \right) \ln \left(\frac{c_2}{c_1} \right), \quad (22.22)$$

with $z = 1$ for AgNO₃. For $T = 300\text{ K}$ and univalent charges $z = +1$, $RT/zF = 0.0258\text{ V}$, and $\ln x = 2.303 \log_{10} x$, so the Nernst equation can be written as (in V)

$$\Delta\psi = 0.0594 \log_{10} \left(\frac{c_2}{c_1} \right). \quad (22.23)$$

According to Equation (22.23), for every factor-of-10 increase in the concentration of AgNO₃ in solution, the electrostatic potential of the liquid increases by 59.4 mV. So, to make a 1 V battery, you would need a difference in salt concentrations of 10^{16} ! This is not a practical way to make batteries, but it is a practical way for biology to power the neurons in your brain, your muscles, and your cells. And, it is the principle used in oxygen sensors in cars. One half-cell of the sensor contains air (which has a constant concentration of oxygen) and the other half-cell contains oxygen from the car's exhaust. The sensor reads the potential difference between the two electrodes. That information is used to regulate the car's fuel : air intake ratio.

Selective Ion Flow Through Membranes Can Create a Voltage

Some biological membranes have protein pores or channels that allow certain ions to pass through, but not others. Figure 22.10 shows a hole in a membrane that represents a potassium channel, through which the permeant K⁺ ions pass freely but other ions cannot. Suppose you put a high concentration of KCl on the left and a low concentration on the right. At first, each side would be electroneutral with no electrical driving forces. Then K⁺ flows from left to right, down its concentration gradient. There is no counterbalancing flow of Cl⁻, because Cl⁻ cannot permeate the membrane. Therefore, a net positive charge builds up on the right side of the membrane. Equilibrium is achieved when the electrochemical potential for K⁺ is the same on both sides:

$$\begin{aligned}\mu'_{\text{K}^+}(L) &= \mu'_{\text{K}^+}(R) \\ \Rightarrow \mu_L^\circ + kT \ln [\text{K}^+]_L + e\psi_L &= \mu_R^\circ + kT \ln [\text{K}^+]_R + e\psi_R \\ \Rightarrow \Delta\psi = \psi_L - \psi_R &= \frac{kT}{e} \ln \frac{[\text{K}^+]_R}{[\text{K}^+]_L},\end{aligned}\quad (22.24)$$

because $\mu_L^\circ = \mu_R^\circ$ and $z = 1$.

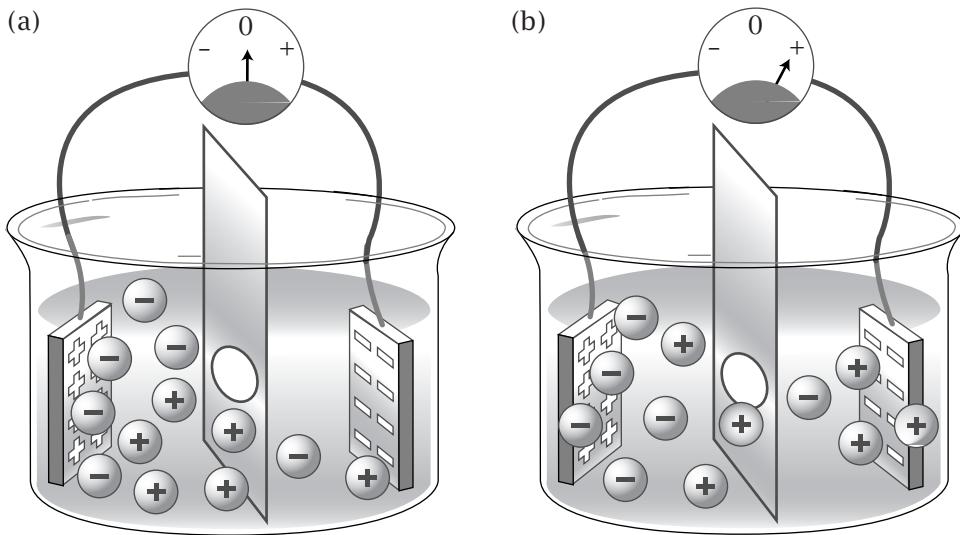


Figure 22.10 Model for an ion channel. (a) The initial concentration of positive ions on the left is high, before the system equilibrates. If positive ions can flow, but negative ions cannot, then positive ions tend to flow down their concentration gradient from left to right. This creates a net positive electrostatic potential on the right that opposes further flow. (b) The final equilibrium state is more positive on the right.

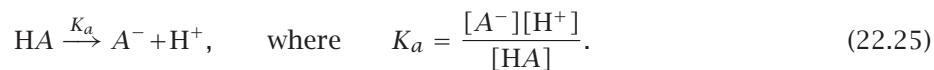
Equation (22.24) defines the *potassium potential*. If Na^+ or Ca^{2+} were the permeant ion instead of K^+ , there would be a different potential. You can think of such potentials in two different ways. First, the potential is the voltage that arises as a counterbalance when an ion flows down its concentration gradient. Second, if you applied a voltage equal to the K^+ potential across a membrane, it would drive K^+ to flow until reaching this particular ratio of concentrations inside and outside.

EXAMPLE 22.5 A potassium potential in skeletal muscle. In mammalian skeletal muscle, the extracellular potassium concentration is $[\text{K}^+]_{\text{out}} = 4 \text{ mM}$ and the intracellular concentration is $[\text{K}^+]_{\text{in}} = 155 \text{ mM}$. The K^+ potential at 37°C is given by Equation (22.24):

$$\psi_{\text{in}} - \psi_{\text{out}} = \frac{(8.314 \text{ JK}^{-1} \text{ mol}^{-1})(310 \text{ K})}{(9.65 \times 10^4 \text{ C mol}^{-1})} \ln\left(\frac{4 \text{ mM}}{155 \text{ mM}}\right) = -98 \text{ mV.}$$

Acid-Base Equilibria: Chemical Reactions Are Affected by Electrostatic Interactions

Acid-base equilibria are proton-exchange chemical equilibria. Consider the dissociation of an acid HA in solution:



The subscript a on K_a indicates that this is an *acid* dissociation constant (not an *association* constant: K_b is the dissociation constant for bases). Taking the logarithm of both sides of Equation (22.25) for K_a gives the **Henderson-Hasselbalch equation**,

$$\log K_a = \log \frac{[A^-]}{[HA]} + \log[H^+], \quad (22.26)$$

which relates the equilibrium constant K_a to the hydrogen ion concentration. Using the notation $p = -\log_{10}$, you have

$$pK_a = pH - \log \frac{[A^-]}{[HA]}. \quad (22.27)$$

Example 22.6 shows how acid-base equilibria can affect the partitioning equilibrium of molecules.

EXAMPLE 22.6 Why does aspirin partition from your stomach into your blood? Acid-base equilibria can help medicines reach their target sites in your body. Drugs will partition into your different body compartments (such as your stomach or your blood) depending on the pH inside the compartments. The *pH-partition theory* of BB Brodie [1] holds that the neutral form of a molecule (HA , for an acidic drug) crosses barrier membranes and reaches the same relative concentration in all aqueous compartments. So $[HA]_1 = [HA]_2$ for compartments 1 and 2. But the drug will ionize to different degrees in the different pH environments. Because the total amount of drug will be the sum of both the neutral and the ionized forms, $[HA] + [A^-]$, the ratio R of concentrations of drug in the two compartments will be

$$R = \frac{[HA]_1 + [A^-]_1}{[HA]_2 + [A^-]_2} = \frac{1 + \alpha_1}{1 + \alpha_2}, \quad (22.28)$$

where

$$\alpha_i = [A^-]_i / [HA]_i = 10^{pH_i - pK_a} \quad (22.29)$$

is the fraction ionized, for each compartment, $i = 1, 2$ in this case, from Equation (22.27). Aspirin has $pK_a = 3.5$. For blood (say, compartment 1), $pH = 7.4$. For the stomach, $pH = 3.0$. So, the aspirin will concentrate in your blood relative to your stomach by a factor of

$$R = \frac{1 + 10^{7.4 - 3.5}}{1 + 10^{3.0 - 3.5}} = 6040. \quad (22.30)$$

Acid-Base Equilibria Are Shifted by the Presence of Charged Surfaces

Acid-base equilibria can be shifted by the presence of electrical fields. For example, the pK_a of an acid group changes if the acid is near a charged surface, such as a membrane or a protein. H^+ ions will be repelled near a positively charged

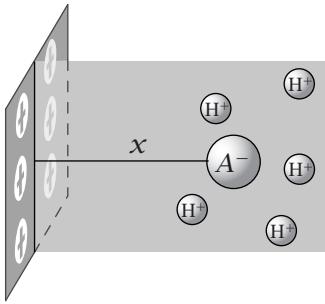


Figure 22.11 An acid group A^- is tethered at a distance x from a charged surface. It undergoes a reaction $HA \rightarrow H^+ + A^-$.

surface, causing an acid attached to the surface to give up its protons more easily than in the absence of the surface. The presence of the charged surface makes the acid appear to be stronger. Consider an acid group A that is rigidly tethered at a fixed distance x from a charged surface (see Figure 22.11). The H^+ ions are distributed throughout the solution, not just at x . For a tether of length x , you can express the equilibrium constant as

$$K_a(x) = \frac{[A^-]_x [H^+]_x}{[HA]_x},$$

where $[H^+]_x$, $[A^-]_x$, and $[HA]_x$ are the concentrations of hydrogen ions and unprotonated and protonated acid groups at a distance x from the surface. But $K_a(x)$ is not a particularly useful quantity, because $[H^+]_x$ is difficult to determine. More useful is the *apparent equilibrium constant*

$$K_{\text{app}}(x) = \frac{[A^-]_x [H^+]_\infty}{[HA]_x},$$

because $-\log[H^+]_\infty$ is the pH of the bulk solution, far away from the surface, which can be readily measured.

$pK_a(x)$ and $pK_{\text{app}}(x)$ are related. The Nernst equation (22.8) shows that the electrochemical potentials of the hydrogen ions must be the same everywhere at equilibrium, $\mu'_{H^+}(x) = \mu'_{H^+}(\infty)$ (see also Equation (22.5)), so Equation (22.8) with $\psi(\infty) = 0$ gives

$$[H^+]_x = [H^+]_\infty e^{-e\psi(x)/kT}. \quad (22.31)$$

Here we show an additional useful relationship is $K_a(x) = K_a(\infty)$. The condition for electrochemical equilibrium is

$$\begin{aligned} \mu_{\text{HA}}(x) &= \mu'_{A^-}(x) + \mu'_{H^+}(x) \\ \Rightarrow \mu_{\text{HA}}^\circ + kT \ln[HA]_x &= \mu_{A^-}^\circ + kT \ln[A^-]_x - e\psi(x) \\ &\quad + \mu_{H^+}^\circ + kT \ln[H^+]_x + e\psi(x). \end{aligned} \quad (22.32)$$

Rearranging Equation (22.32) gives

$$K_a(x) = \frac{[A^-]_x [H^+]_x}{[HA]_x} = e^{-\Delta\mu^\circ/kT}, \quad (22.33)$$

where $\Delta\mu^\circ = \mu_{\text{HA}}^\circ - \mu_{A^-}^\circ - \mu_{H^+}^\circ$ is independent of x . So $K_a(x) = K_a(\infty)$ is independent of x . The positive surface stabilizes A^- to the same degree that it destabilizes H^+ , and the $e\psi(x)$ terms cancel.

Now substitute $K_a(x) = K_a(\infty)$ and Equations (22.31) and (22.33) into the definition of $K_{\text{app}}(x)$ to get

$$K_{\text{app}}(x) = \frac{[A^-]_x [H^+]_x}{[HA]_x} e^{e\psi(x)/kT} = K_a(\infty) e^{e\psi(x)/kT}. \quad (22.34)$$

Use $p = -\log_{10} = -0.4343 \ln$ to simplify Equation (22.34) to

$$pK_{\text{app}}(x) = pK_a(\infty) - \frac{0.4343 e\psi(x)}{kT}. \quad (22.35)$$

A positively charged surface with a potential $\psi > 0$ causes pK_{app} to be lower than the solution value $pK_a(\infty)$ (that is, $K_{\text{app}} > K_a(\infty)$), so an acid near a positively charged surface will seem to be a stronger acid than the same group

in bulk solution. A positive surface leads to greater dissociation and a higher value of $[A^-]/[HA]$ for a given bulk concentration of H^+ ions. Conversely, an acid near a negatively charged surface will seem to be weaker than the corresponding acid in solution.

EXAMPLE 22.7 Histidine in a spherical protein. Histidine is a basic amino acid with $pK_a(\infty) = 6.0$. Suppose that a histidine is attached to the surface of a spherical protein of radius $a = 30\text{ \AA}$ in water. Suppose that the protein has a net charge of $q = -10e$. What is the pK_{app} of the histidine? For a sphere in water, you have the surface potential from Equation (21.43):

$$\frac{e\psi_a}{kT} = \frac{Cqe}{DkTa} = -10\left(\frac{\ell_B}{a}\right) = -10\left(\frac{7.13\text{ \AA}}{30\text{ \AA}}\right) = -2.38.$$

With Equation (22.35), you have

$$pK_{app} = 6.0 + (0.4343 \times 2.38) = 6.0 + 1.03 = 7.03.$$

Acid-base equilibria are affected not only by charged interfaces, but also by dielectric interfaces, as shown in Example 22.8.

EXAMPLE 22.8 Acid dissociation near an oil/water interface. Consider an acid group in water near a dielectric interface. For example, an acidic group may be tethered to an uncharged bilayer membrane in contact with water. Suppose that the intrinsic pK_a of the acid in the bulk solution is $pK_a(\infty) = 4.6$. What is the pK_a of the same acidic group as a function of the tethering distance from the oil interface?

According to the method of image charges in Chapter 21, a charge in water at a distance a perpendicular to an oil interface ‘sees’ an identical (repulsive) charge at a distance a into the oil phase, so Equation (21.53) gives

$$\frac{e\psi}{kT} = -\frac{Ce^2}{2aDkT} = -\frac{\ell_B}{2a} = -\frac{7.13\text{ \AA}}{2a}. \quad (22.36)$$

Substituting Equation (22.36) into Equation (22.35) gives the apparent pK_a of an acid in water as a function of its distance from a planar interface with oil (Table 22.1). This treatment neglects atomic detail, which may be important at such short distances, so it is only approximate.

Table 22.1 pK_{app} , the apparent pK_a of an acid group, as a function of distance a from an oil/water interface into the water phase.

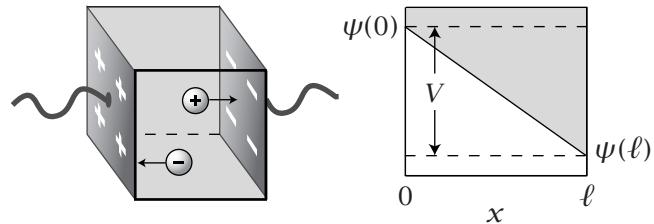
$a(\text{ \AA})$	pK_{app}
∞	4.6
5	4.91
4	4.99
3	5.12
2	5.37

Now we switch from the thermodynamics to the kinetics of moving charges. What are the forces that determine the velocities of ions or electrons in flow?

Gradients of Electrostatic Potential Cause Ion Flows

Concentration gradients can cause particles to flow, whether the particles are neutral or charged. In addition, charged particles flow if they are subject to gradients of *electrostatic* potential. According to Equation (20.13), the force on

Figure 22.12 Applying a voltage $V = \psi(0) - \psi(\ell)$ to plates separated by distance ℓ leads to ion flow and electrical conductance.



a charged particle equals (charge ze) \times (electrostatic field E):

$$f = zeE = -ze\left(\frac{\partial\psi}{\partial x}\right), \quad (22.37)$$

where Equation (21.7) relates E to ψ . Here we treat simple instances where the electrostatic potential $\psi(x)$ varies only in the x direction.

A particle's velocity v is related to the force f acting on it through Equation (17.2): $f = \xi v = (1/u)v$, where $u = 1/\xi$ is the *mobility*, the inverse of the friction coefficient. Substitute this expression into Equation (22.37) and use $J = cv = cuf$ (Equation (17.1)) to get the flux J of ions that results from a gradient of electrostatic potential:

$$J = -zecu\left(\frac{\partial\psi}{\partial x}\right) = -\kappa\left(\frac{\partial\psi}{\partial x}\right), \quad (22.38)$$

where κ is a measure of *electrical conductivity*. Like Fick's law and Fourier's law, Equation (22.38) describes a flow that is proportional to the steepness of a gradient. While the diffusion constant describes the flux of particles per unit of concentration gradient, $\kappa = zecu$ describes the flux of charged particles per unit of electrostatic potential gradient.

Equation (22.38) is an expression of **Ohm's law**, which says that electrical current flow is linearly proportional to the difference in applied electrostatic potential, or voltage. When a voltage $V = \psi(0) - \psi(\ell)$ is applied, charge carriers (ions or electrons) will flow, giving an electrical current I :

$$I = JA = \frac{\kappa A}{\ell}V, \quad (22.39)$$

where A is the cross-sectional area of the conducting medium and the electrical *resistance* is $R = \ell/\kappa A$. Ion conductivities and mobilities can be measured in the type of device shown in Figure 22.12. A voltage is applied, and the resulting electrical current flow is measured.

The Nernst–Planck Equation Describes Ion Flow Due to Both Electrical Gradients and Concentration Gradients

If an ion is driven by both a concentration gradient and a gradient of electrostatic potential, its flux J_p will be the sum of the component fluxes:

$$J_p = -D\left(\frac{\partial c}{\partial x}\right) - zecu\left(\frac{\partial\psi}{\partial x}\right). \quad (22.40)$$

Equation (22.40) is the **Nernst–Planck equation**. The subscript p on the symbol for flux J_p indicates that this is the flux of *particles*. Because each particle carries

a charge ze , the flux of *charge* J_c can be expressed as

$$J_c = zeJ_p. \quad (22.41)$$

Substituting Equation (22.41) and the Einstein relation $D = ukT$ into Equation (22.40) gives

$$\begin{aligned} J_c &= -zecu \left[\frac{kT}{c} \left(\frac{\partial c}{\partial x} \right) + ze \left(\frac{\partial \psi}{\partial x} \right) \right] = -zecu \left[kT \left(\frac{\partial \ln c}{\partial x} \right) + \left(\frac{\partial (ze\psi)}{\partial x} \right) \right] \\ &= -zecu \left(\frac{\partial \mu'}{\partial x} \right). \end{aligned} \quad (22.42)$$

Equation (22.42) is analogous to Equations (17.4)–(17.9), but here the driving force for charged particles is the electrochemical potential, $\mu' = \mu^\circ + kT \ln c(x) + ze\psi(x)$, instead of the chemical potential that drives the flow of neutral particles.

In the next section, we use the Nernst–Planck equation to show that the electrostatic potentials across membranes depend not only on the difference in ion concentrations, but also on the *ion mobilities*.

Ion Flows Through Membranes

Suppose that you start with a high concentration c_ℓ of salt on the left side of a membrane and a lower concentration c_r on the right. Suppose that the membrane is permeable to both the cations and anions. The positive and negative ions both tend to flow from left to right, down their concentration gradients. The flux of positive ions is given by Equation (22.42):

$$J_+ = -z_+ ecu_+ \left[\frac{kT}{c} \left(\frac{\partial c}{\partial x} \right) + z_+ e \left(\frac{\partial \psi}{\partial x} \right) \right], \quad (22.43)$$

and the flux of negative ions is

$$J_- = -z_- ecu_- \left[\frac{kT}{c} \left(\frac{\partial c}{\partial x} \right) + z_- e \left(\frac{\partial \psi}{\partial x} \right) \right], \quad (22.44)$$

where u_+ is the mobility of the positive ions, u_- is the mobility of the negative ions, and z_+ and z_- are the valencies of the positive and negative ions. Elec-troneutrality requires that $J_+ + J_- = 0$. Consider monovalent ions, $z_+ = +1$ and $z_- = -1$. Adding Equation (22.43) to Equation (22.44) gives

$$\begin{aligned} 0 &= (u_- - u_+) ekT \left(\frac{\partial c}{\partial x} \right) - (u_- + u_+) e^2 c \left(\frac{\partial \psi}{\partial x} \right) \\ \Rightarrow \quad \left(\frac{\partial \psi}{\partial x} \right) &= \frac{(u_- - u_+)}{(u_- + u_+)} \frac{kT}{ec} \left(\frac{\partial c}{\partial x} \right). \end{aligned} \quad (22.45)$$

Integrating both sides of Equation (22.45) over x , across the membrane gradients of both concentration and electrostatic potential, gives

$$\Delta\psi = \psi_\ell - \psi_r = \frac{kT}{e} \frac{(u_- - u_+)}{(u_- + u_+)} \ln \left(\frac{c_\ell}{c_r} \right), \quad (22.46)$$

where ψ_ℓ and ψ_r are the electrostatic potentials on the two sides of the membrane.

Table 22.2 Electric mobilities u of ions at 25°C.

Ion	u ($\text{cm}^2 \text{V}^{-1} \text{s}^{-1}$)
H^+	36.25
Li^+	4.01
Na^+	5.19
K^+	7.62
Cs^+	8.01
NH_4^+	7.52
Mg^{2+}	2.75
Ca^{2+}	3.08
F^-	5.74
Cl^-	7.92
Br^-	8.09
I^-	7.96
NO_3^-	7.41
SO_4^{2-}	4.15

Source: B Hille, *Ionic Channels of Excitable Membranes*, Sinauer, Sunderland, MA, 1984.

Let's consider two different situations. In the first, the positive ions can cross the membrane but the negative ions cannot ($u_- = 0$), so Equation (22.46) reduces to the result that we found in the Nernst equation (22.7). Then if $c_\ell = 0.1 \text{ M}$ and $c_r = 0.01 \text{ M}$, the electrostatic potential difference is $\Delta\psi = -60 \text{ mV}$.

But second, the additional implication of Equation (22.46) is that the electrostatic potential across the membrane can depend not only on the salt concentrations on the two sides, but also on the *rates* of ions flowing across the membrane. This is called a *diffusion potential*. Table 22.2 lists ion mobilities u in water at 25°C.

EXAMPLE 22.9 Diffusion potential for NaCl. Suppose NaCl is dissolved in water at concentrations $c_\ell = 0.1 \text{ M}$ on the left and $c_r = 0.01 \text{ M}$ on the right of the membrane. Calculate the diffusion potential at $T = 300 \text{ K}$ using Equation (22.46) and Table 22.2:

$$\begin{aligned}\Delta\psi &= (0.0258 \text{ V}) \left(\frac{7.92 - 5.19}{7.92 + 5.19} \right) \ln\left(\frac{0.1}{0.01}\right) \\ &= 12.4 \text{ mV.}\end{aligned}$$

The Cl^- ions move faster than the Na^+ ions, so the negative ions reach the dilute solution (right side) first, and cause the right side to be 12 mV more negative than the left side. This negative potential pulls the positive ions to the right as long as this concentration gradient is maintained.

Creating a Charge Distribution Costs Free Energy

So far in this chapter we have considered only a limited class of electrostatic processes: moving charges from one point to another within a fixed electrostatic field. Now we consider the other main class of electrostatic problems: computing the free energies for creating the electrostatic fields in the first place. You can view this either as a process of bringing a constellation of charges together into the correct configuration, or as a process of ‘charging up’ an assembly of originally uncharged particles that are already in the correct configuration. Such processes of charging, discharging, and assembly are used to describe how ions partition between different phases, move across membranes, or become buried in proteins, for example.

The Method of Assembling Charges

Imagine assembling a system of charged particles, one at a time. The first particle has charge q_1 . The electrostatic work of bringing a second particle, having charge q_2 , from infinity into the electrostatic field of the first is

$$w_{\text{el}} = q_2 \psi_2, \tag{22.47}$$

where ψ_2 is the electrostatic potential felt by particle 2 due to particle 1. Taking the potential at infinite separation to be zero, the electrostatic potential ψ_2 at

the location of charge q_2 is

$$\psi_2 = \frac{\mathcal{C}q_1}{Dr_{12}}, \quad (22.48)$$

where r_{12} is the distance between the charges. When you have more than two charges, the superposition principle says that the electrostatic potential acting on charge i is the sum of the electrostatic potentials from all other charges j :

$$\psi_i = \sum_{j \neq i} \frac{\mathcal{C}q_j}{Dr_{ij}}. \quad (22.49)$$

The total electrostatic free energy ΔG_{el} for assembling the charges is

$$\Delta G_{\text{el}} = w_{\text{el}} = \frac{1}{2} \sum_i q_i \psi_i = \frac{\mathcal{C}}{2D} \sum_i \sum_{j \neq i} \frac{q_i q_j}{r_{ij}}. \quad (22.50)$$

The factor $1/2$ corrects for the double counting because charge j was included as contributing to the electrostatic potential experienced by i , while i was also counted as contributing to the electrostatic potential experienced by j in the original sum in Equation (22.50) over all charges $i = 1, 2, \dots, N$.

When a system has many charges, it may be more convenient to use integrals rather than sums, and to approximate the distribution of discrete charges as a continuous function $\rho(x, y, z)$ describing charge densities. For a continuous charge distribution with ρ unit charges per unit volume in a total volume V , the free energy is

$$\Delta G_{\text{el}} = \frac{1}{2} \int_V \rho \psi_V dV, \quad (22.51)$$

where, according to Equation (21.11), $\psi_V = (\mathcal{C}/D) \int_V \rho(\mathbf{r}) r^{-1} dV$.

If you have a continuous *surface* distribution of charge, with σ charges per unit *area* on a surface s , the electrostatic free energy is given by a similar integral

$$\Delta G_{\text{el}} = \frac{1}{2} \int_s \sigma \psi_s ds, \quad (22.52)$$

where $\psi_s = (\mathcal{C}/D) \int_s \sigma r^{-1} ds$ is the surface equivalent of Equation (21.11).

These continuum equations contain an error, which is often small. While Equation (22.49) correctly sums over all charges *except* i , the conversion to integrals in Equation (22.51) and (22.52) incorrectly includes the interaction of charge i with itself, called the *self-energy*. This error is small when the number of charges is large enough to warrant the conversion from the sum to the integral. Example 22.10 is an application of Equation (22.52).

EXAMPLE 22.10 The free energy of charging a sphere. What is the free energy required to charge a sphere of radius a to a total charge q uniformly distributed over its surface? The total area of the sphere is $4\pi a^2$. The surface charge density, total charge divided by total area, is $\sigma = q/4\pi a^2$. At the surface of a sphere, Equation (21.43) gives $\psi_s = \mathcal{C}q/Da$. Substitute these quantities

into Equation (22.52):

$$\Delta G_{\text{el}} = \frac{1}{2} \int_s \sigma \psi_s ds = \frac{1}{2} \left(\frac{q}{4\pi a^2} \right) \left(\frac{\mathcal{C}q}{Da} \right) 4\pi a^2 = \frac{\mathcal{C}q^2}{2Da}. \quad (22.53)$$

Charging Processes

Another way to derive Equation (22.53) is based on the work done in a reversible charging process. Start with the uncharged sphere and ‘charge it up’ in small increments $dq = q d\lambda$. The parameter λ is a ‘progress coordinate’ for the charging process. As λ goes from 0 to 1, the surface goes from no charge to full charge. At any intermediate stage λ , the total surface charge is λq and the surface potential is (see Equation (21.43))

$$\psi_s(\lambda) = \frac{\mathcal{C}\lambda q}{Da}. \quad (22.54)$$

The work for the complete charging process is

$$\Delta G_{\text{el}} = \int_0^1 \psi_s(\lambda) q d\lambda = \frac{\mathcal{C}q^2}{Da} \int_0^1 \lambda d\lambda = \frac{\mathcal{C}q^2}{2Da}, \quad (22.55)$$

which is identical to the result in Equation (22.53).

Example 22.11 shows how much free energy is involved in charging up two parallel planes. This is useful when considering the electrostatic repulsion between biological membranes or colloids.

EXAMPLE 22.11 The free energy of charging a parallel-plate capacitor. What is the free energy for charging up two parallel plates? Start with two uncharged plates and transfer charge from one to the other. At the end of this process, the upper plate has a charge density (charge per unit area) $+\sigma$ and the lower plate has charge density $-\sigma$. The area of each plate is A . The dielectric constant of the material between the plates is D . Use Equation (22.52), with surface density σ and $\psi_s = (\sigma d)/(\varepsilon_0 D)$ (see Equation (21.13)), to get

$$\Delta G_{\text{el}} = \frac{1}{2} \sigma \psi_s A = \frac{\sigma^2 Ad}{2\varepsilon_0 D}. \quad (22.56)$$

Let’s use this result to compute the free energy of bringing two charged parallel planes from separation d_1 to d_2 .

EXAMPLE 22.12 Bringing charged planes together. Suppose that two planes with surface charge $\sigma = 10^{-3} \text{ C m}^{-2}$ are separated from each other by $d_1 = 10 \text{ cm}$ (see Figure 22.13). Oil, with a dielectric constant $D = 2$, is between the plates. Equation (22.56) gives the electrostatic free energy per unit area of charging up the planes at this fixed separation:

$$\begin{aligned} \frac{\Delta G_{\text{el}}}{A} &= \frac{\sigma^2 d}{2\varepsilon_0 D} = \frac{(10^{-6} \text{ C}^2 \text{ m}^{-4})(0.1 \text{ m})}{(2)(8.85 \times 10^{-12} \text{ C}^2 \text{ J}^{-1} \text{ m}^{-1})(2)} \\ &= 2.83 \text{ kJ m}^{-2}. \end{aligned} \quad (22.57)$$

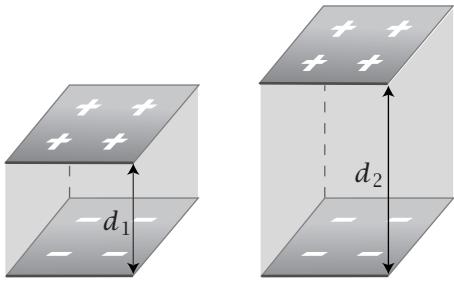


Figure 22.13 Parallel-plate capacitor with two different plate separations, d_1 and d_2 .

Bringing two oppositely charged planes together from $d_1 = 10\text{ cm}$ to $d_2 = 2\text{ cm}$ can be regarded as a process of discharging the planes when they are separated by distance d_1 , moving them together while they are uncharged ($\Delta G_{\text{el}} = 0$ for this step), and charging them up at separation d_2 . This process lowers the free energy by

$$\frac{\Delta G_{\text{el}}}{A} = \frac{\sigma^2(d_2 - d_1)}{2\epsilon_0 D} = \frac{-0.08\sigma^2}{2\epsilon_0 D} = -2.26\text{ kJ m}^{-2}. \quad (22.58)$$

The *Born energy* is the energy cost of transferring an ion from a medium having dielectric constant D_1 to a medium having dielectric constant D_2 .

Ion Solvation: Born Energy

Although the short-ranged interactions between the ion and the molecules of the medium discussed in earlier chapters contribute to solvation energies, here we consider just the electrostatic contribution. According to the *Born model*, this free energy is found by discharging the sphere in the first medium, transferring the neutral sphere, and then charging up the sphere in the second medium (see Figure 22.14).

Suppose the ion is a sphere of radius a with a total charge q uniformly distributed on its surface. The free energy ΔG_1 for *discharging* the sphere in

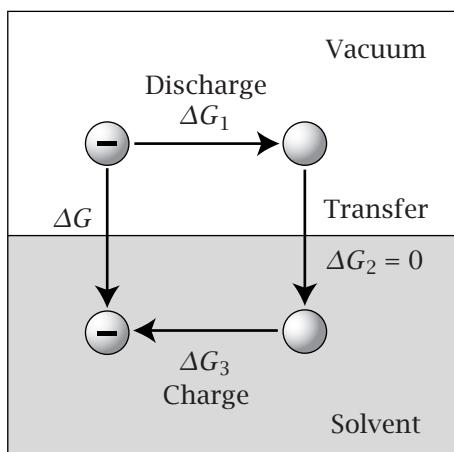


Figure 22.14 Transferring an ion from one medium to another—vacuum to solvent in this case—can be modeled with Born energies in a thermodynamic cycle. We compute here only the electrostatic free energy $\Delta G = \Delta G_{\text{el}}$. ΔG_1 is the free energy of discharging the ion in a vacuum, $\Delta G_2 = 0$ is the electrostatic free energy of transferring the neutral particle to the solvent, and ΔG_3 is the free energy of charging up the ion in the solvent.

the medium with dielectric constant D_1 is given by Equation (22.53):

$$\Delta G_1 = \frac{-\mathcal{C}q^2}{2D_1a}. \quad (22.59)$$

In step 3, the sphere is charged up in the medium with dielectric constant D_2 , so Equation (22.53) gives

$$\Delta G_3 = \frac{\mathcal{C}q^2}{2D_2a}. \quad (22.60)$$

Step 2, transferring the uncharged sphere to the second medium, involves no change in the electrostatic component of the free energy, $\Delta G_2 = 0$. The electrostatic free energy of transfer ΔG_{el} is the sum of the free energies for steps 1 through 3:

$$\Delta G_{el} = \Delta G_1 + \Delta G_2 + \Delta G_3 = \frac{-\mathcal{C}q^2}{2D_1a} + 0 + \frac{\mathcal{C}q^2}{2D_2a} = \frac{\mathcal{C}q^2}{2a} \left(\frac{1}{D_2} - \frac{1}{D_1} \right). \quad (22.61)$$

For the transfer of ions from water, where the dielectric constant is about $D = 80$, to oil, where the dielectric constant is about $D = 2$, ΔG_{el} is positive, as Example 22.13 shows. This is why ion-forming compounds such as salts are less soluble in oil than water.

EXAMPLE 22.13 The cost of partitioning an ion into oil. Compute the free energy of partitioning an ion of radius $a = 2 \text{ \AA}$ and charge $q = e$ from water ($D_{\text{water}} = 80$) into oil ($D_{\text{oil}} = 2$). To simplify, use the Bjerrum length in vacuum $\ell_B = \mathcal{C}e^2/kT = 560 \text{ \AA}$. Equation (22.61) becomes

$$\frac{\Delta G_{el}}{RT} = \frac{\ell_B}{2a} \left(\frac{1}{2} - \frac{1}{80} \right) \approx \frac{\ell_B}{4a} = \frac{1}{4} \left(\frac{560 \text{ \AA}}{2 \text{ \AA}} \right) = 70. \quad (22.62)$$

At $T = 300 \text{ K}$, $\Delta G_{el} = 70 \times (1.987 \text{ cal mol}^{-1} \text{ K}^{-1})(300 \text{ K}) \approx 42 \text{ kcal mol}^{-1}$. This is why ions are seldom found inside nonpolar media, such as lipid bilayers or the cores of globular proteins.

Figure 22.15 shows that polar molecules partition most readily into media with high dielectric constants, according to the $1/D$ dependence predicted for ions by Equation (22.61). The Born model also adequately predicts ion solvation enthalpies, provided that a is taken not as the ionic radius but as the somewhat larger radius of the solvent cavity containing the ion [2].

The Enthalpy of Ion Solvation

You can calculate the enthalpies of ion solvation from the vapor phase to water (dielectric constant D) from the Born model by using the Gibbs-Helmholtz equation (13.39):

$$\Delta H_{el} = \Delta G_{el} - T \frac{d\Delta G_{el}}{dT}. \quad (22.63)$$

Substitute the Born electrostatic free energy ΔG_{el} from Equation (22.61) into Equation (22.63), use $D_{\text{vapor}} = 1$, and recognize that the dielectric constant of

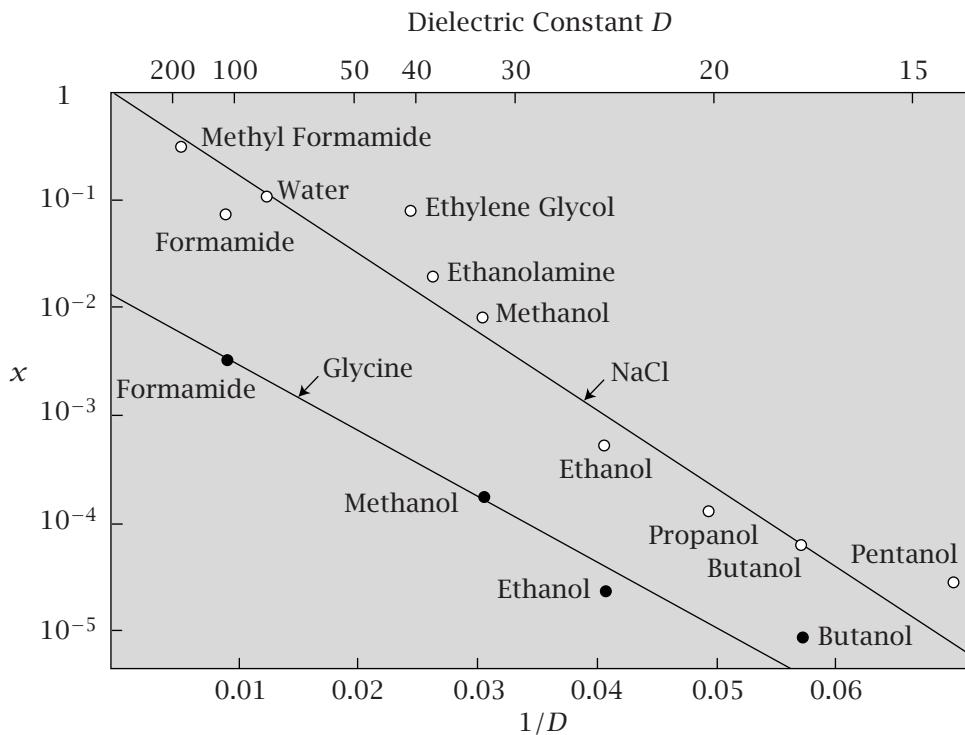


Figure 22.15 Solubilities (mole fraction x) of polar molecules (sodium chloride, \circ ; and glycine, \bullet) in media of different dielectric constants D . Media of high D are the best solvents, and the lines are predicted by Equation (22.61). Source: JN Israelachvili, *Intermolecular and Surface Forces*, Academic Press, New York, 1985. Solubility data are from *Gmelins Handbuch*, Series 21, Volume 7, for NaCl; and DR Lide, ed., *CRC Handbook of Chemistry and Physics*, 81st edition, CRC Press, Boca Raton, FL, 2000, for glycine.

water depends on temperature (Figure 22.16):

$$\Delta H_{\text{el}} = -\frac{Cq^2}{2a} \left[1 - \frac{1}{D} - T \frac{d}{dT} \left(1 - \frac{1}{D} \right) \right]. \quad (22.64)$$

Use the relations $d \ln x = dx/x$ for the term dT/T and $d(1/x) = -dx/x^2 = -(d \ln x)/x$ for the term $d(1/D)$ to rewrite Equation (22.64) as

$$\Delta H_{\text{el}} = -\frac{Cq^2}{2a} \left(1 - \frac{1}{D} - \frac{1}{D} \frac{d \ln D}{d \ln T} \right). \quad (22.65)$$

Table 22.3 and Figure 22.17 show experimental enthalpies of ion solvation in water. Because D depends on temperature, the solvation of ions also involves entropy, which can arise from the degree of alignment of solvent dipoles. Dielectric constants generally decrease with temperature (see Figure 22.16), indicating that the thermal motion weakens the degree of polarization at higher temperatures.

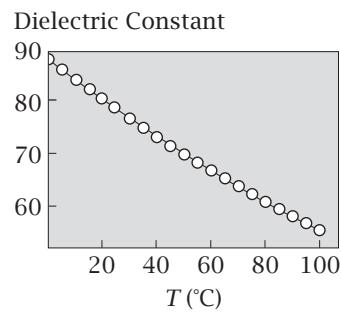


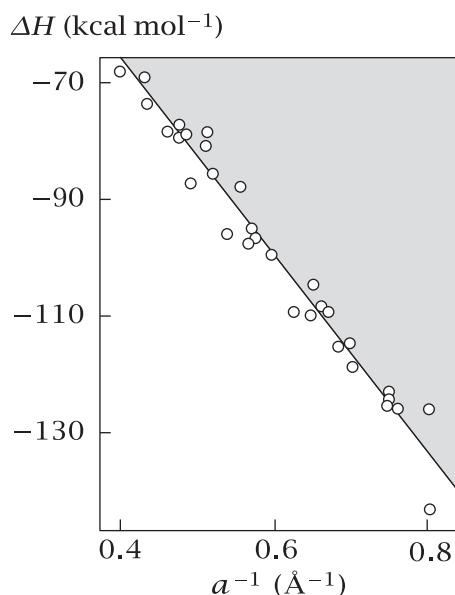
Figure 22.16
Experimental values for the dielectric constant of liquid water as a function of temperature.

Table 22.3 Experimental and theoretical values of the heats of solvation.

Salt	$\Delta H_{\text{experimental}}$ (kcal mol $^{-1}$)	$\Delta H_{\text{calculated}}$ (kcal mol $^{-1}$)
LiF	-245.2	-243.9
LiCl	-211.2	-212.8
LiBr	-204.7	-206.6
LiI	-194.9	-197.9
NaF	-217.8	-216.4
NaCl	-183.8	-185.4
NaBr	-177.3	-179.2
NaI	-197.5	-170.4
KF	-197.8	-194.0
KCl	-163.8	-162.9
KBr	-157.3	-156.7
KI	-147.5	-147.9
RbF	-192.7	-189.3
RbCl	-158.7	-158.3
RbBr	-152.2	-152.4
RbI	-142.4	-143.3
CsF	-186.9	-183.5
CsCl	-152.9	-152.4
CsBr	-146.4	-146.2
CsI	-136.6	-137.5

Source: The calculated enthalpies are from AA Rashin and B Honig, *J Phys Chem* **89**, 5588–5593 (1985). The values of $\Delta H_{\text{experimental}}$ are from JO'M Bockris and AKN Reddy, *Modern Electrochemistry*, Volume 1, Plenum Press, New York, 1977.

Figure 22.17 Enthalpies of hydration versus ion radius a . Source: AA Rashin and B Honig, *J Phys Chem* **89**, 5588–5593 (1985). The radii are from JA Dean, ed., *Lange's Handbook of Chemistry*, 11th edition, McGraw-Hill, New York, 1973; and ES Gold, *Inorganic Reactions and Structure*, Holt, Reinhart & Winston, New York, 1960. The experimental enthalpies are from J O'M Bockris and AKN Reddy, *Modern Electrochemistry*, Volume 1, Plenum Press, New York, 1977.



EXAMPLE 22.14 Calculating the heat of solvation of NaCl. Let's compute the enthalpy of solvation of NaCl using the Born model. If NaCl is sufficiently dilute, the heats of ionization of Na^+ and Cl^- can be treated as independent. Use Equation (22.65) with $q^2 = e^2$, and the Bjerrum length $\ell_B = 7.13 \text{ \AA}$. For water at 298 K, Figure 22.16 shows $d \ln D / d \ln T = -1.357$, and the enthalpy of solvation of NaCl in water is

$$\begin{aligned}\Delta H_{\text{el}} &= \frac{-l_B}{2a} \left[D - 1 - \left(\frac{\partial \ln D}{\partial \ln T} \right) \right] RT \\ &= \frac{-7.13 \text{ \AA}}{2a} (78.54 - 1 + 1.357) (0.592 \text{ kcal mol}^{-1}) \\ &= \left(\frac{-166.5}{a} \right) \text{ kcal mol}^{-1}. \end{aligned}\quad (22.66)$$

For Na^+ , $a = 1.680 \text{ \AA}$, so $\Delta H_{\text{el}} = -166.5 / 1.680 = -99.1 \text{ kcal mol}^{-1}$. For Cl^- , $a = 1.937 \text{ \AA}$, so $\Delta H_{\text{el}} = -166.5 / 1.937 = -86.0 \text{ kcal mol}^{-1}$. Then $\Delta H_{\text{el}}(\text{NaCl}) = \Delta H_{\text{el}}(\text{Na}^+) + \Delta H_{\text{el}}(\text{Cl}^-) = -185.1 \text{ kcal mol}^{-1}$, which agrees well with the experimental value in Table 22.3.

Summary

The electrochemical potential is a generalization of the chemical potential to situations that involve charged molecules. It is useful for predicting the equilibria of particles that are driven by both chemical interactions and electrical fields, including acid-base and redox reactions. Applications of electrochemical potentials include the partitioning of ions from one medium to another, electroplating, batteries, and the binding of ions to surfaces. Electric fields can perturb chemical equilibria such as acid-base interactions. The process of moving charges from one medium to another can be described by charging and discharging cycles. An example is the Born model for the partitioning of ions between different media.

Problems

1. A charged protein. Model a protein as a sphere with a radius of 20 \AA and a charge of $20e$ in water at 25°C and $D_w = 78.54$. Assume the sphere is uniformly charged on its surface.

- In units of kT , what is the potential at a distance 30 \AA from the protein surface?
- What is the electrostatic free energy of the charge distribution on the protein in kcal mol^{-1} ?

2. $\text{p}K_a$'s in a protein. Suppose the protein in problem 1 has an aspartic acid residue and a lysine residue at its surface. In bulk water, the $\text{p}K_a$ of the aspartic acid side group is 3.9 and the $\text{p}K_a$ of the lysine NH_2 side group is 10.8. What are the $\text{p}K_a$'s of these groups in a protein if it has a net charge of $+20e$?

3. Acid dissociation near a protein. Now put the protein of Problems 1 and 2 in a 0.01 M aqueous solution of acetic acid (HAc). In bulk solution, the acetic acid has $\text{p}K_a = 4.0$. It is dissociated to give ion concentrations $[\text{H}^+]_{\text{bulk}}$ and $[\text{Ac}^-]_{\text{bulk}}$.

- What are the concentrations of H^+ and Ac^- ions at a location x near the protein where the potential is $\psi(x)$?
- What is the $\text{p}K_a$ of the acetic acid near the protein at x ? Explain the difference between the situation of this acid and the situation of the aspartic acid in Problem 2.

4. Acids and bases near a charged surface. On a negatively charged surface:

- Are basic groups ionized more or less than in bulk water?
- Are acid groups ionized more or less than in bulk water?

5. Free energy of charging a sphere. A sphere with radius a in a medium with dielectric constant D is uniformly filled with a charge of volume density ρ . Derive the electrical free energy of this sphere in two different ways:

- Derive the potential field and then charge the sphere from charge density 0 to ρ .
- Add fully charged shells of thickness dr from $r = 0$ to $r = a$.

6. Burying a charge in a protein. As an estimate for the free energy of burying a charged amino acid such as aspartic or glutamic acid in protein folding, compute the free energy of transferring an ion of radius 3 \AA and charge $+1$ from water to oil. Assume that water has a dielectric constant $D_w = 80$, and oil has $D_{\text{oil}} = 2$.

7. A solvated charge in a protein. The Born energy in Equation (22.53) allows you to estimate the free energy cost for a charged group to be deep inside a fully folded protein (instead of at the protein surface in contact with

water). Estimate the free energy cost for a charged group to stay inside a *partly* folded protein. Consider the ionic group as a sphere with radius 2 \AA and total surface charge e (see Figure 22.18). This group is first surrounded by a

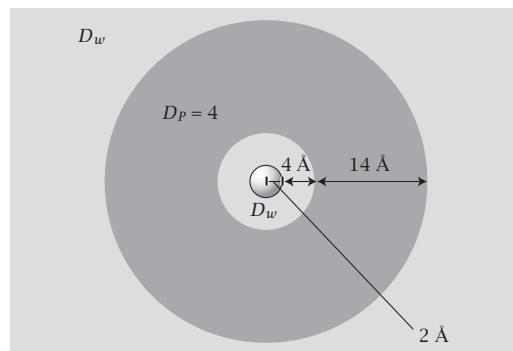


Figure 22.18

shell of bulk water, 4 \AA thick. This in turn is surrounded by a shell of protein, 14 \AA thick with dielectric constant $D_p = 4$, which is in contact with bulk water at 25°C .

Derive the electrostatic free energy of the ionic group:

- in water;
- in the partly folded protein;
- as in (b), but with the water shell replaced by protein, $D_p = 4$, to simulate the completely folded protein.

8. Oil/water interfacial potential. Consider an uncharged oil/water interface. On the aqueous side, the proximity of the oil phase biases the orientation of the water molecules in an unknown way. Calculate the resulting potential across the boundary layer for the maximum bias, the complete line-up of the first layer of water dipoles perpendicular to the interface. Treat this layer as a parallel-plate capacitor, with one water molecule occupying 10 \AA^2 of the interfacial area. The dipole moment is $\mu = 1.85$ debye (1 debye = $3.336 \times 10^{-30}\text{ C m}$) per water molecule. Take two values of the dielectric constant D between the capacitor plates:

- $D = 2$, as for oil, and
- $D = 80$, as for bulk water.

(If water is perfectly oriented—a situation called dielectric saturation—then $D = 2$ is more likely.)

9. Small electrostatic potentials. For a monovalent ion at $T = 300\text{ K}$, what is the value of ψ such that $e\psi = kT$?

10. Sodium potential in a frog muscle. Inside a frog muscle cell, the sodium concentration is $[\text{Na}]_{\text{in}} = 9.2\text{ mM}$. Outside, the sodium concentration is $[\text{Na}]_{\text{out}} = 120\text{ mM}$.

- Compute the sodium potential $\Delta\psi$ across the membrane at $T = 300\text{ K}$.
- Which side of the membrane has the more positive potential, the inside or the outside?

11. Membrane pores. A neutral protein ‘carrier’ may help an ion to transfer into and across a lipid membrane.

- What is the electrostatic free energy change when a monovalent ion is transferred from water at 25°C to a hydrocarbon solvent with dielectric constant $D_{hc} = 2$? The radius of the ion is 2 Å.
- Now wrap the ion in a neutral protein to produce a spherical complex with radius $b = 15\text{ \AA}$. What is the electrostatic free energy of transfer from water to hydrocarbon of the ion-protein complex?

12. Solvating a protein. A spherical protein has a valency of $z = -5$, and a radius of $a = 10\text{ \AA}$. The change in electrostatic free energy, ΔG_{el} , when you transfer the protein from vacuum ($D = 1$) to water ($D \approx 80$) is a part of the solvation free energy. Compute ΔG_{el} .

13. Burying an ion pair in oil. What is the free energy cost of transferring a monovalent anion of radius $a = 2\text{ \AA}$ and a monovalent cation of the same radius a from vacuum into oil ($D = 2$) at an ion-paired separation of $2a$?

14. Temperature-dependent dielectric constants. Show that Coulombic interactions between charges in liquids can have an entropic component if the dielectric constant D depends on temperature.

15. Ligand-protein electrostatic interactions. Consider a ligand with charge of +1e and a spherical protein of +4e and radius of 20 Å in water ($D = 80$), as shown in Figure 22.19.

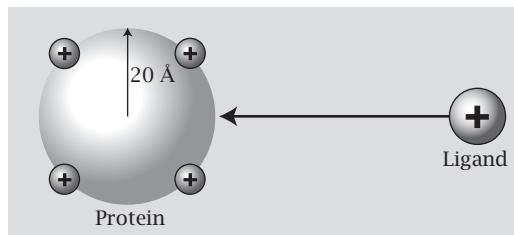


Figure 22.19

- What is the work of bringing the ligand from $r = \infty$ to the surface of the protein?
- Now assume there is a hollow cavity of radius $a = 5\text{ \AA}$ inside the protein, as shown in Figure 22.20. The protein cavity is filled with water, and there is a net charge of +4e on the shell surrounding the cavity. What is the work of transferring the ligand from the bulk water solution far away from the protein to the center of the water-filled cavity?

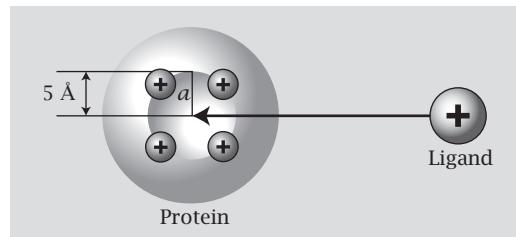


Figure 22.20

- Now consider the ligand to be an uncharged carboxylic acid with $pK_a = 3.0$. If you bring this ligand into the water-filled protein cavity as described in (b), what will be its apparent pK_a ?

References

- [1] PA Shore, BB Brodie, and CAM Hogben, *J Pharmacol Exp Ther* **119**, 361–369 (1957).
- [2] AA Rashin and B Honig, *J Phys Chem* **89**, 5588–5593 (1985).

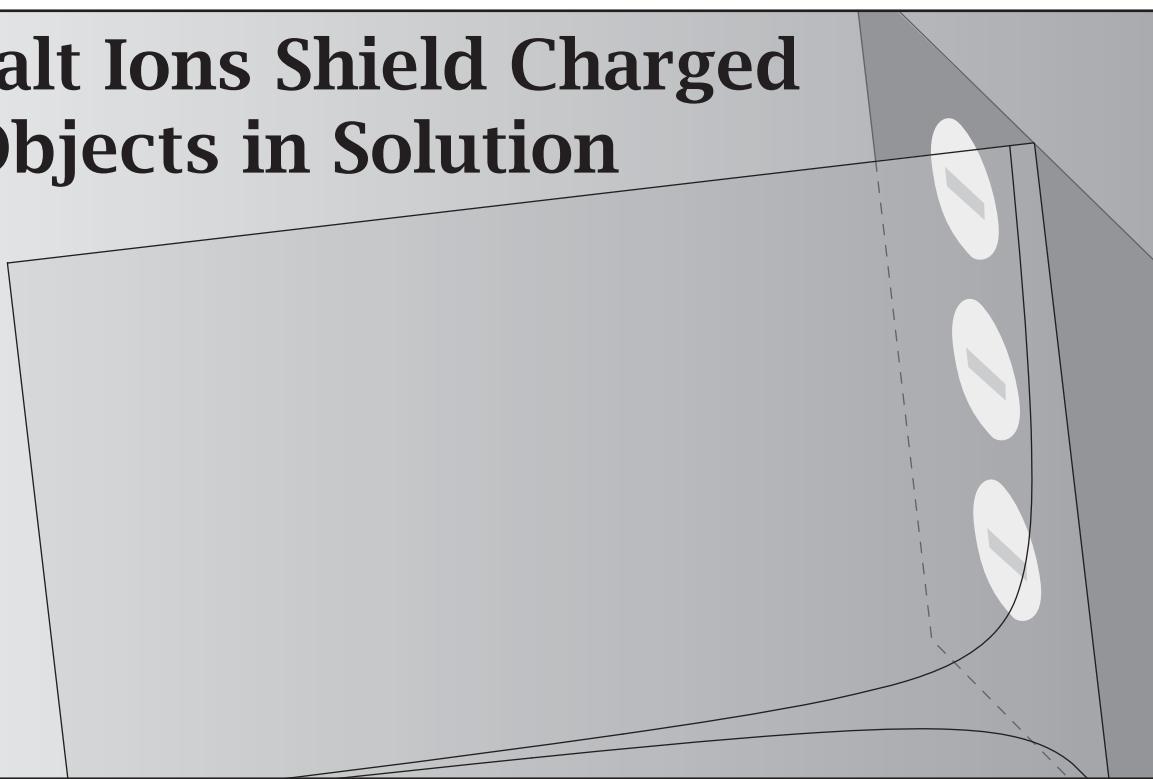
Suggested Reading

Excellent texts and lectures covering electrostatic forces and flows:

- K Boering, *Got Electrons? Redox Reactions*, Available at: <http://academicearth.org/lectures/got-electrons-redox-reactions>. An outstanding chemistry lecture from Kristie Boering at UC Berkeley on redox principles.
- CMA Brett and AMO Brett, *Electrochemistry: Principles, Methods, and Applications*, Oxford University Press, New York, 1993.
- B Hille, *Ionic Channels of Excitable Membranes*, 3rd edition, Sinauer, Sunderland, MA, 2001. Excellent discussion of the dipolar gate model and ion potentials.
- B Katz, *Nerve, Muscle, and Synapse*, McGraw-Hill, New York, 1966. Excellent overview of the Nernst equation and neural physics.
- N Lakshminarayanaiah, *Equations of Membrane Biophysics*, Academic Press, Orlando, 1984.
- SG Schultz, *Basic Principles of Membrane Transport*, Cambridge University Press, Cambridge, 1980.
- TF Weiss, *Cellular Biophysics*, Volume 1: Transport, MIT Press, Cambridge, MA, 1996.

This page is intentionally left blank.

23 Salt Ions Shield Charged Objects in Solution



Salts Dissociate into Mobile Ions and Shield Charged Objects in Water

Put some charged object or molecule into water. Now add salt. The salt will dissolve, dissociate, and swarm around the charged object, *shielding* it, often changing its properties. For example, DNA is a molecular chain of negative phosphate charges. Because its charges repel each other, DNA is elongated in pure water. When NaCl is added to an aqueous solution of DNA, the salt dissociates into positive ions (Na^+) and negative ions (Cl^-). The positive ions seek the DNA's negative charges, surrounding and shielding them. This shielding weakens the intra-DNA charge repulsions, causing the DNA to become more compact. Added salts can precipitate colloids, aggregate proteins, speed up chemical reactions, or allow biological cells to fuse together.

In this chapter, we explore how ions shield charged objects. But first we need some definitions. What do we mean by *added salt*? When you buy a bottle of some charged molecule, such as DNA, it will be electroneutral, so it will already have a stoichiometric complement of counterions. So you may get Na^+ -DNA or Mg^{2+} -DNA, for example. If you add no *additional salt*, this is called the *no-salt* condition. However, we are interested in situations where you add even more salt beyond the minimum required for electroneutrality. The electrostatic interactions of a charged object with added salt can be described by Poisson's equation, and its equilibrium is described by the Boltzmann distribution law.

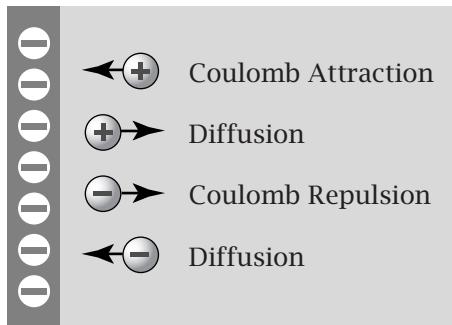


Figure 23.1 Dissociated mobile salt ions move in the presence of a charged surface. The counterions are attracted to the surface and the co-ions are repelled.

Combining them gives the Poisson-Boltzmann model of ionic equilibria and shielding. Molecules that dissociate completely are called *strong electrolytes* and those that dissociate only partly are called *weak electrolytes*.

The Poisson-Boltzmann Model for the Distribution of Mobile Salt Ions Around Charged Objects

Consider a charged molecule or object P . P attracts the mobile salt ions that have the opposite charge to P , called the *counterions*. P repels the mobile ions of the same sign, called the *co-ions* (see Figure 23.1). The counterions distribute around P and act as a sort of electrostatic shield, reducing the electrostatic potential from P that is felt by a charged particle distant from P . The interface between P and the neighboring salt solution is called the *electrical double layer*: the first layer is the charge on P and the second layer is the adjacent diffuse sea of excess counterions.

In the absence of electrolyte, two negatively charged P particles repel each other. But if a salt such as NaCl is added, the small ions intervene to weaken the charge repulsions between the two negative P particles, often to the point that other weak attractions can prevail and cause the P particles to associate. In this way, charged colloids can be induced to aggregate by the addition of salts.

EXAMPLE 23.1 What drives colloids to aggregate in salt water? Colloids are small particles (nanometers to micrometers in diameter) dispersed in gas, liquid, or solid phases. Examples include smoke, particulates in air, fog, mist, milk, foams, the inks in printers and copy machines, soap micelles, and some macromolecules and biological cells. Two colloidal particles in water will often interact through both a chemical attraction and an electrostatic repulsion, due to a net surface charge on each particle. An interesting case of electrostatic shielding is a *river delta*, for example where the Mississippi River meets the ocean water in the Gulf of Mexico. In the upstream fresh river water, fine silt particles become suspended, owing to soil erosion from the surrounding land. The silt particles are small enough to be suspended in the fresh water. And the silt particles repel each other because they carry a net charge. The fresh water carries the silt downstream to the ocean. When the fresh water meets the salty seawater, the salt shields the colloids, reducing their repulsions, causing the silt to precipitate. Silt deposits form the macroscopic landforms known as deltas. This process has grown the coastline of Louisiana by 50 miles over the past 5000 years. Derjaguin, Landau, Verwey, and Overbeek derived a model,

called the DLVO theory, based on the Poisson–Boltzmann theory described below, to explain the forces between colloidal particles [1]. The DLVO model describes how (1) uncharged particles attract each other, precipitating or *flocculating* out of solution, (2) highly charged particles in low-salt solutions repel each other, forming stable suspensions (such as India ink), and (3) charged particles in high-salt solutions attract each other, but with a kinetic barrier.

The simplest model of charge shielding and colloidal stability against aggregation was developed around 1910 independently by L-G Gouy (1854–1926), a French physicist, and DL Chapman (1869–1958), a British chemist. They combined Poisson's equation of electrostatics with the Boltzmann distribution law.

In the simplest case, the surface of a charged colloidal particle is described as a plane. Suppose that a charged particle P , which is fixed in space, produces an electrostatic potential $\psi(x)$, which you take to be a function of a single spatial coordinate x . P is in a salt solution of dissociated mobile ions. To keep the math simple, consider a salt in which the counterion and co-ion have the same valency; for example, a sodium ion has one positive charge and a chloride ion has one negative charge. For NaCl, $z_+ = z = 1$ and $z_- = -z = -1$. Let $n_+(x)$ represent the concentration of mobile positive ions (the number of ions per unit volume) at a distance x from the plane. This quantity is given by the Nernst or Boltzmann equation (22.8):

$$n_+(x) = n_\infty e^{-ze\psi(x)/kT}, \quad (23.1)$$

where n_∞ is the concentration of positive ions in the bulk solution, distant from the particle P . Far from P , the potential is $\psi(\infty) = 0$ (see Figure 23.2). An identical expression applies to the negative ions, except that z is replaced by $-z$, if the electrolyte is ‘symmetrical’ in its charge. Charge neutrality in the bulk requires that the number density of negative ions in the bulk far away from P is also n_∞ , so

$$n_-(x) = n_\infty e^{+ze\psi(x)/kT}. \quad (23.2)$$

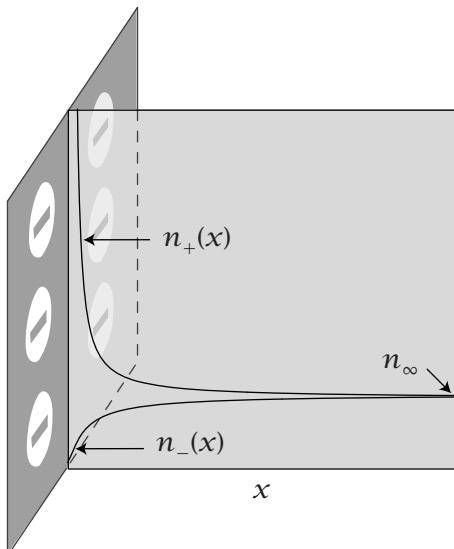


Figure 23.2 The concentration of counterions $n_+(x)$, and concentration of co-ions $n_-(x)$ as functions of distance x away from a negatively charged surface.

In this formulation, the valency z is the number of charges on the ion. So, for a solution of NaCl, you use $z_{\text{Na}} = z_{\text{Cl}} = 1$ in the expressions above.

The mobile ions not only *experience* the electrostatic field surrounding P , they also *contribute to it*. You can compute the electrostatic potential that arises both from the fixed charge of P and from the mobile charges of the dissociated salt ions. First, compute the charge density $\rho(x)$ as a function of the number of ions at position x :

$$\rho(x) = \sum_i z_i e n_i(x) = z e [n_+(x) - n_-(x)], \quad (23.3)$$

where z_i is the valency of an ion of species i and n_i is its concentration at position x . Now use Poisson's equation (21.36) to relate the charge density ρ to the electrostatic potential ψ :

$$\nabla^2 \psi = -\frac{\rho}{D\epsilon_0}, \quad (23.4)$$

where D is the dielectric constant of the solution. D is approximately equal to the dielectric constant of the solvent, because the ions themselves contribute little to D as long as their concentration is low.

Now, to formulate the **Poisson–Boltzmann equation**, substitute $n_+(x)$ and $n_-(x)$ from Equations (23.1) and (23.2) into Equation (23.3) for ρ , and then into Equation (21.36):

$$\nabla^2 \psi = \frac{z e n_\infty}{D\epsilon_0} (e^{z e \psi / kT} - e^{-z e \psi / kT}). \quad (23.5)$$

Equation (23.5) can be expressed more compactly in terms of the hyperbolic sine function, $\sinh x = (e^x - e^{-x})/2$:

$$\nabla^2 \psi = \frac{2 z e n_\infty}{D\epsilon_0} \sinh\left(\frac{z e \psi}{kT}\right). \quad (23.6)$$

The Poisson–Boltzmann equation (23.6) is a nonlinear second-order differential equation from which you can compute ψ if you know both the charge density on P and the bulk salt concentration n_∞ . This equation can be solved numerically by a computer. However, a linear approximation, which is easy to solve without a computer, applies when the electrostatic potential is small. For small potentials, $z e \psi / kT \ll 1$, you can use the approximation $\sinh x \approx [(1+x)-(1-x)]/2 = x$ (which is the first term of the Taylor series expansion for the two exponentials in $\sinh x$ (see Equation (J.1) of Appendix J)). Then Equation (23.6) becomes

$$\nabla^2 \psi = \frac{2 z e n_\infty}{D\epsilon_0} \left(\frac{z e \psi}{kT} \right) = \kappa^2 \psi, \quad (23.7)$$

where κ^2 is defined by

$$\kappa^2 = \frac{2 z^2 e^2 n_\infty}{D\epsilon_0 kT}. \quad (23.8)$$

Equation (23.7) is called the **linearized Poisson–Boltzmann** or the **Debye–Hückel equation**. $1/\kappa$ is called the **Debye length**. The Debye length is a screening, or shielding, distance. A charge that is closer to P than $1/\kappa$ ‘sees’ the

charged plane and interacts with it. A charge that is further than $1/\kappa$ from P is shielded from it by the intervening salt solution, which weakens the attraction or repulsion between the charge and P . Table 23.1 shows that increasing the concentration of the salt decreases the Debye length. Example 23.2 shows how the Debye length is computed.

EXAMPLE 23.2 Computing Debye lengths. For a monovalent salt, $z = 1$, with concentration 0.1 mol L^{-1} , show that the Debye length is 9.62 \AA if $D = 78.54$ (see Table 23.1).

Multiply the numerator and denominator on the right-hand side of Equation (23.8) by Avogadro's number \mathcal{N} to get

$$\kappa^2 = \frac{2(ze)^2 n_\infty \mathcal{N}}{D \varepsilon_0 R T}.$$

From the units given in the box on page 386, $e^2 \mathcal{N}/(4\pi\varepsilon_0) = 1.386 \times 10^{-4} \text{ J m mol}^{-1}$, so

$$\begin{aligned} \kappa^2 &= 2(4\pi) \left(1.386 \times 10^{-4} \text{ J m mol}^{-1} \right) \left(0.1 \text{ mol L}^{-1} \right) \left(10^3 \text{ L m}^{-3} \right) \\ &\quad \times \frac{\left(6.022 \times 10^{23} \text{ molecule mol}^{-1} \right)}{\left[(78.54) \left(8.314 \text{ J K}^{-1} \text{ mol}^{-1} \right) (298 \text{ K}) \right]} \\ &= 0.01078 \times 10^{20} \text{ m}^{-2} \implies \frac{1}{\kappa} = 9.62 \times 10^{-10} \text{ m} = 9.62 \text{ \AA}. \end{aligned}$$

More generally, you may be interested in salts in which the cation valency z_+ is not necessarily equal to the anion valency z_- . The **ionic strength I** of the solution is defined as

$$I = (c_+ z_+^2 + c_- z_-^2)/2, \quad (23.9)$$

where $c = n\mathcal{N}$ is the concentration in moles per liter and \mathcal{N} is Avogadro's number. For NaCl, Equation (23.9) reduces to $I = c_\infty = n_\infty \mathcal{N}$ since $c_+ = c_- = c_\infty$. In this general case, Equation (23.8) generalizes to

$$\kappa^2 = \frac{(z_+^2 + z_-^2) e^2 n_\infty}{D \varepsilon_0 k T} = \frac{2F^2 I}{D \varepsilon_0 R T}, \quad (23.10)$$

where $F = e\mathcal{N} = 23,060 \text{ cal mol}^{-1} \text{ V}^{-1}$ is the molar unit charge. A key result from Equation (23.10) that we will use below is the proportionality $\kappa \propto \sqrt{I}$.

The Poisson-Boltzmann model is used to compute the electrostatic potential ψ if you know the charge on a surface P and the concentration of salt in the solution. In Example 23.3, we compute $\psi(x)$ in the direction normal to a charged plane in a salt solution.

EXAMPLE 23.3 The potential near a uniformly charged plane in a salt solution. We are interested in how ψ depends on x , the perpendicular distance from a plane that has a surface charge density σ . You solve the corresponding

Table 23.1 The Debye length $1/\kappa$ describes the range of the potential, given here for various concentrations c of aqueous monovalent salt solutions at 25°C .

c (mol L $^{-1}$)	$1/\kappa$ (\text{\AA})
0.5	4.30
0.2	6.80
0.1	9.62
0.05	13.6
0.02	21.5
0.01	30.4
0.005	43.0
0.002	68.0
0.001	96.2

Poisson–Boltzmann equation (23.7):

$$\frac{d^2\psi}{dx^2} = \kappa^2\psi. \quad (23.11)$$

You can verify that the function

$$\psi(x) = A_1 e^{\kappa x} + A_2 e^{-\kappa x} \quad (23.12)$$

satisfies Equation (23.11) for any constants A_1 and A_2 by substituting Equation (23.12) into Equation (23.11) and carrying out the two differentiations indicated.

To solve Equation (23.11), you need two boundary conditions. We follow the same conventions that we used in solving Poisson's equation. First, the electrostatic potential is defined to be zero, $\psi(\infty) = 0$, as $x \rightarrow \infty$; this gives $A_1 = 0$. Second, we require that at $x = 0$, $\psi = \psi_0$, where ψ_0 is the potential at the surface. This gives $A_2 = \psi_0$, so the electrostatic potential is

$$\psi(x) = \psi_0 e^{-\kappa x}. \quad (23.13)$$

Equation (23.13) predicts that the electrostatic potential in salt solutions approaches zero *exponentially* as a function of distance from the charged plane. In contrast, Equation (21.13) shows that in the absence of salt, the electrostatic potential varies *linearly* with distance from a charged plane.

To complete Equation (23.13), you need to know the surface potential ψ_0 . You can get ψ_0 by using Equation (21.7), which relates the electrostatic potential to the field E , since we have already computed the field from a plane, $E = \sigma / (\epsilon_0 D)$ (Equation (20.24)). In the language of Equation (20.24), the charged plane of interest here is ‘thick,’ i.e., wider than the screening distance $1/\kappa$, which is typically nanometers or smaller. So you have

$$\frac{d\psi}{dx} = -E_x = -\frac{\sigma}{\epsilon_0 D}. \quad (23.14)$$

You can also compute the derivative of $\psi(x)$ from Equation (23.13):

$$\left(\frac{d\psi}{dx} \right)_{x=0} = -(\psi_0 \kappa e^{-\kappa x})_{x=0} = -\kappa \psi_0. \quad (23.15)$$

Comparing Equations (23.14) and (23.15) gives $\psi_0 = \sigma / (\kappa \epsilon_0 D)$. So the electrostatic potential near a planar surface of charge density σ is

$$\psi(x) = \frac{\sigma}{\kappa \epsilon_0 D} e^{-\kappa x}. \quad (23.16)$$

Equation (23.13) shows that the range of the electrostatic potential from the plane P shortens as κ increases. Because $\kappa \propto \sqrt{c_\infty}$, where c_∞ is the salt concentration (see Equation (23.10)), the range of the potential $\psi(x)$ is shortened by the added salt. At a distance equal to the Debye length, $x = 1/\kappa$, the potential ψ is decreased by a factor of $1/e$, to $\psi(1/\kappa) = \psi_0/e$. The Debye length shortens as the salt concentration increases. For example, for a z - z -valent salt

($+z$ on each cation and $-z$ on each anion) in water at 25°C, where $D = 78.54$, Equation (23.8) gives $1/\kappa = 3.044/(z\sqrt{c}) \text{ \AA}$.

Table 23.1 shows the dependence of the Debye length on the salt concentration in aqueous monovalent salt solutions. When the radius of curvature of particle P is much larger than $1/\kappa$, the double layer can be regarded as planar. Planar double layers have an important role as models in colloid chemistry, where particles often range in length from 10^2 to 10^5 \AA .

Example 23.4 finds the potential near a charged sphere, a model for proteins, small ions, and micelles in salt solutions.

EXAMPLE 23.4 The spherical double layer. Let's compute the electrostatic potential $\psi(r)$ as a function of the radial distance from a charged sphere in a salt solution. The sphere has radius a , net charge Q , and a uniform surface charge density $\sigma = Q/(4\pi a^2)$. The sphere is in a monovalent salt solution that has a Debye length $1/\kappa$. The potential changes only in the radial direction, so $(\partial\psi/\partial\theta) = 0$ and $(\partial^2\psi/\partial\Phi^2) = 0$. For spherical coordinates (see Equation (G.24) in Appendix G), the linearized Poisson-Boltzmann equation (23.7) becomes

$$\nabla^2\psi = \frac{1}{r} \frac{d^2(r\psi)}{dr^2} = \kappa^2\psi. \quad (23.17)$$

Multiply both sides of Equation (23.17) by r to get

$$\frac{d^2(r\psi)}{dr^2} = \kappa^2 r\psi. \quad (23.18)$$

Comparison of Equation (23.18) with Equation (23.11) shows that $r\psi(r)$ has the same role as $\psi(x)$ in the planar problem. So the general solution of Equation (23.18) is

$$r\psi = A_1 e^{\kappa r} + A_2 e^{-\kappa r} \implies \psi = A_1 \frac{e^{\kappa r}}{r} + A_2 \frac{e^{-\kappa r}}{r}. \quad (23.19)$$

Again, you need two boundary conditions to establish A_1 and A_2 . It is conventional to choose $\psi(\infty) = 0$, so $A_1 = 0$. The second boundary condition is at $r = a$, where $\psi = \psi_a = A_2 e^{-\kappa a}/a$. Inverting this expression gives $A_2 = a\psi_a e^{\kappa a}$. Then Equation (23.19) becomes

$$\psi = \frac{a\psi_a}{r} e^{-\kappa(r-a)} \quad \text{for } r > a. \quad (23.20)$$

Now determine the surface potential ψ_a from the total charge Q . You can find this by taking the derivative $d\psi/dr$ of Equation (23.20) at $r = a$:

$$\left(\frac{d\psi}{dr} \right)_{r=a} = -a\psi_a e^{\kappa a} \left[\frac{\kappa}{r} e^{-\kappa r} + \frac{1}{r^2} e^{-\kappa r} \right]_{r=a} = -\psi_a \frac{1+\kappa a}{a}. \quad (23.21)$$

Now use Equation (21.7) to relate ψ and E , and Gauss's law for $E(r = a)$ around a sphere (Equation (21.43)) to get

$$\left(\frac{d\psi}{dr} \right)_{r=a} = -E_{r=a} = \frac{-\mathcal{C}Q}{Da^2}. \quad (23.22)$$

Setting Equation (23.22) equal to Equation (23.21) gives ψ_a in terms of Q and κ :

$$\psi_a = \frac{\mathcal{C}Q}{D\alpha(1+\kappa\alpha)}. \quad (23.23)$$

In the absence of added salt, $\kappa = 0$ and $\psi_a = \mathcal{C}Q/D\alpha$. Equation (23.23) shows that the effect of added salt is to shield the charge on the sphere and reduce the surface potential by a factor $(1+\kappa\alpha)$. Substituting Equation (23.23) into Equation (23.20) gives the potential we seek:

$$\psi(r) = \frac{\mathcal{C}Q}{Dr(1+\kappa\alpha)} e^{-\kappa(r-\alpha)}. \quad (23.24)$$

In the absence of added salt, $\kappa = 0$ so $\psi \propto r^{-1}$, but the presence of salt introduces shielding that causes ψ to decrease with an additional exponential dependence on r .

Equation (23.24) gives the electrostatic potential from a point charge Q ($\alpha = 0$) in a salt solution as

$$\psi(r) = \frac{\mathcal{C}Q}{Dr} e^{-\kappa r}. \quad (23.25)$$

The Poisson–Boltzmann equation describes spatial variations in salt concentrations over microscopic length scales near the surfaces of molecules and particles. Equation (23.24) holds for a large spherical particle having a uniform surface charge Q , and also for a small ion, such as sodium or chloride, that has a single charge.

Equations (23.24) and (23.25) come from the *linearized* Poisson–Boltzmann Equation (23.7). It is a general property of linear differential equations that you can sum their solutions. For example, if you have two point charges in a uniform salt solution, q_1 at distance r_1 from P and q_2 at distance r_2 from P , the Debye–Hückel potential at P is

$$\psi_P(r_1, r_2) = \frac{\mathcal{C}q_1}{D} \frac{e^{-\kappa r_1}}{r_1} + \frac{\mathcal{C}q_2}{D} \frac{e^{-\kappa r_2}}{r_2}. \quad (23.26)$$

Summing potentials this way is sometimes the easiest way to solve the linear Poisson–Boltzmann equation for an arbitrary constellation of charges.

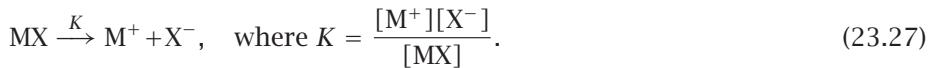
Adding salt to a solution can drive charged particles to aggregate, as shown in Example 23.1.

Before electrolytes can shield charged objects, they must first dissociate in solution. Some electrolytes dissociate fully and others only partly. One of the main uses of the Poisson–Boltzmann equation has been to understand the degree of dissociation in electrolyte solutions.

Electrolytes Are Strong or Weak, Depending on Whether They Fully Dissociate in Water

Salts and acids and bases are *electrolytes*. In aqueous solutions, electrolytes dissolve and dissociate into ions. Molecules that dissociate completely into individual ions are called *strong electrolytes*, and those that dissociate only partly

are called *weak electrolytes*. The degree of dissociation can be represented by the equilibrium dissociation constant K :



A strong electrolyte has a large K , typically $K > 0.1$. Using the notation $pK = -\log_{10} K$, this means that a strong electrolyte has $pK < 1$ (the pK can be negative, which occurs if $K > 1$). For a weak electrolyte, $K < 0.1$, so $pK > 1$.

Much of what is known about ion dissociation comes from two types of experiment: measurements of colligative properties and ion conductivities. Chapter 16 shows that the freezing point depression ΔT_f should increase with the concentration of a solute species m according to $\Delta T_f = K_f m$, where K_f is a constant that is independent of the solute (see page 292). But what value of m should you use for a solute, like NaCl, that can dissociate? Each mole of NaCl dissociates into two moles of ions, Na^+ and Cl^- . Should m be the concentration of NaCl or twice the concentration of NaCl? If you let m be the concentration of the neutral species then the observed value i in the colligative expression $\Delta T_f = imK_f$ gives a measure of the degree of dissociation. For example, if you observe that $i = 2$ for NaCl, it implies full dissociation.

The degree of ion dissociation can also be measured by *ion conductivities*. Ion conductivities are found by dissolving salts in water and applying an electric field, which causes the ions to flow, resulting in a measurable electrical current. The conductivity is the observed current flow (number of ions per unit time) divided by the applied voltage. Increasing the electrolyte concentration increases the number of charge carriers and thus the conductivity. For example, at low concentrations, KCl fully dissociates into K^+ and Cl^- ions, so twice the KCl concentration gives the number of ions that carry current.

A strong electrolyte is defined by two observations: (1) the value of i measured by colligative properties equals the number of ion types that would be expected from complete dissociation (for NaCl, $i = 2$; for Na_2SO_4 , $i = 3$, etc.), and (2) doubling the electrolyte concentration leads to a doubling of the conductivity of the solution, implying that the ions dissociate completely and that each ion acts as an independent carrier of electrical current. The principle that ions contribute independently to conductivity is called *Kohlrausch's law*, named after the German chemist F Kohlrausch (1840–1910). NaCl, HCl, KCl, and K_2SO_4 are strong electrolytes when they are at low concentrations.

Figure 23.3 shows that electrolyte dissociation can deviate in two different ways from the ideal behavior predicted by Kohlrausch's law. First, weak electrolytes, such as acetic acid and ammonia, do not fully dissociate at any concentration, so their conductivities and colligative i values are smaller than predicted by Kohlrausch's law.

Second, even strong electrolytes that fully dissociate at low electrolyte concentrations may become nonideal at higher concentrations because the positive ions tend to be surrounded by negative ions, and vice versa. Table 23.2 shows that a strong electrolyte, K_2SO_4 , appears to dissociate completely into three ions ($i \rightarrow 3$, where i is the dissociation parameter) as the electrolyte concentration approaches zero. We want to explain why strong electrolytes don't fully dissociate. At low concentrations, nonidealities depend only on the valence and electrolyte concentration, and not on the chemical identity of the ions. Each ion

Table 23.2 Freezing point depression of water by K_2SO_4 salt. At low concentrations, the salt approaches full dissociation, $i \rightarrow 3$.

Molality	i
0.10	2.32
0.05	2.45
0.01	2.67
0.005	2.77
0.001	2.84

Source: CH Langford and RA Beebe, *The Development of Chemical Principles*, Addison-Wesley, Reading, MA, 1969 (reprinted by Dover Publications, New York, 1995).

Conductance

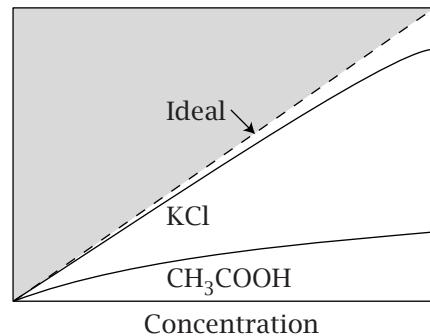


Figure 23.3 For strong electrolytes like KCl, the conductance is proportional to the electrolyte concentration (except at very high concentrations). For weak electrolytes like acetic acid (CH₃COOH), it is not. Source: CH Langford and RA Beebe, *The Development of Chemical Principles*, Addison-Wesley, Reading, MA, 1969 (reprinted by Dover Publications, New York, 1995).

is partly *shielded* by its neighboring ions, leading to an apparently incomplete dissociation. This is described by the Debye-Hückel model.

Debye-Hückel Theory for the Nonidealities of Strong Electrolytes

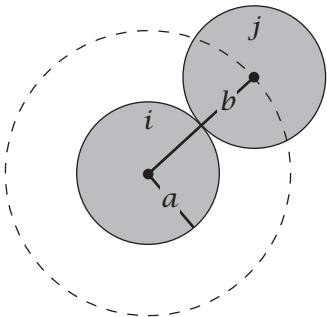


Figure 23.4 In the Debye-Hückel theory, a central ion of type *i* and radius *a* is surrounded by counterions of type *j*. The center-to-center distance of closest approach is *b*.

In 1923, E Hückel and P Debye, winner of the 1936 Nobel Prize in Chemistry, adapted the Poisson-Boltzmann theory to explain the nonidealities of dilute solutions of strong electrolytes. To visualize Na⁺ ions surrounded by Cl⁻ ions and Cl⁻ surrounded by Na⁺ at the same time, think of a NaCl crystal that is expanded uniformly. Now add fluctuations. Debye and Hückel focused on one ion as a charged sphere, and used the linear approximation to the Poisson-Boltzmann equation to compute the electrostatic free energy of creating the *nonuniform distribution* of its surrounding counterions and co-ions: more Cl⁻ ions are close to a Na⁺, and vice versa.

We focus on one ion, called the central ion (Figure 23.4), in a salt solution. Let *b* be the distance of closest approach between the centers of two oppositely charged ions. The charge *Q* on the central ion is given in terms of its valence: *Q* = *ze*. The nonideality of the solution can be described in terms of an activity coefficient *y* in the expression for the chemical potential of the salt:

$$\begin{aligned}\mu_{\text{NaCl}} &= \mu_{\text{Na}}^\circ + kT \ln c_{\text{Na}} + kT \ln y_+ + \mu_{\text{Cl}}^\circ + kT \ln c_{\text{Cl}} + kT \ln y_- \\ &= \mu^\circ + 2kT \ln c_{\text{NaCl}} + 2kT \ln y,\end{aligned}\quad (23.28)$$

where *c_{NaCl}* = *c_{Na}* = *c_{Cl}* and *y* = $\sqrt{y_+ y_-}$ is the mean activity, since the anion and cation cannot be measured separately.

Debye and Hückel regarded all the nonideality as arising from the electrostatics, and none from short-ranged interactions. In their model, the nonideality reflected in the activity coefficient arises because of the nonuniformity of nearby charges. Around a central ion is an excess of counterions and a depletion of co-ions, in a spherically symmetrical way.

The activity coefficient for one of the ions in the pair, say Na⁺, can be computed by assuming a reversible process of charging up the central ion from *Q* = 0 to *Q* = *ze*, using a parameter $0 \leq \lambda \leq 1$ to describe the degree of charging:

$$kT \ln y_+ = \int_0^1 \psi_{\text{dist}}(\lambda) z_+ e d\lambda. \quad (23.29)$$

The quantity $kT \ln \gamma$ in Equation (23.29) is the electrostatic free energy that accounts for the ion-induced distribution of neighboring ions.

How do we compute ψ_{dist} ? Charging up the central ion results in two kinds of work. First, even if there were no added salt present, charging the central ion would give rise to a Born energy for the central ion in the pure solvent. The Born energy, Equation (22.60), is not dependent on the salt concentration, so it will be included in μ° in Equation (23.28), rather than in the activity coefficient. Second, charging the central ion also leads to a redistribution of the surrounding salt ions, drawing counterions nearer to the central ion and pushing co-ions further away. We regard ψ_{dist} , the electrostatic potential responsible for the nonideality, as the difference between the total potential ψ_b for charging up the central ion, and the potential $\psi_{\text{no salt}}$ for charging up the central ion in the pure solvent where $\kappa = 0$. Inside the radius b , the total charge is ze . So you can use Equation (23.23) with the radius b , instead of the ion radius a , to get ψ_b . Then

$$\begin{aligned}\psi_{\text{dist}} &= \psi_b - \psi_{\text{no salt}} = \frac{\mathcal{C}z_+e}{Db(1+\kappa b)} - \frac{\mathcal{C}z_+e}{Db} \\ &= -\frac{\mathcal{C}z_+e}{D} \frac{\kappa}{1+\kappa b}.\end{aligned}\quad (23.30)$$

At any intermediate state of charging, $0 \leq \lambda \leq 1$, the potential is

$$\psi_{\text{dist}}(\lambda) = -\lambda \frac{\mathcal{C}z_+e}{D} \frac{\kappa}{1+\kappa b}. \quad (23.31)$$

Substitute $\psi_{\text{dist}}(\lambda)$ from Equation (23.31) into Equation (23.29) and integrate:

$$kT \ln \gamma_+ = -\frac{\mathcal{C}z_+^2 e^2}{D} \left(\frac{\kappa}{1+\kappa b} \right) \int_0^1 \lambda d\lambda = -\frac{\mathcal{C}z_+^2 e^2}{2D} \left(\frac{\kappa}{1+\kappa b} \right). \quad (23.32)$$

Equation (23.32) predicts that $\ln \gamma = \ln(\gamma_+ \gamma_-) \propto \kappa \propto \sqrt{I}$ is proportional to the square root of the ionic strength at low salt concentrations I (see Equation (23.9)), where $\kappa b \ll 1$. Figure 23.5 shows $\log \gamma$ versus \sqrt{I} for several salts. The straight lines are calculated from the Debye–Hückel equation (23.32) for $b = 0$. This is called a limiting law because all ionic solutions are predicted to obey it in the limit of sufficiently low concentrations. For $b = 0$, the Debye–Hückel theory fits the limiting law slopes very well. Furthermore, with

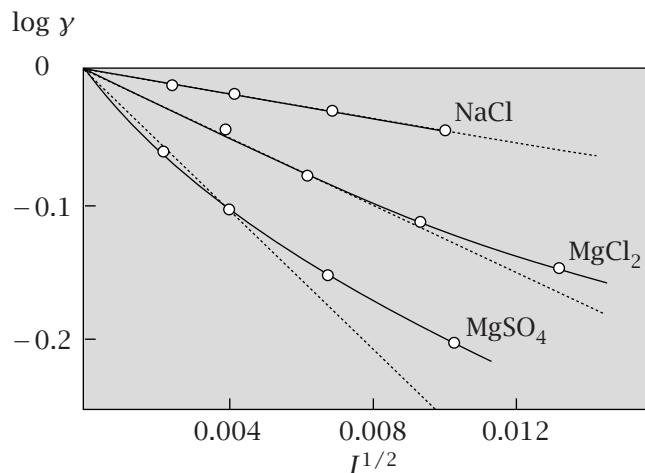


Figure 23.5 Experimental activity coefficients for various electrolytes (—) compared with the Debye–Hückel limiting law (····). Source: P Atkins, *Physical Chemistry*, 6th edition, WH Freeman, New York, 1998.

reasonable values of the distance of closest approach b , Equation (23.32) gives excellent agreement for higher concentrations, up to $c = 0.1 \text{ M}$ for monovalent salts at 25°C .

The Debye–Hückel model is useful for explaining how the rates of chemical reactions depend on salt concentration.

EXAMPLE 23.5 Salts affect the rates of chemical reactions. Consider the chemical reaction in Equation (19.13):



The rate k_0 of the reaction in the absence of salt (subscript 0) is proportional to the concentration of activated species:

$$k_0 \propto [(AB)^{\ddagger}]_0 = \text{constant } [A][B]. \quad (23.33)$$

The rate k_s of the reaction in the presence of salt (subscript s) depends, in addition, on electrostatic nonidealities, which are accounted for by activity coefficients:

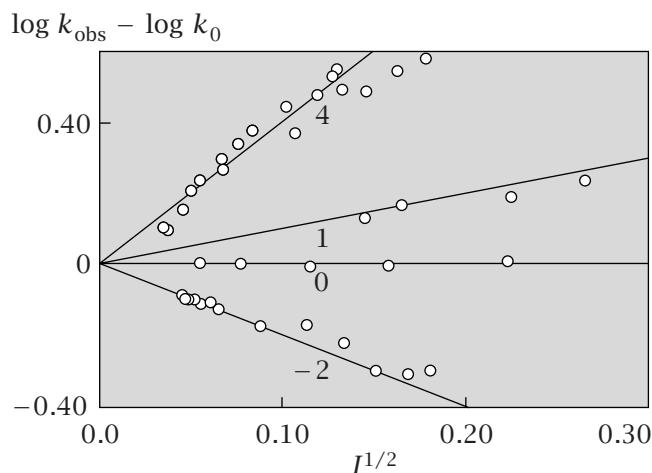
$$k_s \propto [(AB)^{\ddagger}]_s = \text{constant } [A][B] \frac{\gamma_A \gamma_B}{\gamma_{AB^{\ddagger}}}. \quad (23.34)$$

Dividing Equation (23.34) by Equation (23.33) and taking logarithms gives

$$\ln\left(\frac{k_s}{k_0}\right) = \ln \frac{\gamma_A \gamma_B}{\gamma_{AB^{\ddagger}}}. \quad (23.35)$$

If the total charge is conserved in the transition state, you have $z_{\text{TS}} = z_A + z_B$, where z_{TS} is the valence of the transition-state structure and z_A and z_B are the valences on the reactants A and B , respectively. Substituting the Debye–Hückel

Figure 23.6 Chemical reaction rates of ions as a function of the square root of the ionic strength $I^{1/2}$, for various reactions. The top curve, for example, is $2[\text{Co}(\text{NH}_3)_5\text{Br}]^{2+} + \text{Hg}^{2+} + 2\text{H}_2\text{O} \rightarrow 2[\text{Co}(\text{NH}_3)_5\text{H}_2\text{O}]^{3+} + \text{HgBr}_2$. The numbers on the curves are the values of $z_A z_B$, the signed product of the valencies of the reacting ions. The lines show predictions from the Debye–Hückel model. Source: ML Bender, RJ Bergeron, and M Komiyama, *The Bioorganic Chemistry of Enzymatic Catalysis*, Wiley, New York, 1984.



equation (23.32) for the activity coefficients into Equation (23.35) gives

$$\begin{aligned}\ln\left(\frac{k_s}{k_0}\right) &= \frac{-Ce^2}{2DkT} \left(\frac{\kappa}{1+\kappa b}\right) [z_A^2 + z_B^2 - (z_A + z_B)^2] \\ &= \frac{Cz_A z_B e^2}{DkT} \left(\frac{z_A z_B \kappa}{1+\kappa b}\right) \approx 2A_0 z_A z_B \sqrt{I},\end{aligned}\quad (23.36)$$

where we simplified $[z_A^2 + z_B^2 - (z_A + z_B)^2] = -2z_A z_B$. A_0 is a constant. The approximation on the right side comes from considering the limit of low salt concentrations, for which $\kappa b \ll 1$, and from recognizing that the Debye length is defined by $\kappa \propto \sqrt{I}$; see Equation (23.10). Figure 23.6 shows experimental evidence that reaction rates depend on the product of the valences of the ions involved and the square root of the ionic strength of the medium.

The Poisson-Boltzmann and Debye-Hückel models are relatively successful predictors of the shielding of charged objects by dissociated salt ions. But these models are approximate, and they have limitations. The Debye-Hückel model shows that when salts dissociate into ions, the ions are not distributed uniformly, but are clumpy. There is an enhanced concentration of mobile negative ions around each mobile positive ion, and vice versa. The main nonideality of salt solutions is attributed to this clumpiness.

But the Poisson-Boltzmann theory treats only the distribution of mobile ions as a function of distance from the macro-ion P , and does not also treat the clumpiness of the counterion/co-ion distribution in the solution. The Poisson-Boltzmann approach involves an implicit averaging over the distributions of the mobile ions, which eliminates the discreteness of the small ions [2]. The Poisson-Boltzmann model works best for low ion concentrations, and for monovalent mobile ions because the clumpiness is greatest for multivalent ions, $z = 2$ or $z = 3$, etc. Better approaches, such as integral equation treatments or Monte Carlo simulations, take into account all the charge interactions, but at the expense of greater complexity [3].

Summary

When ions are dissolved in solution, they dissociate. The Poisson-Boltzmann equation predicts that mobile ions form electrostatic shields around charged objects. Sometimes the dissociation of strong electrolytes is complete. In other cases, dissociation is incomplete. The apparent incomplete dissociation of strong electrolytes is described in terms of the long-range electrostatic interactions between the ions. Strong electrolyte behavior can be described by the Debye-Hückel theory, a linear approximation to the Poisson-Boltzmann equation.

Problems

1. The potential around colloidal spheres. What is the dimensionless surface potential $\Phi = e\psi/kT$ at a distance of 50 Å from

- a colloidal sphere with a radius of 20 Å and charge 20e in pure water?
- the same sphere in 0.1 M NaCl at 25°C?
- What are the potentials in V?

2. The potential near a protein in salt solution. Consider a protein sphere with a radius of 18 Å, and charge $Q = -10e$, in an aqueous solution of 0.05 M NaCl at 25°C. Consider the small ions as point charges and use the Debye-Hückel linear approximation of the Poisson-Boltzmann equation.

- What is the dimensionless surface potential $e\psi_a/kT$ of the protein?
- What are the concentrations of Na^+ and Cl^- ions at the surface of the protein?
- What are the concentrations of Na^+ and Cl^- ions at a distance of 3 Å from the protein surface?

3. Surface potentials and Debye lengths. You have a uniformly charged sphere with radius $a = 50 \text{ \AA}$ in a 0.02 M NaCl solution. At a distance of 30 Å from the surface of the sphere, the potential $\psi = 20 \text{ mV}$.

- What is the Debye length $1/\kappa$ in the solution?
- What is the surface potential ψ_a of the sphere? (Assume that the potential field in the solution around the sphere can be derived from the linear Poisson-Boltzmann equation.)
- What is the charge Q on the sphere?
- Sketch the potential as a function of distance from the sphere.

4. The potential near a plane. You have a uniformly charged flat plate in contact with a 0.02 M NaCl solution at 25°C. At a distance of 3 nm from the plate, the potential is 30 mV.

- What is the Debye length $1/\kappa$ in the solution?
- What is the surface potential of the charged plane in mV, and in units of kT/e ? Use the linear Poisson-Boltzmann equation.
- If you had used the nonlinear Poisson-Boltzmann equation, would you find the surface potential to be larger or smaller than you found under (b)? Why?
- Use the surface potential from (b) to find the surface charge density σ of the plane.

5. A charged protein near a chromatographic surface. Consider a spherical protein in water with charge $-q$ ($q > 0$) at a distance R_0 from a planar ion chromatography column surface with net positive charge, shown in Figure 23.7.

- Draw the field lines approximately. Include arrows to indicate the field direction.
- Draw the equipotential contours.

(c) If the solvent water were replaced by a water/methanol mixture with a lower dielectric constant, would it weaken or strengthen the attraction between the protein and the planar surface?

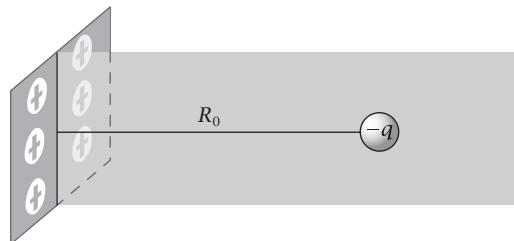


Figure 23.7

6. The potential of a membrane. Consider a phospholipid bilayer membrane consisting of 90% uncharged lipid (zwitterionic phosphatidylcholine) and 10% acid lipid (singly charged phosphatidylserine or phosphatidylglycerol). Assume 68 \AA^2 surface area per lipid head group. The membrane is in contact with an aqueous solution of NaCl, concentration c_{NaCl} at 25°C. Calculate the surface potential of the membrane for $c_{\text{NaCl}} = 0.05 \text{ M}$ and 0.1 M .

7. Binding to a membrane. What is the electrostatic free energy of binding to the membrane in Problem 6 of a trivalent positive ion such as spermidine (a biologically active polyamine), assuming that

- binding occurs at the membrane surface, and
- owing to steric factors, the charges of the bound spermidine stay in the water 5 Å distant from the membrane surface?

8. Electrostatic potential near a protein. A protein in aqueous solution with 0.1 M monovalent salt has 9 positive and 22 negative charges at 25°C. It is modeled as a sphere with radius 20 Å and uniform surface charge. What is the potential in units of kT/e :

- at the surface of the protein?
- at a distance 10 Å from the protein surface into the solution?

Use (i) Coulomb's equation (no shielding by small ions) and (ii) the linear Poisson-Boltzmann equation.

9. Debye-Hückel model. Apply the Debye-Hückel theory to a 0.01 M monovalent salt solution at 25°C. Treat the ions as hard spheres with a radius of 2 Å. Assume $D = 80$. What is the change in chemical potential in cal mol^{-1} due to the ion interactions? What is the activity coefficient of the ions?

10. Electrostatic potential around a line charge. A line charge with density λ and length ℓ is in a salt solution with Debye length $1/\kappa$. Use Equation (23.26) to find an integral expression for the potential in the plane bisecting the line charge.

11. Electrostatic potential near a vesicle. A simple model for a spherical vesicle is a spherical shell with radius a and charge density σ , permeated by a salt solution with Debye length $1/\kappa$. Use Equation (23.26) to derive the potential outside the shell at distance r from the center.

- Derive and discuss the limit of $\psi(r)$ for $\kappa \rightarrow 0$.
- For the situation in (a), derive the potential inside the spherical shell. What is the limit for $\kappa \rightarrow 0$?
- From the results of (a) and (b), derive the expression for the limit $\kappa a \rightarrow \infty$ and compare these with the result of Example 23.3.

12. Debye lengths. In the Debye-Hückel theory of monovalent salt solutions, there is a characteristic length quantity κ , defined by $\kappa^2 = (2e^2 n_\infty)/(\epsilon_0 D kT)$, where n_∞ is the salt concentration.

- Express κ in terms of the Bjerrum length ℓ_B .
- For water at room temperature, $\ell_B = 7.13 \text{ \AA}$. Compute the Debye length $1/\kappa$ in \AA for a solution of $n_\infty = 1 \text{ mol L}^{-1}$.

13. Ion binding to a sphere. Compute the free energy of bringing a divalent ion into contact with a spherical particle of radius $a = 14 \text{ \AA}$, from far away. The ion has valency $z = +2$ and the particle has valency $Z = +20$. For water at room temperature, compute the free energy:

- in a solution having monovalent salt concentration 0.1 M ;
- for a solution having monovalent salt concentration 0.01 M .

14. Polymer on charged surface. A polymer with a single positive charge at one end is attached to a uniformly charged sheet in a 0.1 M NaCl solution at the other end. It can take on the conformations shown in Figure 23.8. The distance d between model monomer beads is 2 \AA .

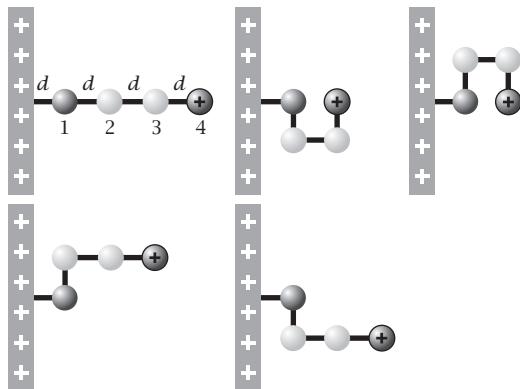


Figure 23.8

The electric potential at bead 4 due to the shielded sheet is

$$\psi(x) = \frac{\sigma}{\kappa \epsilon_0 D} e^{-\kappa x},$$

where $\sigma = 4 \times 10^{-4} \text{ C m}^{-2}$, $D = 80$, and $\kappa = 10 \text{ \AA}$. Beads 1 and 4 have a favorable contact energy of $\varepsilon = -2kT$, where

$T = 300 \text{ K}$. If the beads are not adjacent to each other, there is no contact energy.

- Compute the energy and density of states for each energy level of the ensemble shown in Figure 23.8.
- What is the partition function for this polymer? Express your answers in terms of the energies computed above.
- Based on your partition function, what is the fraction of polymer in each state?
- Now assume that the system is moved to a lower dielectric medium, with $D = 20$, but still with 0.1 M NaCl. What is the fraction of polymer in the linear state when immersed in the new solvent? Explain why this fraction increased or decreased from the previous system.

15. Salt can fold or unfold a protein. A particular protein has a large net charge at low pH. Does adding salt stabilize the folded or the unfolded state?

16. Another Debye length. What is the Debye length for a 0.01 M solution of magnesium sulfate (MgSO_4) at $T = 300 \text{ K}$?

References

- [1] RJ Hunter, *Foundations of Colloid Science*, 2nd edition, Oxford University Press, Oxford, 2000.
- [2] R Kjellander, *Ber Bunsen Ges Phys Chem* **100**, 894–904 (1996).
- [3] V Vlachy, *Annu Rev Phys Chem* **50**, 145–165 (1999).

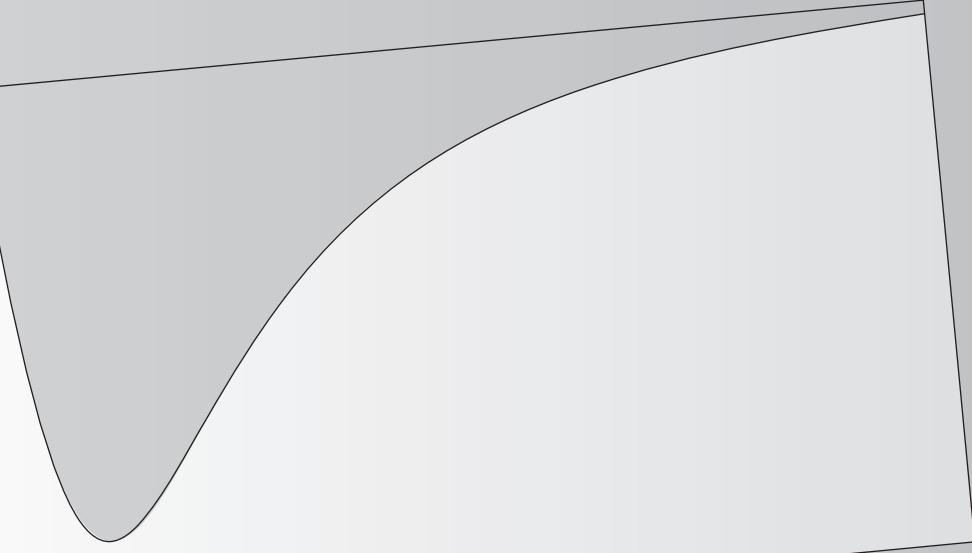
Suggested Reading

Excellent treatments of the Poisson-Boltzmann and Debye-Hückel theories:

- AW Adamson and AP Gast, *Physical Chemistry of Surfaces*, 6th edition, Wiley, New York, 1997.
- CMA Brett and AMO Brett, *Electrochemistry: Principles, Methods and Applications*, Oxford University Press, New York, 1993.
- PC Hiemenz and R Rajagopalan, *Principles of Colloid and Surface Chemistry*, 3rd edition, Marcel Dekker, New York, 1997.
- RJ Hunter, *Foundations of Colloid Science*, 2nd edition, Oxford University Press, Oxford, 2000.
- RJ Hunter, *Introduction to Modern Colloid Science*, Oxford University Press, Oxford, 1993.
- J Lyklema, *Fundamentals of Interface and Colloid Science, Volume 1: Fundamentals*, Academic Press, San Diego, 1991.
- WB Russel, DA Saville, and WR Schowalter, *Colloidal Dispersions*, Cambridge University Press, New York, 1989.

This page is intentionally left blank.

24 Intermolecular Interactions



Atoms and molecules, even uncharged ones, are attracted to each other. They can form noncovalent as well as covalent bonds. We know this because they condense into liquids and solids at low temperatures. Noncovalent interactions between uncharged particles are relatively weak and short-ranged, but they are the fundamental driving forces for much of chemistry, physics, and biology. These intermolecular interactions can be understood through measurements of the pressures of nonideal gases. The laws of electrostatics explain the attractions between charged atoms, say Na^+ and Cl^- . But what forces cause neutral atoms to bond together noncovalently? To a first approximation, the bonding between neutral molecules can also be explained by electrostatic interactions. Even when the atoms have no net charge, they have *charge distributions* and *polarizabilities*, which lead to weak attractions.

Molecules Repel Each Other at Very Short Range and Attract at Longer Distances

A bond between two particles is described by a *pair potential* $u(r)$, the energy as a function of the separation r between the particles.

The force $f(r)$ between two particles is the derivative of the pair potential:

$$f(r) = -\frac{du(r)}{dr}. \quad (24.1)$$

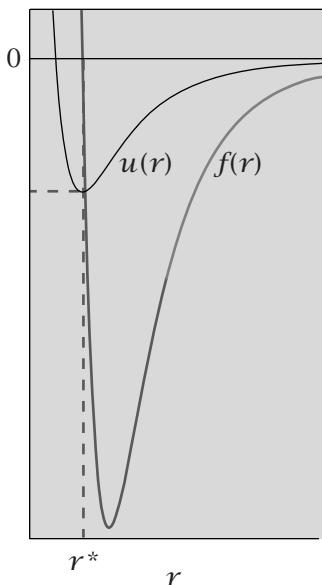


Figure 24.1 The energy of interaction $u(r)$ between two particles as a function of their separation r , and the corresponding force $f = -du/dr$; r^* is the equilibrium bond length.

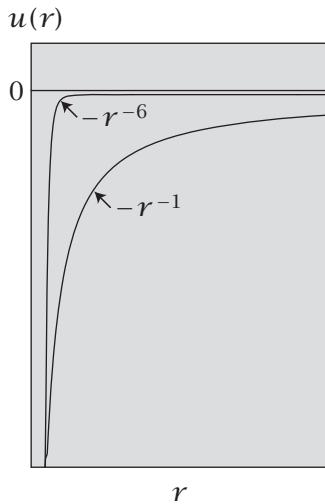


Figure 24.2 The range of an interaction. r^{-1} is long-ranged and r^{-6} is short-ranged.

Pair potentials have three main features (see Figure 24.1). First, particles do not interact if they are far apart ($u \rightarrow 0$ and $f \rightarrow 0$ as $r \rightarrow \infty$). Second, if two particles are close enough, they attract each other ($f < 0$ indicates that a positive force acts along the $-r$ direction to bring the particles together). Third, particles repel if they are too close together ($f > 0$ indicates that a positive force acts along the $+r$ direction to separate the particles).

At $r = r^*$, the attraction balances the repulsion, and the pair potential is at a minimum, so the force between two particles is zero. r^* represents the average length of an intermolecular ‘bond.’ Although the functional form for $u(r)$ is different for covalent bonds than for weaker interactions, all bonds have these main features: no interaction at very long range, attraction at short range, and repulsion at very short range, leading to an average equilibrium bond length where the net force is zero.

Intermolecular forces can be divided into two classes: those that are *long-ranged* and those that are *short-ranged*. The *range* is defined by the dependence of u on the separation r . Intermolecular interactions are commonly modeled as a power law:

$$u(r) = (\text{constant})r^{-p}, \quad (24.2)$$

where p is a positive integer. Interactions are called short-ranged if $p > 3$ and long-ranged if $p \leq 3$. The Coulombic interaction $u(r) \propto \pm 1/r$ is long-ranged, while $u(r) \propto \pm 1/r^6$ is short-ranged, whether the interactions are attractive ($u(r)$ is negative) or repulsive ($u(r)$ is positive) (see Figure 24.2).

When two atoms are very close together in space, they repel, owing to the Pauli principle that electrons in the same state can't occupy the same space. Quantum mechanical calculations show that these very short-ranged repulsions can best be modeled as exponential functions or power laws, typically with $p = 9, 12$, or 14 .

Short-Ranged Attractions Can Be Explained as Electrostatic Interactions

Electrostatic interactions are long-ranged. The exponent in Equation (24.2) is $p = 1$ in Coulomb's law, $u(r) = C/r$. Yet Coulomb's law can also explain weak intermolecular attractions, which are short-ranged. Even neutral atoms and molecules, which have no net charge, have *charge distributions*: some positive charge here and some negative charge there.

There are three different ways in which a shorter-ranged exponent ($p > 1$) can arise from Coulombic interactions between neutral molecules. (1) Atoms and molecules may be *multipoles*, such as *dipoles* or *quadrupoles*, in which the charge density varies throughout the molecule, even when there is no net charge. The interactions between multipoles are shorter-ranged than the interactions between monopoles. (2) When dipolar atoms or molecules are free to orient, orientational averaging reduces the range of interaction. (3) Molecules are *polarizable*. They can be induced to have charge distributions by the electric field from nearby atoms or molecules or other sources. The interactions between polarizable molecules are shorter-ranged than the interactions between molecules having fixed dipoles. Here are the details.

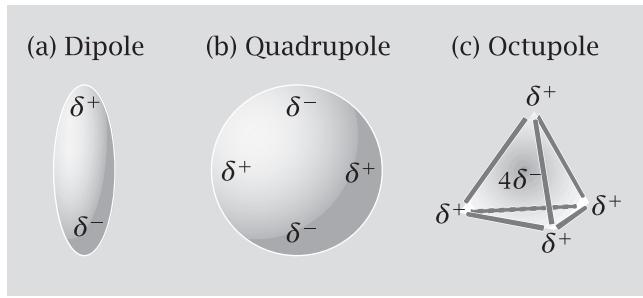


Figure 24.3 The moments of charge distributions are monopoles (not shown), (a) dipoles, (b) quadrupoles, (c) octupoles, etc.; δ^- and δ^+ indicate partial negative and positive charges.

A Charge Distribution Is Characterized by Its Multipole Expansion

Any spatial distribution of charge $\rho(x, y, z) = \rho(\mathbf{r})$ can be described by a *multipole expansion*, a series in which the first term is called the *monopole*, the second is the *dipole*, the third is the *quadrupole*, then the *octupole*, etc. The various terms are *moments* of the distribution in the same way that the mean and standard deviation are related to the first two moments of a probability distribution (see Chapter 1).

The zeroth moment, the monopole, is the total charge of the distribution, $\int \rho(\mathbf{r}) dV$. The first moment $\int \mathbf{r}\rho(\mathbf{r}) dV$, the dipole, describes a symmetric arrangement of equal amounts of positive and negative charge separated by a vector \mathbf{r} . The dipole is the lowest-order distribution for a molecule that has no net charge (see Figure 24.3). A quadrupole is the next lowest-order distribution for a molecule that has no net charge and no net dipole moment. An octupole has no net charge, or dipole, or quadrupole moment. There is no tripole or pentapole, etc., because these distributions can always be expressed more simply as linear combinations of some amount of net charge plus some amount of dipole moment, plus some amount of quadrupole moment, etc. In this way, any charge distribution can be described by a multipole expansion, a sum of moments. For example, a molecule having a dipole moment in addition to a net charge can be described by a sum of the first two terms in a multipole expansion.

Let's consider dipoles. There are two types: permanent and induced. A *permanent dipole* occurs when a charge separation is always present in a molecule, even in the absence of any external electric fields. An *induced dipole* is a charge separation that arises only in the presence of an applied electric field. Some molecules have both permanent and induced dipole moments.

A permanent dipole is characterized by its dipole moment $\mu = q\ell$ (see Equation (21.16)). Dipole moments are measured in units of debyes (D). One debye equals $1 \times 10^{-18} \text{ esu cm} = 3.33564 \times 10^{-30} \text{ coulomb meter (C m)}$. When one positive charge (such as a proton) is separated from one negative charge (such as an electron) by 1 Å, the magnitude of the dipole moment is $q\ell = (1.6022 \times 10^{-19} \text{ C}) \times (1 \times 10^{-10} \text{ m}) \times [1 / (3.33564 \times 10^{-30} \text{ C m})] = 4.8033 \text{ D}$.

Table 24.1 gives some dipole moments. For example, water, which has no net charge, has a permanent dipole moment of 1.85 D, due to the partial negative charge on the oxygen and the partial positive charges on the two hydrogens. Figure 24.4 shows how dipole moments can arise from molecular asymmetry.

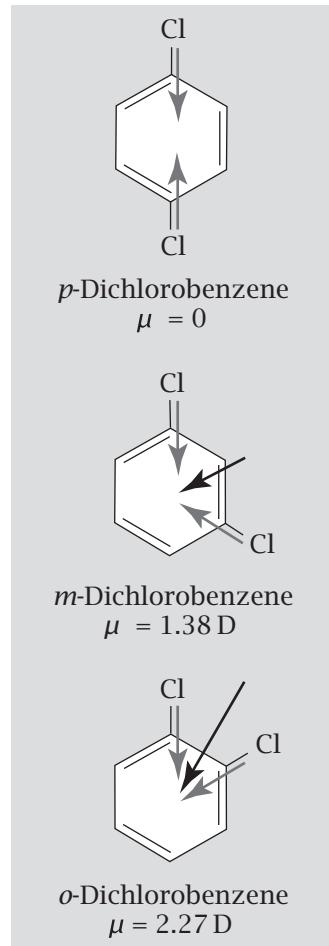


Figure 24.4 Permanent dipole moments can depend on molecular symmetry. The dipole moment (\leftrightarrow) is the vector sum of unit dipoles (\leftarrow). Source: PW Atkins, *Physical Chemistry*, 6th edition, WH Freeman, New York, 1998.

Table 24.1 Dipole moments μ and polarizability volumes, $\alpha' = \alpha/4\pi\epsilon_0 = \alpha\mathcal{C}$, where α is the polarizability.

Molecule	$\mu (10^{-30} \text{ C m})$	$\mu (\text{D})$	$\alpha' (10^{-24} \text{ cm}^3)$
H ₂	0	0	0.819
N ₂	0	0	1.77
CO ₂	0	0	2.63
CO	0.390	0.117	1.98
HF	6.37	1.91	0.51
HCl	3.60	1.08	2.63
HBr	2.67	0.80	3.61
HI	1.40	0.42	5.45
H ₂ O	6.17	1.85	1.48
NH ₃	4.90	1.47	2.22
CCl ₄	0	0	10.5
CHCl ₃	3.37	1.01	8.50
CH ₂ Cl ₂	5.24	1.57	6.80
CH ₃ Cl	6.24	1.87	4.53
CH ₄	0	0	2.60
CH ₃ OH	5.70	1.71	3.23
CH ₃ CH ₂ OH	5.64	1.69	
C ₆ H ₆	0	0	10.4
C ₆ H ₅ CH ₃	1.20	0.36	
<i>o</i> -C ₆ H ₄ (CH ₃) ₂	2.07	0.62	
He	0	0	0.20
Ar	0	0	1.66

Source: PW Atkins, *Physical Chemistry*, 6th edition, WH Freeman, New York, 1998; *Handbook of Chemistry and Physics*, CJF Böttcher and P Bordewijk, *Theory of Electrical Polarization*, Elsevier, Amsterdam, 1978.

Table 24.2 Some covalent bond energies.

Bond	Energy (kcal mol ⁻¹)
C–C	80.6
C=C	145.2
C≡C	198.1
C–H	98.3
O=O (in O ₂)	118.1
F–F (in F ₂)	37.0

Source: RS Berry, SA Rice, and J Ross, *Physical Chemistry*, Wiley, New York, 1980.

Neutral Molecules Attract Because of Charge Asymmetries

Interactions become shorter-ranged and weaker as higher multipole moments become involved. When a monopole interacts with a monopole, Coulomb's law says $u(r) \propto r^{-1}$. But when a monopole interacts with a distant dipole, Coulombic interactions lead to $u(r) \propto r^{-2}$ (see Equation (21.31)). Continuing up the multipole series, two permanent dipoles that are far apart interact as $u(r) \propto r^{-3}$. Such interactions can be either attractive or repulsive, depending on the orientations of the dipoles. Table 24.2 gives typical energies of some covalent bonds, and Table 24.3 compares covalent with noncovalent bond strengths.

Orientational Averaging Shortens the Range of Interactions

Compare two situations: (1) a charge interacts with a dipole that has a fixed orientation, and (2) a charge interacts with a dipole that orients freely over all possible angles (see Figure 24.5). The energy of a charge interacting with a rotating dipole is shorter-ranged than the energy of interacting with a fixed dipole. Here is the explanation.

Table 24.3 Various types of energy u (at 5 Å) and the dependence on distance r .

Type of Interaction	u (kcal mol ⁻¹)	Distance r Dependence
Ionic	66	$1/r$
Ion/dipole	4	$1/r^2$
Dipole/dipole	0.5	$1/r^3$
Dipole/induced dipole	0.012	$1/r^6$

Source: RS Berry, SA Rice, and J Ross, *Physical Chemistry*, Wiley, New York, 1980.

Equation (21.31) shows that when a charge Q is separated by a distance r from a dipole having a dipole moment of magnitude $\mu = q\ell$, constrained to a fixed angle θ , the pair energy is

$$u(r, \theta) = u_0 \cos \theta, \quad \text{where} \quad u_0 = \frac{eQ\mu}{Dr^2}. \quad (24.3)$$

However, if a dipole is free to rotate over all angles θ , the *average* interaction energy is

$$\langle u(r) \rangle = \frac{\int_0^\pi u(r, \theta) e^{-u(r, \theta)/kT} \sin \theta d\theta}{\int_0^\pi e^{-u(r, \theta)/kT} \sin \theta d\theta}. \quad (24.4)$$

The factor of $\sin \theta$ accounts for the different numbers of dipoles that point in the different directions θ (see Example 1.26 and Figure 24.5). Relatively few dipole orientations point in the directions $\theta = 0^\circ$ and $\theta = 180^\circ$ where the area elements are small. More dipole orientations point toward $\theta = 90^\circ$.

If the energy is small, $u/kT \ll 1$, you can use a Taylor series expansion for the exponential terms (see Equation (J.1) in Appendix J), $e^{-u/kT} \approx 1 - u/kT + \dots$, and Equation (24.4) becomes

$$\langle u(r) \rangle \approx \frac{\int_0^\pi u_0 \cos \theta \left[1 - \left(\frac{u_0 \cos \theta}{kT} \right) \right] \sin \theta d\theta}{\int_0^\pi \left[1 - \left(\frac{u_0 \cos \theta}{kT} \right) \right] \sin \theta d\theta}. \quad (24.5)$$

Equation (24.5) has two terms in the numerator and two terms in the denominator. The first term in the denominator is

$$\int_0^\pi \sin \theta d\theta = -\cos \theta \Big|_0^\pi = 2.$$

To evaluate the other three terms, let $x = \cos \theta$. Then $dx = -\sin \theta d\theta$. The first term in the numerator of Equation (24.5) and the second term in the denominator are zero because

$$\int_0^\pi \cos \theta \sin \theta d\theta = - \int_1^{-1} x dx = 0. \quad (24.6)$$

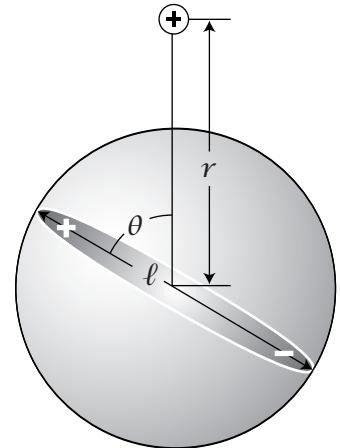


Figure 24.5 A charge is at a distance r from a dipole of length ℓ that is oriented at an angle θ from the axis between them.

The second term in the numerator of Equation (24.5) becomes

$$\begin{aligned}-\frac{u_0^2}{kT} \int_0^\pi \cos^2 \theta \sin \theta d\theta &= \frac{u_0^2}{kT} \int_0^\pi \cos^2 \theta d\cos \theta = \frac{u_0^2}{kT} \int_1^{-1} x^2 dx \\ &= \left(\frac{u_0^2}{kT} \right) \frac{x^3}{3} \Big|_1^{-1} = -\frac{2}{3} \frac{u_0^2}{kT}.\end{aligned}$$

Combining all the terms in Equation (24.5) gives

$$\begin{aligned}\langle u(r) \rangle &= -\frac{u_0^2}{3kT} \\ &= -\frac{1}{3kT} \left(\frac{CQ\mu}{D} \right)^2 \frac{1}{r^4}.\end{aligned}\tag{24.7}$$

Compare Equations (24.7) and (24.3). If the dipole angle is fixed, $u(r) \propto (-1/r^2)$. If the dipole rotates, the orientational average gives $\langle u(r) \rangle \propto (-1/r^4)$. The range of interaction is shorter when the dipole tumbles freely. This argument readily generalizes: because the energy of interaction of two permanent dipoles with fixed orientations is $u(r) \propto -1/r^3$, the interaction between two tumbling permanent dipoles is $u(r) \propto -(1/r^6)$. The interaction energy between two permanent dipoles that are free to rotate, that have moments μ_A and μ_B , and that have a center-to-center separation of r is

$$\langle u(r) \rangle = -\frac{2}{3kT} \left(\frac{\mu_A \mu_B}{4\pi \epsilon_0 D} \right)^2 \frac{1}{r^6},\tag{24.8}$$

when averaged over all the possible angles of each dipole. Example 24.1 computes the magnitude of a dipole-dipole interaction.

EXAMPLE 24.1 The dipole-dipole interaction of ethanol molecules. If two ethanol molecules are $r = 10 \text{ \AA}$ apart in the gas phase, what is their average interaction energy $\langle u(r) \rangle$? Table 24.1 gives the dipole moment of ethanol as $\mu = 5.70 \times 10^{-30} \text{ C m}$ per molecule. Equation (24.8) gives

$$\begin{aligned}\langle u(r) \rangle &= -\frac{2}{3kT} \left(\frac{\mu^2}{4\pi \epsilon_0} \right)^2 \frac{1}{r^6} \\ &= -\frac{2}{3(1.38 \times 10^{-23} \text{ J K}^{-1} \text{ per molecule})(300 \text{ K})} \\ &\quad \times \left(\frac{(5.70 \times 10^{-30} \text{ C m per molecule})^2}{4\pi \times 8.85 \times 10^{-12} \text{ C}^2 \text{ J}^{-1} \text{ m}^{-1}} \right)^2 \left(\frac{1}{10^{-9} \text{ m}} \right)^6 \\ &\quad \times 6.02 \times 10^{23} \text{ molecules mol}^{-1} \\ &= -8.27 \text{ J mol}^{-1}.\end{aligned}$$

At $r = 10 \text{ \AA}$, this attraction is very weak owing to the r^{-6} distance dependence. At $r = 5 \text{ \AA}$, the interaction is $2^6 = 64$ times stronger, -529 J mol^{-1} .

London Dispersion Forces Are Due to the Polarizabilities of Atoms

Attractive interactions are universal. Molecules need not have net charge, or internal charge asymmetry, or even an ability to orient, to experience attractions. Even spherical uncharged inert gas atoms condense into liquids at very low temperatures. Such attractions were first described in 1937 by F London (1900–1957), an American physicist. They are called *London forces*, or *dispersion forces*. Two molecules can induce an attraction in each other because they are polarizable.

A *polarizable* atom or medium is one that responds to an applied electric field by redistributing its internal charge. In the simplest case, when an electrostatic field is applied to a polarizable atom, charge inside the atom redistributes to form a dipole pointing in the direction opposite to the applied field (see Figure 24.6). The dipole moment μ_{ind} that is induced by the field E is often found to be proportional to the applied field, if the field is sufficiently small:

$$\mu_{\text{ind}} = \alpha E, \quad (24.9)$$

where α is called the *polarizability* of the atom or medium. (If the field is large, the dipole moment will not be linear in E but may depend on higher powers, $\mu = \alpha E + \alpha_2 E^2 + \alpha_3 E^3, \dots$) The polarizability is the induced dipole moment per unit of applied electric field. Table 24.1 lists some polarizabilities. Polarizabilities have units of volume, and they are typically about the size of a molecular volume, a few cubic ångströms.

A Charge Will Polarize a Neutral Atom and Attract It

A neutral atom will be attracted to an electrostatic charge. How? The neutral atom becomes polarized by proximity to the electrical charge, causing the neutral atom to act like a dipole, which orients and is then attracted to the charge. Here's a simple model. Suppose a charge Q is at a distance r from the center of a neutral atom, as shown in Figure 24.7. The electric field from Q causes a redistribution of the charge inside the neutral atom, into amounts q and $-q$ that are separated by a distance Δr , giving an induced dipole moment $\mu_{\text{ind}} = q \Delta r$. Induction of the dipole results in an attractive force because Q is closer to the partial charge of opposite sign ($-q$ in this case), which it attracts, than to the partial charge of the same sign ($+q$), which it repels. The net force on the neutral atom is $f = (\text{charge}) \times E$ (see Equation (20.13)). It is the sum of the force

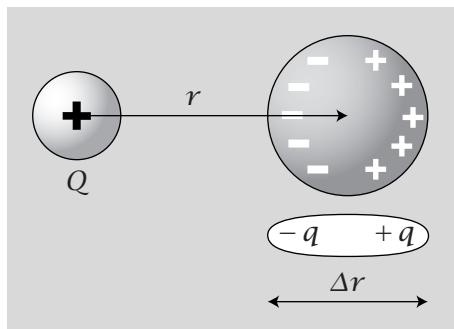


Figure 24.7 When a neutral atom (on the right) is in the electrostatic field of a charge Q , a charge separation is induced in the atom, modeled as a dipole. This attracts the neutral atom to the charge.

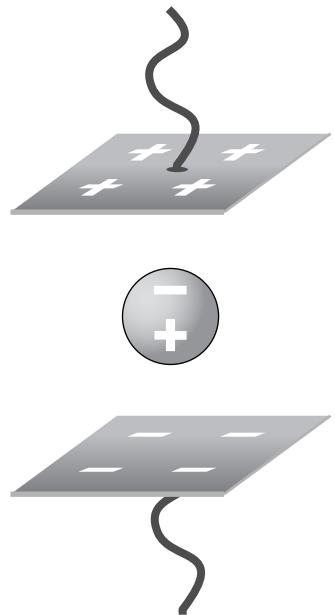


Figure 24.6 An applied field can induce a dipolar charge distribution in a neutral atom.

$(-q)E[r - (\Delta r/2)]$ of Q on the nearby charge $-q$, plus the force $qE[r + (\Delta r/2)]$ of Q on the more distant charge q :

$$\begin{aligned} f &= -q \left[E \left(r - \frac{\Delta r}{2} \right) - E \left(r + \frac{\Delta r}{2} \right) \right] \approx q \Delta r \frac{dE}{dr} = \mu_{\text{ind}} \frac{dE}{dr} \\ &= \alpha E \frac{dE}{dr}. \end{aligned} \quad (24.10)$$

At a distance r from a charge Q , the field is $E = CQ/Dr^2$ (see Equation (20.14)). Substituting $E = CQ/Dr^2$ into Equation (24.10) and taking the derivative gives

$$f = -2\alpha \left(\frac{CQ}{D} \right)^2 \frac{1}{r^5}. \quad (24.11)$$

Integrating to express Equation (24.11) instead as an intermolecular potential yields

$$u(r) = - \int f(r) dr = -\frac{\alpha}{2} \left(\frac{CQ}{D} \right)^2 \frac{1}{r^4}. \quad (24.12)$$

The minus signs in Equations (24.11) and (24.12) show that the interaction of a charge with a neutral atom is attractive, irrespective of the sign of Q . Equations (24.11) and (24.12) show that the interaction is short-ranged ($p > 3$) and that the attraction increases with the polarizability of the neutral atom.

Similarly, when two neutral atoms interact, they induce dipoles in each other, with a pair interaction $u(r) \propto (-1/r^6)$ that is also proportional to the polarizabilities α of the two dipoles.

All the short-ranged attractions we have described, involving multipoles, induced polarization, and orientational averaging, lead to power laws with $p \geq 3$, and are collectively called *van der Waals* forces.

The fact that dipoles interact with a distance dependence r^{-6} gives a physical basis for the form of the attractive interaction in a widely used energy function called the **Lennard-Jones potential**,

$$u(r) = \frac{a}{r^{12}} - \frac{b}{r^6},$$

where a and b are parameters that depend on the types of interacting atoms [1, 2]. The positive sign in the first term implies a repulsion, and the minus sign in the second term implies an attraction. The repulsive part of this potential, a/r^{12} , was originally chosen because it can be calculated rapidly by computers using the square of r^{-6} . The virtue of this model is that it captures the universal features of a short-ranged attraction and even shorter-ranged repulsion, and the two parameters a and b give enough flexibility for the model to predict experimental data fairly accurately.

Hydrogen Bonds

Hydrogen bonds are weak interactions (typically a few kcal mol⁻¹) that occur when a hydrogen is situated between two other atoms. For example, a hydrogen bond can form between an amide and carbonyl group, N–H · · · O=C. In this case, the N–H group is called the *hydrogen bond donor* and the C=O group is called the *hydrogen bond acceptor*. To a first approximation, a hydrogen bond

can be described as an electrostatic interaction between the N–H dipole and the O=C dipole. Hydrogen bonds act to align the N–H and O=C bonds, and to stretch the N–H bonds. To describe hydrogen bonds more accurately requires a quantum mechanical treatment of the charge distribution.

Empirical Energy Functions

A popular approach to modeling the interactions and conformations of large molecules, particularly in solution, is to assume an energy U that is a sum of terms: (1) Coulombic interactions between charged atoms, (2) spring forces that stretch and bend bonds, (3) periodic potentials for torsional rotations around bonds, and (4) a Lennard-Jones potential for nonbonded interactions [3–5]. For example,

$$U = \sum_{\text{bond lengths } b} \frac{K_b}{2} (b - b_0)^2 + \sum_{\text{bond angles } \theta} \frac{K_\theta}{2} (\theta - \theta_0)^2 \\ + \sum_{\text{dihedral angles } \phi} K_\phi [1 + \cos(n\phi - \delta)] \\ + \sum_{\text{nonbonded pairs } i < j} \left(\frac{a_{ij}}{r_{ij}^{12}} - \frac{b_{ij}}{r_{ij}^6} + \frac{c q_i q_j}{D r_{ij}} \right), \quad (24.13)$$

where K_b is the spring constant for stretching bonds, b_0 is the equilibrium bond length, K_θ is the spring constant for bending bonds, and θ_0 is the equilibrium bond angle. K_ϕ and δ are torsional constants, and a_{ij} and b_{ij} are Lennard-Jones constants for atoms i and j separated by a distance r_{ij} . The summation over the index $i < j$ is a convenient way to count every interaction only once. Summing the pair interaction energies u_{ij} over all indices, $\sum_{\text{all } i} \sum_{\text{all } j} u_{ij}$, would count every interaction exactly twice. For example, the interaction of particle 2 with particle 3, u_{23} , is equal to u_{32} . Thus,

$$\sum_{i < j} u_{ij} = \frac{1}{2} \sum_{\text{all } i} \sum_{\text{all } j} u_{ij}. \quad (24.14)$$

Much of what is known about intermolecular interactions comes from measuring the pressures of nonideal gases. The simplest model that relates intermolecular interactions to the pressures of gases is the van der Waals model.

The van der Waals Gas Model Accounts for Intermolecular Interactions

Real gases are more complex than ideal gases. The ideal gas model predicts that the pressure depends only on the temperature and the gas density ($\rho = N/V = p/kT$). The pressure of an ideal gas does not depend on the types or atomic structures of the gas molecules. But ideal behavior applies only at low densities where molecules don't interact much with each other. For denser gases, intermolecular interactions affect the pressures, and gases differ from each other

depending on their atomic structures. The ideal gas law does not predict that gases condense to liquids. The van der Waals gas law does.

The **van der Waals equation of state** for the pressure $p(T, V, N)$ is

$$p = \frac{NkT}{V-Nb} - \frac{aN^2}{V^2} = \frac{\rho RT}{1-b\rho} - a\rho^2, \quad (24.15)$$

where a and b are parameters (not to be confused with the Lennard-Jones parameters). Equation (24.15) was developed in 1873 in the PhD thesis of the Dutch physicist JD van der Waals (1837–1923), who won the Nobel prize in 1910 for this work.

The van der Waals gas law can be derived in various ways from an underlying model of intermolecular interactions. The next two sections give a simple derivation.

We will find the pressure as a sum of energy and entropy components (combining Equations (8.12) and (8.22)):

$$p = -\left(\frac{\partial F}{\partial V}\right)_{T,N} = -\left(\frac{\partial U}{\partial V}\right)_{T,N} + T\left(\frac{\partial S}{\partial V}\right)_{T,N}. \quad (24.16)$$

The lattice model in Example 6.1 gives the second term, the entropic component, as $T(\partial S/\partial V)_{T,N} = NkT/(V-Nb)$. The next section shows that intermolecular attractions lead to $-(\partial U/\partial V)_{T,N} = -aN^2/V^2$, the energetic component.

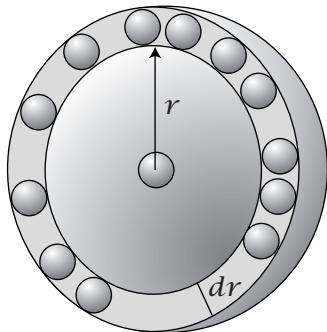


Figure 24.8 To compute the internal energy of a gas, identify one particle in a gas as a test particle. A spherical shell at a distance r from that particle has volume $4\pi r^2 dr$. The number of particles in the shell is $\rho 4\pi r^2 dr$.

The Energetic Component of the van der Waals Gas Pressure

To describe the energetic component of the pressure, you need a model for the energy as a function of volume, $U(V)$. The total interaction energy U of the gas is the sum of all the interparticle interactions. You can treat only the *interactions* between the particles and ignore the contributions from the internal partition functions of the particles because they don't change with V , and won't contribute to the pressure $(\partial U/\partial V)_{T,N}$. (The translational partition function does depend on V , but this contribution is treated in the entropy component.)

Choose one particle, and call it the *test particle*. Divide the space around the test particle into spherical shells. Figure 24.8 shows a shell at radius r from a test particle. What is the energy U' of interaction between the test particle and all the other particles in the system? To compute U' , multiply the number of particles in the shell at radius r by the interaction energy $u(r)$ between the test particle and all the particles in that shell, then integrate over all shells.

How many particles are in the shell at radius r ? A gas of density $\rho = N/V$ has N particles in a volume V . If there is a uniform spatial distribution, the number of particles in the shell at radius r will be (density) \times (volume of the shell) $= \rho \times 4\pi r^2 dr$. So the interaction energy of the test particle with all the other molecules is $U' = \int u(r)\rho 4\pi r^2 dr$.

To get the total energy U , sum the energy U' over all N particles (each taken once as the test particle), and divide by 2 to correct the double-counting. (If you multiplied U' by N , you would count every interaction exactly twice—once when a particle was the test particle at the center, and once when it was in a

shell at radius r from the other test particle.) The total energy is

$$U = \frac{N}{2} U' = \frac{N}{2} \int_0^\infty u(r) \rho 4\pi r^2 dr. \quad (24.17)$$

Equation (24.17) holds when particles are distributed uniformly and when the energies are pairwise-additive, i.e., when the pair interaction is not influenced by the presence of a third particle.

To compute U from Equation (24.17), you need to know the pair interaction energy function $u(r)$. Because repulsions between atoms and molecules are very steep short-ranged functions for small r , it is simplest to assume ‘hard-core’ repulsions approximately like those between billiard balls. If you assume that the attractions fall off as r^{-6} (see Figure 24.9), the pair potential is

$$u(r) = \begin{cases} \infty & \text{if } r < r^*, \\ -u_0(r^*/r)^6 & \text{otherwise.} \end{cases} \quad (24.18)$$

where r^* is the minimum distance between the pair of particles. Substituting the pair potential, Equation (24.18), into Equation (24.17) gives

$$\begin{aligned} U &= \frac{N}{2} \int_{r^*}^\infty \left[-u_0 \left(\frac{r^*}{r} \right)^6 \right] \left(\frac{N}{V} \right) 4\pi r^2 dr \\ &= \frac{-u_0 N^2 2\pi (r^*)^6}{V} \int_{r^*}^\infty \frac{dr}{r^4}. \end{aligned} \quad (24.19)$$

In principle, the upper limit of the integration in Equation (24.19) should be some dimension of the gas container, but because the integral converges to its final value for as few as 10–20 shells, you can define the upper limit as $r = \infty$. No particle is located at a distance between $r = 0$ and $r = r^*$ from the center of another particle, so you have $\int_0^\infty = \int_{r^*}^\infty$.

Because the integral $\int_{r^*}^\infty r^{-4} dr = -[1/(3r^3)]_{r^*}^\infty$ equals $1/(3r^{*3})$, Equation (24.19) reduces to

$$U = -\frac{aN^2}{V}, \quad \text{where } a = \frac{2\pi(r^*)^3}{3} u_0. \quad (24.20)$$

The minus sign in Equation (24.20) indicates that attractive interactions reduce the energy relative to that of an ideal gas.

Now you can complete the free energy by considering the entropic component.

The Entropic Component of the Pressure

The entropy of a van der Waals gas can be derived in different ways. Let’s use a simple lattice model. Recall from the lattice model of gases (Equation (6.10)) that the entropy S of distributing N particles onto a lattice of M sites is

$$\frac{S}{k} = -N \ln \left(\frac{N}{M} \right) - (M-N) \ln \left(\frac{M-N}{M} \right).$$

Let V equal the total volume of the system and let $b_0 = V/M$ be the volume per lattice site. Substituting $M = V/b_0$ into Equation (6.10) and using $F = U - TS$

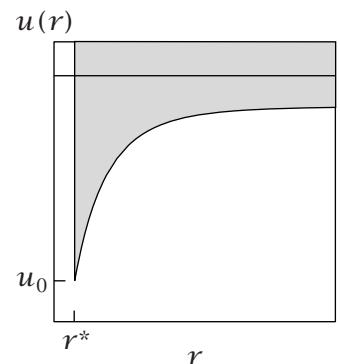


Figure 24.9 A model for the attractive energy between two particles, used to derive the van der Waals gas law.

with Equation (24.20) for U gives

$$F = -\frac{aN^2}{V} + kT \left[N \ln\left(\frac{Nb_0}{V}\right) + \left(\frac{V}{b_0} - N\right) \ln\left(1 - \frac{Nb_0}{V}\right) \right]. \quad (24.21)$$

To get the pressure p from the free energy F , use Equation (8.12):

$$\begin{aligned} p &= -\left(\frac{\partial F}{\partial V}\right)_{T,N} \\ &= -\frac{aN^2}{V^2} - \frac{kT}{b_0} \ln\left(1 - \frac{Nb_0}{V}\right). \end{aligned} \quad (24.22)$$

For low densities, $\ln(1-x) \approx -x - x^2/2 - \dots$ (Equation (J.4) in Appendix J), so for $x = Nb_0/V$ you have

$$-\frac{kT}{b_0} \ln\left(1 - \frac{Nb_0}{V}\right) \approx \frac{NkT}{V} \left[1 + \frac{1}{2} \left(\frac{Nb_0}{V} \right) + \dots \right]. \quad (24.23)$$

The first term on the right-hand side of Equation (24.23) gives the ideal gas law. For densities that are higher than those of ideal gases, include the next higher-order term in Equation (24.23). Using the approximation $(1+x/2) \approx 1/(1-x/2)$ (Equation (J.6) in Appendix J) gives the **van der Waals equation**,

$$\begin{aligned} p &= \frac{NkT}{V(1-Nb/V)} - \frac{aN^2}{V^2} \\ &= \frac{NkT}{V-Nb} - \frac{aN^2}{V^2}. \end{aligned}$$

The van der Waals constant $b = b_0/2$ represents half the volume of each particle. In practice, a and b in the van der Waals equation are taken as adjustable parameters for fitting pressure/volume/temperature data.

If you had substituted any attractive short-ranged potential (one that falls off more rapidly than r^{-3}) into Equation (24.17), it would have led to the pressure $p \propto V^{-2}$, just as in Equation (24.22). For this reason, many different types of interaction are collectively called van der Waals forces. They all lead to the same contribution to $p(V)$. For longer-ranged forces, such as Coulombic interactions, the integral in Equation (24.17) becomes infinite, and Equation (24.20) would not apply. For long-range interactions, you need to resort to the methods of Chapters 20–23, such as Poisson's equation.

The van der Waals equation is just one of many different models for the pressures of real gases. It is perhaps the simplest model that can describe the phase transition between gas and liquid states (boiling) (see Chapter 25). Another model that accounts for phase transitions is the **Redlich-Kwong equation** for the pressure [6, 7],

$$p = \frac{kT}{V-b_1} - \frac{a_1}{T^{1/2}V(V+b_1)}$$

in terms of parameters a_1 and b_1 .

One assumption that we made for computing the energy of the van der Waals model is that the gas is distributed uniformly in space. In reality, if particles have attractions, they tend to cluster. The next section shows how you can correct for this.

Radial Distribution Functions Describe Structural Correlations in Liquids

You can express the true density in the shell at radius r from a test particle as $\rho g(r)$, where $\rho = N/V$ is the average density and $g(r)$ is called the *radial distribution function* or sometimes the *pair correlation function*. $g(r)$ is the ratio of the actual density of particles at r to the mean density ρ . The pair correlation function $g(r)$ equals 1 when the local density in a shell is the same as the mean density averaged over the whole volume. Where $g(r) > 1$, this means the density of particles in that shell is higher than mean density ρ . $g(r)$ can be measured in experiments.

Figure 24.10 shows three examples of correlation functions. Figure 24.10(a) shows the correlation function that we assumed for the van der Waals gas: particles cannot be closer together than a distance r^* , and they are randomly distributed otherwise. Figure 24.10(b) shows a correlation function for a typical liquid: there is a depletion $g(r) = 0$ for small r because particles cannot overlap, then a peak indicates there is an excess of particles that make up a first-neighbor shell, then a slight depletion, followed by a second-neighbor shell, etc. Beyond the third or fourth shell, there is little correlation and $g(r) = 1$. Figure 24.10(c) illustrates the *long-ranged order* in crystalline solids.

Now we show how the number of neighbors and pairwise energies of interaction are computed from pair correlation functions.

Using Pair Correlation Functions

A pair correlation function $g(r)$ can be found from experiments using the scattering of electromagnetic radiation. From this function, you can get other quantities, such as the average number of nearest neighbors of any molecule. Because the volume of a spherical shell at radius r is $4\pi r^2 dr$, the number of molecules in the first-neighbor shell is density \times (volume of the shell) = $\rho g(r) 4\pi r^2 dr$. The total number of molecules in the first shell of a molecule

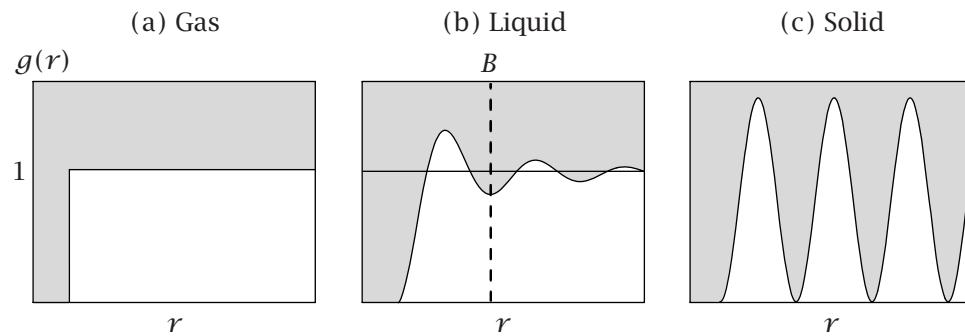


Figure 24.10 Examples of pair correlation functions: (a) the model we have used for the van der Waals gas—hard-core repulsions, and uniform distributions otherwise; (b) typical liquids (B is the limit used to define the first shell of neighbors); (c) solids, which have long-range order.

is the integral

$$\text{number of nearest neighbors} = \int_0^B \rho g(r) 4\pi r^2 dr,$$

where B is the location shown in Figure 24.10(b). If you integrate over all the shells, you get n , the total number of particles (minus the test particle):

$$\int_0^\infty \rho g(r) 4\pi r^2 dr = n - 1 \approx n.$$

In a system where particles are not distributed uniformly, such as a liquid, the energy U' of interaction between one particle and all the other particles requires that you include a factor of $g(r)$ in the integral:

$$U' = \int_0^\infty \rho u(r) g(r) 4\pi r^2 dr. \quad (24.24)$$

Potentials of Mean Force Describe Interactions in Fluids

Imagine putting two solute molecules in water. Brownian motion separates them by different distances r at different times. Suppose you plot how often the solute molecules are found at separation r , for all possible values $r = 0$ to $r \rightarrow \infty$. This is the radial distribution function $g(r)$ (see page 483). $w(r)$, the *potential of mean force* (pmf), is defined as the corresponding free energy:

$$w(r) = -kT \ln g(r).$$

The pmf $w(r)$ applies to any system having a pair distribution function $g(r)$, including pure liquids and solids. Figure 24.11 shows $g(r)$ and $w(r)$ for two spherical solutes in a solvent.

If the same two molecules were brought together in the gas phase, $w(r)$ would simply be the pair potential $u(r)$ (see Figure 24.1), which has only a single minimum. But the pmf between two particles in liquids—even simple liquids—oscillates, with maxima and minima. For a given separation r between the two molecules of interest, the pmf is a potential of *mean* force because it is averaged over all the conformations of the surrounding solvent molecules. It reflects the apparent attractions and repulsions between solute molecules that arise from the surrounding solvent.

Figure 24.11 illustrates the main features of the pmf. First, the test molecules may come into contact if they are attracted to each other. This separation can be the position of the deepest well in the pmf. The second well in the pmf is called the *solvent-separated minimum*. This is where the two test particles are separated by a single layer of solvent. Each test particle protects the near side of the other particle from collisions with the solvent that would otherwise drive the two solute molecules apart. So it is more probable that solvent collisions will drive two nearby solute particles together than that they will drive them apart. This means that there is an apparent solvent-driven attraction of the test particles even when they are *solvent-separated* by one solvation shell. The system is unstable when two particles are too close together for a solvent molecule to fit between them.

In Chapters 14–16, we developed lattice models of liquids and solids composed of uncharged particles. We assumed that each molecule interacts only

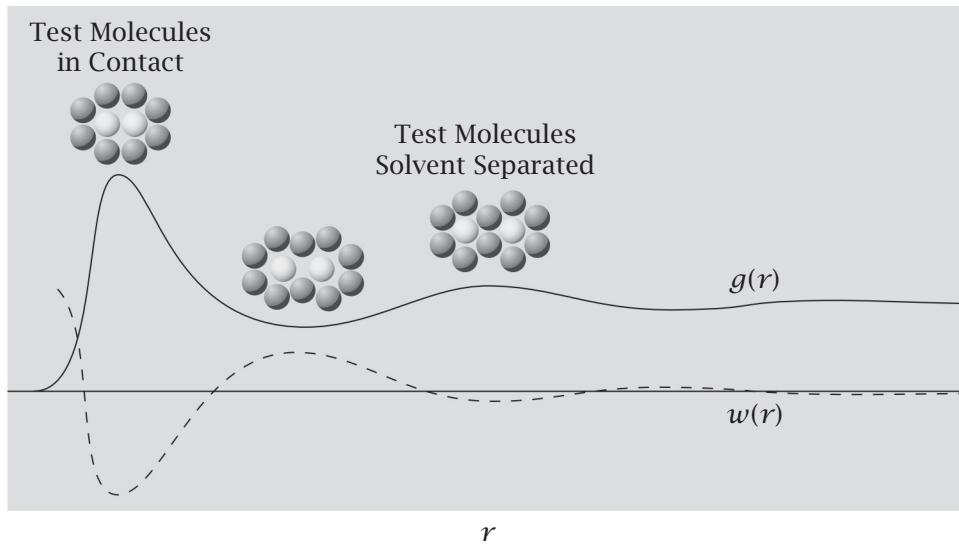


Figure 24.11 Two white spheres represent solute test particles in a solvent of darker-colored spheres. The x axis shows the separation r between the test particles. $g(r)$ (—) is a typical distribution of the different separations of the test particles. In this case, the particles prefer to be in contact. Slightly greater separation is unfavorable. A second small peak indicates a small preference of the solutes to be separated by a single layer of solvent. The potential of mean force $w(r)$ (---) is the corresponding free energy.

with its nearest neighbor through a contact energy w . The next section gives the justification for using such contact energies and for neglecting more distant interactions.

The Lattice Model Contact Energy w Approximates Intermolecular Interactions

The lattice model quantity w (e.g., in Equations (14.6) and (15.5)) is $u(r^*)$, the pair potential evaluated at the equilibrium bond separation distance $r = r^*$. In the lattice model, the sum over all pairs in Equation (24.17) is treated as a sum over only nearest neighbors:

$$U = \frac{N}{2} \sum_{r=0}^{\infty} u(r) \rho g(r) 4\pi r^2 \approx \frac{N}{2} u(r^*) z = \frac{Nwz}{2}. \quad (24.25)$$

The justification for neglecting all other shells in the lattice model of liquids and solids is that nearest-neighbor pair interactions contribute most of the energy for dense phases of particles having short-ranged interactions. For London interactions, the energy of interaction diminishes as r^{-6} , but the number of molecules increases with distance as $4\pi r^2$. The product of these two factors gives $r^{-6} \times r^2 = r^{-4}$. So, even though the number of particles increases with distance from a test particle, the total energy contribution diminishes with distance.

Let's see what fraction of the total energy comes from the second shell and beyond. On the lattice, you can sum over shells numbered by integers $r = 1, 2, 3, \dots$. You can look up these sums in math tables [8]:

$$\frac{\sum_{r=2}^{\infty} (1/r)^4}{\sum_{r=1}^{\infty} (1/r)^4} = \frac{(\pi^4/90) - 1}{(\pi^4/90)} = 0.076. \quad (24.26)$$

The total energy due to all neighbors beyond the first-neighbor shell is only about 7.6% of the total. This is the size of the error that we make in neglecting all except nearest-neighbor interactions.

In Chapters 15 and 16, we noted that two uncharged atoms or molecules A and B generally prefer to self-associate than to form AB interactions, leading normally to a positive value of the exchange energy, $\chi_{AB} > 0$. Example 24.2 shows how this arises from dipole interactions.

EXAMPLE 24.2 Like dissolves like: a consequence of induced dipole interactions. For uncharged molecules having permanent dipoles, the interaction energy is proportional to the product $\mu_A \mu_B$ (see Equation (24.8)). For molecules having induced dipoles, a similar derivation would show that the interaction energy is proportional to the product of polarizabilities $\alpha_A \alpha_B$. Such products imply the following pair interactions:

$$w_{AA} = -c \alpha_A^2, \quad w_{BB} = -c \alpha_B^2, \quad w_{AB} = -c \alpha_A \alpha_B, \quad (24.27)$$

where c is a positive constant that is approximately the same for atoms and for molecules of similar size. Thus, χ_{AB} , defined in Equation (15.13), can be given in terms of the dipole moments or polarizabilities:

$$\chi_{AB} = \left(\frac{z}{kT} \right) (-c) \left(\alpha_A \alpha_B - \frac{\alpha_A^2 + \alpha_B^2}{2} \right) = \left(\frac{cz}{2kT} \right) (\alpha_A - \alpha_B)^2 > 0. \quad (24.28)$$

The quantity χ_{AB} is positive because $(\alpha_A - \alpha_B)^2$ is always positive. This is the basis for the general rule that like dissolves like.

Summary

Intermolecular bonds are the weak interactions that hold liquids and solids together. Atoms and molecules repel each other when they are too close, attract when they are further apart, and are stable at the equilibrium *bond length* in between. Intermolecular attractions make the pressures of van der Waals gases lower than the pressures of ideal gases at low densities. To a first approximation, intermolecular attractions can be explained as electrostatic interactions between charge distributions, due to internal charge asymmetries in the molecules, due to the freedom of molecules to rotate, or due to molecular polarizabilities.

Problems

1. Interpreting Lennard-Jones parameters. Intermolecular interactions are often described by the Lennard-Jones potential $u(r)$, which gives the internal energy of interaction between two molecules as a function of intermolecular separation:

$$u(r) = 4\epsilon \left[\left(\frac{\sigma}{r} \right)^{12} - \left(\frac{\sigma}{r} \right)^6 \right],$$

where ϵ and σ are characteristic energy and bond length parameters.

- (a) At low temperatures, entropy is relatively unimportant, and the free energy is minimized at the intermolecular separation of the molecules at which the potential energy is minimized. At what separation does that occur? What is the energy of that state?
- (b) It's often convenient to divide energies ϵ by Boltzmann's constant k to give them in units of temperature (K). In these units, the constants in Table 24.4 have been found. What physical properties of these systems can you deduce from this information?

Table 24.4

Molecule	ϵ/k (K)	$\sigma(\text{\AA})$
He	10.22	2.58
Ethane	205	4.23
Benzene	440	5.27

Source: PW Atkins, *Physical Chemistry*, 6th edition, WH Freeman, New York, 1998.

2. A model for angular electrostatic bonding. Consider two fixed unit positive charges separated by a distance r_0 , and a unit negative charge that is free to move along a dividing line (see Figure 24.12).

- (a) For fixed r_0 , how does the Coulombic energy depend on r_0 and θ ?
- (b) For what angle θ is the Coulombic energy a minimum?
- (c) What is the energy at its minimum value?
- (d) At what angle θ is the energy equal to zero?

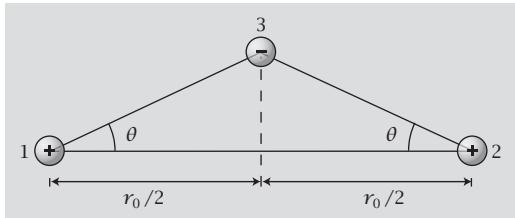


Figure 24.12 Two positive charges (1 and 2) at a fixed separation r_0 , with a negative charge (3) that is free to move vertically between them.

3. How the dielectric constant affects electrostatic binding. An anesthesiologist has isolated two proteins A and B that bind to each other by electrostatic attraction. A has net positive charge and B has net negative charge. In water, which has a dielectric constant of $D = 80$ at room temperature, their binding constant is $K_1 = 1000$. What is their binding constant if the medium is trifluoroethanol with a dielectric constant of $D = 27$? (Assume that the charges on each protein and the distance between them are the same in water and trifluoroethanol.)

4. Direction and distance in a dipolar interaction.

- (a) Which dipole pair in Figure 24.13 has the lower energy: the parallel pair in Figure 24.13(a) or the antiparallel pair in Figure 24.13(b)?

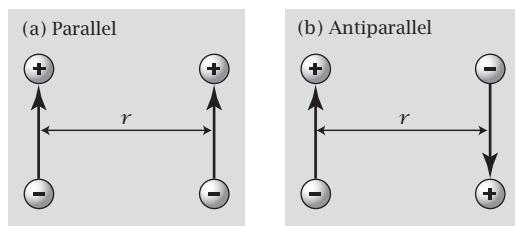


Figure 24.13 Dipoles separated by a distance r : (a) parallel and (b) antiparallel.

- (b) If you double the distance between the dipole centers, by what factor does each pair interaction change? Does it increase or decrease?

5. The r dependence of pair potential functions. You have two different pair potential functions:

$$u_1(r) = -\frac{1}{r} \quad \text{and} \quad u_2(r) = -\frac{1}{r^6}.$$

- (a) Plot both functions.
- (b) At $r = 1$, which pair potential has the stronger attraction?
- (c) At $r = 2$, which pair potential has the stronger attraction?

6. Relating force and equilibria to a pair potential. Consider the interaction potential between two particles shown in Figure 24.14.

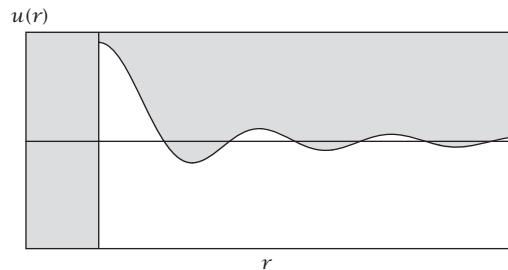


Figure 24.14 Energy $u(r)$ and axes for drawing the force curve $f(r)$.

- (a) Draw the corresponding force curve $f(r)$.
 (b) Identify the points of equilibrium.

7. Predicting electrostatic attraction. The system shown in Figure 24.15 has two fixed positive charges separated by a distance $3r$. The negative charge is closer to the left positive charge. Which way will the negative charge move?



Figure 24.15 Two fixed positive charges and a movable negative charge on the line between them.

References

- [1] U Burkert and NL Allinger, *Molecular Mechanics*, ACS Monograph 177, American Chemical Society, Washington, DC, 1982.
- [2] M Rigby, EB Smith, WA Wakeham, and GC Maitland, *The Forces between Molecules*, Oxford University Press, New York, 1986.
- [3] JA McCammon and SC Harvey, *Dynamics of Proteins and Nucleic Acids*, Cambridge University Press, New York, 1987.
- [4] WD Cornell, AE Howard, and P Kollman, *Curr Opin Struct Biol* **1**, 201 (1991).
- [5] CL Brooks III, M Karplus, and BM Pettitt, *Proteins: A Theoretical Perspective of Dynamics, Structure, and Thermodynamics*, Advances in Chemical Physics, Volume 71, Wiley, New York, 1988.
- [6] M Modell and RC Reid, *Thermodynamics and Its Applications*, 2nd edition, Prentice-Hall, Englewood Cliffs, NJ, 1983.
- [7] O Redlich and JNS Kwong, *Chem Rev* **44**, 233 (1949).
- [8] MR Siegel, S Lipschutz, and J Liu, *Schaum's Outline of Mathematical Handbook of Formulas and Tables*, 3rd edition, McGraw-Hill, New York, 2009.

Suggested Reading

Detailed discussions of intermolecular interactions and polarizability are found in:

RS Berry, SA Rice, and J Ross, *Physical Chemistry*, 2nd edition, Oxford University Press, New York, 2000, Chapter 10.

JN Israelachvili, *Intermolecular and Surface Forces*, 2nd edition, Academic Press, London, 1992.

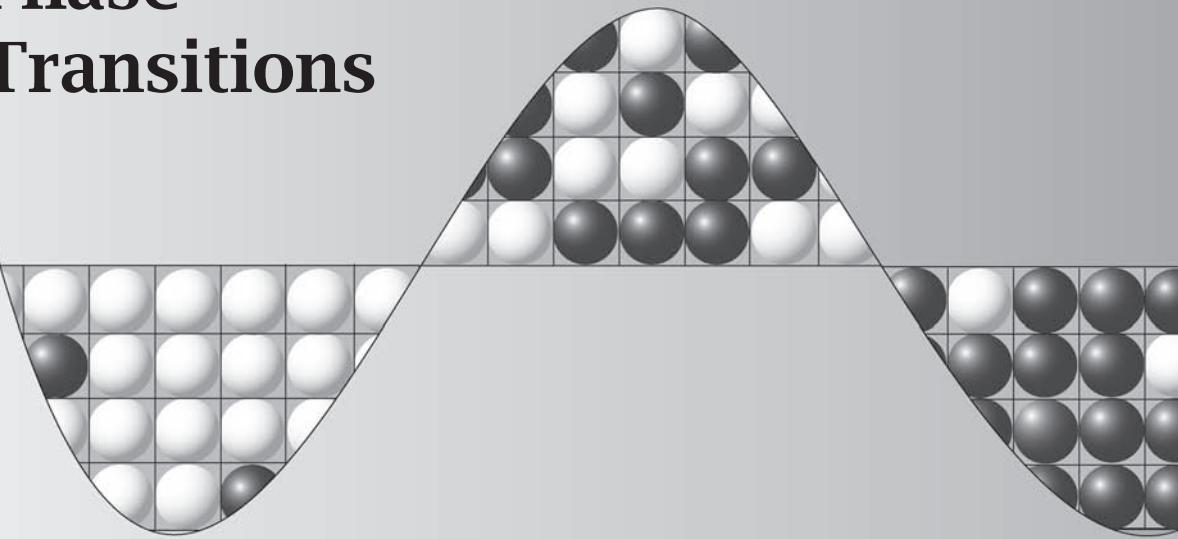
Equations of state are discussed in:

HT Davis, *Statistical Mechanics of Phases, Interfaces, and Thin Films*, Wiley-VCH, New York, 1996.

JC Slater, *Introduction to Chemical Physics*, McGraw-Hill, New York, 1939.

25

Phase Transitions



Two States Can Be Stable at the Same Time

At water's boiling point, two different phases are stable at the same time: liquid water is in equilibrium with steam. A *phase transition* is a dramatic change in a system's physical properties. When you heat water by 2°C at 99°C, it boils. Water transforms from a liquid to a gas, decreasing its density 1600-fold, losing its liquid properties. A small change in temperature drives a huge change in density. In contrast, if you heat water by 2°C at 25°C, water remains a liquid and its density decreases by only 0.2%. The dramatic change in density at 99°C is a phase transition. Other phase transitions among different states of matter include freezing, the magnetization of metals, solubilization and precipitation in liquids, the formation of membranes and micelles, and sharp transitions in the alignments of liquid crystal molecules (used in watches and computer displays). While the term 'phase transition' usually applies to macroscopic systems, 'cooperativity' more broadly also refers to sharp changes of properties of single molecules, for example in protein folding, helix-coil transitions, or ligand binding, described in Chapter 26. In this chapter we focus on phase transitions.

To understand how multiple states can be stable at the same time, first recall the principle of stability of a single state. If a system at fixed temperature has a degree of freedom x , the stable state occurs at $x = x^*$, where the free energy $F(x)$ is a minimum; see Figure 25.1. Using the metaphor of Chapter 2, a ball rolls along x to reach the bottom of the valley, $x = x^*$. For the boiling

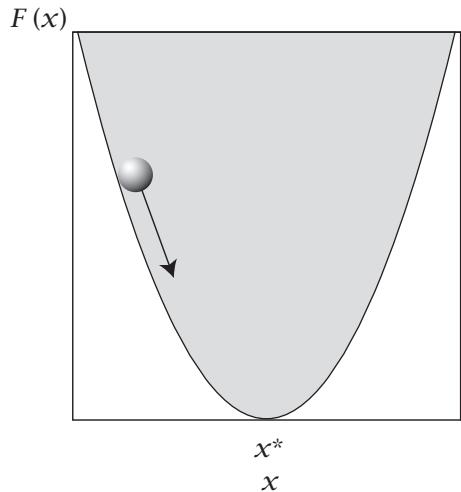


Figure 25.1 Systems at constant temperature are stable in states having minimum free energy $F(x)$.

of water, take the degree of freedom to be the density ρ of water. At $T = 25^\circ\text{C}$, $\rho = \rho^* = 1 \text{ g cm}^{-3}$. The free energy increases in either direction away from this minimum. Water gets no denser than $\rho = \rho^*$ because steric repulsions among the molecules would increase the free energy. Water gets no looser than $\rho = \rho^*$ because the loss of water–water attractions would increase the free energy. This is why liquid water is a stable state. At higher temperatures, say $T = 150^\circ\text{C}$, steam is the single stable state of water. Again, the free energy function has a single minimum. When two or more phases are in equilibrium, it implies that the free energy function of a system has multiple minima. At the boiling temperature, $T = 100^\circ\text{C}$, some water is stable as a liquid and some water is stable as steam, indicating two minima in free energy.

Energy Landscapes Describe Stabilities and Transitions

Figure 25.2 shows different *energy landscapes*. It shows how changing the temperature shifts a system from one stable state to another through the course of a phase transition. There are two types of phase transitions: *first-order* and *higher-order*. Different types of transitions happen for different systems under different conditions. The three panels down the right column of Figure 25.2 show a *first-order* (also called *two-state*) phase transition between one state (steam) and the other state (liquid water). At the midpoint temperature of a first-order transition (i.e., exactly at the boiling temperature), the landscape has two equal minima in free energy, representing two equally stable states. The three panels down the left column of Figure 25.2 show a *higher-order* (also called *continuous*) phase transition. At the midpoint of this type of transition, the landscape has one broad minimum in the free energy instead.

What is the difference between these types of transitions? Suppose you could metaphorically reach in and grab a small cluster of molecules, one after another. At the midpoint of a first-order transition, you would find some clusters having low density, like steam, and other clusters having high density, like liquid water, but you would rarely find clusters having properties in between. In

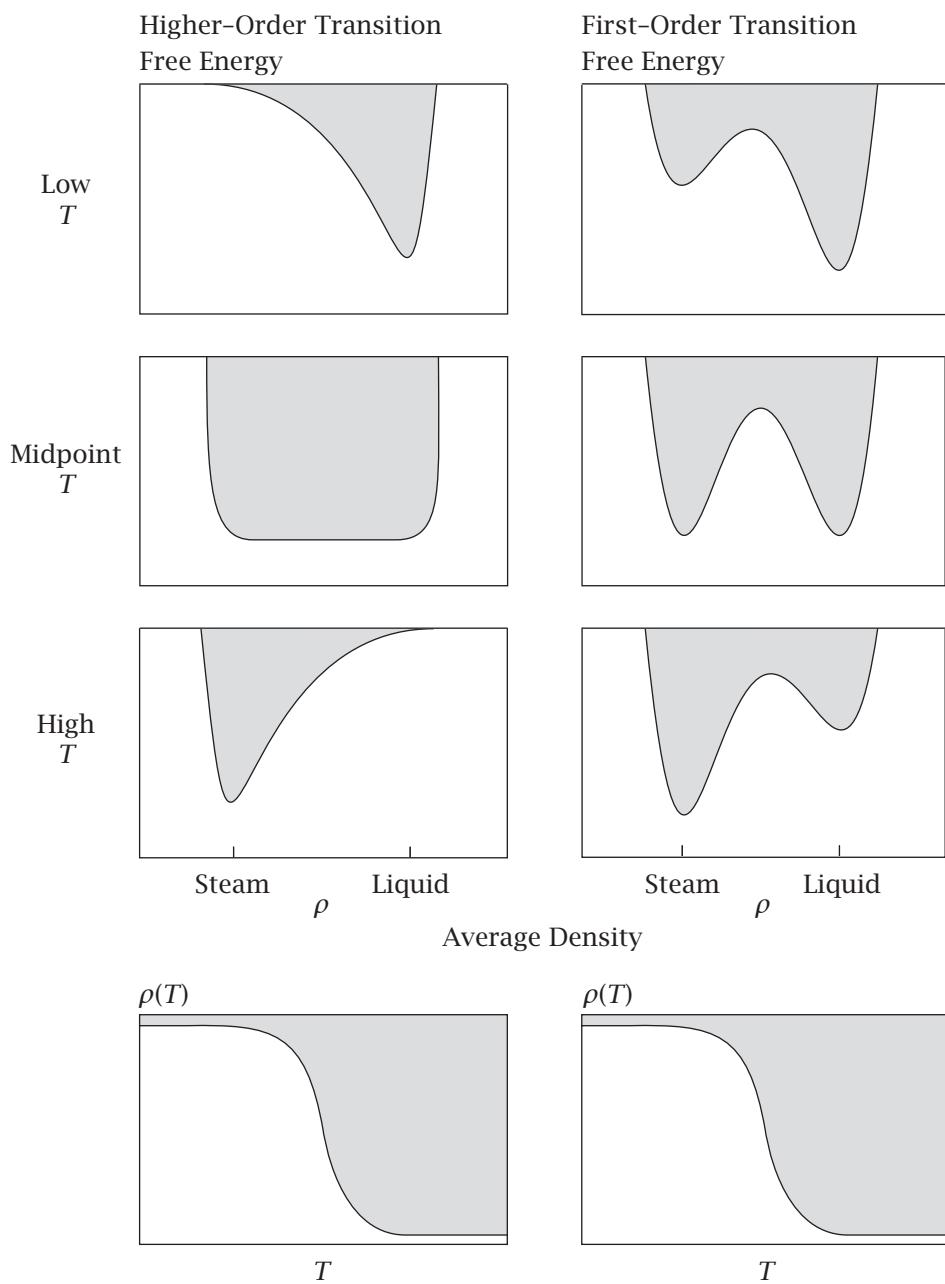


Figure 25.2 Free energy profiles (landscapes) showing how the stable state of water shifts from liquid (at low T) to steam (at high T). Left column: at the midpoint of a higher-order transition, states are stable that have a wide range of densities. Right column: at the midpoint of a first-order transition, only two states are stable. The bottom figure shows that the average density versus temperature can be the same for these two different types of transformations. Quantities such as the density do not give the landscape shape, because they are averaged over the different landscape states.

contrast, at the midpoint of a higher-order transition, some clusters would have intermediate densities between those of steam and liquid water. At $T = 100^\circ\text{C}$ and $p = 1 \text{ atm}$, the boiling of water is a first-order transition: water molecules

will be clustered in either of the two states, steam or liquid. Energy landscapes represent free energy functions of the degrees of freedom of a system and give a graphical way to think about phase transitions.

Simple experimental observables typically do not tell you whether a transition is first-order or higher-order. For boiling, you could measure the average density of water molecules, $\rho(T)$, as a function of temperature. You would find that $\rho(T)$ drops sharply at the boiling temperature of water (Figure 25.2). This tells you that there is a transition, but not what type. The quantity $\rho(T)$ gives you only the *average* density of all the clusters. To determine the type of phase transition, you also need to know the *distribution* of densities of the clusters.

We want to understand phase equilibria at different levels, from the top, the *phase diagram*, which you can measure but which doesn't tell you about the underlying forces and interactions, to the bottom, microscopic models. First, let's consider the basic observable, the phase diagram.

Phase Diagrams Describe What Phases Are Stable Under Which Conditions

In the boiling example above, we considered temperature as our *control variable*: changing the temperature switches the stable state from liquid water to steam. You can also boil water by changing the pressure. Or, you can boil water by different combinations of pressure and temperature. With two control variables, such as pressure and temperature, you can make a two-dimensional plot, called a *phase diagram*. A phase diagram tells you which phases are stable under what conditions. For example, Figures 14.4 and 30.15 show the pressure-temperature diagram of the phases of water: ice, liquid, and steam. Below, we explore a different type of phase diagram, namely for the mixing or demixing of two liquids (see Figures 25.3 and 25.4). In this case, the two control variables are the temperature and composition (concentration of one of the components).

Liquids or Solids May Mix at High Temperatures but Not at Low Temperatures

Sometimes water and oil don't mix. But, sometimes they do. Figure 25.4 gives the phase diagram. Let's follow the type of experiment that creates this diagram. Make six solutions, numbered 1 to 6, having different compositions of oil and water. Beaker 1 has mostly water and a little oil. Beaker 6 has mostly oil and a little water. All of the solutions are at the same temperature T_0 . The vertical axes of Figures 25.3 and 25.4 represent the temperatures of the solutions, and the horizontal axes represent the B concentrations x_B of those solutions. If B in Figure 25.3 is oil and A is water, x_B represents the mole fraction of the oil in the water. The left end of the horizontal axis ($x_B = 0$) represents pure water and the right end ($x_B = 1$) represents pure oil.

1. **A small amount of oil dissolved in water.** Test tube 1 has only a small amount of oil in a large volume of water. The oil fully dissolves in the water, meaning that most of the individual oil molecules are dispersed and surrounded by water molecules. This is called a *single-phase* solution.

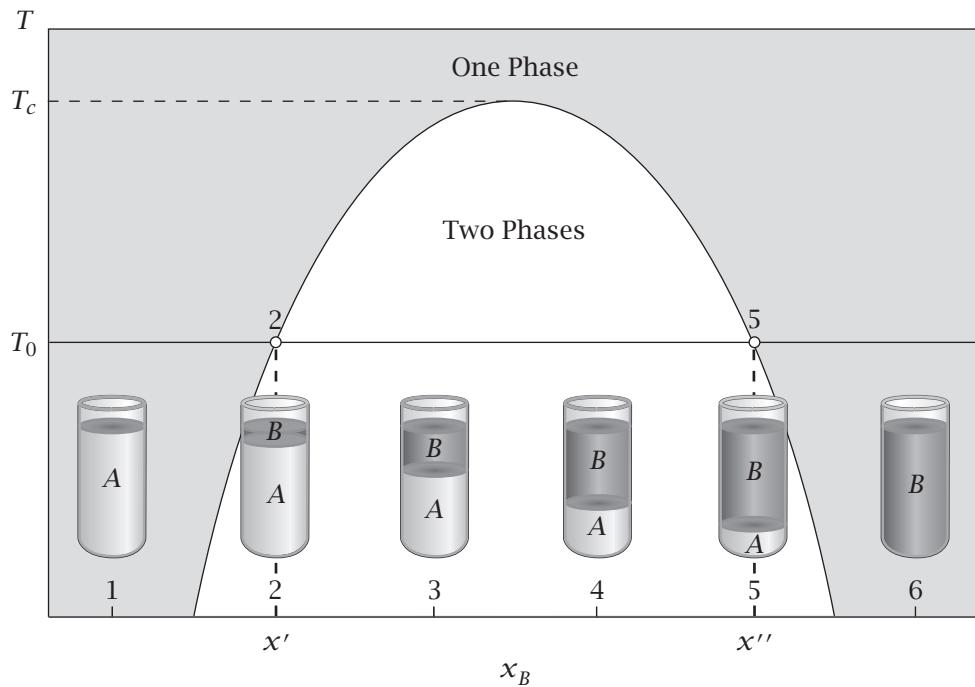


Figure 25.3 Six different solutions help to map out the phase diagram. T_0 is the temperature of the solution, x' is the concentration of B that saturates liquid A , and x'' is the concentration of B at which A saturates B . T_c is the critical temperature.

2. The solubility limit of oil in water. The oil concentration in test tube 2 is greater than in tube 1. This is the lowest oil concentration at which you can detect a second phase, a thin layer of oil on the water. Tube 2 has enough oil to *saturate* the capacity of the water to hold oil. To draw a phase diagram, you mark this temperature and composition at which you first see the appearance of an additional phase. This mark defines one point on the *phase boundary* in Figure 25.3.

3. An oil phase on top of a water phase. Test tube 3 has a higher oil concentration than tube 2. This is a *two-phase solution*. The top phase is mostly oil (but contains some water) and the bottom phase is mostly water (but contains some oil). The water-rich phase on the bottom is saturated with oil, having oil concentration x' . The oil-rich phase on top is saturated with water, having oil concentration x'' .

4. More oil phase on top of the water phase. Tube 4 has a larger volume of the oil-rich phase than tube 3, and a smaller volume of the water-rich phase. But the *compositions* of the oil-rich and water-rich phases remain the same as they are in test tube 3: the bottom phase still has oil concentration x' , and the top phase still has oil concentration x'' .

5. The solubility limit of water in oil. Beyond this point, the system becomes a single phase again. At this point, the water-rich phase disappears and you

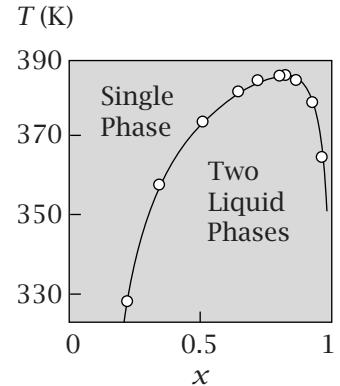


Figure 25.4 A phase diagram for liquid benzene (mole fraction x) in perfluoroheptane. Source: JN Murrell and AD Jenkins, *Properties of Liquids and Solutions*, Wiley, New York, 1982. Data are from JH Hildebrand, BB Fisher, and HA Benesi, *J Am Chem Soc* **72**, 4348–4351 (1950).

have only an oil-rich phase saturated with water. As you draw the phase diagram, mark this second point to indicate the changeover from a two-phase to a one-phase system.

6. A small amount of water dissolved in oil. Test tube 6 contains a single phase, in which a very small amount of water is fully dissolved in a large amount of oil.

Solutions 2 and 5 each define a point on the *phase boundary*, also called the *solubility curve* or *coexistence curve*. A phase is a region in which the composition and properties are uniform. Inside the phase boundary is the *coexistence region* where two phases coexist in equilibrium: an oil-rich phase sits on a water-rich phase (as in test tubes 3 and 4). Outside the phase boundary, you have a one-phase solution: either oil dissolved in water or water dissolved in oil (tubes 1 and 6). From the six experiments at temperature $T = T_0$, the phase boundary is defined by two compositions: x' and x'' .

Now perform the same experiment with another series of tubes of different compositions at a higher temperature $T = T_1$. Again, mark the two phase-boundary points $x'(T_1)$ and $x''(T_1)$. At higher temperatures, the phase boundaries are closer together: more oil dissolves in water and more water dissolves in oil. Continue increasing the temperature and finding the two phase-boundary points for each temperature. This is how you map out liquid-liquid phase diagrams such as the one shown in Figure 25.4. For any temperature higher than the *critical temperature* T_c (see Figure 25.3), the two components are *miscible in all proportions*. That is, above T_c , there are not two distinct phases. Any proportion of the two components form a single mixed system.

Focus on solution 3 at temperature T_0 . Suppose you want to know the solubility limits x' and x'' . Follow the isotherm (that is, the horizontal line of constant temperature, called the *tie line*) to the left, where it intersects the phase boundary at composition x' at point 2. Also, follow the tie line to the right where it intersects the phase boundary at composition x'' (point 5). Points 2 and 5 define the two compositions of the coexisting phases. For the oil phase, the mole fraction of oil is x'' and that of water is $1-x''$. For the water phase, the mole fraction of oil is x' and that of water is $1-x'$.

In this series of test tubes, we fixed the temperature and varied the composition. Now, instead, fix the composition and change the temperature, say of solution 4. Read the results from the same phase diagram. At low temperatures, the solution has two phases. At a higher temperature, above the phase boundary, the two components dissolve in each other. This is the standard lore of bench chemistry: heating helps liquids to mix and dissolve. It also works for solids; this is how you make metal alloys, like brass or bronze.

If you are not skilled with a pipette, phase equilibria can help. Suppose you need a highly reproducible concentration of an oil-water mixture. Your mixing skills need only be good enough to produce any composition between solutions 2 and 5. The phase equilibrium will regulate the concentrations to be precisely x' or x'' .

EXAMPLE 25.1 Cataracts in your eye. Figure 25.5 shows the phase diagram for mixtures of water with a protein called crystallin, which is a component of the lens of your eye.

What is the solubility limit for the protein in water at 25°C? The horizontal line at 25°C intersects the left side of the coexistence curve at about 50 mg mL⁻¹. This is the solubility limit, the most crystallin that can dissolve in water (at 25°C).

What is the concentration of protein in the ‘mostly protein’ phase at 25°C? The same horizontal line intersects the right side of the coexistence curve at about 600 mg mL⁻¹.

What is the solubility limit at body temperature ($T = 37^\circ\text{C}$)? This temperature is near the critical point, so the solubility limit varies strongly with small changes in temperature. Slightly below 37°C, the solubility limit is 100–200 mg mL⁻¹. Slightly above 37°C, the protein is miscible in water at all concentrations. Such broad fluctuations around biological temperatures may contribute to forming cataracts.

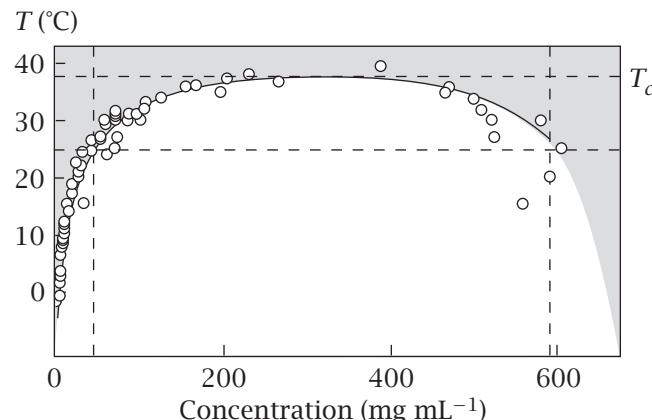


Figure 25.5 The phase diagram for γ -crystallin in water. The solubility limits at 25°C are the points where the horizontal dashed line intersects the experimental curve. The higher horizontal line shows that the critical solution temperature is 37°C, which coincides with body temperature. Source: ML Broide, CR Berland, J Pande, et al., *Proc Natl Acad Sci USA* **88**, 5660–5664 (1991).

Not only does a phase diagram tell you the two *compositions* x' and x'' that define the solubility curve. It also tells you the *amounts*, i.e., how much of the total volume of solution are in each of the top and bottom phases.

The Lever Rule: Computing the Amounts of the Stable Phases

You can determine the amounts (the volumes) of the phases from the phase diagram by using the *lever rule*. For example, what are the amounts of the oil-rich and water-rich phases in tube 3 in Figure 25.3? The ‘lever’ is the tie-line from point 2 to point 5. Imagine that the line ‘pivots’ at point 3 (see also Example 25.2). The relative length of the line from point 3 to point 2 compared with the whole length, from 5 to 2, gives the fraction of the volume of solution that is in the “ phase. Similarly the fractional distance from point 3 to point 5 gives the fraction of the volume that is in the ’ phase.

The lever rule has the following basis. The amounts of *B* in each phase must sum to the total amount that were put into the solution. Let the relative amount of material in the *A*-rich phase be represented by the fraction

$$f = \frac{\text{number of molecules in } A\text{-rich phase}}{\text{total number of molecules in both phases}},$$

so $(1-f)$ is the fraction of all molecules that are in the *B*-rich phase. You also have

$$x' = \frac{\text{number of } B \text{ molecules in } A\text{-rich phase}}{\text{number of molecules in } A\text{-rich phase}}$$

and

$$x'' = \frac{\text{number of } B \text{ molecules in } B\text{-rich phase}}{\text{number of molecules in } B\text{-rich phase}}.$$

Combining these expressions gives

$$fx' = \frac{\text{number of } B \text{ molecules in } A\text{-rich phase}}{\text{total number of molecules in both phases}}.$$

In addition,

$$(1-f)x'' = \frac{\text{number of } B \text{ molecules in } B\text{-rich phase}}{\text{total number of molecules in both phases}}.$$

Combining the last two expressions gives

$$\begin{aligned} fx' + (1-f)x'' &= x_0 \\ &= \frac{\text{number of } B \text{ molecules in both phases}}{\text{total number of molecules in both phases}}. \end{aligned} \quad (25.1)$$

Rearranging Equation (25.1) leads to the **lever rule**:

$$f = \frac{x_0 - x''}{x' - x''}. \quad (25.2)$$

EXAMPLE 25.2 The lever rule. Use the phase diagram in Figure 25.6 to compute the fraction f of molecules that are in the *A*-rich and *B*-rich phases at temperature T_0 . From the given phase diagram, you have $x' = 0.1$ and $x'' = 0.9$. You also know the composition of your solution: the overall *B* concentration is $x_0 = 0.3$. So, the total length of the lever arm is $x'' - x' = 0.9 - 0.1 = 0.8$. Substitute these quantities into Equation (25.2) to get $f = (x_0 - x'')/0.8 = 3/4$; this is the fraction of all molecules that are in the *A*-rich phase. The fraction of all molecules that are in the *B*-rich phase is $1-f = 1/4$.

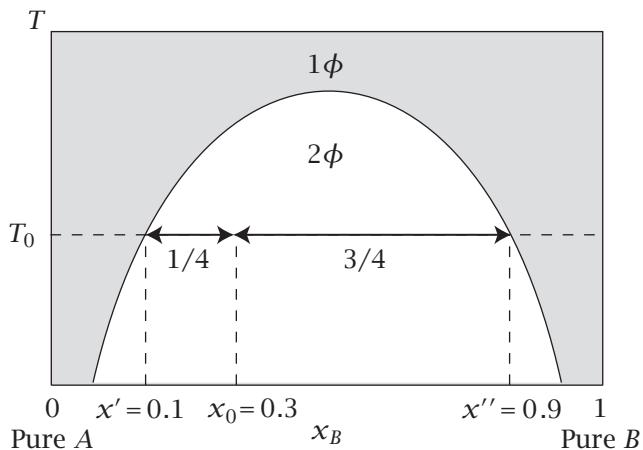


Figure 25.6 The lever rule. Suppose you are given the phase diagram points $x' = 0.1$ and $x'' = 0.9$. Now you mix together two components to have a B composition $x_0 = 0.3$. To find the relative amounts of the two phases, compute f using Equation (25.2). This gives $f = (0.9 - 0.3)/(0.9 - 0.1) = 3/4$, the relative length of the right side of the lever arm. So, the left side of the lever arm has relative length $(1 - f) = 1/4$. Three-quarters of the system is in the x' (mostly A) phase. The other one quarter of the system is in the mostly- B phase.

These types of such phase diagrams tell you what values of temperature and composition define the boundaries between single-phase and two-phase regions. And, they tell you the compositions and amounts of the coexisting phases. Now, let's use energy landscapes to understand when a system will be stable as a single phase or when it be unstable and split into multiple phases in equilibrium.

Peaks on Energy Landscapes Are Points of Instability

What drives a mixture of oil and water to separate into two phases? Compare the energy landscapes in Figures 25.7. The horizontal axes represent the mole fraction concentration of oil, x . The vertical axes give the free energy $F(x)$ of the combined system of oil plus water at composition x . (We will develop a microscopic model for $F(x)$ later in this chapter. For now, we just focus on the general principles that would predict phase stabilities if we knew the shape of the energy landscape.) Suppose you put into a beaker two components A and B , with B concentration $x_0 = 1/2$. The landscape has two stable states (the minima). The system has two options: two balls metaphorically stay at the top of the hill (the system mixes, Figure 25.7(a)) or one ball rolls down to the left energy well and the other ball rolls down to the right well (the system forms separate phases, Figure 25.7(b)). Of these two options, the system can achieve its lowest possible total free energy by forming separate phases. That is, $F_{\text{hilltop}} = 2F_0$ and $F_{\text{valleys}} = 2F_1$. Since $F_1 < F_0$, you see that $F_{\text{valleys}} < F_{\text{hilltop}}$; see Figure 25.8.

Why roll two balls on this hill, rather than one? Our system has a constraint. If one ball rolled from the top of the hill to the valley on the left, it would represent changing the composition of the system from x_0 to x' . But this is not possible. The composition of the system is fixed to be x_0 by the total amount of oil and water that you put into the beaker in the first place. The system does not have the freedom to change its overall composition. But, the system *can* divide into two component subsystems having compositions x' and x'' , where $x' + x'' = 2x_0$ satisfies the constraint. (For example, imagine that you start with 100 A 's and 100 B 's, $x_0 = 1/2$. A possible phase-separated state is

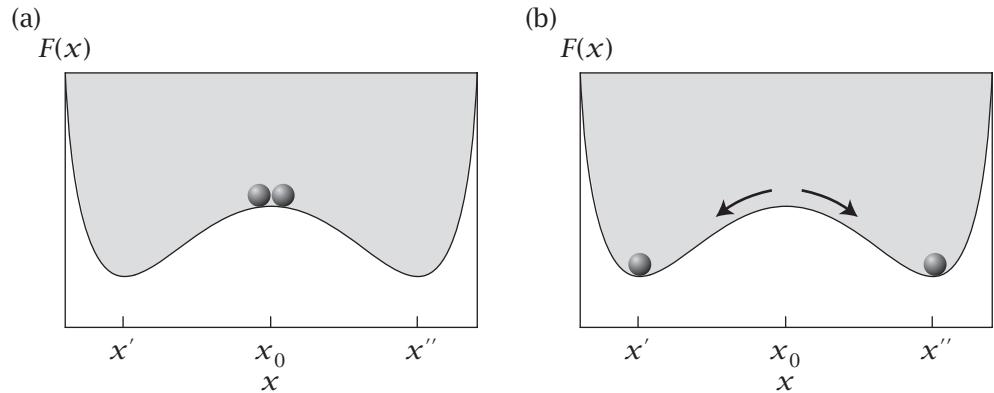


Figure 25.7 Unstable system has a peak in $F(x)$. If a system has composition x_0 , it could either (a) remain mixed (two balls at the top of the hill) or (b) undergo a separation (two balls in the valleys) into two phases having compositions x' and x'' . The latter option has lower total free energy.

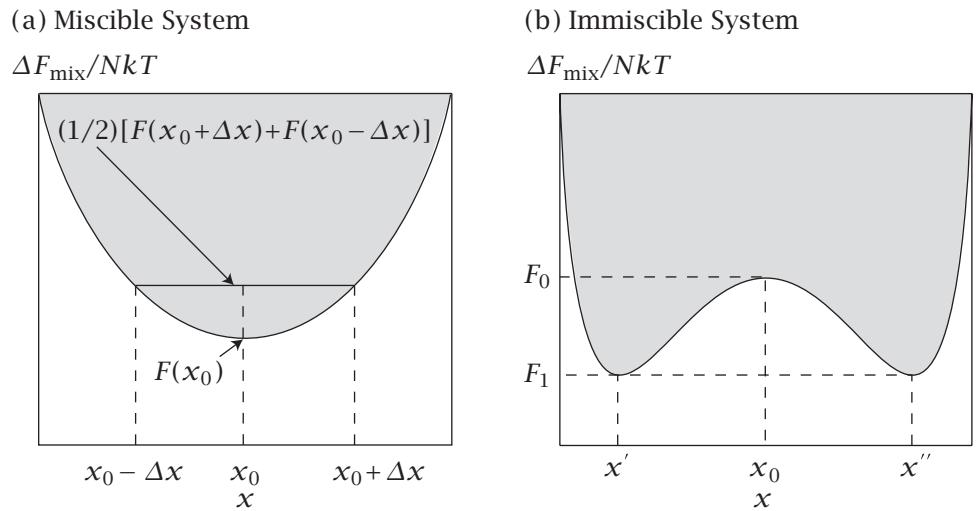


Figure 25.8 (a) If a system's free energy $\Delta F_{\text{mix}}(x)$ is concave upwards, the system is miscible. (b) For compositions (near $x = x_0$, in this case) where $\Delta F_{\text{mix}}(x)$ is concave downwards, systems are immiscible.

an *A*-rich phase having 75 *A* and 25 *B* molecules and a *B*-rich phase having 25 *A* molecules and 75 *B* molecules. This phase separation would fully satisfy composition constraints.)

From Figure 25.7, it is clear how to compute the coexistence curve. The phase compositions x' and x'' are points at which the free energy $F(x)$ is a minimum,

$$\left(\frac{\partial F}{\partial x}\right)_{x=x'} = \left(\frac{\partial F}{\partial x}\right)_{x=x''} = 0. \quad (25.3)$$

If you have a model of the mixing free energy versus composition for different temperatures, $F(x, T)$, then finding these two minima at each temperature will give you the full coexistence curve, as we'll see later.

The Common Tangent Predicts the Compositions of Phases

Here is a way that you can predict the compositions of the two phases in general if you know the function $F(x)$, even if it is not symmetrical around $x = 0.5$ (see Figure 25.9). You can find the compositions x' and x'' by drawing a line that is tangent to $F(x)$ at two points. The two points of tangency predict the compositions of the two phases. Here is why the common tangent line identifies the compositions of the phases in equilibrium.

To get the free energy $F(x)$ at fixed T and V , start with Equation (8.18) or Equation (8.10). For a system having N_A particles of type A with chemical potential μ_A , and N_B particles of type B with chemical potential μ_B ,

$$\begin{aligned} F &= \mu_A N_A + \mu_B N_B \\ &= N(\mu_A x_A + \mu_B x_B). \end{aligned} \quad (25.4)$$

Use $x_A = 1 - x_B$, and take the derivative $(\partial F / \partial x_B)$ to get

$$\left(\frac{\partial F}{\partial x_B} \right) = N(\mu_B - \mu_A). \quad (25.5)$$

In a two-phase equilibrium, the chemical potential of each component, A or B, must be the same in both phases, so $\mu'_A = \mu''_A$ and $\mu'_B = \mu''_B$. Subtract $\mu'_A = \mu''_A$ from $\mu'_B = \mu''_B$, and multiply by N to get

$$N(\mu'_B - \mu'_A) = N(\mu''_B - \mu''_A). \quad (25.6)$$

Substitute Equation (25.5) into Equation (25.6) for each phase. This leads to the common tangency condition that the slope of the line must be the same at x' and x'' ,

$$\left(\frac{\partial F}{\partial x} \right)_{x=x'} = \left(\frac{\partial F}{\partial x} \right)_{x=x''}. \quad (25.7)$$

If the function $F(x)$ has minima of two different depths, the common tangency condition means that the points x' and x'' are not exactly the points at which the free energy is a local minimum. However, when $F(x)$ has two minima of equal depths, the common tangent line has zero slope, $(\partial F / \partial x) = 0$. Now, let's model the underlying driving forces for phase equilibria.

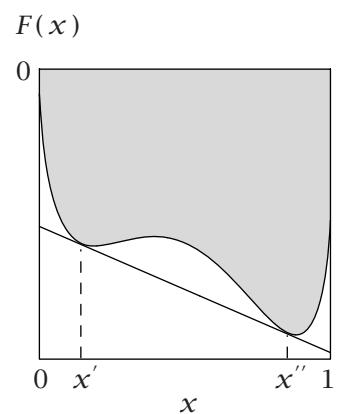


Figure 25.9 The common tangent to $F(x)$, the free energy as a function of composition, identifies the stable phases having compositions x' and x'' .

Thermal Phase Equilibria Arise from a Balance between Energies and Entropies

Why don't oil and water mix? At low temperatures, *AA* and *BB* attractions are strong, so the system develops an *A*-rich phase and a *B*-rich phase to maximize these favorable interactions. However, at high temperatures, the mixing entropy dominates, so *A* molecules mix with *B* molecules. Let's make this more quantitative using the lattice model of solutions.

Begin with two liquid phases, one of pure *A* and one of pure *B*. Combine them to give a mole fraction x of *B* in *A*. The free energy of mixing ΔF_{mix} for that process is given in the lattice model by Equation (15.16):

$$\frac{\Delta F_{\text{mix}}}{NkT} = x \ln x + (1-x) \ln(1-x) + \chi_{AB}x(1-x).$$

where $N = N_A + N_B$ is the total number of molecules and kT is Boltzmann's constant \times temperature. Recall that $\chi_{AB} = c_1/T$ is a dimensionless free energy quantity (see Equation (15.13)) that is inversely proportional to temperature and describes the interactions between the *A* and *B* molecules, where

$$c_1 = \frac{z}{k} \left(w_{AB} - \frac{w_{AA} + w_{BB}}{2} \right) \quad (25.8)$$

from the simple lattice model is a constant, independent of temperature. Changing the temperature changes the balance of forces. Higher temperatures correspond to smaller values of χ_{AB} in Equation (15.16).

Here's how to predict the phase diagram and the energy landscapes using this lattice model. Choose a temperature of interest, $T = T_1$. For two given materials *A* and *B*, you know c_1 . Now compute ΔF_{mix} as a function of x . Then, calculate also the corresponding curves for other temperatures T_2, T_3, \dots . These should look like the six curves in Figure 25.10(a). Figure 25.10(b) shows all six curves superimposed on a single figure. If *A* and *B* dislike each other enough at a given temperature ($\chi_{AB} > 2$), then the free energy function has two minima, and the system will form separate phases at appropriate concentrations. If *A* and *B* are relatively attracted to each other ($\chi_{AB} < 2$), then *A* and *B* are miscible in all proportions. The critical point is at $\chi_{AB} = 2$. These different values of χ_{AB} are obtained by changing the temperature. Now, find the one or two minima of each curve, where $\partial \Delta F_{\text{mix}} / \partial x = 0$. The dashed curve in Figure 25.10(b) connects all these minima, defining the coexistence curve predicted by the model. Finally, put the coexistence curve on a new figure (Figure 25.10(c)), where the vertical axis is now temperature, not free energy. This gives you the model's phase diagram; see Example 25.3.

The coexistence curve obtained from this simple lattice theory captures the main feature of typical experiments: the phase diagram is concave downward, with a two-phase region inside, a one-phase region outside, and a critical point at the top. The theory predicts a phase diagram that is symmetrical around $x = 0.5$. Experiments, however, are often not so symmetrical (see Figure 25.4); in those cases, you need a better model.

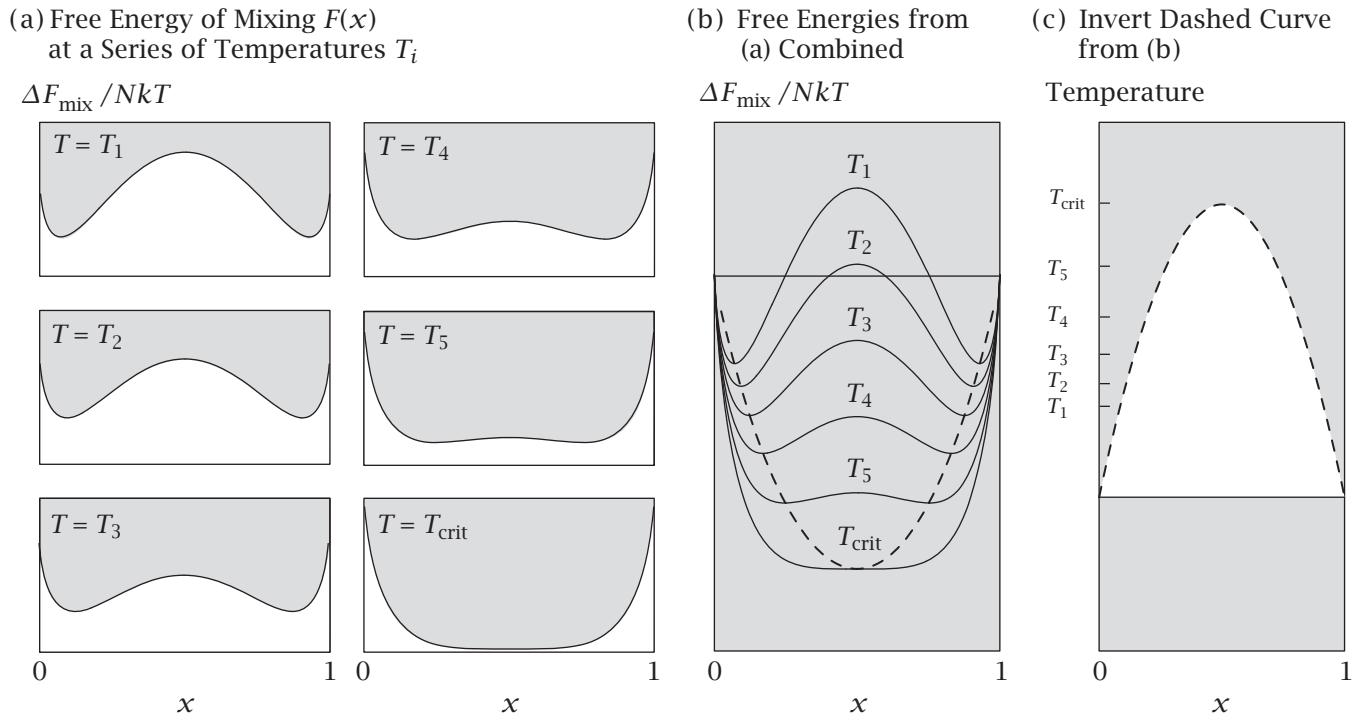


Figure 25.10 Use the lattice model to compute the mixing phase diagram. (a) Compute $\Delta F_{\text{mix}}(x)/NkT$ for a series of temperatures given c_1 , using $\chi_{AB} = c_1/T$ in Equation (15.16). (b) Put those curves onto a single figure. The dashed curve represents the two minima in $\Delta F_{\text{mix}}(x)/NkT$ against x for each temperature. Inside the dashed curve is the two-phase region. (c) Replot the dashed curve of (b) so that the temperatures of the free energy minima are on the vertical axis.

EXAMPLE 25.3 The phase diagram for the lattice model of solutions. We want to locate the minima in the free energy function. We seek the points where the derivatives are zero. Substitute the lattice model free energy, Equation (15.16), into Equation (25.3):

$$\left(\frac{\partial \Delta F}{\partial x}\right)_{x=x' \text{ or } x''} = NkT \left[\ln\left(\frac{x}{1-x}\right) + \chi_{AB}(1-2x) \right]_{x=x' \text{ or } x''} = 0.$$

You can compute x' and x'' for a series of different values of χ_{AB} by rearranging and solving the equation

$$\ln\left(\frac{x'}{1-x'}\right) = -\chi_{AB}(T)(1-2x'). \quad (25.9)$$

This is a *transcendental* equation. You can solve for $x'(T)$ by iterative methods. Because the lattice model free energy is symmetrical around $x = 1/2$, you have $x'' = 1 - x'$. If the solute is only sparingly soluble (as oil would be in water), then $x' \ll 1$. In that case, Equation (25.9) simplifies, and predicts that the solubility is

$$\ln x' \approx -\chi_{AB} \quad \text{or} \quad x' = e^{-\chi_{AB}}. \quad (25.10)$$

If χ_{AB} is large (A has very little affinity for B), then the solubility is small, according to Equation (25.10).

In summary, when $\chi_{AB} < 2$, which happens at high temperatures, the disaffinity of A for B is small, and the entropic tendency to mix is greater than the energetic tendency to separate. The free energy function is concave upward for all compositions, so the system is stable as a single mixed phase. However, if the temperature is low enough, the disaffinity of A for B will be strong ($\chi_{AB} \gg 0$), so the energetic affinities overwhelm the mixing entropy, resulting in a separation of the system into two phases. Very dilute solutions are miscible, despite the strong disaffinities, because there is a large mixing entropy at those compositions. Phases in equilibrium are seldom perfectly pure, because perfect purity is strongly opposed by the mixing entropy.

The Spinodal Curve Describes the Limit of Metastability

So far, we have discussed the coexistence curve, also called the *binodal* curve. It describes the *global stability* of a system against phase separation. Here we describe another property, called *metastability* or *local stability* or the *spinodal curve*. Look at Figure 25.11. In the sections above, we found that a solution having overall composition x_0 that is inside the two-phase region should form separate phases having compositions x' and x'' in order to lower its total free energy. But, sometimes a system that is inside its two-phase region, if handled very gently and not shaken or stirred, does not actually form separate phases.

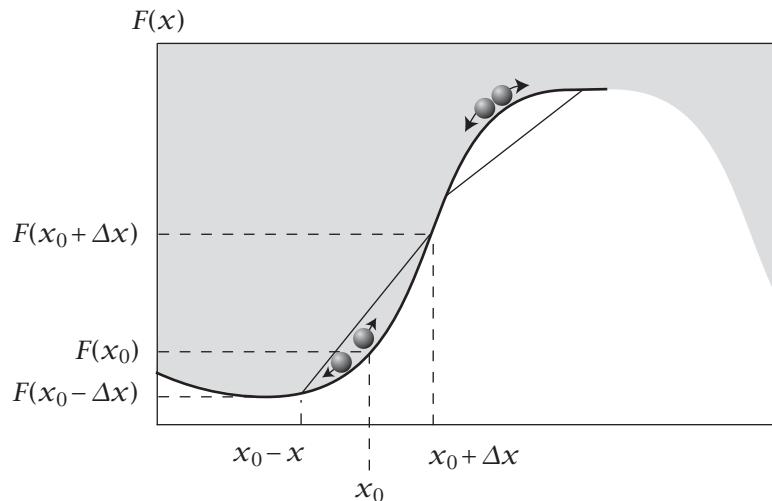


Figure 25.11 Metastability. Where the free energy is concave upwards (balls on the bottom left), the system at x_0 does not undergo local phase separation into phases having compositions $x_0 + \Delta x$ and $x_0 - \Delta x$ because a local phase separation would increase the total free energy (indicated by the straight sloping line on the left). (If the system were shaken vigorously, so that it could explore the whole energy landscape more broadly, the system would undergo a global separation to phases of compositions x' and x'' ; see Figure 25.8). But, where the free energy is concave downwards, the system will undergo a local phase separation to lower its net free energy.

Such solutions are called *metastable solutions*. For example, a solution can be *supersaturated*. It can have slightly more oil dissolved in the water than would be predicted by the coexistence curve: x can be larger than x' for a given temperature. Or, a solution can be *supercooled*: it can remain mixed even at a lower temperature T for a given x than would be predicted by its coexistence curve. Gentle handling can postpone phase separation.

Figure 25.11 shows that a system has *local stability* where the energy landscape is concave upwards ($\partial^2 F / \partial x^2 > 0$) and *local instability* where the landscape is concave downwards ($\partial^2 F / \partial x^2 < 0$). To see this, compare the free energy $2F(x_0)$ of the mixed system with the free energy $F(x_0 + \Delta x) + F(x_0 - \Delta x)$ of a system that locally separates into slightly different compositions $(x_0 + \Delta x)$ and $(x_0 - \Delta x)$. This system is locally stable as a single phase when

$$F(x_0 + \Delta x) + F(x_0 - \Delta x) > 2F(x_0). \quad (25.11)$$

Rearrange Equation (25.11) to get the condition for stability against small fluctuations:

$$F(x_0 + \Delta x) - F(x_0) - [F(x_0) - F(x_0 - \Delta x)] > 0. \quad (25.12)$$

Equation (25.12) is a difference of a difference:

$$\Delta\Delta F > 0.$$

As $\Delta x \rightarrow 0$, this double difference can be expressed as a second derivative:

$$\left(\frac{\partial^2 F}{\partial x^2} \right)_{x=x_0} > 0. \quad (25.13)$$

When this second derivative is positive, the mixed system is locally stable and does not form separate phases. When $(\partial^2 F / \partial x^2)$ is negative, the system will locally divide into separate phases, and then will also proceed further to global instability and full phase separation.

EXAMPLE 25.4 The spinodal (metastability) curve for the lattice model. You can compute the spinodal curve for the lattice model by taking the second derivative of Equation (15.16) and setting it equal to zero:

$$\left(\frac{\partial^2 F}{\partial x^2} \right) = NkT \left(\frac{1}{x} + \frac{1}{1-x} - 2\chi_{AB} \right) = 0. \quad (25.14)$$

Figure 25.12 shows that the spinodal curve falls inside the binodal curve. For systems that are sparingly soluble ($x \approx 0$), the spinodal composition x'_s is given by $x'_s = 1/(2\chi_{AB} - 1)$.

Figure 25.12 shows the metastable region. The spinodal curve is inside the binodal curve. The binodal curve is where $(\partial F / \partial x) = 0$. The spinodal curve is where $(\partial^2 F / \partial x^2) = 0$. In the region M between the two curves, the system is metastable. In the M region, if you don't shake the system, it won't form separate phases. However, if you shake the system sufficiently, the system will

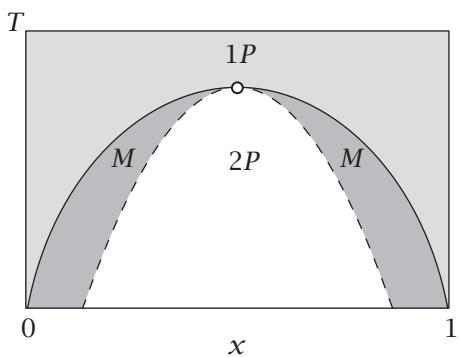


Figure 25.12 In region $1P$, the system is always in one phase. M is the metastable region: the system is a single phase if perturbations are small, or two phases if perturbations are large enough to cause demixing. In region $2P$, the system is always in two phases. (—) binodal curve; (---) spinodal curve. The spinodal and binodal curves meet at the critical point (○).

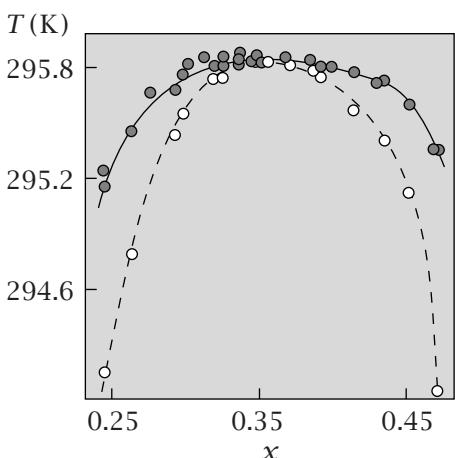


Figure 25.13 Binodal (—○—) and spinodal (---○---) curves for a binary liquid of *n*-hexane in *n*-tetradecafluorohexane. Source: GI Pozharskaya, NL Kasapova, VP Skripov, and YD Kolpakov, *J Chem Thermodyn* **16**, 267 (1984); WJ Gaw and RL Scott, *J Chem Thermodyn* **3**, 335 (1971).

form separate phases. The system is stable to small perturbations, but unstable to large perturbations. However, inside the spinodal curve, in the two-phase region labeled $2P$, the system will always separate into two phases, no matter how gently you handle it. Careful experiments are required if you want to find both binodal and spinodal curves, because it is often hard to tell whether your perturbations are small enough to define the true limits of metastability. Binodal and spinodal curves are shown for hexane-tetradecafluorohexane mixtures in Figure 25.13.

The Critical Point Is Where Coexisting Phases Merge

The *critical point* is where the binodal and spinodal curves intersect. For our mixing processes, the critical point is the top peak point of the phase diagram. Above the critical temperature T_c , the two components are miscible in all proportions (see Figure 25.12). The critical point $x = x_c$ is where the binodal and spinodal curves merge, where both the first and second derivatives of the free energy are zero. The *A*-rich and *B*-rich phases merge at the critical point, so

the two compositions x' and x'' approach each other. The spinodal points also merge. So, the critical point is also where the third derivative is zero:

$$\lim_{x' \rightarrow x''} \left[\left(\frac{\partial^2 F}{\partial x^2} \right)_{x=x'} - \left(\frac{\partial^2 F}{\partial x^2} \right)_{x=x''} \right] = 0 \implies \left(\frac{\partial^3 F}{\partial x^3} \right) = 0. \quad (25.15)$$

EXAMPLE 25.5 Lattice model critical point. To find the critical point (x_c, χ_c, T_c) for the lattice mixture model, determine the point where both the second and third derivatives of the free energy (given by Equation (15.16)) equal zero:

$$\left(\frac{\partial^3 F}{\partial x^3} \right)_{x=x_c} = NkT \left[\frac{1}{(1-x_c)^2} - \frac{1}{x_c^2} \right] = 0 \implies x_c = \frac{1}{2}. \quad (25.16)$$

Now use $x_c = 1/2$ and Equation (25.14) to get the second derivative:

$$\begin{aligned} \left(\frac{\partial^2 F}{\partial x^2} \right)_{x=x_c} &= \frac{1}{(1/2)} + \frac{1}{(1-\frac{1}{2})} - 2\chi_c \\ &= 4 - 2\chi_c = 0 \implies \chi_c = 2, \end{aligned} \quad (25.17)$$

where χ_c is the value of χ at the critical point.

From Equations (25.16) and (25.17), you can compute the critical point x_c and T_c for the lattice model. Because $\chi_c = c_1/T_c = 2$, you can compute the intermolecular interaction quantity $c_1 = 2T_c$ if you know the critical temperature. If $c_1 = 0$, the system is fully miscible at all temperatures, and there is no two-phase region at all. Figure 25.14 shows the lattice model phase diagram.

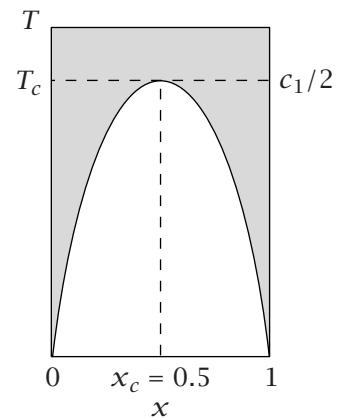


Figure 25.14 The lattice model phase diagram for mixing. The *critical point* is the peak, with critical temperature T_c and composition x_c .

The Principles of Boiling Are Related to the Principles of Immiscibility

The boiling of water would seem to be a very different type of phase equilibrium from the mixing of oil and water. But in both cases, phase boundaries are concave downward with a two-phase region inside and a critical point at the top. Remarkably, both can be described, to a first approximation, by the same physical model.

To create the phase diagram for boiling, put N particles in a container of volume V at temperature T . Follow the system pressure $p(T, V, N)$ as you decrease the volume, starting with low densities ρ , that is, large volumes per molecule, $v = V/N = \rho^{-1}$.

At low densities (right side of Figure 25.15), the system is a single phase: a gas. As you compress the gas, decreasing the volume per molecule to v_2 , the pressure increases. At volumes between v_1 and v_2 , a second phase begins to emerge in the container: some liquid forms in equilibrium with the gas phase. Compress the system more, and you get more liquid and less gas. When the volume per molecule becomes smaller than v_1 , the system becomes a single phase again, but now it is a liquid. Reducing the volume further leads to a very steep increase in the pressure because liquids are not very compressible.

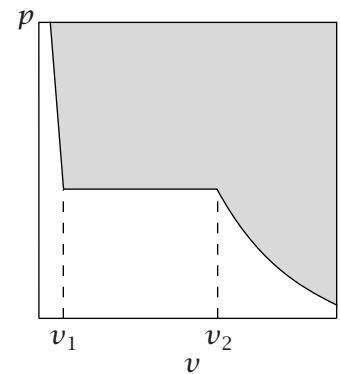
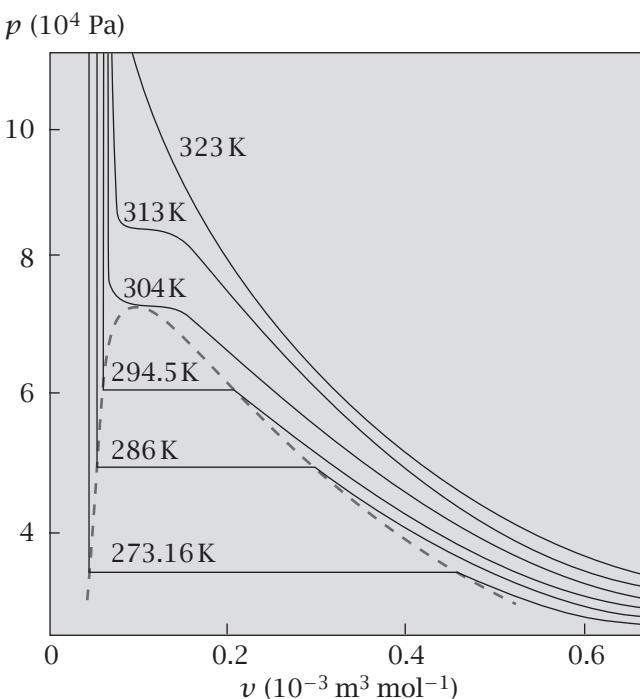


Figure 25.15 A pressure–volume isotherm for a gas: v is the volume per molecule and v_1 and v_2 are the volumes of the coexistent vapor and liquid phases.

Figure 25.16 The isotherms of pressure p versus molar volume v for carbon dioxide. The broken line is the phase diagram. Like a miscibility phase diagram, this phase diagram is concave downward and has a critical point at the top. Inside the dashed region, the system is boiling, and has two stable phases. Outside the dashed curve, the system is in a single phase: liquid or gas.

Source: AJ Walton, *Three Phases of Matter*, 2nd edition. Oxford University Press, New York, 1983.



Volumes v_1 and v_2 define two phase boundaries. To the right of v_2 , the system is a single phase (gas). To the left of v_1 , the system is a single phase (liquid). For any intermediate volume per molecule between v_1 and v_2 , the system is two phases in equilibrium: it is boiling. In the boiling region, the gas has volume per molecule v_2 and the liquid has volume per molecule v_1 . For constant temperature, the pressure–volume curve in Figure 25.15 is called an *isotherm*. The total amounts of liquid and gas phases can be found by using the lever rule (see pages 495–497).

Compress again, but now at a higher temperature. You’ll get two more points on the phase boundary. Just as with the liquid solubilization transition, the phase boundaries are closer together at higher temperatures. Another similarity between boiling and liquid immiscibility is the existence of a critical point. Above the critical temperature, you can continuously compress the gas to make it denser, ultimately reaching liquid densities, without ever passing through a two-phase region: the fluid never boils. The dashed line in Figure 25.16 shows the phase diagram for CO₂. The critical point for CO₂ is 304 K. Above the critical temperature, CO₂ is called a *supercritical fluid*, rather than a supercritical gas or liquid. The supercritical fluid does not have two free energy minima that would distinguish liquid and gas states. Supercritical fluids have practical applications, such as decaffeinating coffee (see Figure 25.19 in the box on page 509).

Because the phase diagram for boiling is concave downward with a two-phase region inside and a critical point at the top, it must arise from a free energy function with two minima. For boiling, the degree of freedom is the volume per molecule (or its inverse, the density), rather than the composition x_B that we used for liquid mixtures. In general, the degree of freedom that

changes dramatically in a phase transition is called the *order parameter*. Different types of phase transition have different order parameters, but the common thread among different phase transitions is the shape of the free energy function versus its order parameter. For both boiling and binary mixture equilibria, the plot of free energy versus the order parameter has two minima that merge together as the temperature increases through the critical point (see Figure 25.17).

The similarities between boiling and immiscibility transitions can be traced to similar underlying principles. Remarkably, the same model can be used for both. For liquid mixtures, we used a lattice to count arrangements of *A*'s and *B*'s to get the entropy, and we summed the *AA*, *BB*, and *AB* interactions to get the energy. For gases, the same lattice-counting procedure gives the number of arrangements of filled sites (particles) and empty sites (free space). For lattice gases, the energy is just the sum of particle pairwise interactions. For the liquid–vapor equilibrium, the lattice model gives the free energy per unit volume (see Equation (24.21)). You can express this free energy in terms of the density $\rho = N/V$, where N is the number of particles and V is the volume of the system,

$$\frac{F}{V} = -a\rho^2 + kT \left[\rho \ln(b_0\rho) + \left(\frac{1}{b_0} - \rho \right) \ln(1 - b_0\rho) \right], \quad (25.18)$$

where a represents the model attractive energy and b_0 represents the excluded volume between two particles. The free energy function versus ρ (Equation (25.18)) has double minima of the type shown in Figure 25.17; see also Figure 25.18.

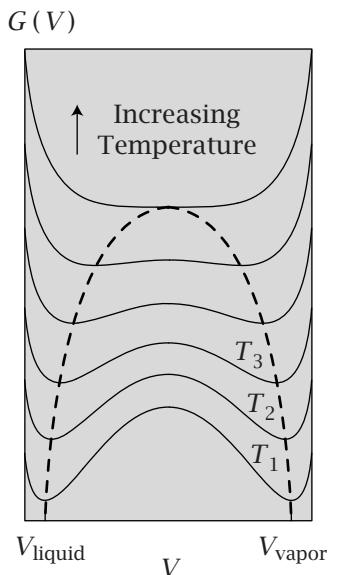


Figure 25.17 The free energy $G(V)$ for a gas–liquid equilibrium. The dashed line encloses the two-phase (boiling) region.

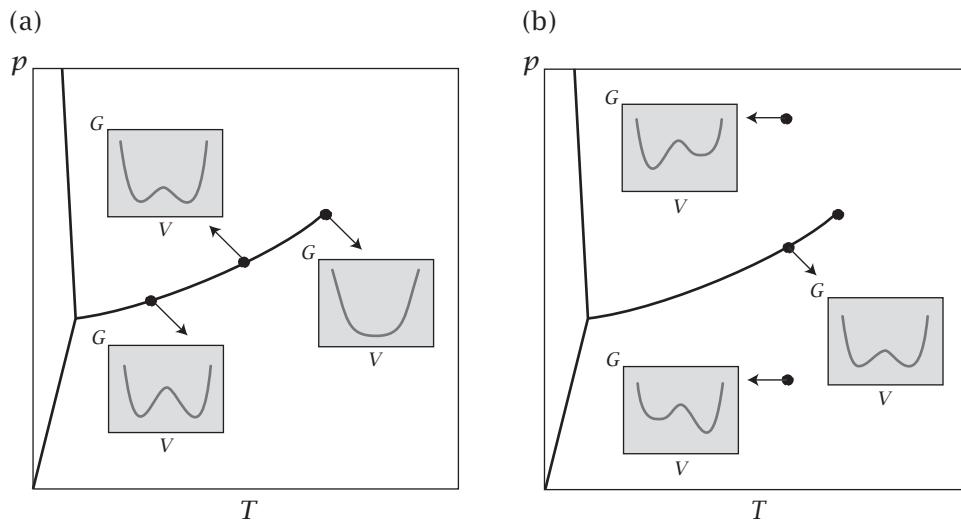


Figure 25.18 (a) $G(V)$ has two minima that become one minimum as the critical point is approached. (b) On the liquid–gas coexistence curve, the free energy has two minima of equal depth. Above it, the liquid phase is more stable. Below it, the gas phase is more stable. Source: Redrawn from HB Callen, *Thermodynamics and an Introduction to Thermostatistics*, 2nd edition, Wiley, New York, 1985.

As with liquid-liquid immiscibility, the coexistence curve for boiling a simple lattice model liquid is the set of points for which $(\partial F/\partial V) = 0$. The metastability curve is where $(\partial^2 F/\partial V^2) = 0$ and the critical point is where you also have $(\partial^3 F/\partial V^3) = 0$. Because the pressure is defined as $p = -(\partial F/\partial V)$, it is often more convenient to express these derivatives for metastability and the critical point in terms of p , respectively, as

$$\left(\frac{\partial p}{\partial V}\right) = 0 \quad \text{and} \quad \left(\frac{\partial^2 p}{\partial V^2}\right) = 0. \quad (25.19)$$

Example 25.6 shows how to compute the critical properties for a van der Waals gas.

EXAMPLE 25.6 The critical point of the van der Waals gas. The van der Waals equation (24.15) is

$$p = \frac{NRT}{V-Nb} - \frac{aN^2}{V^2}.$$

To compute the critical temperature T_c , pressure P_c , and volume V_c , substitute Equation (24.15) into Equations (25.19), taking $N = 1$:

$$\left(\frac{\partial p}{\partial V}\right) = \frac{-RT_c}{(V_c-b)^2} + \frac{2a}{V_c^3} = 0 \quad (25.20)$$

and

$$\left(\frac{\partial^2 p}{\partial V^2}\right) = \frac{2RT_c}{(V_c-b)^3} - \frac{6a}{V_c^4} = 0. \quad (25.21)$$

Rearrange Equation (25.20) to find RT_c ,

$$RT_c = \frac{2a}{V_c^3}(V_c-b)^2, \quad (25.22)$$

and substitute this expression for RT_c into Equation (25.21) to get

$$V_c = 3b. \quad (25.23)$$

Substitute Equation (25.23) for V_c into Equation (25.22) to get

$$T_c = \frac{8a}{27Rb}. \quad (25.24)$$

Substitute Equations (25.23) and (25.24) into Equation (24.15) to get

$$p_c = \frac{a}{27b^2}. \quad (25.25)$$

These equations allow you to find the parameters a and b of a van der Waals gas from its experimental critical point T_c and p_c .

Supercritical Fluids Are Used to Decaffeinate Coffee

Figure 25.19 shows an application of supercritical fluids. It shows how caffeine is removed from coffee. Coffee is mixed with CO₂ at a temperature and pressure above the critical point of the CO₂. The supercritical CO₂ dissolves the caffeine (small black dots), decaffeinating the coffee beans. The fluid mixture of CO₂ with caffeine then flows into a chamber where the pressure is lowered below the critical point so caffeine partitions into water. The carrier CO₂ is recaptured and the caffeine is dumped in the aqueous phase.

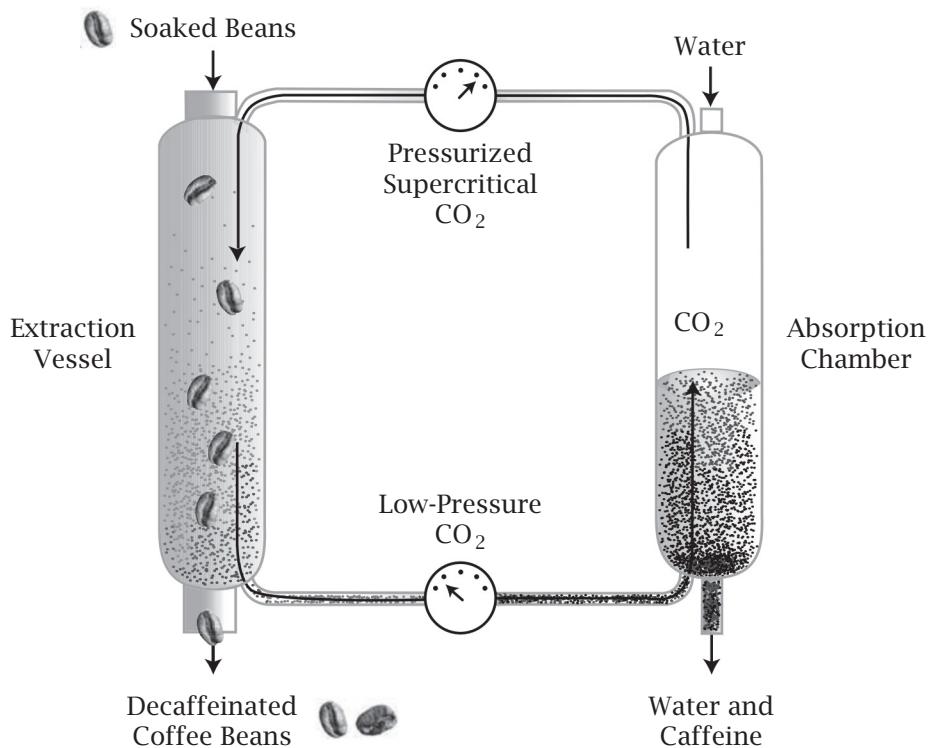


Figure 25.19

The Law of Corresponding States: a Universal Description of Gases and Boiling

A revolution in understanding phase transitions and critical phenomena has occurred since the 1940s. A remarkable underlying unity was found among different types of transition, and for different types of material undergoing the same type of transition. An important step in this unification was the *law of corresponding states*. Different fluids have different critical temperatures and pressures, and different parameters a and b when modeled by the van der Waals equation. However, when equations of state are expressed in terms of dimensionless variables that are normalized by their values at the critical point (critical temperature T_c , pressure p_c , and volume V_c), universal behavior is often

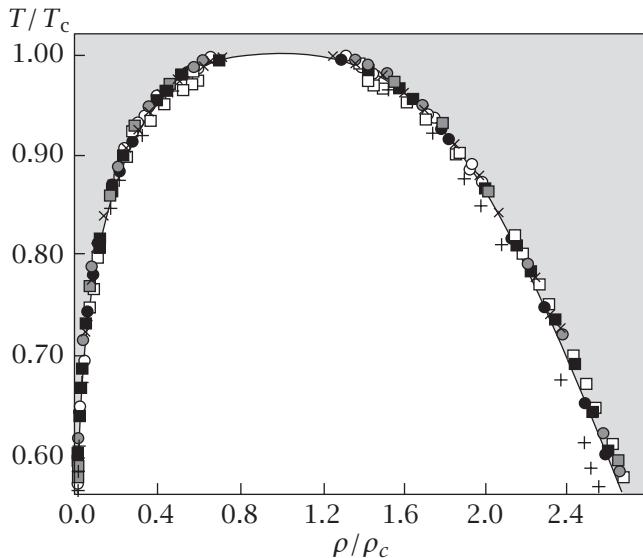


Figure 25.20 Reduced temperature T/T_c plotted against reduced density ρ/ρ_c for eight different gases, illustrating corresponding states: (+) Ne; (■) Kr; (×) Xe; (●) N₂; (□) O₂; (□) CO; (○) CH₄; (●) Ar. Source: EA Guggenheim, *J Chem Phys* **13**, 253–261 (1945).

observed. To see this, express the van der Waals equation in terms of reduced variables, $\pi = p/p_c$, $\theta = T/T_c$, and $\phi = V/V_c$. For $N = 1$, rearrange Equation (24.15) to

$$\left(p + \frac{a}{V^2} \right) (V - b) = RT. \quad (25.26)$$

Divide both sides of Equation (25.26) by p_c and use Equation (25.25) to get

$$\left[\frac{p}{p_c} + \frac{a}{V^2} \left(\frac{27b^2}{a} \right) \right] (V - b) = \left(\frac{27b^2}{a} \right) RT. \quad (25.27)$$

On the left side of Equation (25.27), replace each factor of b with $V_c/3$ from Equation (25.23) and then replace each factor of V/V_c with ϕ . On the right side of Equation (25.27), replace T with θT_c and use Equation (25.24) to get

$$\left(\pi + \frac{3}{\phi^2} \right) \left(\phi - \frac{1}{3} \right) = \frac{8}{3} \theta. \quad (25.28)$$

Figure 25.20 shows the advantage of using reduced variables: it illuminates the underlying similarities between different materials. By plotting the reduced density ρ/ρ_c against the reduced temperature T/T_c , it is found that the data for eight different gases fall on the same line. This is called the law of corresponding states. The importance of this universality is discussed in Chapter 26. Now we consider multicomponent boiling.

Boiling a Liquid Mixture Involves Two Types of Phase Transition

Now let's consider combined phase equilibria: a system that undergoes both liquid-liquid immiscibility and boiling. Consider a liquid having two components *A* and *B*. The system may have a miscibility phase transition. But it will also boil at a sufficiently high temperature. Typically, each component boils

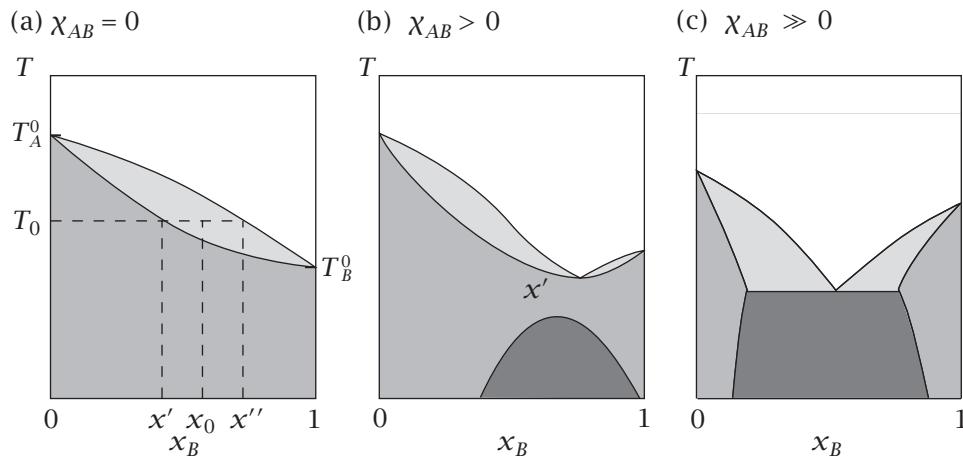


Figure 25.21 Phase diagrams for boiling liquid mixtures: (white) both components are in the vapor phase; (gray) a vapor phase enriched in one component is in equilibrium with a liquid phase enriched in the other component; (dark gray) two liquid phases in equilibrium; (black) a single liquid-phase solution. Temperature T_0 is between the boiling point of pure A , T_A^0 , and the boiling temperature of pure B , T_B^0 . Source: JH Brophy, RM Rose and J Wulff, *The Structure and Properties of Materials*, Volume II: *Thermodynamics of Structure*, Wiley, New York, 1964.

at a different temperature. Boiling a mixture, say of alcohol and water, is the basis for *fractional distillation* (see below). Figure 25.21 shows three typical temperature-composition phase diagrams for boiling a mixture. Figure 25.21(a) shows an ideal mixture ($\chi_{AB} = 0$), having no liquid-liquid two-phase region. The progression from Figure 25.21(a) to (c) represents decreased affinity of A for B (increasing χ_{AB} , in the terminology of the lattice model): the two-phase region grows and the one-phase liquid region shrinks.

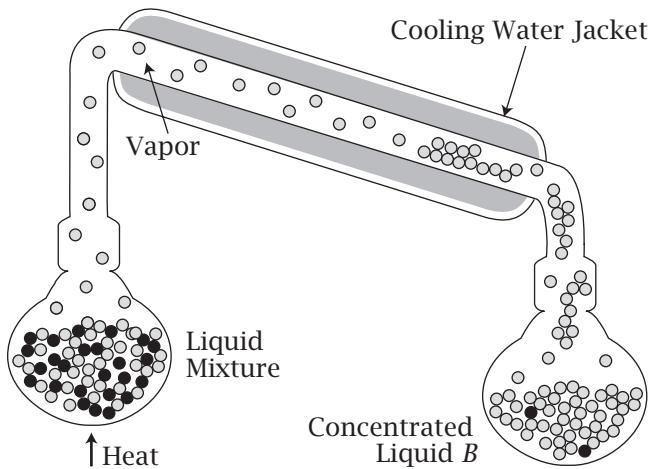
Figure 25.21(a) is a phase diagram for the boiling equilibrium of an ideal two-component liquid. At low temperatures, below the canoe-shaped region, both components are liquids. Their liquids are miscible in all proportions. At high temperatures, above the canoe, both components are gases. Inside the canoe is a two-phase region: a liquid and a gas. Pure B boils at temperature T_B^0 . Pure A boils at temperature T_A^0 . If B is more volatile than A , then $T_B^0 < T_A^0$. For example, alcohol is more volatile than water (see Figure 25.30).

Mix alcohol in water at concentration x_0 (Figure 25.21(a)). Heat the mixture to a temperature T_0 above the boiling point T_B^0 , but below the boiling point T_A^0 , $T_B^0 < T_0 < T_A^0$. Follow the tie-line to the left to see that the liquid that remains behind is purified water (alcohol concentration $x' < x_0$). Follow the tie-line to the right to see that the vapor phase is purified alcohol (concentration $x'' > x_0$). This is the basis for the distillation and purification of liquids.

Distillation Can Result from Boiling a Liquid Mixture

Figure 25.22 shows how this principle is used to distill liquids. A mixture of A and B is heated in the left flask to a temperature above the boiling temperature of the more volatile component B . The volatile component B concentrates in the vapor phase. For example in alcohol-water mixtures, alcohol

Figure 25.22 Distillation apparatus. The left-hand flask contains a mixture of A (●) and B (○), with B being more volatile than A. It is heated to a temperature above the boiling temperature of B.



is more volatile (*B*). The vapor is cooled and condenses into the right flask. *B* is more concentrated in the right flask than in the left flask. By repeating the process, heating the liquid in the right flask at a lower temperature than before (but above the boiling temperature of pure *B*), *B* can be concentrated further.

Phase diagrams for solid mixtures and their melting behavior often resemble those of liquid mixtures and their boiling behavior. In the same way that repeated boiling can purify a liquid mixture, repeated melting can purify a solid mixture. This is called *zone refining*: it is used to purify metals.

Figure 25.23 shows an example of phase equilibria of ‘solid-like’ and ‘liquid-like’ lipid bilayer membranes containing two kinds of phospholipid. In this case, solid-like refers to a high degree of orientational order in the hydrocarbon chains of the lipids, and liquid-like refers to a more disordered state of the chains. A sharp transition between the solid-like and liquid-like phases happens at certain temperatures and compositions.

Phase transitions are not limited to boiling, freezing, and mixing. Several different forms of molecular organization of surfactant molecules as a function of their concentration in water are shown in Figure 25.24. At very low temperatures, they crystallize. At higher temperatures, low concentrations dissolve in water. Increasing their concentration to beyond the critical micelle concentration leads to spherical micelles, then cylindrical micelles, then lamellar phases such as bilayers.

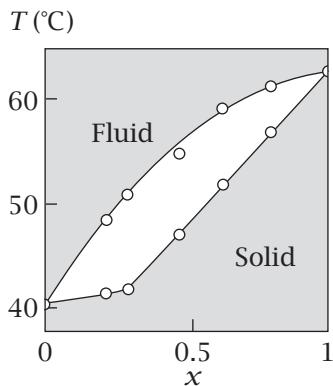


Figure 25.23 A fluid–solid phase diagram for lipid bilayer membranes containing two different phospholipids (DPPC–DPPE), showing that at some compositions *x* and temperatures *T*, a bilayer is disordered (fluid), at others it is ordered (solid), and at intermediate compositions and temperatures a bilayer has coexisting solid and liquid phases. Source: EJ Shimshick and HM McConnell, *Biochemistry* **12**, 2351–2360 (1973).

Summary

Phase transitions occur at the point where two or more states are stable at the same time. For example, when liquids *A* and *B* are mixed, an *A*-rich phase can be in equilibrium with a *B*-rich phase. Boiling is also an equilibrium between two phases: a vapor phase and a liquid phase. Changing the temperature or other variables can shift the equilibrium from the point of transition toward one phase or the other. The common feature is that the free energy as a function of a degree of freedom (or order parameter) has two minima at the phase transition points. Heating causes the two phases to become more similar until they ultimately merge at the *critical point*. The compositions of the stable phases

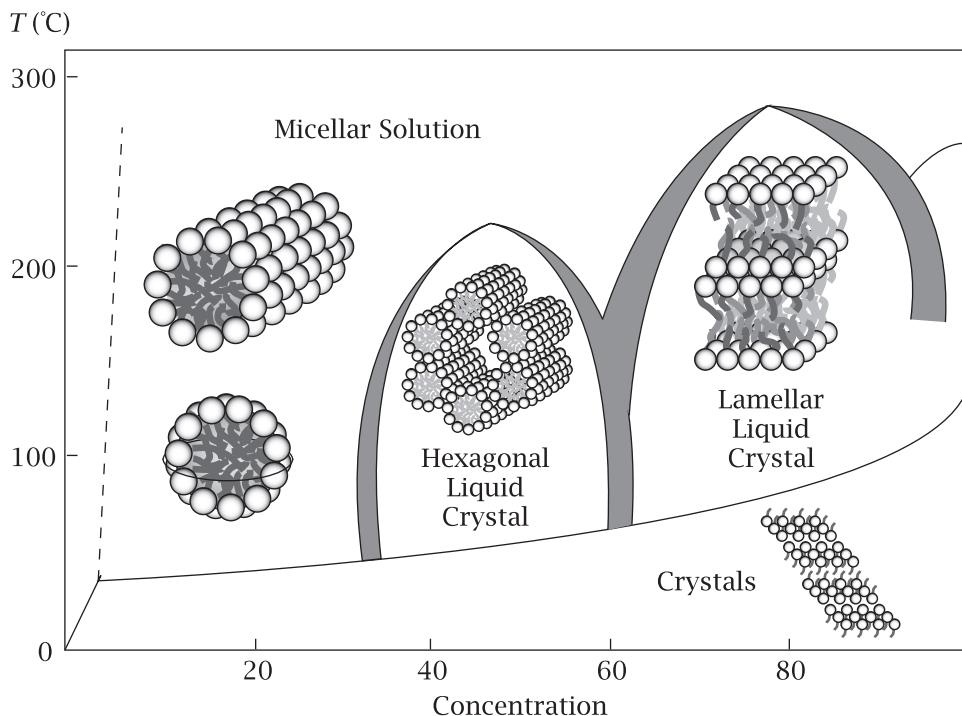


Figure 25.24 Phase diagram for surfactant molecules. Surfactants can form micelles that are spheres, cylinders, or planar bilayers, with increasing surfactant concentration. Critical micelle concentration (—). Source: HT Davis, *Statistical Mechanics of Phases, Interfaces, and Thin Films*, Wiley-VCH, New York, 1996.

are found by identifying the points of common tangency on the free energy surface—the ends of tie-lines on phase diagrams. The volumes of each phase are determined by using the lever rule. The common physical basis for such transitions is the gain in translational or mixing entropy at high temperatures, compared with the attractions that favor self-association at low temperatures. A virtue of models that are as simplified as the lattice model is that they show the underlying unity of such processes by leaving out other details.

Problems

1. Protein aggregation. You have a solution with mole fraction x of proteins. The proteins can aggregate and thus change their local concentration. The entropy of mixing is

$$\Delta S = -k[x \ln x + (1-x) \ln(1-x)]$$

and the internal energy is $\Delta U = \chi x(1-x)$. Draw curves for the free energy of the system as a function of x (between 0 and 1), for $\chi = 0, 1, 2$, and $4kT$. Identify the stable states of the system. For what values of χ does the system prefer to phase separate?

2. Closed-loop phase diagram. Figure 25.25 is a *closed-loop* phase diagram for mixing molecules A and B as a function of temperature T , and mole fraction composition of B , x_B . Suppose you mix the two liquids at temperature T_1 and composition x_B .

- (a) By drawing lines on the figure, determine the compositions x' and x'' of the two phases that will be in equilibrium under those conditions.
- (b) If there is 10 mL of solution, estimate the amounts of A-rich and B-rich components.

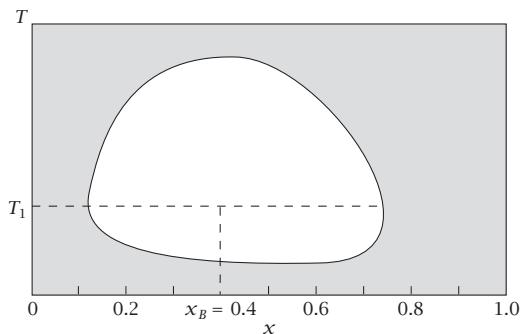


Figure 25.25 Closed-loop phase diagram. White region: two phases. Gray region: one phase.

3. Inverted mixing phase diagrams. In the free energy of mixing of a regular solution, the pair potential energies w are assumed to be constants independent of temperature. Therefore the exchange energy χ_{AB} depends inversely on temperature. Plot the miscibility phase diagram of temperature against mole fraction:

- (a) When χ_{AB} is a constant independent of temperature. (This occurs when local orientational entropy of at least one component is important.)
- (b) When χ_{AB} increases linearly with temperature. (This occurs for solutions with large heat capacities: nicotine in water shows this behavior.)

4. Phase diagram for uric acid in gout. Old Ma Kettle has gout. Gout is a disease of crystallization of uric acid in the blood. Figure 25.26 is the phase diagram.

- (a) What is the solubility limit according to this diagram for uric acid at 10°C? At 37°C?
- (b) Ma Kettle claims she can always tell when winter's here. How does she do it?

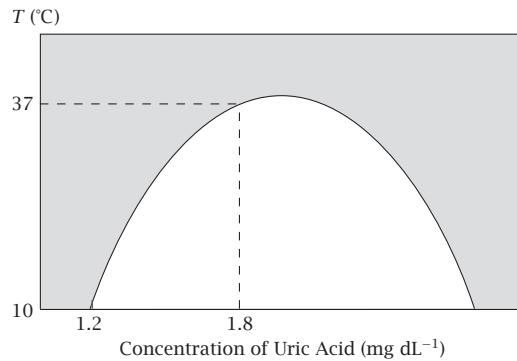


Figure 25.26 Phase diagram of uric acid. Gray region: one phase. White region: two phases.

5. Underlying free energy functions. Draw a diagram of free energy versus composition that would give rise to the phase diagram in Figure 25.27.

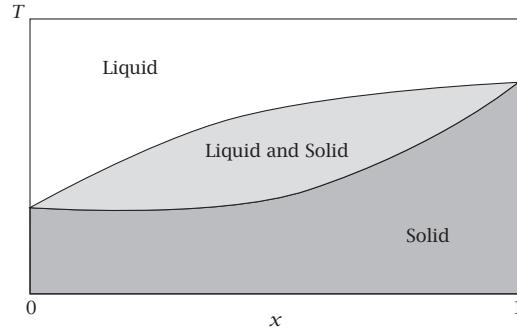


Figure 25.27 General phase diagram of a two-component freezing process.

6. Phase diagram for solder. The phase diagram in Figure 25.28 describes an alloy of tin (Sn) and lead (Pb) that can be melted at relatively low temperature.

- (a) If you wish to design solder to have the lowest possible melting temperature, what composition x_{Sn} would you choose?
- (b) Using graphical methods for the solid mixture (x_0, T_0), determine the compositions of the two phases x' and x'' into which the solid will partition.
- (c) Compute the relative weight percents of the two phases.

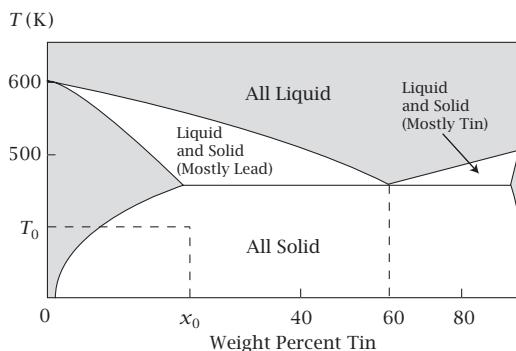


Figure 25.28 Phase diagram for tin and lead (solder).

7. Lattice model mixture. You have a solution with molecules of type A mixed with molecules of type B at $T = 300\text{ K}$. Suppose that every two AB molecular contacts are less stable than one AA plus one BB contact, by an energy kT , and that every molecule has an average of six neighbor contacts. For $x_A = x_B = 1/2$, to what temperature do you have to heat the system to get A to dissolve fully in B ?

8. Phase diagram for lattice model mixing. You have a lattice solution mixture at a temperature for which $\chi_{AB} = 5.0$.

- At what concentrations $x_b = x'$ and $x_t = x''$ are the binodal points?
- At what concentrations are the spinodal points?
- Over what concentration range of x_B is the solution metastable?
- If $\chi_{AB} = 5$ at temperature $T = 300\text{ K}$, what is the critical temperature T_c of the solution?

9. Boiling acetone-cyclohexane. Using the liquid-vapor binary phase diagram in Figure 25.29, describe what is qualitatively different about the boiling of a liquid that has $x = 0.4$ acetone in comparison with a liquid that has $x = 0.9$ acetone.

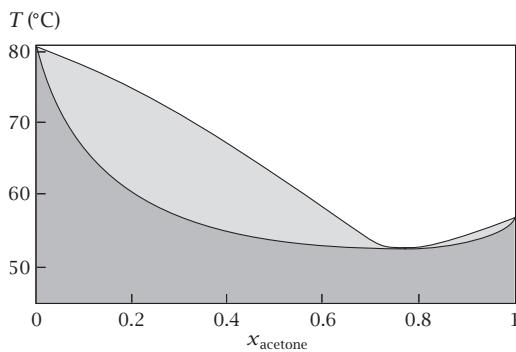


Figure 25.29 Liquid-vapor equilibria for mixtures of acetone in cyclohexane. Source: WE Acree Jr, *Thermodynamic Properties of Nonelectrolyte Solutions*, Academic Press, Orlando, 1984. From: KV Rao and CV Rao, *Chem Eng Sci* **7**, 97 (1957).

10. Azeotrope for alcohol/water mixtures. The azeotrope is the point on Figure 25.30 where $x = 0.105$ and $T = 78.2^\circ\text{C}$. Describe what the azeotrope point represents and what it means for distilling alcohol.

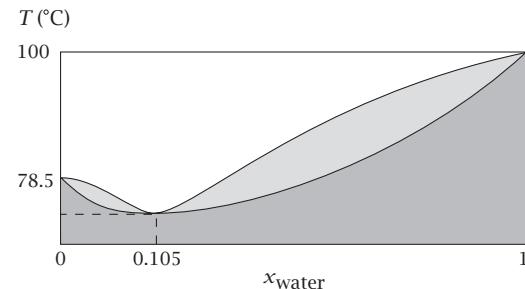


Figure 25.30 Phase diagram for boiling alcohol/water mixtures. Source: J deHeer, *Phenomenological Thermodynamics with Applications to Chemistry*, Prentice-Hall, Englewood Cliffs, NJ, 1986.

11. Boiling of a van der Waals fluid. The first region of the pressure-volume isotherm of a given van der Waals fluid extends from $V' = 38\text{ cm}^3\text{ mol}^{-1}$ to $V'' = 58\text{ cm}^3\text{ mol}^{-1}$. If this fluid is compressed to $41\text{ cm}^3\text{ mol}^{-1}$, is it a gas, or liquid, or mixture of both? If it is a mixture, what is its composition?

12. Pressure-induced protein denaturation. Figure 25.31 shows a (p, T) phase diagram for the denaturation of proteins (1 bar = 1 atm).

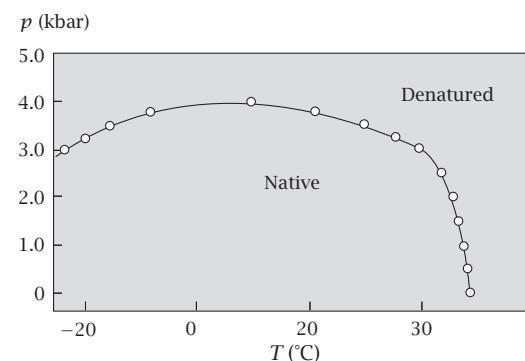


Figure 25.31 Pressure can denature proteins, ribonuclease A in this case. Source: J Zhang, X Peng, A Jones, and J Jones, *Biochemistry* **34**, 8631–8641 (1995).

- What is the denaturation temperature at 1 atm applied pressure?
- What pressure is needed to denature a protein at 25°C ?
- Does denaturation involve an increase or decrease of volume? (Hint: Use $(\partial \ln K / \partial p) = (-\Delta v / RT)$) (see

Equation 13.46), where K is the ratio of folded to unfolded protein molecules.) Explain what volume is changing.

- 13. Measuring solubilities to determine χ_{AB} .** The solubility of ethylbenzene in water is $0.00162 \text{ mol L}^{-1}$ at 25°C , and the solubility of propylbenzene in water at the same temperature is $0.000392 \text{ mol L}^{-1}$.

- Compute the χ interaction parameter for ethylbenzene in water.
- Compute the χ parameter for propylbenzene in water.
- Because propylbenzene has one more CH_2 group than ethylbenzene, compute the free energy of transferring the CH_2 group from hydrocarbon to water at 25°C by using the data given here.

- 14. Lower and upper critical solution temperatures.** Figure 25.32 shows three different liquid-liquid phase diagrams. Explain what they tell us about intermolecular interactions in these systems.

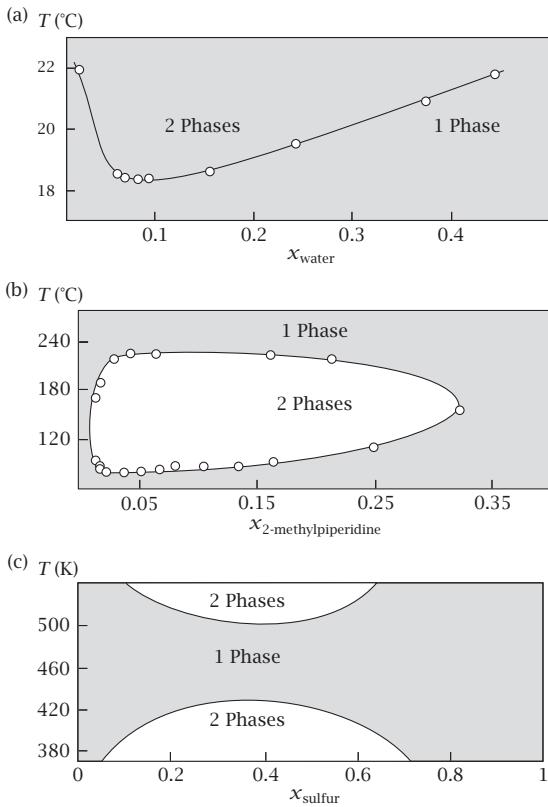


Figure 25.32 Phase diagram for mixtures: (a) water+triethylamine, (b) water+2-methylpiperidine, (c) sulfur+benzene. Source: WE Acree Jr, *Thermodynamic Properties of Nonelectrolyte Solutions*, Academic Press, Orlando, 1984. From: (a) F Kohler and OK Rice, *J Chem Phys* **26**, 1614 (1957); (b) O Flaschner and B Marewen, *J Chem Soc* **93**, 1000 (1908).

- 15. Critical points of water and CO_2 .** The van der Waals gas constants are $a = 3.592 \text{ L}^2 \text{ atm mol}^{-2}$ and $b = 0.043 \text{ L mol}^{-1}$ for CO_2 , and $a = 5.464 \text{ L}^2 \text{ atm mol}^{-2}$ and $b = 0.0305 \text{ L mol}^{-1}$ for H_2O . Compute the critical temperature T_c and pressure p_c for both CO_2 and water, and suggest why CO_2 is a more practical working fluid for supercritical fluid chromatography.

- 16. Solubilities of organic liquids.** From Table 25.1:

- Compute $\chi_{\text{benzene/water}}$.
- Describe the relative affinities for water of aliphatic hydrocarbons, aromatic hydrocarbons (benzenes), unsaturated hydrocarbons, and alcohols.

Table 25.1 Mole fraction of some common nonpolar organic liquids saturated with water.

Organic Liquid	x_0
Heptane ^a	0.99916
Octane ^a	0.99911
Benzene ^a	0.9977
Chlorobenzene ^b	0.9975
Trichloroethylene ^b	0.9977
Tetrachloroethylene ^b	0.99913
Chloroform ^b	0.9946
1,1,1-Trichloroethane ^b	0.9974
Diethyl ether ^c	0.942
Methyl acetate ^c	0.89
2-Butanone ^c	0.69
Pentanol ^d	0.64
Octanol ^d	0.79

^aW Gerrard, *Gas Solubilities, Widespread Applications*, Pergamon, New York, 1980.

^bAL Horvath, *Halogenated Hydrocarbons. Solubility-Miscibility with Water*, Marcel Dekker, New York, 1982. ^cJA Riddick and WB Bunker, *Organic Solvents*, Wiley, New York, 1970. ^dR Stephenson, J Stuart, and M Tabak, *J Chem Eng Data* **29**, 287-290 (1984).

Source: RP Schwartzenbach, PM Gschwend, and DN Imoboden, *Environmental Organic Chemistry: Illustrative Examples, Problems, and Case Studies*, Wiley, New York, 1995.

- 17. Lattice model spinodal curve.** Using the lattice binary solution theory:

- Express the *spinodal* curve concentration x' as a function of temperature T , where $\chi_{AB} = \theta/T$, and θ is a constant.
- Plot this spinodal function $x'(T)$.
- What is x' when $T = \theta/4$?

- 18. Liquid-vapor instability in a van der Waals fluid.**

Figure 25.33 is a plot of pressure versus volume for a van der Waals fluid at a temperature below the critical temperature. What region of this plot is nonphysical? How could you use this to predict a liquid-vapor phase transition?

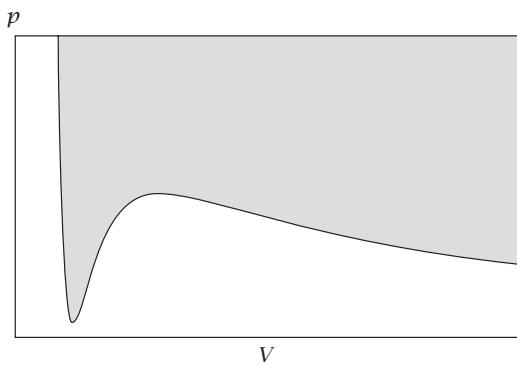


Figure 25.33

19. Applying the lever rule. You have a mixture of molecules of *A* and *B* with the phase diagram shown in Figure 25.34, at temperature T_0 . The mole-fraction composition of molecules of *B* in the container is $x_0 = 0.5$. From the phase diagram, you are given $x' = 0.10$ and $x'' = 0.70$.

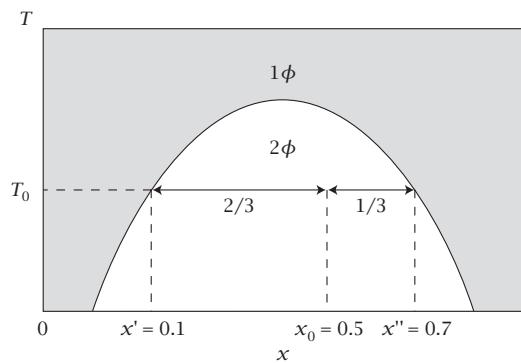


Figure 25.34 You are given $x_0 = 0.5$, $x' = 0.10$, and $x'' = 0.70$.

- What is the fraction of the *B*-rich phase that are molecules of *B*?
- What is the fraction of the *A*-rich phase that are molecules of *A*?
- What fraction of all the molecules are in the *B*-rich phase?

Suggested Reading

There are several outstanding general texts on phase equilibria:

RS Berry, SA Rice, and J Ross, *Physical Chemistry*, 2nd edition, Oxford University Press, New York, 2000.

HB Callen, *Thermodynamics and an Introduction to Thermostatics*, 2nd edition, Wiley, New York, 1985.

G Careri, *Order and Disorder in Matter*. Benjamin-Cummings, Reading, MA, 1984.

PG Debenedetti, *Metastable Liquids: Concepts and Principles*, Princeton University Press, Princeton, NJ, 1966.

JH Hildebrand and RL Scott, *The Solubility of Nonelectrolytes*, 3rd edition, Reinhold, New York, 1950.

R Kubo, with H Ichimura, T Uzui, and N Itashitsumo, *Statistical Mechanics*, North-Holland, New York, 1965.

JS Rowlinson and FL Swinton, *Liquids and Liquid Mixtures*, 3rd edition, Butterworth Scientific, Boston, 1982.

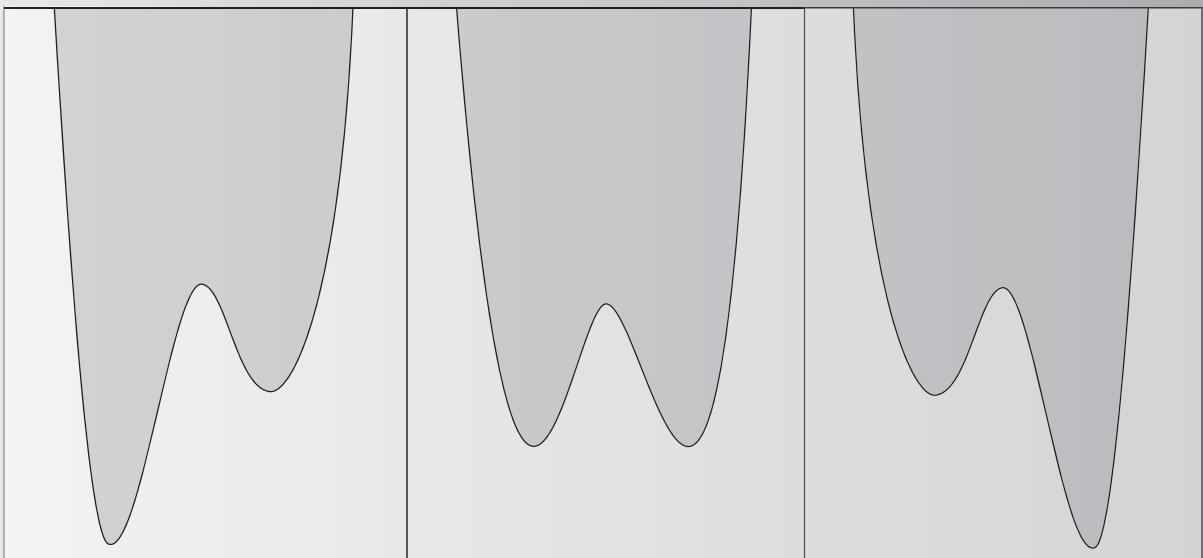
EKH Salje, *Phase Transitions in Ferroelastic and Co-elastic Crystals*, Cambridge University Press, Cambridge, 1990.

HE Stanley, *Introduction to Phase Transitions and Critical Phenomena*, Oxford University Press, New York, 1971.

JM Yeomans, *Statistical Mechanics of Phase Transitions*, Oxford University Press, Oxford, 1992.

This page is intentionally left blank.

26 Cooperativity: the Helix-Coil, Ising, & Landau Models



Abrupt Transitions Are Common in Nature

The principles that govern phase transitions in macroscopic systems also underlie the dramatic transformations that take place within individual molecules. Small changes in the temperature or the solvent can cause a polymer molecule to undergo a helix-coil transition, cause the two strands of DNA to zip together, or cause a protein or RNA molecule to fold. A small change in oxygen concentration in the blood can cause a large change in how much oxygen is picked up or released by hemoglobin. Small changes in a cell can cause sickle-cell hemoglobin to polymerize into rods, disrupting blood cells and causing anemia. In all cases—macroscopic phase transitions or single-molecule conformational changes—the cooperativity is predicted from the shape of a *free energy landscape*. We first describe the history of the revolution that led to our current models of cooperativity.

Transitions and Critical Points Are Universal

In Chapter 25, we noted that the shapes of the coexistence curves for liquid-liquid immiscibility are about the same as for boiling. Both display critical points, and both are described approximately by the same simple model. Advances in theory and experiments since the 1940s led to a revolution in understanding critical phenomena.

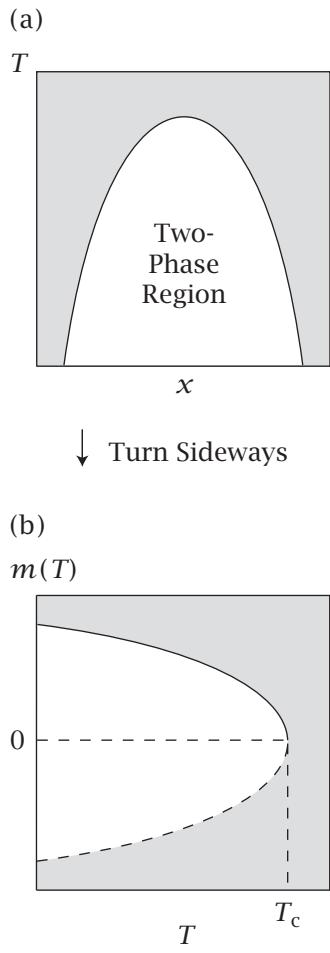


Figure 26.1 (a) A coexistence curve for solubilization is temperature versus composition, $T(x)$. (b) Turn it sideways to get $x(T)$. The order parameter in this case is $m(T) = 2x(T) - 1$.

This revolution started with two important developments. In 1945, EA Guggenheim demonstrated that the ‘corresponding states’ liquid-gas coexistence curve (Figure 25.20) was slightly flatter than you would expect from the van der Waals and lattice models of Chapter 25. And in 1944, L Onsager derived an exact solution to the two-dimensional version of a better model, called the *Ising model*. The Ising model treats nearest-neighbor interactions, and captures the shape of the coexistence curves more accurately. During the 1960s and 1970s, a highly successful theory of critical phenomena emerged, called renormalization group theory [1]. Remarkably, renormalization group theory shows that the flaws in the simpler models do not lie in their structural simplifications—the use of lattices or beads to represent atoms and molecules, or the use of simplified energies. Rather, the flaws in simpler models are due to their mean-field approximations and their neglect of large fluctuations near critical points. Renormalization group theory also shows that behavior near critical points can be described by *universality classes*: many different types of transition share the same behaviors, depending mainly on an order parameter and the dimensionality of the system.

How can you mathematically express the shape of a coexistence curve near the critical point? First, you need to define the *order parameter* of the system. An order parameter is a quantity m on which the free energy depends. The order parameter m is zero above the critical temperature, indicating that the system is disordered or randomly mixed. m is nonzero below the critical temperature, indicating that the system is ordered or phase-separated in some way. For liquid-liquid immiscibility, a good choice of order parameter is the difference in phase compositions $m = x'' - x'$. At the critical temperature, $x' = x'' = 0.5$, so the quantity $m(T) = 2x'' - 1$ is a conveniently normalized order parameter: it equals 1 at low temperature and 0 at the critical point $T = T_c$. At low temperatures, $m \rightarrow 1$ indicates that the two coexisting phases are very different. As $T \rightarrow T_c$, $m \rightarrow 0$ indicates that the two states of the system become indistinguishable from each other. To see the function $m(T)$ for liquid immiscibility, turn the phase diagram on its side (see Figure 26.1(a)) to get $x(T)$ rather than $T(x)$, then multiply by 2 and subtract 1: $m(T) = 2x(T) - 1$ (see Figure 26.1(b)). For boiling, an appropriate order parameter is the difference in density between the liquid and the gas phases (see Figure 26.2). A function with a similar shape near the critical point is the enthalpy of vaporization as a function of temperature (see Figure 26.3).

Metal alloys provide another example. Brass is an alloy made from copper and zinc. Some brass has two interpenetrating body-centered cubic lattices, one lattice containing the copper atoms and the other the zinc atoms. Figure 26.4(a) shows a two-dimensional representation of the low-temperature structure, where the system is perfectly ordered. Figure 26.4(b) shows a higher-temperature structure, where the alloy is more disordered.

The disordering with temperature is called an order-disorder transition, and the process is called *symmetry-breaking*. The order parameter is the fraction of atoms located on correct lattice sites minus the fraction located on incorrect sites. The order parameter ranges from zero in the fully disordered state to one in the fully ordered state. The mixing of metals in alloys can be treated with the lattice model of previous chapters. Brass with A’s and B’s alternating throughout the structure is a stable state of a system that has a negative value

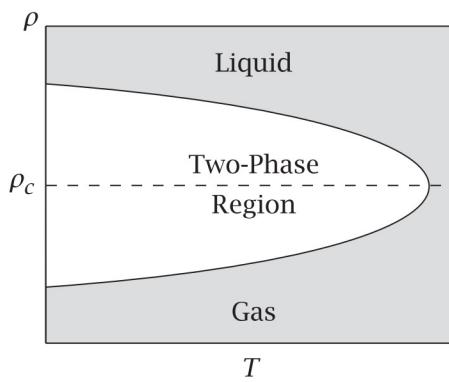


Figure 26.2 For boiling, a useful order parameter is the difference between liquid and gas densities $\Delta\rho = \rho_{\text{liquid}} - \rho_{\text{gas}}$. ρ_c is the density of the fluid at the critical point.

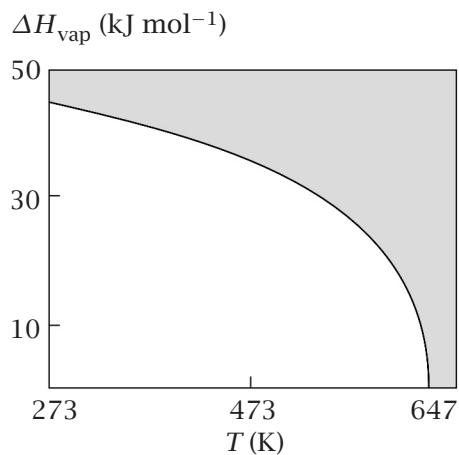
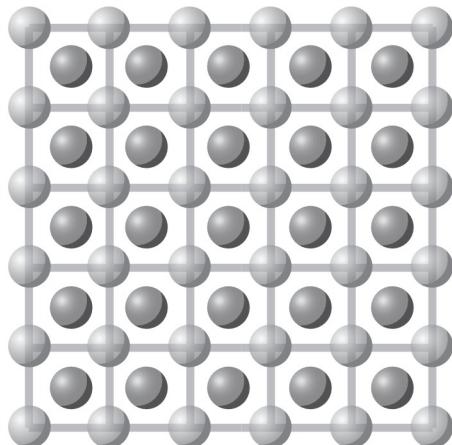


Figure 26.3 The enthalpy of vaporization of water as a function of temperature has a functional form similar to the order parameter, becoming zero at the critical point.
Source: JN Murrell and AD Jenkins, *Properties of Liquids and Solutions*, Wiley, New York, 1982.

(a) Ordered



(b) Disordered

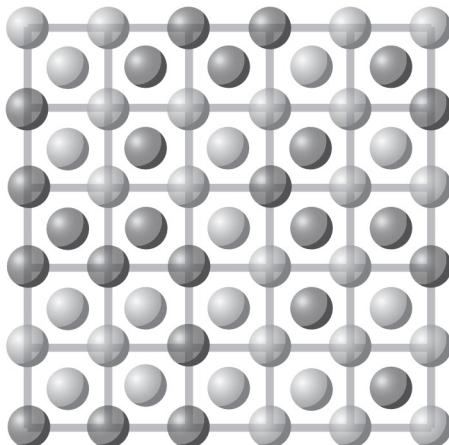


Figure 26.4 Ordering in metal alloys. Two different atom types are interspersed: (a) ordered at low temperature and (b) disordered at high temperature. Source: G Careri, *Order and Disorder in Matter*, Benjamin-Cummings, Reading, MA, 1984.

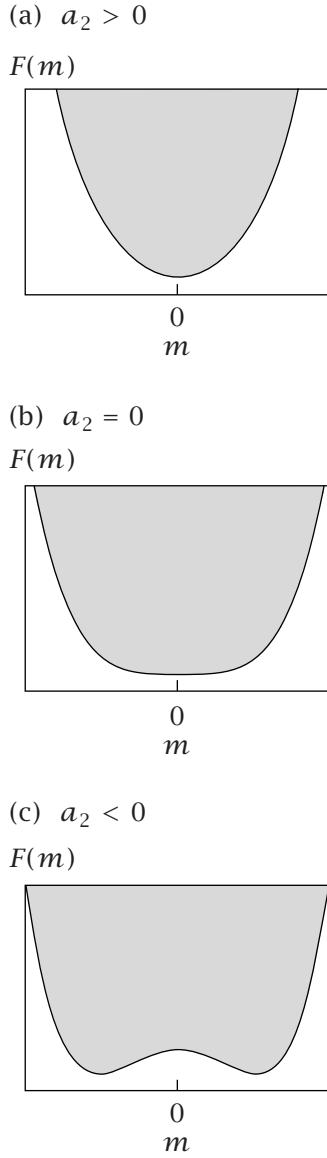
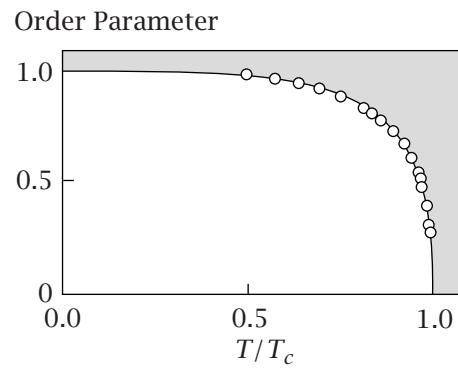


Figure 26.6 (a) The Landau model of one stable state. (b) The Landau model of a critical point. The free energy is flat and broad at the minimum. This implies large fluctuations in m at equilibrium. (c) The Landau model of two stable states.

Figure 26.5 The dependence of the order parameter of β -brass on temperature T . The points are neutron-scattering results, and the line is the theoretical result for a compressible Ising model. T_c is the critical temperature. Source: J Als-Nielsen, *Phase Transitions and Critical Phenomena*, Volume 5A, C Domb and MS Green, eds, Academic Press, London, 1976, pages 87–164.



of the mixing parameter, $\chi_{AB} < 0$, indicating that AB interactions are more favorable than AA and BB interactions. This contrasts with liquid solutions, for which χ_{AB} is usually positive. Figure 26.5 shows how the order parameter changes with temperature for brass; $m(T)$ for this solid resembles $m(T)$ for liquid–liquid and boiling equilibria. Many types of system have transitions and critical points with similar shapes for $m(T)$.

To predict the experiments, we want to compute the mathematical form of $m(T)$ near the critical temperature $T \approx T_c$, where $m(T) \rightarrow 0$. We could get this function from the lattice or van der Waals gas models, but instead we will find the same result from a model that is simpler and more general, called the *Landau model*.

The Landau Model Is a General Description of Phase Transitions and Critical Exponents

We now describe the Landau model, named for the Russian physicist LD Landau (1908–1968), who won the 1962 Nobel Prize in Physics for his work on condensed phases of matter. The Landau model is a generic treatment of phase transitions and critical points. It is based on the idea that coexistence curves are simply mathematical functions that have two minima at low temperatures, merging into a single minimum at higher temperatures. You can capture that mathematically by expressing the free energy $F(m)$ as a polynomial function,

$$F(m) = F(0) + a_1 m + a_2 m^2 + a_3 m^3 + a_4 m^4 + \dots, \quad (26.1)$$

of an order parameter m . The a_i 's are coefficients that depend on the physical problem at hand. For the simplest problems, we want this function to be symmetrical around $m = 0$, so we eliminate the terms involving odd powers. To allow for the possibilities of either one or two minima, we must keep terms at least up to fourth order. So, a minimal model is

$$F(m) = F(0) + a_2 m^2 + a_4 m^4. \quad (26.2)$$

Figure 26.6 shows that this function gives the shapes we want. In the figure, we fix $a_4 > 0$. We then choose a_2 to be negative, zero, or positive. $F(m)$ has a single minimum when $a_2 > 0$ (Figure 26.6(a)). $F(m)$ has a broad flat minimum, implying large fluctuations in m , when $a_2 = 0$ (Figure 26.6(b)). $F(m)$ has two

minima, indicating two stable states, when $a_2 < 0$ (Figure 26.6(c)). Lowering a_2 affects the polynomial function $F(m)$ in the same way that reducing the temperature affects phase equilibria free energy functions. At high temperatures, $T > T_c$, the free energy has a single minimum; at the critical temperature $T = T_c$, the free energy minimum is broad; and below the critical temperature, $T < T_c$, the free energy function has two minima.

A phase transition is called *first-order* when the free energy function has two minima separated by a hill (Figure 25.8), and is called *higher-order* when the transition involves a broad flat minimum without a hill, as in Figure 26.6(b).

To capture the correspondence between a_2 and temperature, let t represent the fractional deviation of the temperature T away from the critical temperature T_c :

$$t = \frac{T - T_c}{T_c}. \quad (26.3)$$

At $T = T_c$, this dimensionless temperature $t = 0$. If you take a_2 to be proportional to t ,

$$a_2 = at, \quad (26.4)$$

and choose a to be a positive constant, then the function $F(m)$ will have one narrow minimum when $t > 0$ (a single phase above the critical point), a broad minimum when $t = 0$ (at the critical point), and two minima when $t < 0$ (two phases in equilibrium below the critical point). Equation (26.2) becomes

$$F(m) = F(0) + atm^2 + a_4m^4. \quad (26.5)$$

To determine how the equilibrium value of the order parameter m^* depends on T near T_c , find the value $m = m^*$ that minimizes $F(m)$,

$$\left(\frac{dF}{dm}\right)_{m=m^*} = 2at(m^*) + 4a_4(m^*)^3 = 0. \quad (26.6)$$

Solving Equation (26.6) for m^* gives

$$(m^*)^2 = -\frac{at}{2a_4}. \quad (26.7)$$

The main result is the dependence of the equilibrium value of the order parameter m^* on temperature t . Equation (26.7) gives a *power law* ($f(x) \propto x^{\text{constant}}$):

$$m^* = \text{constant} \times t^\beta, \quad (26.8)$$

where β is called the *critical exponent*. Equation (26.7) shows that the critical exponent in the Landau model is $\beta = 1/2$.

In general, independently of the model, the critical exponent λ for a function $g(x)$ is

$$\lambda = \lim_{x \rightarrow 0} \frac{\ln|g(x)|}{\ln|x|}. \quad (26.9)$$

Table 26.1 Critical exponents α , β , γ , ν , and η for boiling. The correlation length ξ gives roughly the sizes of droplets.

Property	Critical Exponent
Specific heat at constant volume	$c_V \propto t ^{-\alpha}$
Liquid-gas density difference	$(\rho_\ell - \rho_g) \propto (-t)^\beta$
Isothermal compressibility	$\kappa \propto t ^{-\gamma}$
Correlation length	$\xi \propto t ^{-\nu}$
Pair correlation function at T_c	$g(r) \propto 1/r^{1+\eta}$

Source: JM Yeomans, *Statistical Mechanics of Phase Transitions*, Oxford University Press, Oxford, 1992.

EXAMPLE 26.1 Finding a critical exponent. What is the critical exponent of the free energy function $g(t) = t^2 + t^{1/2}$? Equation (26.9) gives

$$\lambda = \lim_{t \rightarrow 0} \left[\frac{\ln(t^2 + t^{1/2})}{\ln t} \right] \approx \frac{\ln(t^{1/2})}{\ln t} = \frac{1}{2}.$$

This approximation applies because $t^{1/2}$ becomes much larger than t^2 as $t \rightarrow 0$.

Different physical properties of a system, which are different functions of the free energy, have different critical exponents (see Table 26.1). Because physical properties are often related through thermodynamic relationships, there are algebraic relationships among critical exponents [1-3].

The importance of critical exponents is their *universality*. While the *critical temperature* depends on the details of the interatomic interactions, the *critical exponent* does not. For example, $\beta = 1/2$ arises in the Landau model just from the general notion that two minima in the free energy function merge to a single minimum as the temperature approaches the critical point, irrespective of whether the transition involves boiling, mixing, metal alloy order-disorder, or magnetization, and irrespective of microscopic parameters such as χ_{AB} . It can be shown [4] that the van der Waals and lattice mean-field models also give a critical exponent of $\beta = 1/2$. Guggenheim's data in Figure 25.20 show that eight different materials have identical coexistence curves. The shapes of these coexistence curves do not depend on the types of atoms that are interacting, or on their interaction energies.

The revolution in understanding critical phenomena began with Guggenheim's experiments. The critical exponent for those data and others is found to be $\beta \approx 1/3$ (see Figure 26.7), in disagreement with the value $\beta = 1/2$ predicted by the Landau model [4]. The discrepancy is evidently not in the neglect of atomic detail, because the coexistence curves for the eight different materials superimpose upon each other. Rather, the problem is that very near the critical point, higher-order terms beyond m^4 are needed. The true free energy surfaces are flatter near critical points—so the fluctuations are broader—than would be predicted by mean-field or Landau models. A type of model in which $F(m)$ can be expressed exactly, without mean-field approximations or low-order expansions in m , is the *Ising model*, also called the *nearest-neighbor model*. It handles critical behavior more accurately. It was originally derived to treat the magnetization of metals.

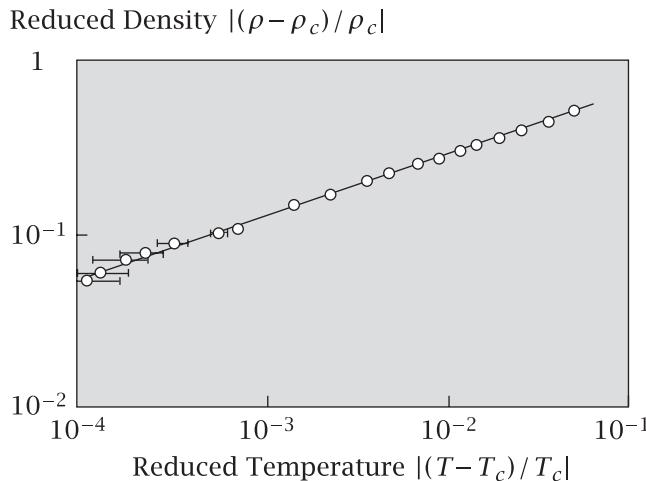


Figure 26.7 Measurements of the helium coexistence curve near the critical point, on a log–log scale, indicating that the critical exponent is $\beta = 0.354$. Source: HE Stanley, *Introduction to Phase Transitions and Critical Phenomena*, Oxford University Press, New York, 1971. Data are from PR Roach, *Phys Rev* **170**, 287 (1968).

The Ising Model Describes Magnetization

You can magnetize certain materials, such as iron, by putting them into a magnetic field. Below a certain critical temperature, such materials have a net magnetic moment. They are magnetized. But if you heat them above that temperature, they lose magnetization. What is the microscopic basis for this behavior?

Atomic spins act like small magnets: they are magnetic dipoles with north and south poles that become aligned in applied magnetic fields. At high temperatures, the spins are randomly oriented (see Figure 26.8(b) and (c)), but below a critical temperature $T = T_c$, the spins align with each other (Figure 26.8(a)).

The net magnetic moment in the up direction is represented by the number of upward arrows (\uparrow) minus the number of downward arrows (\downarrow).

Magnetic alignment is cooperative because of *nearest-neighbor* interactions. At low temperatures, the spins in ferromagnets line up to satisfy the favorable nearest-neighbor interactions; but at high temperatures, the arrow directions randomize to maximize the entropy. In a *ferromagnet*, such as iron, certain iron oxides, or cerium antimonide, each spin prefers to orient in the same direction as its neighbors. In *antiferromagnets*, like MnO, each spin prefers to orient in the direction opposite to its neighbors.

In the *Ising model*, each atomic magnet occupies one lattice site with a spin that is either up or down. This model is named for physicist E Ising, who solved it in his PhD thesis work with W Lenz in 1925. If you model each spin as an \uparrow or \downarrow on a one-dimensional lattice, the partition function is simple to compute. The Boltzmann weights are $q(\uparrow\uparrow) = q(\downarrow\downarrow) = e^J$ and $q(\uparrow\downarrow) = q(\downarrow\uparrow) = e^{-J}$, where $J = J_0/kT$ is a dimensionless energy. Ferromagnets are modeled by $J > 0$ and antiferromagnets by $J < 0$.

For a single site, the partition function is $Q_1 = 2$, because there are two states: \uparrow or \downarrow . To get the two-site partition function, you multiply Q_1 by $(e^J + e^{-J})$, because the second arrow can point either in the same direction as the first or in the opposite direction. In terms of the hyperbolic cosine function, $\cosh x = (e^x + e^{-x})/2$, you have

$$Q_2 = Q_1(e^J + e^{-J}) = Q_1(2 \cosh J). \quad (26.10)$$

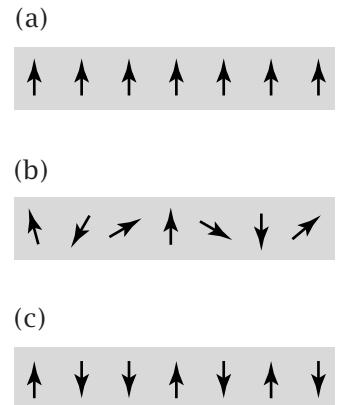


Figure 26.8 Magnetic spins in a one-dimensional system: (a) fully aligned, for example at low temperatures; (b) disordered, at high temperatures; (c) a simple ‘binary’ model of the disordered states in which spins are either up or down.

Every site that you add to the one-dimensional array of spins multiplies the partition function by a factor of $(e^J + e^{-J})$, so the partition function for a linear lattice of N magnetic spins is

$$Q_N = Q_1 (2 \cosh J)^{N-1} = 2^N \cosh^{N-1} J. \quad (26.11)$$

The thermodynamic properties of the one-dimensional Ising model are computed from the partition function in the standard way (see Table 10.1). Rather than elaborating further here on spins and magnets, we instead explore the Ising model in more detail below as it applies to helix-coil transitions.

Real magnets are three-dimensional. The importance of the one-dimensional Ising model is that it illustrates the principles of the nearest-neighbor model and that it can be solved exactly. However, computing the properties of the Ising model in two or three dimensions is much more challenging. The three-dimensional Ising model must be solved by computer simulations. The two-dimensional Ising model has an analytical solution that was found by the Norwegian-American physicist L Onsager (1903–1976; Nobel Prize in Chemistry, 1968) through a mathematical *tour de force* in 1944. It serves as an exact standard with which other approximations are often compared. Figure 26.9 shows a plot of the predicted order parameter for the two-dimensional Ising model versus temperature, compared with the Bragg–Williams approximation, Equation (15.11), that we used in the lattice model in Chapter 25. Near the critical point, these two models give results that are quite different.

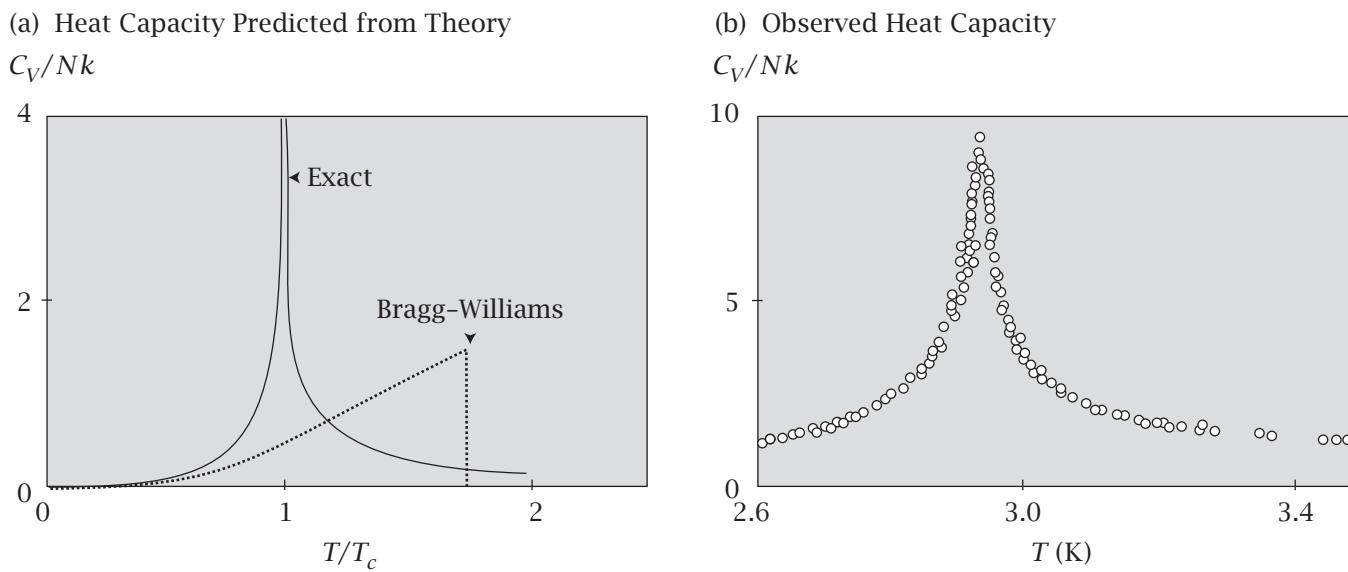


Figure 26.9 (a) Theoretical predictions for the heat capacity C_V near the critical temperature for two-dimensional systems. The Bragg–Williams mean-field lattice model of Chapter 25 leads to a triangular function, while the exact solution of the two-dimensional Ising model shows a sharp peak. Source: R Kubo, with H Ichimura, T Usui, and N Hashitsome, *Statistical Mechanics*, New York, 1965. (b) Experimental data for helium on graphite closely resembles the Ising model prediction. Source: RE Ecke and JG Dash, *Phys Rev B* **28**, 3738–3752 (1983).

The nearest-neighbor model has become an important model for cooperative processes throughout physics, chemistry, and biology. Below, we apply it to helix-coil transitions in polymers and biomolecules.

Polymers Undergo Helix-Coil Transitions

Figure 26.10 illustrates a transformation that occurs in some polymers, including RNA, DNA, and polypeptides. In aqueous solution, at high temperatures, the chain has a large ensemble of disordered conformations. This ensemble is collectively called the *coil state*. When the temperature is lowered, each molecule undergoes a transition to a helix (see Figure 26.10). These transitions are called *cooperative*, which is sometimes taken to mean the following. (1) Some property (the amount of helix, in this case) changes in a sigmoidal way as a function of an external variable (such as temperature or solvent). (2) The sigmoidal curve steepens with increasing size of the system (the chain length, in this case). Interestingly, some of the most important protein helix-coil cooperativity data (see Figure 26.11) were taken in a solvent (80% dichloroacetic acid, 20% 1,2-dichloroethane) that inverts the transition: increasing the temperature increases the helix content. Cooperativity in either type of solvent is treated by the same models.

The simplest models of this cooperative behavior suppose that each monomeric unit along the chain has one of two possible states: *H* (helix) or *C* (coil). The chain conformations are represented as one-dimensional sequences of *H* and *C* units. An example conformation is *HHCCCCCHCHHHHC*.

Our aim is to compute the partition function, based on the combinatorics for counting the numbers of arrangements of *H* and *C* units in the chain. We will consider three different models for the partition function. To show the nature of cooperativity, it is useful to start with a model that has no cooperativity.

Noncooperative Model: Neighboring Units Are Independent of Each Other

First, consider a model that has no cooperativity. The chain has N monomer units. Assume that the state (either *H* or *C*) of each monomer $j = 1, 2, 3, \dots, N$

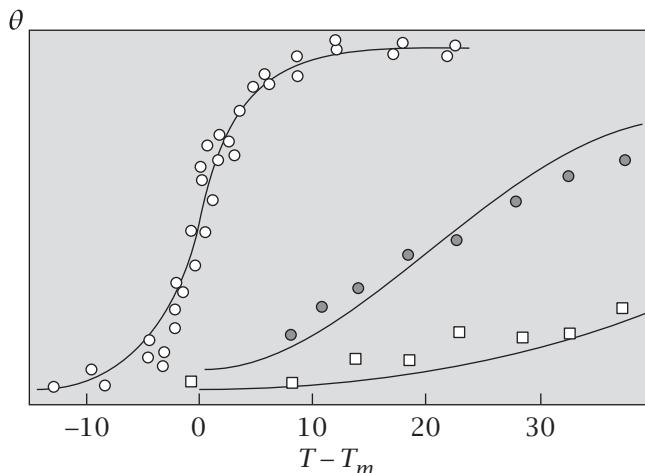


Figure 26.11 The helix–coil transition of poly- γ -benzyl-L glutamate chains: (□) 26-mers, (○) 46-mers, (○) 1500-mers. Fractional helicity $f_H = \theta$ versus temperature T , relative to the midpoint T_m for the long chains, from optical-rotation data. Curves are calculated from Zimm–Bragg theory by using $\sigma = 2 \times 10^{-4}$ and $\Delta h = 890 \text{ cal mol}^{-1}$, where $s = \exp(-\Delta h/kT)$. Source: CR Cantor and PR Schimmel, *Biophysical Chemistry*, WH Freeman, San Francisco, 1980. After BH Zimm, P Doty, and K Iso, *Proc Natl Acad Sci USA* **45**, 1601–1607 (1959).

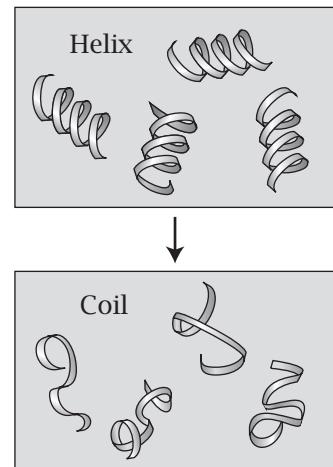


Figure 26.10 Some polymers undergo a helix–coil transition. In the single-helical conformation, the polymer molecule is ordered. The coil state is a collection of disordered conformations.

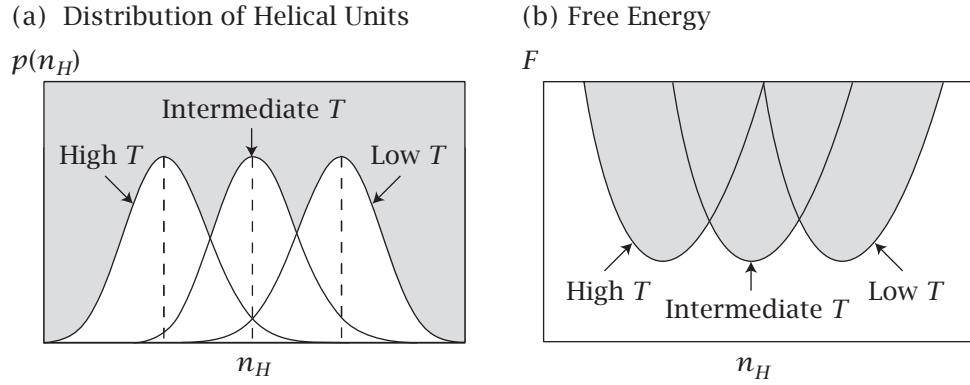


Figure 26.12 For the noncooperative model, lowering the temperature increases the helicity gradually, not sharply. n_H is the number of helical units.

in the chain is independent of the state of every other monomer. Let q_C be the partition function for each C unit. Let q_H be the partition function for each H unit. Because we are interested only in relative populations among different states, you are free to choose $q_C = 1$, which is equivalent to defining the free energy of a coil unit as zero. Let $q_H = \exp(-\beta\Delta\varepsilon)$, where $\Delta\varepsilon < 0$ represents the energy of forming a helical unit relative to a coil unit, and $\beta = 1/kT$. The ratio

$$s = \frac{q_H}{q_C} = e^{-\beta\Delta\varepsilon} \quad (26.12)$$

defines the relative probabilities of H and C states for a unit of the chain and represents its helix-coil equilibrium constant. The probability $p(n_H)$ that an N -mer chain has n_H of its units in the helical state (and $n_C = N - n_H$ in the coil state) is given by the statistics of independent coin flips (think of N independent flips of a coin labeled H and C):

$$p(n_H) = \frac{N!}{n_H!n_C!} q_H^{n_H} q_C^{n_C} = \frac{N!}{n_H!(N-n_H)!} s^{n_H}, \quad (26.13)$$

since $q_C = 1$. The partition function Q is the sum over all the possible sequences of H and C :

$$Q = \sum_{n_H=0}^N p(n_H) = (q_C + q_H)^N = q_C^N (1+s)^N = (1+s)^N. \quad (26.14)$$

This model predicts a gradual transition between coil and helical states, not a sharp transition. This is illustrated in Figure 26.12(a), which plots the probability $p(n_H)$ versus n_H from Equation (26.13) for three different values of s , corresponding to three different temperatures. (Equation (26.12) shows that $s(T)$ increases as T decreases.) Or, to convert from the probability distributions to the energy landscapes, you can compute the free energy function $F(n_H) = -kT \ln p(n_H)$, shown in Figure 26.12(b). The figure illustrates that the population distribution shifts smoothly as the temperature is changed.

Now compute $f_H = \langle n_H \rangle / N$, the equilibrium fraction of the N monomers that are H 's, as a function of the chain length N and the temperature T . To calculate $f_H(T) = \langle n_H \rangle / N$, you could compute the free energy for each T , find

the value of $n = \langle n_H \rangle$ that causes F to be at a minimum, and then plot $\langle n_H \rangle$ as a function of T . An easier way is to take the derivative of Equation (26.14), as indicated by Equation (C.5) in Appendix C. You have

$$\left(\frac{dQ}{ds} \right) = N(1+s)^{N-1},$$

which leads to

$$\langle n_H \rangle = \frac{s}{Q} \left(\frac{dQ}{ds} \right) = \frac{Ns(1+s)^{N-1}}{(1+s)^N} = \frac{Ns}{1+s}.$$

So the average helicity is

$$f_H = \frac{\langle n_H \rangle}{N} = \frac{s}{1+s}, \quad (26.15)$$

which depends on temperature through the definition of s in Equation (26.12).

The noncooperative model predicts that the helicity changes gradually, not sharply, with temperature. Cooperativity requires some interdependence of the interactions among the monomers. Equation (26.15) fails to predict a sufficiently sharp dependence of f_H on chain length that is indicated by experiments (in Figure 26.11).

Gradual noncooperative changes are characterized by free energy functions in which a single minimum $m^*(T)$ shifts with T or with some other external variable (see Figure 26.12). Here's a different model that defines the opposite extreme: maximum cooperativity.

The Two-State Model: Maximum Cooperativity

Now we consider a model that describes the highest possible degree of cooperativity: the two-state model. Suppose we model the helix-coil system as having only two possible states: $CCCC\dots C$ or $HHH\dots H$. Assume that all other states have zero probability. Maximum cooperativity means that if one monomer is H , all are H ; if one monomer is C , all are C . Now the partition function for each of the two states is q_C^N or q_H^N , so the partition function for the system is

$$Q = q_C^N + q_H^N = q_C^N(1+s^N) = 1+s^N. \quad (26.16)$$

In this case, the populations of the only two states of the system (all C and all H) are

$$p_{\text{all } C} = \frac{1}{1+s^N}$$

and

$$p_{\text{all } H} = \frac{s^N}{1+s^N}.$$

Figure 26.13(b) shows two population peaks, separated by a trough. At low temperatures, all molecules are fully helical. At high temperatures, all molecules are in coil conformations. At the midpoint of the transition, the average helicity is $1/2$, not because the individual molecules are half helix and half

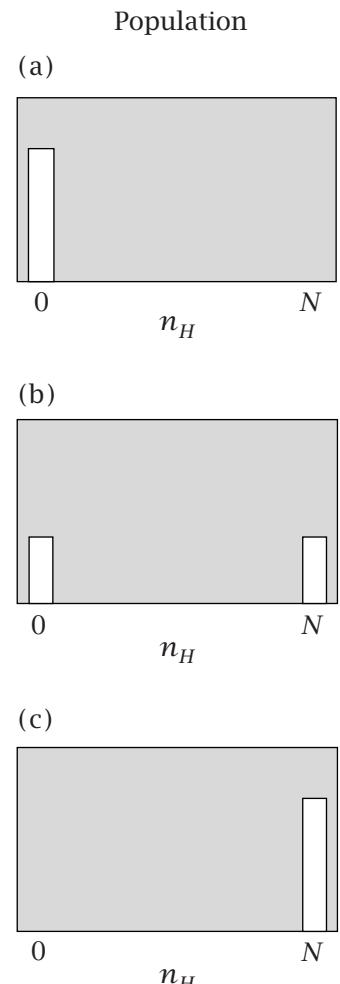


Figure 26.13 The two-state model: populations are helix ($n_H = N$) and coil ($n_H = 0$), and there are no ‘intermediate’ states. (a) $s < 1$ (high temperature); (b) $s = 1$; (c) $s > 1$ (low temperature).

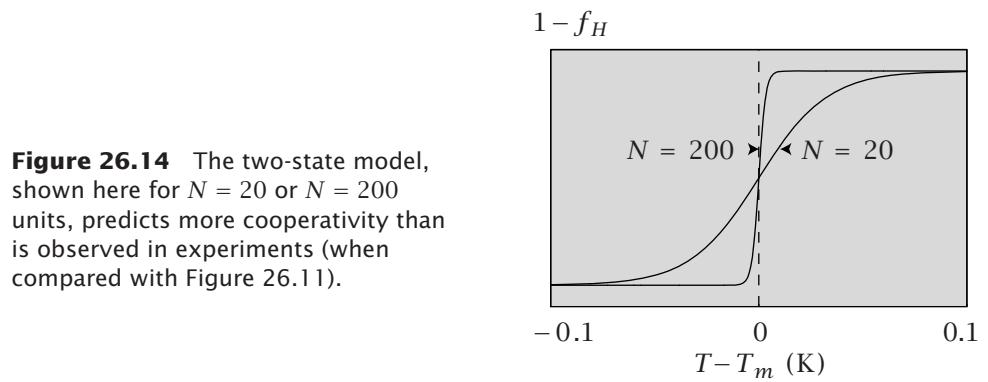


Figure 26.14 The two-state model, shown here for $N = 20$ or $N = 200$ units, predicts more cooperativity than is observed in experiments (when compared with Figure 26.11).

coil, but because half of the molecules are all-helix, and half of the molecules are all-coil. In the two-state model, no molecule is in an *intermediate* state. In this regard, the two-state model differs markedly from the noncooperative model, which predicts substantial populations of intermediate states.

Get the average helicity for the two-state model by taking the derivative of Q (Equation (C.5)):

$$\langle n \rangle = \frac{s}{Q} \left(\frac{dQ}{ds} \right) = \frac{sNs^{N-1}}{1+s^N} = N \frac{s^N}{1+s^N}, \quad \text{so} \quad f_H = \frac{s^N}{1+s^N}.$$

Comparing Figure 26.14 with Figure 26.11 shows that the two-state model for the helix-coil transition is too cooperative to explain the experimental data.

In summary, one-state and two-state transformations are distinguished by the shapes of their free energy functions $F(m)$, where m represents the value of an order parameter such as the fractional helicity. At the midpoint of a two-state transition, the free energy function has two minima (see Figure 26.15). Under those conditions, half the molecules are in state m_1^* and half are in state m_2^* . There is little or no population of ‘intermediate states’ having intermediate values of the order parameter. In contrast, at the midpoint of a one-state transition, the free energy function has a single minimum, and the molecules in those systems have a distribution of order parameter values, $m_1^* \leq m^* \leq m_2^*$.

How can experiments determine whether a transition is one-state or two-state? Both models can give sigmoidal curves of $f_H(T)$, or of an order parameter $m(x)$ as a function of an externally controlled variable x , such as temperature. So a sigmoidal shape alone is not sufficient to establish whether the free energy has one or two minima. The strongest proof of a two-state transition would be direct evidence of two populations. Figure 26.16 shows an example of direct evidence of a two-state process: the unfolding of a protein with increasing concentrations of urea, a denaturing agent. Three other methods are commonly accepted evidence for two-state transitions: (1) when the enthalpy measured by calorimetry equals the enthalpy based on a two-state model (see Problem 11 at the end of this chapter), (2) when different experimental methods give superimposable sigmoidal curves (see Problem 10), and (3) when there is an *isodichroic* or *isosbestic point* in the spectra (as shown in Example 26.2).

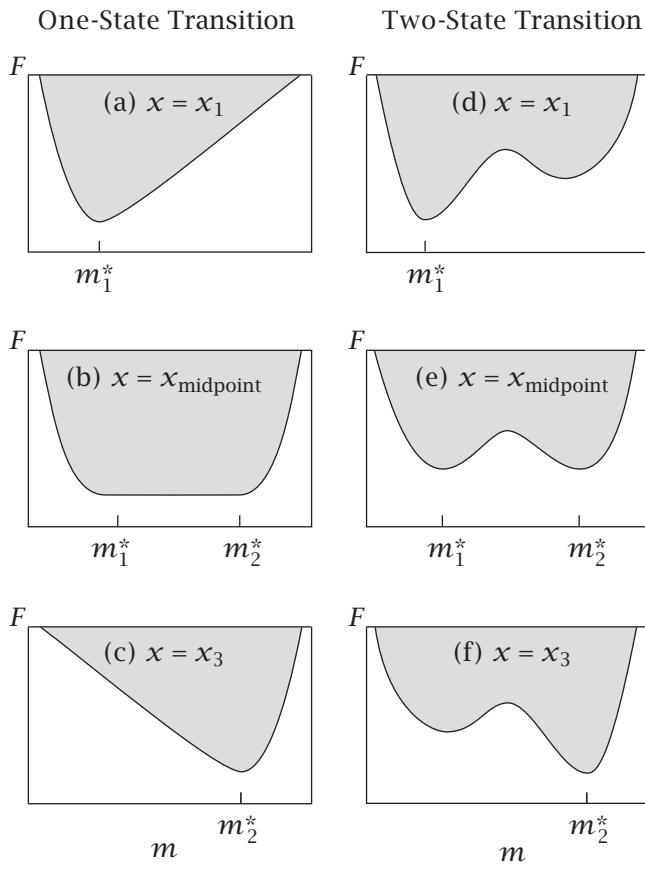


Figure 26.15 Transitions are *two-state* if there are two minima in the free energy and *one-state* if there is one minimum. In both types of system, the system changes from state m_1^* when $x = x_1$ to m_2^* when $x = x_2$. At the midpoint $x = x_{\text{midpoint}}$ of a two-state transition, there are two stable states: m_1^* and m_2^* . At the midpoint of a one-state transition, there is a broad distribution of stable states from m_1^* to m_2^* .

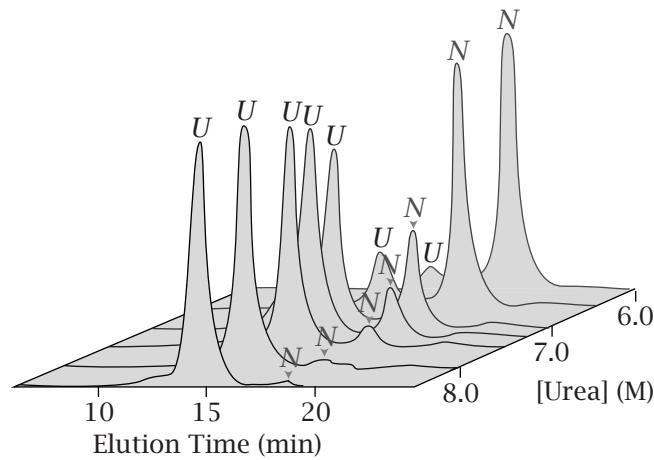


Figure 26.16 This chromatography experiment resolves the two states of a protein. The native state (N) is compact and has a long elution time. The unfolded state (U) is expanded and has a short elution time. Urea is a denaturing agent. It shifts the equilibrium from N to U . The peak areas represent the relative fractions of the two species. Source: RJ Corbett and RS Roche, *Biochemistry* **23**, 1888–1894 (1984).

EXAMPLE 26.2 Identifying a two-state transition by a spectroscopic isosbestic point. Suppose that a state A can be identified by its absorbance spectrum $f(\lambda)$ as a function of the wavelength λ . Suppose that state B is identified by its spectrum $g(\lambda)$. m is the order parameter that describes the fractional amount of B that is present in the system, so $1-m$ is the amount of A . In a two-state process, the observed spectrum of a mixture of states A and B is

Fluorescence Intensity

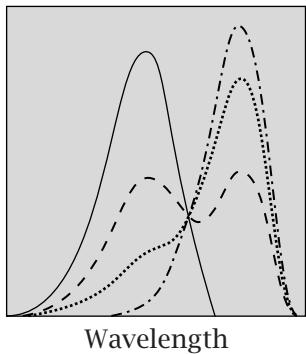


Figure 26.17
Fluorescence spectra of naphthol at various pH values. The common intersection point is an *isosbestic point*, and indicates a two-state transition, in this case $A + H^+ \rightarrow AH^+$. Source: W Hoppe, W Lohmann, and H Markl, eds, *Biophysics*, Springer-Verlag, New York, 1983.

$h(\lambda)$, which is the weighted sum over states A and B :

$$\begin{aligned} h(\lambda) &= mg(\lambda) + (1-m)f(\lambda) \\ &= f(\lambda) + m[g(\lambda) - f(\lambda)]. \end{aligned} \quad (26.17)$$

Suppose that the two spectra intersect at an *isosbestic point* λ_0 :

$$g(\lambda_0) = f(\lambda_0). \quad (26.18)$$

Substituting Equation (26.18) into Equation (26.17) gives

$$h(\lambda_0) = f(\lambda_0),$$

which is independent of m . This means that if you vary a parameter, such as the temperature or the concentration of a chemical agent that changes m , and changes the relative amounts of A and B , all the spectra from those experiments intersect at the wavelength $\lambda = \lambda_0$ (if the physical process is adequately modeled with two states). So if you observe a single point where all the spectra for different stages of the transition superimpose, you have evidence for a two-state transition (see Figure 26.17).

Now we explore a third model, in which the monomers are neither fully independent of each other nor totally cooperative. This Ising-like nearest-neighbor model gives deeper insights into cooperativity.

The Zimm-Bragg Model of the Helix-Coil Transition: Nearest-Neighbor Cooperativity

In the this model, developed by BH Zimm and JK Bragg [5], the probability that monomer j is in state C or H depends on whether its neighboring monomer $j-1$ is in state C or H . The statistical weight q for monomer j is *conditional* upon the state of its neighboring monomer. There are four different *monomer conditional statistical weights*: $q(C|C)$, $q(C|H)$, $q(H|C)$, and $q(H|H)$. Statistical weights are quantities that describe relative populations or probabilities. For example $q(C|H)$ is the relative amount of monomer j that is in state C if monomer $j-1$ is in state H . You can multiply together appropriate monomer statistical weights to get the *sequence statistical weight* for a chain of H and C monomers of length N . Then the partition function is the sum over the statistical weights of all the possible sequences. Table 26.2 lists all four possible monomer statistical weights, in the Zimm-Bragg notation.

Every coil unit C is assigned a statistical weight of 1, by convention, whether the preceding monomer is an H or a C . Any helical unit H is assigned a statistical weight s . You can think of s as an equilibrium constant for converting a C to an H :

$$s = \frac{[H]}{[C]}.$$

Assign a statistical weight σs if an H follows a C or if the H is the first monomer in the chain. Think of σ as a *nucleation parameter* for initiating a helix. For most

Table 26.2 Statistical weights for the Zimm–Bragg model. The statistical weight q for having a monomer type H or C at position j in the chain is *conditional* upon what monomer type is at position $j-1$.

Monomer State		Zimm–Bragg Statistical Weight for j
$j-1$	j	
C	C	$q(C C) = 1$
H	C	$q(C H) = 1$
C	H	$q(H C) = \sigma s$
H	H	$q(H H) = s$

amino acids in proteins, s is slightly greater than 1 and $\sigma \approx 10^{-3}$ to 10^{-4} [6]. Propagation is easy but nucleation is difficult.

Using the rules of conditional probability, Equation (1.11), you can compute the statistical weight for a sequence, say CHH , as the product $q(CHH) = q(H|H)q(H|C)q(C \text{ first}) = s \times \sigma s \times 1 = \sigma s^2$. Table 26.3 lists the statistical weights of all sequences of C 's and H 's that have three monomers. Each is a product of a factor of 1 for every C , a factor of s for every H , and a factor of σ for each beginning of a helix. For the first monomer of a chain, you have $q(C) = 1$ or $q(H) = \sigma s$.

Because each chain of length N can have any possible configuration—any sequence of H 's and C 's—the partition function is the sum over all the possible chain configurations. You can compute the partition function using a matrix method. First, express the statistical weights for the two possible states of the first monomer as a statistical weight vector \mathbf{q}_1 :

$$\mathbf{q}_1 = [q(C), \quad q(H)].$$

Now express the monomer conditional statistical weights in terms of a matrix \mathbf{G} :

$$\mathbf{G} = \begin{bmatrix} q(C|C) & q(H|C) \\ q(C|H) & q(H|H) \end{bmatrix} = \begin{bmatrix} 1 & \sigma s \\ 1 & s \end{bmatrix}. \quad (26.19)$$

To get the statistical weight vector \mathbf{q}_2 for all the two-monomer sequences, multiply \mathbf{q}_1 by \mathbf{G} (see the box on page 346 in Chapter 18):

$$\begin{aligned} \mathbf{q}_2 &= \mathbf{q}_1 \mathbf{G} \\ &= [q(C)q(C|C) + q(H)q(C|H), \quad q(C)q(H|C) + q(H)q(H|H)] \\ &= [q(CC) + q(CH), \quad q(HC) + q(HH)] \\ &= [1 + \sigma s, \quad \sigma s + \sigma s^2]. \end{aligned} \quad (26.20)$$

The leftmost term in the vector \mathbf{q}_2 is the sum of the statistical weights of all of the sequences that end in C . The rightmost term counts all of the sequences that end in H . To get the partition function Q_2 , which is the sum over the statistical weights of all sequences for a two-monomer chain, multiply \mathbf{q}_2 by a

Table 26.3 The Zimm–Bragg partition functions for all eight possible trimer sequences of H 's and C 's.

Trimer Sequence	Statistical Weight
CCC	1
CCH	σs
CHC	σs
CHH	σs^2
HCC	σs
HCH	$\sigma^2 s^2$
HHC	σs^2
HHH	σs^3

column vector of 1's:

$$\begin{aligned} Q_2 &= \mathbf{q}_2 \begin{bmatrix} 1 \\ 1 \end{bmatrix} = q(CC) + q(CH) + q(HC) + q(HH) \\ &= 1 + 2\sigma s + \sigma s^2. \end{aligned} \quad (26.21)$$

This procedure readily generalizes for a chain of any length N . Every multiplication by the matrix \mathbf{G} 'grows' the chain by one unit. For example,

$$\mathbf{q}_3 = \mathbf{q}_2 \mathbf{G} = \mathbf{q}_1 \mathbf{G}^2.$$

The partition function Q_N for a chain of N monomers is

$$Q_N = [1, \sigma s] \mathbf{G}^{N-1} \begin{bmatrix} 1 \\ 1 \end{bmatrix}. \quad (26.22)$$

Equation (26.22) gives the sum of the statistical weights for all of the possible sequences of H 's and C 's.

EXAMPLE 26.3 Compute the partition function for a trimer. Substitute $N = 3$ into Equation (26.22):

$$\begin{aligned} Q_3 &= [1, \sigma s] \mathbf{G}^2 \begin{bmatrix} 1 \\ 1 \end{bmatrix} \\ &= [1, \sigma s] \begin{bmatrix} 1 & \sigma s \\ 1 & s \end{bmatrix} \begin{bmatrix} 1 & \sigma s \\ 1 & s \end{bmatrix} \begin{bmatrix} 1 \\ 1 \end{bmatrix} \\ &= [1 + \sigma s, \sigma s + \sigma s^2] \begin{bmatrix} 1 & \sigma s \\ 1 & s \end{bmatrix} \begin{bmatrix} 1 \\ 1 \end{bmatrix} \\ &= [1 + \sigma s + \sigma s + \sigma s^2, \sigma s + \sigma s^2 + \sigma^2 s^2 + \sigma s^3] \begin{bmatrix} 1 \\ 1 \end{bmatrix} \\ &= 1 + 3\sigma s + 2\sigma s^2 + \sigma^2 s^2 + \sigma s^3. \end{aligned} \quad (26.23)$$

The population of a given state is the ratio of that particular term to the whole partition function. For example, the population of fully helical molecules is

$$p_{\text{all } H} = \sigma s^3 / Q. \quad (26.24)$$

You can check Equation (26.23) by noting that it gives the same sum as adding all the statistical weights in Table 26.3.

From the partition function, Equation (26.22), you can compute the properties of the model, such as the fractional helicity.

EXAMPLE 26.4 Compute the fractional helicity of the trimer. Following Equation (C.5), take the derivative of Equation (26.23) to get the fractional helicity:

$$f_H = \frac{1}{N} \left(\frac{\partial \ln Q}{\partial \ln s} \right) = \left(\frac{1}{3} \right) \left(\frac{3\sigma s + 4\sigma s^2 + 2\sigma^2 s^2 + 3\sigma s^3}{1 + 3\sigma s + 2\sigma s^2 + \sigma^2 s^2 + \sigma s^3} \right). \quad (26.25)$$

In the Zimm-Bragg model, the cooperativity is controlled by the parameter σ . Helix formation is cooperative when σ is small. When $\sigma = 1$ in Equation (26.25), the Zimm-Bragg model reduces to the noncooperative independent units model:

$$f_H = \frac{s}{1+s}.$$

But if nucleation is unfavorable ($\sigma \ll 1$) and if propagation is favorable ($s > 1$), then helix formation is cooperative. That is, you will see either no helices (only coils) or long helices; you won't see short helices. This is because the population of helices is approximately proportional to σs^j , where j is the number of helical units in the chain. This population will only be high if s^j is sufficiently large to compensate for a small value of σ . As σ becomes small and s becomes sufficiently large, this model becomes identical to the two-state model.

Figure 26.11 shows that the Zimm-Bragg model captures the chain-length dependence of the cooperativity in the helix-coil transitions of polypeptides. This model simplifies the problem of the many different three-dimensional polymer chain configurations into a one-dimensional problem of enumerating the strings of arrangements of nearest neighbors.

Now we consider one of the simplest models for the kinetics of phase transitions.

The Kinetics of Phase Transitions Can Be Controlled by Nucleation

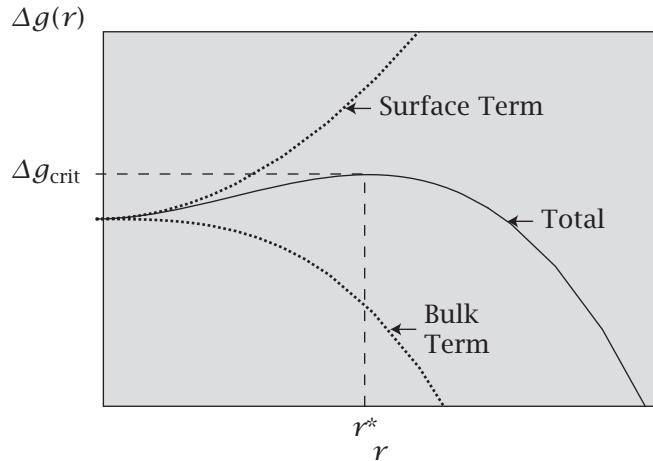
How fast is a phase transition? If you cool a binary liquid mixture, what limits the rate of phase separation? How fast do crystals form in a liquid? How fast do micelles form in a surfactant solution? For such processes, there are two different mechanisms. First, if you supercool a single-phase solution into the always-unstable part of the two-phase region (see Figure 25.12), the process is spontaneous. It is called *spinodal decomposition*. But a different mechanism applies if you supercool a single-phase solution into the metastable region between the binodal and spinodal curves. Then phase separation can be slow. To form a *B*-rich phase in a metastable solution of components *A* and *B* requires that enough molecules of *B* come together to form a nucleus that is so large that further growth of the *B*-rich phase will be spontaneous. The process of *nucleation* is the formation of this *critical nucleus*, which requires that the system reach the top of an energy barrier. Here's a simple model of the activation energy for this process.

What is the free energy Δg of forming a spherical nucleus of radius r for a *B*-rich phase in a background sea of mostly *A* molecules? There are two contributions to the free energy of droplet formation (see Figure 26.18):

$$\Delta g = 4\pi r^2 \gamma - \frac{4}{3}\pi r^3 \varepsilon_0. \quad (26.26)$$

Growth is driven by ε_0 , the energy per unit volume for forming *BB* contacts inside the droplet, multiplied by the droplet volume $(4/3)\pi r^3$. Growth is opposed by the surface tension γ at the interface between the *B*-rich droplet

Figure 26.18 The free energy Δg of forming a spherical droplet that nucleates a phase transition. A droplet must reach a minimum radius r^* to grow into a macroscopic phase. The droplet growth is favored by a ‘bulk’ term ε_0 and opposed by a surface-area term γ .



and the A -rich solution, multiplied by the surface area $4\pi r^2$. The radius r^* of the critical nucleus is the point at which $\Delta g(r^*) = \Delta g_{\text{crit}}$ is a maximum:

$$\left(\frac{d\Delta g}{dr}\right)_{r=r^*} = 0 \implies r^* = \frac{2\gamma}{\varepsilon_0}. \quad (26.27)$$

A large surface tension γ implies that large droplets are required to nucleate the phase separation. Substitute Equation (26.27) into Equation (26.26) to compute the free energy,

$$\begin{aligned} \Delta g_{\text{crit}} &= 4\pi(r^*)^2 \left(\gamma - \frac{r^* \varepsilon_0}{3} \right) \\ &= \frac{4\pi(r^*)^2 \gamma}{3} = \frac{16\pi\gamma^3}{3\varepsilon_0^2}. \end{aligned} \quad (26.28)$$

Droplet formation is an activated process. The rate of forming a droplet is proportional to $\exp(-\Delta g_{\text{crit}}/kT)$, where kT is Boltzmann’s constant multiplied by temperature.

Summary

Both macroscopic matter and individual molecules can undergo dramatic transitions. Changing an external variable x , such as the temperature, can cause a change in an order parameter $m(x)$, such as the helix content of a polymer, the density of a liquid, the composition of a binary mixture or alloy, or the magnetization of a metal. Widely different physical processes can often be described by the same mathematical models. If the underlying free energy surface has two minima at the transition point, it is a *two-state* or *first-order* transition. If there is only a single minimum, it is a *one-state* or *higher-order* transition. Such transitions can be modeled with the nearest-neighbor model, which takes correlations between neighboring molecules into account. This model has played an important role in understanding critical exponents. Very close to critical points, fluctuations can occur over large spatial scales, and models based on low-order expansions predict incorrect critical exponents.

Problems

1. A micellization model. You have developed a model for the formation of micelles, based on expressions for the chemical potential $\mu(\text{mono})$ of the free monomeric molecules in solution and the chemical potential $\mu(\text{mic})$ of the aggregated state of the monomers in micelles as a function of the mole fraction x of monomers in solution:

$$\mu(\text{mono}) = 0.5 + 0.1x^2$$

and

$$\mu(\text{mic}) = 1.3 - 25x^2.$$

- (a) What is the state of the system at low concentration?
 - (b) What is the state of the system at high concentration?
 - (c) What is the concentration at which the micelle transition occurs?
- 2. Stabilities of droplets.** Certain molecular aggregates (like surfactant micelles and oil droplets) shrink owing to surface tension and expand owing to volumetric forces. Suppose the free energy $g(r)$ as a function of the radius r of the aggregate is $g(r) = 2r^2 - r^3$. Such a system has two equilibrium radii.
- (a) Calculate the two radii.
 - (b) Identify each radius as stable, unstable, neutral, or metastable.
- 3. The energy of the Ising model.** Derive an expression for the energy of the Ising model from the partition function.
- 4. The Landau model for the critical exponent of $C_V(T)$.** In the Landau model, show that $C_V \propto T$ as T increases toward the critical temperature T_c .

5. Nucleation droplet size. What is the number of particles n^* in a critical nucleus for phase separation?

6. Broad energy wells imply large fluctuations. A stable state of a thermodynamic system can be described by the free energy $G(x)$ as a function of the degree of freedom x . Suppose G obeys a square law with spring constant k_s : $G(x)/kT = k_s x^2$. A small spring constant k_s implies a broad shallow well (see Figure 26.19).

- (a) Show that the fluctuations $\langle x^2 \rangle$ are larger when k_s is small.
- (b) Suppose two phases in equilibrium have narrow energy wells, each with spring constant k_1 , and that the critical point of the system has a broad well with spring constant $k_2 = (1/4)k_1$. What is the ratio of the fluctuations, $\langle x_2^2 \rangle / \langle x_1^2 \rangle$?

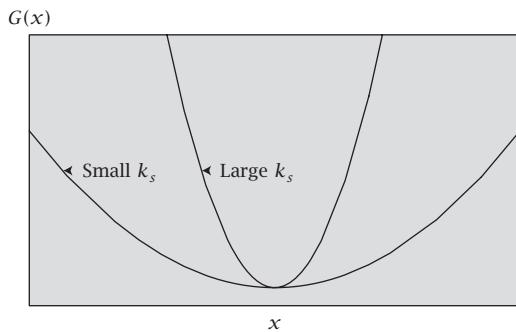


Figure 26.19 Two different square-law free energies for Problem 6: one minimum is broad and the other is narrow.

7. Critical exponents. What are the critical exponents for the following free energy functions? A is a constant and $t = |(T - T_c)/T_c|$ is the reduced temperature.

- (a) $g(t) = At^2 e^{-t}$.
- (b) $g(t) = At^{5/3}$.
- (c) $g(t) = A$.

8. Zimm-Bragg helix-coil theory for $N = 4$ chain units.

- (a) Write the Zimm-Bragg partition function Q_4 in terms of σ and s for a four-unit chain, where HHHH is the helical state.
- (b) Write an expression for $f_H(s)$ for this transition.

9. The Schellman helix-coil model. A helix-coil model developed by JA Schellman [7] is simpler than the Zimm-Bragg model, and works well for short chains. Consider a chain having N units.

- (a) Write an expression for Ω_k , the number of configurations of a chain that has all its H units in a single helix k units long, as a function of N and k .
- (b) If σ is the parameter for nucleating a helix and s is the propagation parameter, write an expression for the partition function Q_N over all possible helix lengths k .
- (c) Write an expression for $p_k(N)$, the probability of finding a k -unit helix in the N -mer.

10. Overlapping transition curves as evidence for a two-state transition. Consider a two-state transition $A \rightarrow B$. m is an order parameter: when $m = 0$, the system is fully in state A ; when $m = 1$, the system is in state B . Suppose some agent x (which could be the temperature or the concentration of ligands or salts) shifts the equilibrium as indicated in Figure 26.20. An experiment measures a property g (which might be fluorescence, light scattering, circular dichroism, heat, etc.):

$$g(m) = g_B m + g_A (1-m),$$

where g_B and g_A are the measured values for pure B and A , respectively. A different type of experiment measures another property h :

$$h(m) = h_B m + h_A (1-m),$$

where h_B and h_A are also the measured values for the pure states. Show that if the baselines of the two types of experiment are superimposed ($g_A = h_A$) and the amplitudes are scaled to be the same ($h_B - h_A = g_B - g_A$), then the curve $h(x)$ must superimpose on $g(x)$.

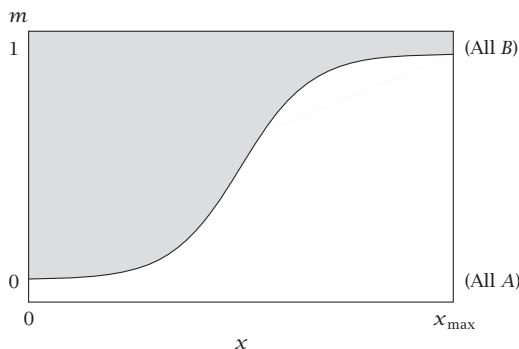


Figure 26.20 A transition curve for folding, binding, or conformational change, in which the order parameter m is a function of some agent x that shifts the equilibrium.

11. Calorimetric evidence for a two-state transition: the relationship between ΔH_{cal} and $\Delta H_{\text{van't Hoff}}$. A calorimeter measures the heat capacity $C_p(T)$ as a function of temperature T (see Figure 26.21). The area under the full curve describes an absorption of heat

$$q_0 = \Delta H_{\text{cal}} = \int_{T_A}^{T_B} C_p dT.$$

Suppose the absorption of heat results from a two-state process $A \xrightarrow{K} B$, where $K = [B]/[A]$.

- (a) Express the fraction $f = [B]/([A]+[B])$ in terms of K .
- (b) Express $C_p(T)$ as a function of the heat q_0 and the equilibrium constant $K(T)$.

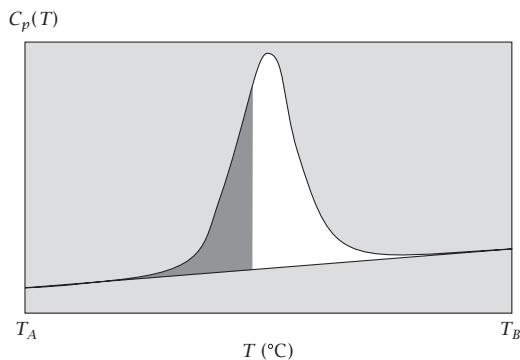


Figure 26.21 The heat capacity change in a two-state transition.

- (c) Assume a two-state model (Equation (13.35)):

$$\left(\frac{\partial \ln K}{\partial T} \right) = \frac{\Delta H_{\text{van't Hoff}}}{kT^2},$$

and $K = 1$ at $T = T_{\max}$, the point at which C_p is a maximum.

Derive a relationship between $\Delta H_{\text{van't Hoff}}$, ΔH_{cal} , and $C_p(T_{\max})$.

12. Cooperativity in a three-state system. Perhaps the simplest statistical mechanical system having cooperativity is the three-level system shown in Figure 26.22.

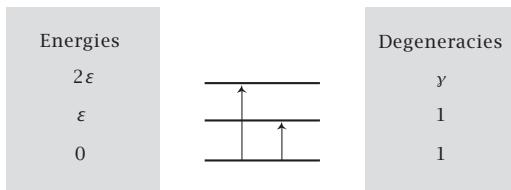


Figure 26.22 Diagram of three energy levels and degeneracies.

- (a) Write an expression for the partition function q as a function of energy ϵ , degeneracy γ , and temperature T .
- (b) Write an expression for the average energy $\langle \epsilon \rangle$ versus T .
- (c) For $\epsilon/kT = 1$ and $\gamma = 1$, compute the populations, or probabilities p_1 , p_2 , and p_3 , of the three energy levels.
- (d) Now if $\epsilon = 2 \text{ kcal mol}^{-1}$ and $\gamma = 1000$, find the temperature T_0 at which $p_1 = p_3$.
- (e) Under the condition of (d), compute p_1 , p_2 , and p_3 at temperature T_0 . In what sense is this system cooperative?

References

- [1] JJ Binney, NJ Dowrick, AF Fisher, and MEJ Newman, *The Theory of Critical Phenomena: An Introduction to the Renormalization Group*. Oxford University Press, Oxford, 1992.
- [2] HE Stanley, *Introduction to Phase Transitions and Critical Phenomena*. Oxford University Press, New York, 1971.
- [3] JM Yeomans, *Statistical Mechanics of Phase Transitions*. Oxford University Press, Oxford, 1992.
- [4] R Kubo, with H Ichimura, T Usui, and N Hashitsome, *Statistical Mechanics*, North-Holland, New York, 1965.
- [5] BH Zimm and JK Bragg, *J Chem Phys* **31**, 526–531 (1959).
- [6] BH Zimm, P Doty, and K Iso, *Proc Natl Acad Sci USA* **45**, 1601–1607 (1959).
- [7] JA Schellman, *J Phys Chem* **62**, 1485–1494 (1958).

Suggested Reading

Detailed treatments of Zimm-Bragg helix-coil theory:

CR Cantor and PR Schimmel, *Biophysical Chemistry*. WH Freeman, San Francisco, 1980.

BH Zimm and JK Bragg, *J Chem Phys* **31**, 526–531 (1959).

Excellent treatments of phase transitions, critical phenomena, and Ising models:

JJ Binney, NJ Dowrick, AF Fisher, and MEJ Newman, *The Theory of Critical Phenomena: An Introduction to the Renormalization Group*, Oxford University Press, Oxford, 1992.

N Goldenfeld, *Lectures on Phase Transitions and the Renormalization Group*, Addison-Wesley, Reading, 1992.

R Kubo, with H Ichimura, T Usui, and N Hashitsome, *Statistical Mechanics*. North-Holland, New York, 1965.

HE Stanley, *Introduction to Phase Transitions and Critical Phenomena*, Oxford University Press, New York, 1971.

JM Yeomans, *Statistical Mechanics of Phase Transitions*. Oxford University Press, Oxford, 1992.

Excellent treatments of nucleation theory:

PG Debenedetti, *Metastable Liquids: Concepts and Principles*, Princeton University Press, Princeton, NJ, 1996.

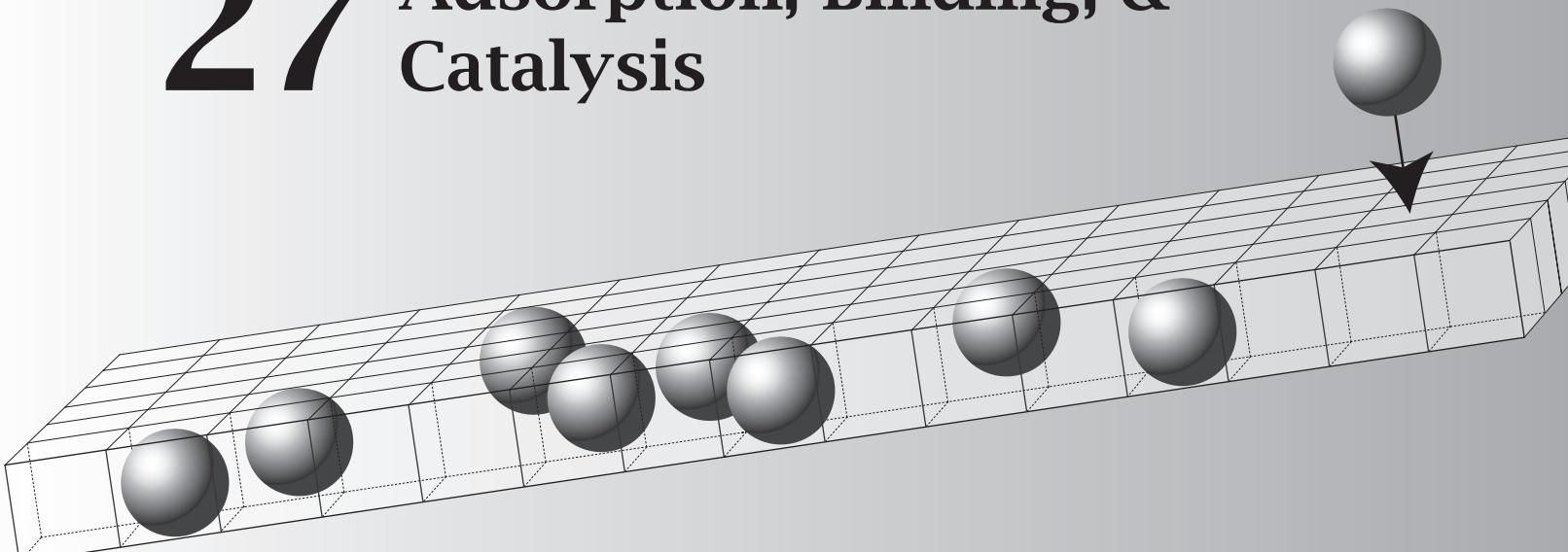
DV Ragone, *Thermodynamics of Materials*, Volume 2, Wiley, New York, 1995.

JW Mullin, *Crystallization*, 4th edition, Butterworth-Heinemann, Oxford, 2001.

G Wilemski, *J Phys Chem* **91**, 2492–2498 (1987).

This page is intentionally left blank.

27 Adsorption, Binding, & Catalysis



Binding and Adsorption Processes Are Saturable

Atoms and molecules bind, or adsorb, to surfaces. Adsorption plays a key role in catalysis, filtration, oxidation, corrosion, and chromatographic separations. And, elementary binding processes are at the root of most biological mechanisms of action, and underpin the modern pharmacological approach to treating diseases. For example, Figure 27.1 shows the potencies of different schizophrenia drugs. The potencies are proportional to how well the drug inhibits the binding of a molecule called haloperidol to a protein dopamine receptor.

Atoms or molecules that bind to surfaces are called *adsorbates* or *ligands*. Increasing the concentration of an adsorbate will ultimately *saturate* a surface, filling all the available space. The simplest model of binding and saturation is the Langmuir model. It applies not only to adsorbates on macroscopic solid surfaces but also to pH titration, the kinetics of enzymatic reactions, and the transport of particles through biological membranes. The Langmuir model describes the balance between the energetic tendency of the particles to stick to the surfaces and the entropic tendency of the particles to gain translational freedom by escaping from them.

The Langmuir Model Describes Adsorption of Gas Molecules on a Surface

If a solid surface is available to a gas, some of the gas atoms or molecules may adsorb to it. Increasing the pressure of the gas will increase the number of gas

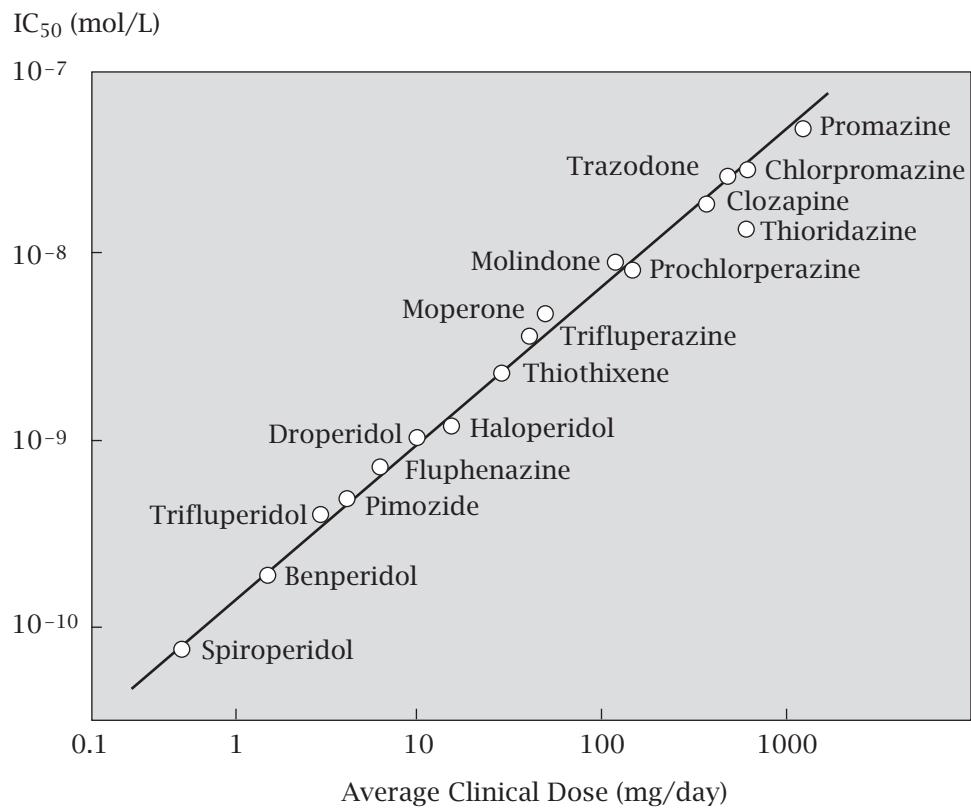
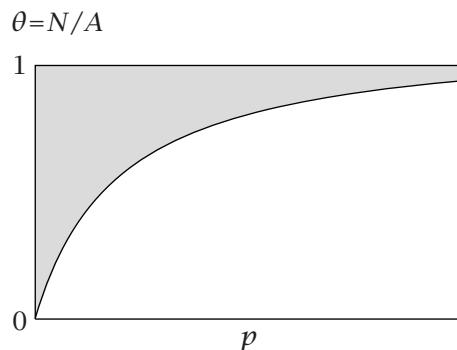


Figure 27.1 Binding processes are important in biology and medicine. The clinical potencies of schizophrenia drugs (in average clinical dosage, mg/day) correlates with their binding activities (given as the concentration that causes 50% inhibition, called the IC_{50} , of haloperidol binding) to dopamine D_2 receptors. (From Figure 37.1, page 527 from HP Rang et al., *Pharmacology*, 5th edition, Churchill Livingstone, Edinburgh, 2003; data from P Seeman et al., *Nature* **261**, 717–719 (1976).)

Figure 27.2 Increasing the gas pressure p increases the amount of gas adsorbed until it saturates the surface. θ is the fraction of surface sites filled by adsorbate; N is the number of ligands bound; A is the number of binding sites available.



molecules bound to the surface (see Figure 27.2). The simplest treatment of adsorption and binding is the Langmuir model, named after I Langmuir (1881–1957), an American chemist who won the 1932 Nobel Prize in Chemistry for his work in surface science. Our aim is to compute how N , the number of ligand molecules adsorbed to the surface, depends on the pressure p of the gas. At equilibrium, the chemical potential μ_{gas} of the adsorbate in the gas phase equals

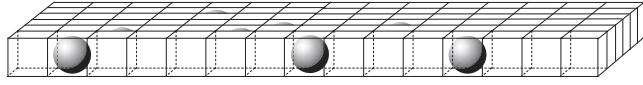


Figure 27.3 Two-dimensional lattice for the Langmuir model. N particles occupy the lattice. The lattice has A sites.

the chemical potential μ_{bound} of the adsorbate on the surface:

$$\mu_{\text{bound}} = \mu_{\text{gas}}. \quad (27.1)$$

To model μ_{bound} , we begin with the translational entropy of the adsorbate distributed on the surface. A lattice approach is simplest. We consider the surface of the solid to be a two-dimensional lattice having A sites (Figure 27.3). Each of the N ligand molecules occupies one site. The density θ of the adsorbate on the surface is

$$\theta = \frac{N}{A}, \quad (27.2)$$

the fraction of surface sites that are filled with ligand. The number of arrangements of the particles on the surface lattice is given by Equation (1.22):

$$W = \frac{A!}{N!(A-N)!}. \quad (27.3)$$

To compute the translational entropy, apply the Boltzmann relation and Stirling's approximation, Equation (B.3) in Appendix B, to Equation (27.3):

$$\begin{aligned} \frac{S}{k} &= \ln W = -N \ln \left(\frac{N}{A} \right) - (A-N) \ln \left(\frac{A-N}{A} \right) \\ \Rightarrow \quad \frac{S}{Ak} &= -\theta \ln \theta - (1-\theta) \ln (1-\theta). \end{aligned} \quad (27.4)$$

Equation (27.4) has appeared before as the translational entropy of a lattice gas (Equation (6.10)) and as the entropy of mixing in a three-dimensional system (Equation (15.3)). Remarkably, this distributional entropy does not depend on whether a system is one-, two-, or three-dimensional. Nor does it depend on the arrangement of the binding sites. Binding sites could form a contiguous two-dimensional plane on a metal surface, or they could be found individually on different protein or polymer molecules in solution.

Now we compute the adsorption free energy. If N particles stick to the surface, each with energy $w < 0$, then binding contributes an amount

$$U = Nw \quad (27.5)$$

to the total energy U of the system. If the internal degrees of freedom of the particle also change because particle orientations become restricted upon binding, or because new modes of vibrations are created in the surface, there is an additional contribution to the free energy, $-NkT \ln q_{\text{bound}}$. The free energy F of the adsorbed gas is

$$\frac{F}{AkT} = \frac{U - TS}{AkT} = \theta \ln \theta + (1-\theta) \ln (1-\theta) + \left(\frac{w}{kT} \right) \theta - \theta \ln q_{\text{bound}}. \quad (27.6)$$

The degrees of freedom are (N, A, T) . They correspond to (N, V, T) in a single-component three-dimensional system of size V . To get the chemical

potential μ_{bound} of the adsorbate at the surface, take the derivative with respect to N , holding A and T constant:

$$\begin{aligned}\frac{\mu_{\text{bound}}}{kT} &= \left(\frac{\partial(F/kT)}{\partial N} \right)_{A,T} \\ &= \ln\left(\frac{N}{A-N}\right) + \frac{w}{kT} - \ln q_{\text{bound}} = \ln\left(\frac{\theta}{1-\theta}\right) + \frac{w}{kT} - \ln q_{\text{bound}}.\end{aligned}\quad (27.7)$$

For the adsorbate molecules in the gas phase, the chemical potential is given by Equation (14.5):

$$\mu_{\text{gas}} = kT \ln\left(\frac{p}{p_{\text{int}}^{\circ}}\right), \quad (27.8)$$

where $p_{\text{int}}^{\circ} = q'_{\text{gas}} kT$, and q'_{gas} is the partition function of the gas molecules with the volume V factored out ($q_{\text{gas}} = q'_{\text{gas}} V$; see Equation (11.48)).

Substitute Equation (27.7) for μ_{bound} and Equation (14.5) for μ_{gas} into the condition for equilibrium given in Equation (27.1):

$$p = \left(\frac{q'_{\text{gas}} kT}{q_{\text{bound}}} \right) \left(\frac{\theta}{1-\theta} \right) e^{w/kT}. \quad (27.9)$$

Collect terms together to form a *binding constant*

$$K = \frac{q_{\text{bound}}}{q'_{\text{gas}} kT} e^{-w/kT} \quad (27.10)$$

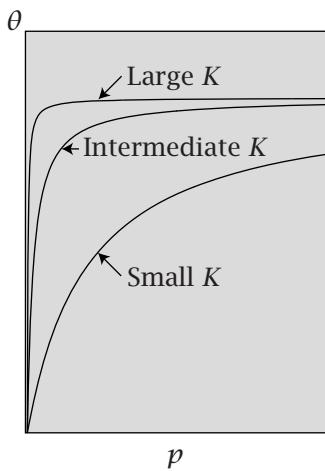


Figure 27.4 Increasing the binding affinity of the adsorbate for the surface increases the surface coverage, at fixed pressure.

(see Equation (13.10)). Substituting K into Equation (27.9) gives

$$Kp = \frac{\theta}{1-\theta}, \quad (27.11)$$

which rearranges to the **Langmuir adsorption equation**,

$$\theta = \frac{Kp}{1+Kp}. \quad (27.12)$$

Figure 27.4 shows this function. For low gas pressures, Equation (27.12) predicts that the ligand coverage of the surface is linearly proportional to the gas pressure ($\theta = Kp$). This is called the Henry's law region, by analogy with the vapor pressure of a mixture (see page 286). For adsorbates that have a strong affinity for the surface ($K \gg 1$), little pressure is required to saturate the surface (see Figure 27.4). The Langmuir equation predicts that the surface will saturate ($\theta \rightarrow 1$) at high pressures ($Kp \gg 1$) as all the sites become filled.

If you want to test whether your data follow the Langmuir model, there is a better functional form than Equation (27.12). Equation (27.12) requires data over a wide range of pressures to ensure that you reach saturation, and this can often be difficult to obtain. Instead, you can rearrange Equation (27.12) into a linearized form:

$$\frac{1}{\theta} = \frac{1+Kp}{Kp} \implies \frac{p}{\theta} = \frac{1}{K} + p. \quad (27.13)$$

A plot of p/θ against p will yield a straight line with intercept $1/K$ when the Langmuir model applies. Figure 27.5(a) shows that the Langmuir model predicts the adsorption of propane in zeolite pores.

To determine how binding depends on temperature, take the derivative of Equation (27.10) (neglecting the small dependence $q(T)$):

$$\left(\frac{\partial \ln K}{\partial T}\right) = \frac{1}{T} \left(\frac{w}{kT} - 1\right). \quad (27.14)$$

Or the temperature dependence can be expressed approximately with the van't Hoff form, Equation (13.35):

$$\left(\frac{\partial \ln K}{\partial T}\right) = \frac{\Delta h}{kT^2}. \quad (27.15)$$

The energetic tendency of the particles to stick to the surface dominates at low temperatures, while the entropic tendency of the ligands to escape dominates at high temperatures (see Figure 27.5(b)).

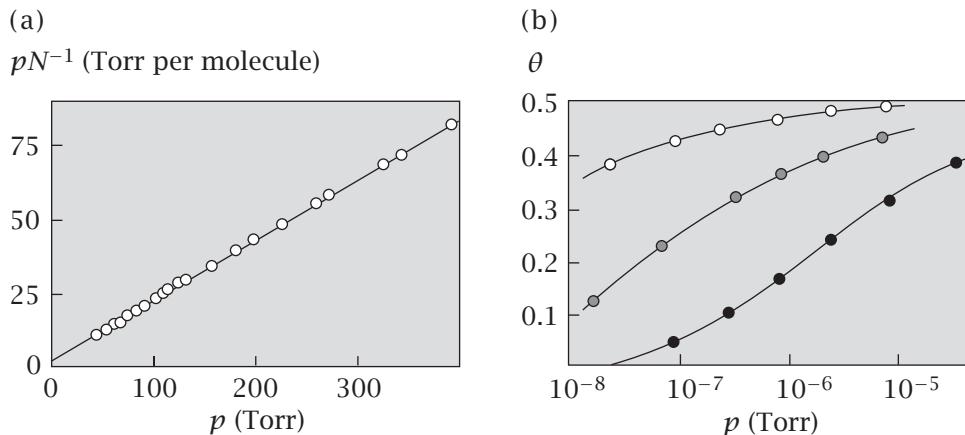


Figure 27.5 (a) A plot of $p/N (= 1/(KA)+p/A)$ versus p for propane in zeolite pores at 273 K is linear, following the Langmuir model (see Equation (27.13)). (Note that $\theta = N/A$.) Source: DM Ruthven, *Principles of Adsorption and Adsorption Processes*, Wiley, New York, 1984. Data are from DM Ruthven and KF Loughlin, *J Chem Soc Faraday Trans* **68**, 696–708 (1972). (b) Increasing temperature desorbs CO from palladium surfaces: (○) 383 K; (◐) 425 K; (●) 493 K. Note the logarithmic horizontal axis. Source: A Zangwill, *Physics at Surfaces*, Cambridge University Press, New York, 1988. Data are from G Ertl and J Koch, *Z Naturforsch* **25A**, 1906 (1970).

The Langmuir Model Also Treats Binding and Saturation in Solution

Put ligand molecules of type X into a solution in equilibrium with some type of particle, polymer, or surface, P . Ligands can bind the particles to form a ‘bound complex’ PX :



$[X]$ is the concentration of free ligand molecules, $[P]$ is the concentration of free particles (each of which has a single binding site), and $[PX]$ is the concentration of the bound complex. The binding, or association, equilibrium constant is

$$K = \frac{[PX]}{[P][X]}. \quad (27.17)$$

We want to know how θ , the fraction of binding sites on P that are filled by ligand, depends on the solution concentration of the ligand $[X]$. θ is the ratio (filled sites on P)/(empty + filled sites on P):

$$\begin{aligned} \theta &= \frac{[PX]}{[P] + [PX]} \\ &= \frac{K[P][X]}{[P] + K[P][X]} = \frac{Kx}{1 + Kx}. \end{aligned} \quad (27.18)$$

To simplify the notation slightly, we replaced $[X]$ by x . Equation (27.18) resembles the Langmuir equation (27.12), but with the pressure replaced by the ligand concentration. Equation (27.18) is general, and applies whether the concentration x refers to mole fractions, molarities, molalities, or any other units. The units that you use for x determine the units of K .

For binding in solution, the equilibrium constant K in Equation (27.18) results from both the direct ligand–surface interactions and the interactions of the ligand and surface with the solvent.

A simple way to determine K from a binding curve $\theta(x)$ is to find the point $\theta = 1/2$. At that point, $x = 1/K$, as shown in Examples 27.1 and 27.2.

EXAMPLE 27.1 Determining a binding affinity. Figure 27.6 shows the curve θ for binding *Escherichia coli* cyclic AMP receptor protein to a 32-base-pair DNA molecule as a function of DNA concentration. If you assume Langmuir binding, the dashed lines give $K = 1/(16 \text{ nM}) \approx 6 \times 10^7 \text{ M}^{-1}$. Note that in general the true Langmuir binding does not reach a plateau until very large values of x (compare Figure 27.7, in which the concentration of ligand is hundreds of μM to ensure saturation). The preferred method for determining a binding constant is to linearize the plot with Equations (27.13) or (28.19).

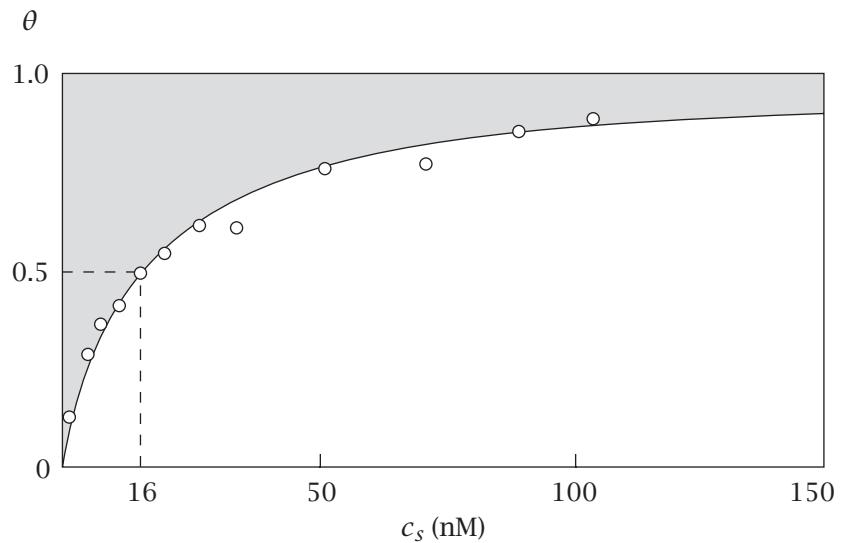


Figure 27.6 DNA molecule ligands bind to the *E. coli* cyclic AMP receptor protein. The fraction of filled sites θ is shown as a function of the DNA ligand concentration c_s . The dashed line indicates the ligand concentration at half-saturation of the binding sites (see Example 27.1). Source: T Heyduk and JC Lee, *Proc Natl Acad Sci USA* **87**, 1744–1748 (1990).

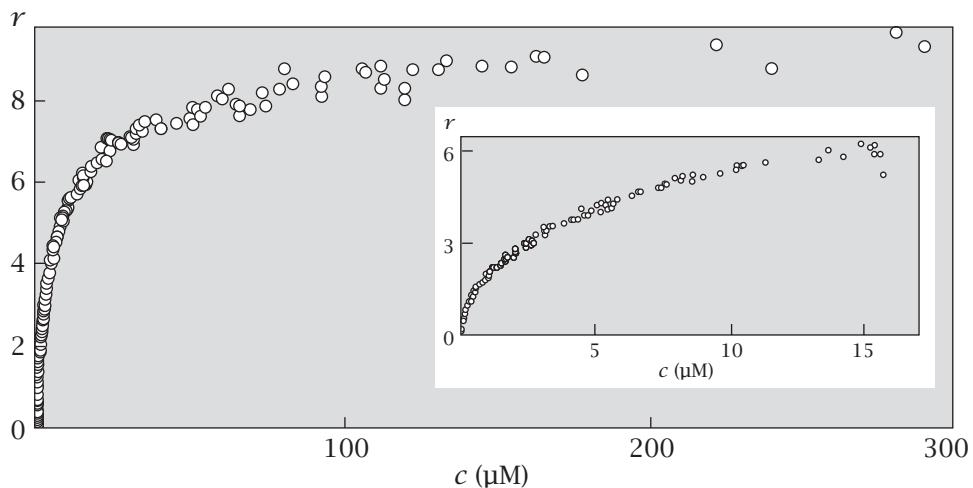


Figure 27.7 Laurate ion binding r to human serum albumin protein follows the Langmuir model. c is the laurate ion concentration. Note that to reach saturation requires a wide range of ligand concentrations. Source: AO Pedersen, B Hust, S Andersen, et al., *Eur J Biochem* **154**, 545–552 (1986).

The Langmuir model is useful for treating oxidation and reduction and also pH titration (see Figure 27.8). In oxidation-reduction processes, the ligand is an electron. In pH titration processes, the ligands are protons and the binding sites are acidic or basic groups on molecules.

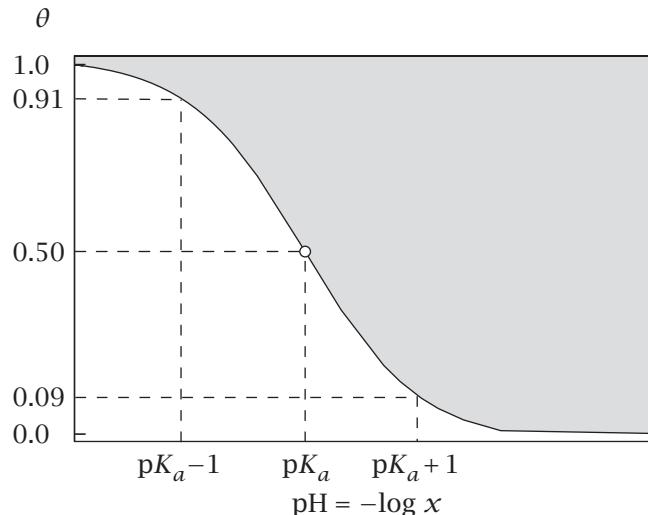


Figure 27.8 A pH titration curve is an example of Langmuir binding. The fractional saturation curve θ is sigmoidal because the pH on the x axis is proportional to the *logarithm* of the ligand concentration. When the pH is low, the concentration of hydrogen ions (the ligand) in solution is high, so the fractional binding θ to acid groups is high. As the available ligands (protons) decrease (pH increases), fewer protons bind to acid groups.

EXAMPLE 27.2 pH titration is an example of Langmuir binding. For an acid A^- , the equilibrium is $HA \rightleftharpoons H^+ + A^-$, and the acid dissociation constant is $K_a = [H^+][A^-]/[HA]$. Varying the hydrogen ion concentration (pH) in the solution can change the protonation state of acids and bases according to $\theta = [H^+]K_a^{-1}/(1+[H^+]K_a^{-1})$, which is Langmuir binding. If you plot this function using a *logarithmic* x axis, the Langmuir function has a sigmoidal shape. This is the standard x axis for a *pH titration curve*: $-\log x = -\log[H^+] = \text{pH}$. At the midpoint of titration, $\theta = 1/2$ and $Kx = 1$: half the acid or base groups have a proton on them, and the other half do not.

The Langmuir model also applies to chemical separation processes such as chromatography.

The Independent-Site Model Describes the Principle of Adsorption Chromatography

Gas and liquid chromatographies are based on the principle of selective adsorption (see Figure 27.9). Solutes are carried along in the flow of a *mobile phase*, which can be a gas or a liquid. Assume that the flow is slow enough for the solute to be in exchange equilibrium between the mobile phase and adsorption sites on a *stationary phase*. We approximate the fluid flow profile by using two velocities: v_m is the velocity of the mobile phase fluid down the center of the channel and v_s is the velocity of flow very close to the wall surface. To a first approximation, the particles that are close to the surface are stuck to the wall and don't move very quickly, so $v_s \approx 0$. The average flow velocity $\langle v \rangle$ for the solute passing through the chromatographic column is

$$\langle v \rangle = f_m v_m + f_s v_s \approx f_m v_m, \quad (27.19)$$

where f_m and $f_s = 1 - f_m$ are the fractions of the solute molecules that are in the mobile and stationary phases, respectively. The fraction f_m is given in terms of the concentration of solute c_m in the mobile phase, the concentration

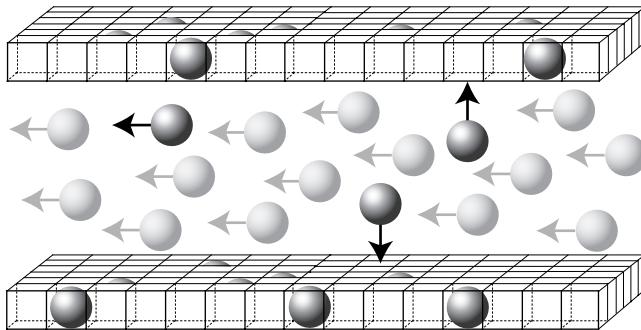


Figure 27.9 Model for chromatographic separations. Molecule types ● and ○ are carried by a mobile phase that flows through the system; ● adsorbs more strongly to the walls, so those molecules have a lower average flow velocity. Arrows indicate the flow direction of the mobile phase or toward the walls.

of solute c_s in the stationary phase, the volume V_m of the mobile phase, and the volume V_s of the stationary phase:

$$f_m = \frac{c_m V_m}{c_m V_m + c_s V_s} = \frac{1}{1 + (c_s V_s)/(c_m V_m)} = \frac{1}{1 + K\phi}. \quad (27.20)$$

$K = c_s/c_m$ is the partition coefficient of the solute from the mobile phase into the stationary phase (see Equation (16.42)). $\phi = V_s/V_m$ is a property of the chromatography apparatus called the *phase ratio*. Langmuir binding is reflected in the expression $f_s = 1 - f_m = K\phi/(1+K\phi)$. One purpose of chromatography experiments is to measure K (see Chapter 16). Substituting Equation (27.20) into Equation (27.19) gives the average flow velocity of the solute:

$$\langle v \rangle = \frac{v_m}{1 + K\phi}. \quad (27.21)$$

Solute molecules that have a greater affinity for the stationary phase have larger K , so they have a smaller average flow velocity $\langle v \rangle$. Molecules that have no affinity for the stationary phase, $K = 0$, do not adsorb, and so they flow faster, at the rate of the mobile phase, $\langle v \rangle = v_m$.

Rearranging Equation (27.21) shows how chromatography can be used to measure partition coefficients. If the phase ratio is known, the partition coefficient is found from measuring the flow delay of a solute relative to the velocity of the carrier fluid:

$$K\phi = \frac{v_m - \langle v \rangle}{\langle v \rangle}. \quad (27.22)$$

So far, we have applied the Langmuir model to equilibria. Now we switch our attention to kinetics. Some *rate* processes saturate as a function of ligand concentration, resembling Langmuir behavior. This is typically because rate processes often have an underlying saturable equilibrium binding step. An example is Michaelis-Menten kinetics, in which the rate of enzymatic catalysis of biochemical reactions reaches a maximum speed as the substrate concentration reaches levels that saturate the available enzyme molecules (see Figure 27.10).

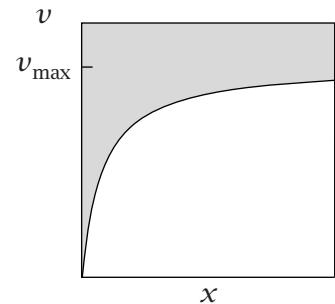


Figure 27.10 The rate v of enzyme catalysis saturates ($v = v_{\max}$) at high substrate concentrations x .

The Michaelis-Menten Model Describes Saturation in Rate Processes

Among the most important reactions in biology is



where E is the enzyme, S is the substrate, ES is the enzyme with substrate bound, and P is the product. Assume that the rate indicated by k_2 is small enough for the binding of substrate to enzyme to be considered to be at equilibrium. The binding equilibrium constant is

$$K = \frac{[ES]}{[E]x}, \quad (27.24)$$

where $x = [S]$ is the substrate concentration, $[ES]$ is the concentration of the enzyme-substrate complex ES , and $[E]$ is the concentration of free enzyme. The velocity v of the reaction is defined in terms of the rate constant k_2 for the reaction step shown in Equation (27.23):

$$v = \frac{dP}{dt} = k_2[ES] = k_2K[E]x. \quad (27.25)$$

Because the total enzyme concentration is

$$E_T = [E] + [ES] = [E](1 + Kx), \quad (27.26)$$

you can get the velocity per enzyme molecule by dividing Equation (27.25) by Equation (27.26):

$$\frac{v}{E_T} = \frac{k_2K[E]x}{[E](1 + Kx)} = \frac{k_2Kx}{1 + Kx}. \quad (27.27)$$

Product formation is fastest when the enzyme is fully saturated by substrate, $Kx/(1 + Kx) = 1$. Then the maximum rate is $v_{\max} = k_2E_T$. Expressed in terms of this maximum rate, the velocity is

$$\frac{v}{v_{\max}} = \frac{Kx}{1 + Kx}. \quad (27.28)$$

Equation (27.28) is a Langmuir-like expression (see Equation (27.12)) because of an underlying saturable binding step—the substrate must bind the enzyme before the reaction can be catalyzed.

The **Michaelis-Menten** Equation (27.28) is often expressed in a slightly different way, in terms of the *dissociation constant* or *Michaelis constant* $K_m = 1/K$. K_m is the inverse of the binding constant when the rate constant k_2 is small enough for the binding to be in equilibrium, as we have assumed. In those terms, you have

$$\frac{v}{v_{\max}} = \frac{x/K_m}{1 + x/K_m} = \frac{x}{x + K_m}. \quad (27.29)$$

You can linearize Equation (27.29) in the same way as the Langmuir equation (Equation (27.13)):

$$\frac{v_{\max}}{v} = \frac{K_m + x}{x} \implies \frac{1}{v} = \frac{K_m}{v_{\max}} \frac{1}{x} + \frac{1}{v_{\max}}. \quad (27.30)$$

If a plot of $1/v$ against $1/x$ is linear, it supports the use of the Michaelis-Menten model. Then the slope will be K_m/v_{\max} and the intercept will be $1/v_{\max}$. Such plots are called *Lineweaver-Burk* or *double-reciprocal* plots. Figure 27.11 shows this type of plot for the phosphorylation of (*S*)-1-amino-2,3-propanediol by

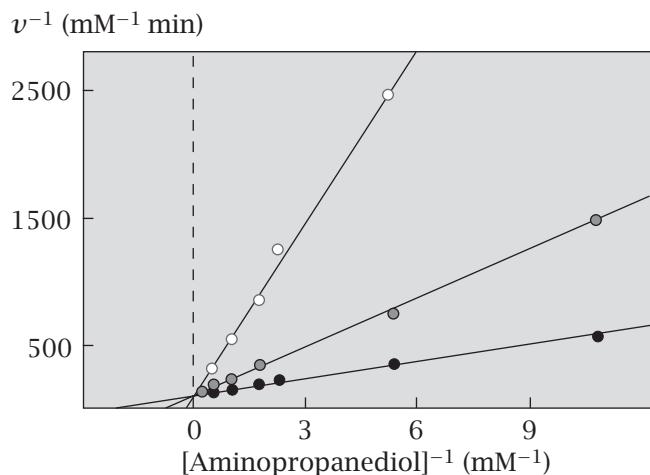


Figure 27.11 A double-reciprocal plot for the initial phosphorylation rate v of glycerol kinase against the concentration of the substrate aminopropanediol, at various concentrations of MgATP: (○) 0.126 mM; (●) 0.504 mM; (•) 2.52 mM. Source: J Kyte, *Mechanism in Protein Chemistry*, Garland, New York, 1995. Adapted from WB Knight and WW Cleland, *Biochemistry* **28**, 5728–5734 (1989).

glycerol kinase. Another example of saturable kinetics is the carrier-mediated transport of sugars and other molecules across cell membranes (see Figure 27.12). When sugar molecules are in sufficiently high concentration, the carrier proteins become saturated and reach full speed, and cannot transport sugars any faster.

The Langmuir model also explains how surfaces that bind reactants can catalyze reactions. A catalyst binds a reactant and activates it in some way for conversion to product. Figure 27.13 illustrates some of the possible mechanisms in which binding to a catalyst increases a reaction rate. When an appropriate surface binds a reactant, surface forces can help to pull the reactant into a structure resembling the product. This mechanism is called *strain*. Or, for catalyzing a bond-breaking reaction, surface forces can help to pull the atoms apart. For catalyzing bond-forming reactions, $A + B \rightarrow AB$, the surface can provide a good binding site for both A and B , bringing them into positions and/or orientations that are favorable for reaction.

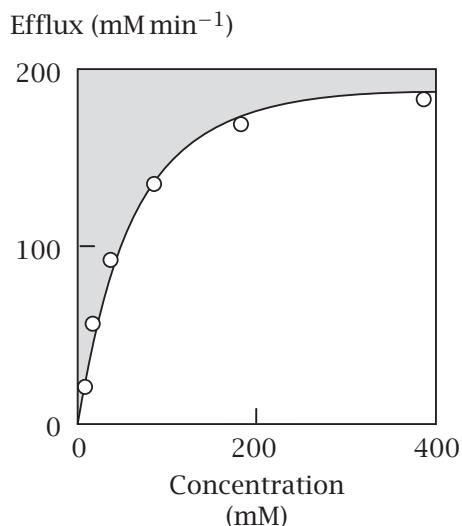


Figure 27.12 The rate of flow of galactose out of erythrocyte membranes as a function of the intracellular galactose concentration, relative to zero outside. Source: TF Weiss, *Cellular Biophysics*, Volume 1: *Transport*, MIT Press, Cambridge, MA, 1996. From: H Ginsburg, *Biochim Biophys Acta* **506**, 119–135 (1978).

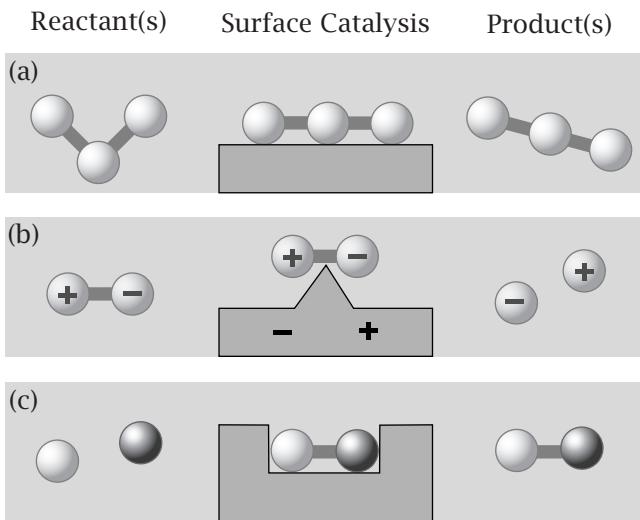


Figure 27.13 Some mechanisms of surface catalysis:
 (a) a surface might help to distort a reactant into a product-like form; (b) a charged surface can help to pull apart a charged molecule; (c) two reactants can be drawn together into a confined space to react.

Sabatier's Principle: Catalysts Should Bind Neither Too Tightly Nor Too Weakly

We now explore *Sabatier's principle*, named for the French chemist P Sabatier (1854–1941), who won the 1912 Nobel Prize in Chemistry for his work on catalysis. This principle explains why the best catalysts are surfaces that bind reactants neither too weakly nor too strongly.

Consider the reaction



Suppose S is a surface that catalyzes this reaction. A adsorbs to the surface. The binding of A to the surface 'activates' the molecules of A in some way, making them more ' B -like.' We want to know how the overall reaction rate depends on how tightly the catalyst binds to the reactants and products. Here's a simple model, due to AA Balandin [1–3]. We use the Michaelis–Menten model, Equation (27.23). Divide the reaction into two steps:



Step 1 is the adsorption of A to the surface to reach an activated state AS . We use the Langmuir model for this step. Step 2 involves the conversion of A to B and the desorption of B from the catalyst. This second step involves crossing an energy barrier (see Figure 27.14), which occurs with rate constant k_2 .

Equations (27.25) and (19.9) give the overall rate r of the reaction as the product (rate k_2 of conversion plus desorption) \times (number of molecules in state AS):

$$\frac{r}{AS} = k_2 \theta = \frac{k_2 K p}{1 + K p}, \quad (27.31)$$

where K is the affinity of ligand A for the surface, p is the gas pressure of A , A_s is the number of surface sites, and θ is the fraction of the surface that is covered by molecules of A . Now use the Arrhenius rate law, Equation (19.11),

Energy

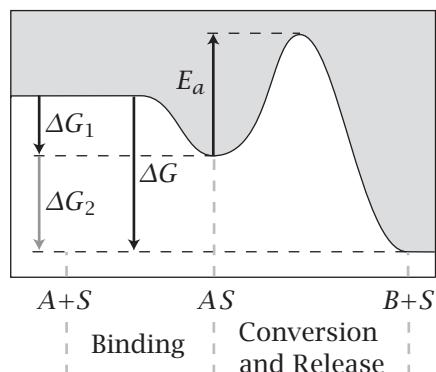


Figure 27.14 Definitions of the energies used in the model of Sabatier's principle. ΔG_1 is the free energy of binding A to the surface; ΔG_2 is the free energy of converting the bound complex AS to B and releasing it. $\Delta G = \Delta G_1 + \Delta G_2$ is the free energy of converting reactant A to product B . E_a is the activation energy of conversion.

to express the reaction rate k_2 in terms of the activation energy E_a for the conversion of A to B and the desorption of B :

$$k_2 = c_1 e^{-E_a/kT}, \quad (27.32)$$

where c_1 is a constant. If the Brønsted equation (19.40) holds, relating rates to equilibria, you have $E_a = a\Delta G_2 + b$ (see Figure 19.16), where $a > 0$ and $b > 0$ are constants. ΔG_2 is the free energy change for converting AS to B and releasing B from the surface (see Figure 27.14).

Now, relate the rate k_2 to the affinity ΔG_1 for the binding of A to S ; see Figure 27.14. Because the free energy difference ΔG for the conversion of A to B is independent of the catalyst, the free energies are related by $\Delta G_2 = \Delta G - \Delta G_1$. When this is expressed in terms of the binding free energy ΔG_1 , you have $E_a = a(\Delta G - \Delta G_1) + b = (\text{constant} - a\Delta G_1)$, and Equation (27.32) becomes

$$k_2 = c_2 e^{a\Delta G_1/kT} = c_2 K^{-a}, \quad (27.33)$$

where $K = e^{-\Delta G_1/kT}$ and $c_2 = c_1 e^{-(a\Delta G_1+b)/kT}$ is a constant, independent of affinity K .

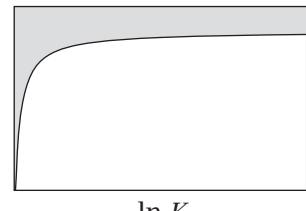
Equation (27.33) shows that the desorption rate k_2 decreases as the binding affinity of A for the surface increases (K increases), because $a > 0$. Substituting Equation (27.33) for k_2 into Equation (27.31) gives the overall rate of reaction in the presence of the surface:

$$\frac{r}{AS} = \frac{c_2 K^{1-a} p}{1+Kp}. \quad (27.34)$$

Figure 27.15 shows that there is an optimal binding affinity for catalytic effectiveness. When A binds the catalyst too weakly, the reaction is slow because there is too little reactant on the surface. When B binds the catalyst too strongly, the reaction is slow because the product is slow to leave the surface, occupying sites where new A molecules could bind. Such plots are sometimes called *volcano curves* because of their inverted 'V' shapes. Figure 27.16 shows that the rate of formic acid dehydrogenation depends on the type of metal surface that is used as a catalyst. The catalytic power has a volcano shape as a function of the catalyst binding enthalpy ΔH , corresponding roughly to $\ln K$ in our simple model. The main idea expressed by this model is that the best catalysts bind neither too weakly nor too strongly. Similarly, the most effective enzymes have

(a) Step 1

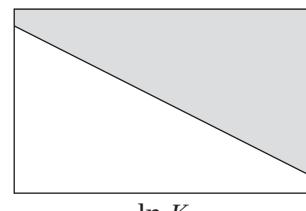
$\ln(\text{amount bound})$



$\ln K$

(b) Step 2

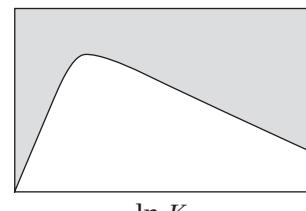
$\ln(\text{rate})$



$\ln K$

(c) Overall Reaction

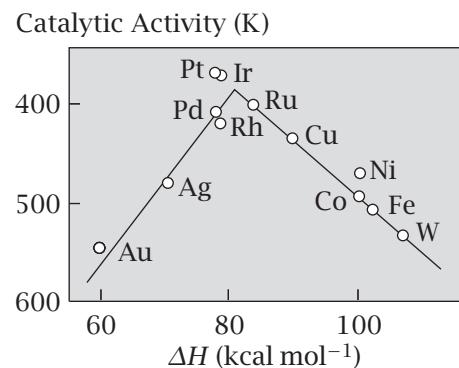
$\ln(\text{rate})$



$\ln K$

Figure 27.15 In a transformation of A to B catalyzed by surface S : (a) the amount of A bound to S increases with binding affinity K ; (b) the rate of release of B decreases with binding affinity K (Equation (27.33)); (c) the net result of both steps, binding and release, is an increase then a decrease of overall rate with K (Equation (27.34)).

Figure 27.16 A volcano curve shows the catalytic activities of metal surfaces in formic acid dehydrogenation, as a function of the enthalpy ΔH of formation of the metal formates. The enthalpies represent the strength of the substrate–surface interaction. Catalytic activity is given in kelvins, the temperature required to achieve 50% conversion. Source: BC Gates, *Catalytic Chemistry*, Wiley, New York, 1992. Data are from WJM Rootsart and WMH Sachtler, *Z Phys Chem NF* **26**, 16 (1960).



the two properties (a) that they have a relatively high affinity for the transition state of the reaction and (b) they do not bind reactants too weakly or products too tightly [4].

Summary

Ligands bind to surfaces and to sites on molecules. Increasing the concentration of a ligand increases the coverage of the available binding sites and leads to saturation. The Langmuir model treats binding and saturation when binding sites are independent. It accounts for gases sticking to metals, pH titration, rates of catalysis, and the saturable rates of transport through cell membranes. A catalyst should bind strongly enough to capture the reactant, but weakly enough to release the product. In the next chapter, we go beyond the model of binding to independent sites, and consider cooperative binding processes.

Problems

1. Adsorption in gas masks. You have a gas mask that adsorbs a certain toxic compound with adsorption constant $K = 1000 \text{ atm}^{-1}$ at $T = 25^\circ\text{C}$, and binding energy $w = -5 \text{ kcal mol}^{-1}$. In the desert, where the temperature is $T = 35^\circ\text{C}$, what is the adsorption constant?

2. Binding in solution. A ligand A in solvent S binds to a cavity in protein P , as shown in Figure 27.17.

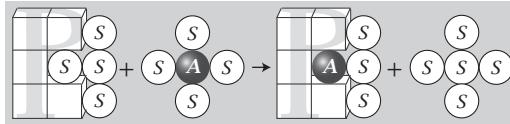


Figure 27.17 Ligand A displaces solvent S when it binds to a cavity in protein P .

- Write an expression for the log of the association constant, $\ln K_{\text{assoc}}$, for the case in which the binding pocket is originally occupied by a solvent molecule, which is then displaced by the ligand.
- Write an expression for the association constant, $\ln K_{\text{assoc}}$, for the case in which the binding pocket is originally an empty cavity, not occupied by a solvent molecule, before occupation by the ligand.
- If the magnitude of the solvent-solvent attraction dominates all the other interactions, and is equal to $w_{ss} = -4 \text{ kcal mol}^{-1}$ at room temperature, which model (a or b), leads to the stronger ligand binding, and by how much?

3. Polymer adsorption. You are designing a new drug delivery polymer to which you plan to covalently attach an active compound. You have three polymers, A , B , and C , with different properties, but they all have the drawback that they adsorb to some unidentified site in the body. Adsorption is of the Langmuir type. At a concentration of 1 mM , half of the body's binding sites are filled with polymer A . At a concentration of 5 mM , half of the sites are filled with B . At 10 mM , half of the sites are filled with C . You have the further problem that you can attach only one-tenth as much drug to polymer A as to either B or C .

- What is the desorption (dissociation) constant for removing polymer B from the body's binding sites?
- What is the best approach for delivering the maximum amount of drug with the minimum amount of adsorption?

4. Chromatographic separations. A typical chromatography column is 15 cm long, has a phase ratio $\phi = 0.2$, and has a mobile phase velocity of 2 mm s^{-1} .

- What is the velocity of a solute that has a partition coefficient $K = 20$ for binding to the stationary phase?

(b) Suppose you can distinguish two peaks as close as 2 mm together when they reach the end of the column. If one peak corresponds to $K = 20$, what is the partition coefficient for the other peak?

5. How binding affects catalysis. When a catalyst binds a substrate with binding constant K , the reaction rate r is given by

$$r = \frac{k_o K^{1-a} p}{1 + K p},$$

where p is the gas pressure of the substrate, and k_o and a are constants. Plot $r(K)$ for $a < 0$ and explain what it means.

6. The binding affinity of a ligand to a protein. Figure 27.18 shows the binding constant K for binding cytidine monophosphate to ribonuclease A , as a function of $1/T$ where T is temperature. For $T = 300 \text{ K}$, compute

- the enthalpy;
- the Gibbs free energy;
- the entropy of binding.

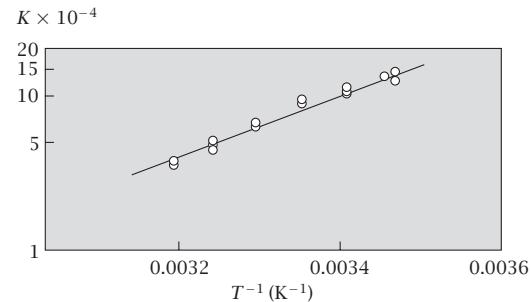


Figure 27.18 Source: IM Klotz, *Ligand-Receptor Energetics: A Guide for the Perplexed*, Wiley, New York, 1997. Data are from H Naghibi, A Tamura, and JM Sturtevant, *Proc Natl Acad Sci USA* **92**, 5597–5599 (1995).

7. Ligand-protein binding thermodynamics. A drug D binds to a protein P with binding constant $K_1 = 10^6$ at $T_1 = 300 \text{ K}$ and with $K_2 = 1.58 \times 10^6$ at $T_2 = 310 \text{ K}$.

- Calculate ΔG° for binding at $T = T_1 = 300 \text{ K}$.
- Assuming that the binding enthalpy ΔH° is independent of temperature, calculate ΔH° .
- Assuming that the binding entropy ΔS° is independent of temperature, calculate ΔS° at $T_1 = 300 \text{ K}$.

8. Binding of insulin to its receptor. Figure 27.19 shows the binding of insulin to the insulin receptor in a detergent solution. Use the Langmuir model to estimate the binding constant.

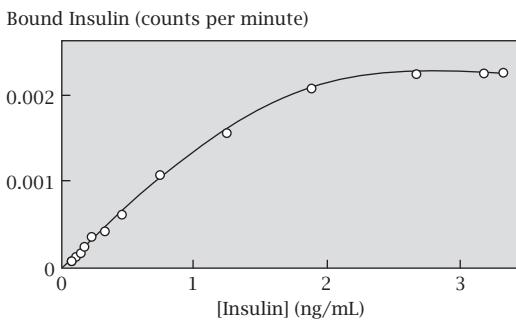


Figure 27.19 Source: J Kyte, *Mechanism in Protein Chemistry*, Garland, New York, 1995. Adapted from WB Knight and WW Cleland, *Biochemistry* **28**, 5728–5734 (1989).

9. Ion flow through membrane pores. Figure 27.20 shows the passive flow of sodium ion current through a membrane channel.

- (a) Make a model for the ion current $i(x)$ as a function of sodium ion concentration $x = [\text{Na}^+]$.
- (b) Write an expression that might linearize the data.

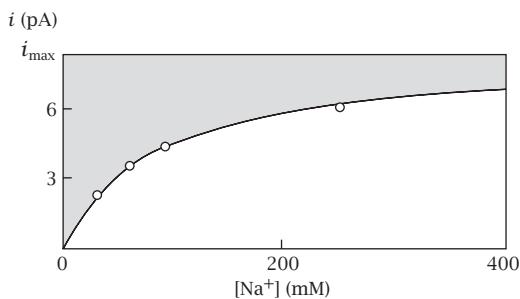
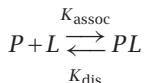


Figure 27.20 Source: B Hille, *Ionic Channels of Excitable Membranes*, Sinauer Associates, Sunderland, MA, 1984. Data are from R Coronado, RL Rosenberg, and C Miller, *J Genl Physiol* **76**, 425 (1980).

10. Saturation ligand concentrations. What concentration x of a ligand L do you need in solution in order to saturate three-quarters of the binding sites for a protein P if the dissociation constant for the reaction below is $K = 10 \text{ nM}$?



11. Binding an acidic drug to a protein. A spherical ligand L of radius $a = 3.5 \text{ \AA}$ binds to a protein P (see Figure 27.21), where $\ln K_{\text{bind}} = -\Delta\mu^\circ/(RT)$ at equilibrium at $T = 300 \text{ K}$. The ligand binding involves both a hydrophobic interaction and an electrostatic interaction $\Delta\mu^\circ = \Delta\mu_{\text{H}\phi}^\circ + \Delta\mu_{\text{estat}}$, where $\Delta\mu_{\text{H}\phi}^\circ = -2.4 \text{ kcal mol}^{-1}$.

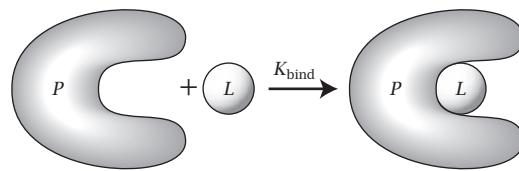


Figure 27.21 Ligand binding to a protein in Problem 10.

- (a) What is the value of $\Delta\mu_{\text{H}\phi}^\circ/RT$?
- (b) The ligand is a derivative of lactic acid $HL \xrightarrow{K_a} L^- + \text{H}^+$ with a $pK_a = 4.0$. Derive an expression for the fractional charge $f = [L^-]/[L^-] + [\text{HL}]$ on the ligand, in terms of K_a and $[\text{H}^+]$.
- (c) If the charge is $z = -1$ on the ligand, make a rough plot of the net charge fz versus pH over the range from pH 2 to pH 7.
- (d) Write an expression for $\Delta\mu_{\text{estat}}^\circ/(RT)$ as a function of f , the radius of the ligand a , the dielectric constants $D_{\text{water}} \approx 80$ and $D_{\text{protein}} \approx 4$, and the Bjerrum length l_{Bv} in a vacuum.
- (e) At roughly what pH will the ligand no longer bind to the protein?

References

- [1] RI Masel, *Principles of Adsorption and Reaction on Solid Surfaces*, Wiley, New York, 1996.
- [2] AA Balandin, *Adv Catal*, **10**, 96–129 (1958).
- [3] AA Balandin, *Adv Catal*, **19**, 1–210 (1969).
- [4] A Fersht, *Enzyme Structure and Mechanism*, 2nd edition, WH Freeman, San Francisco, 1985.

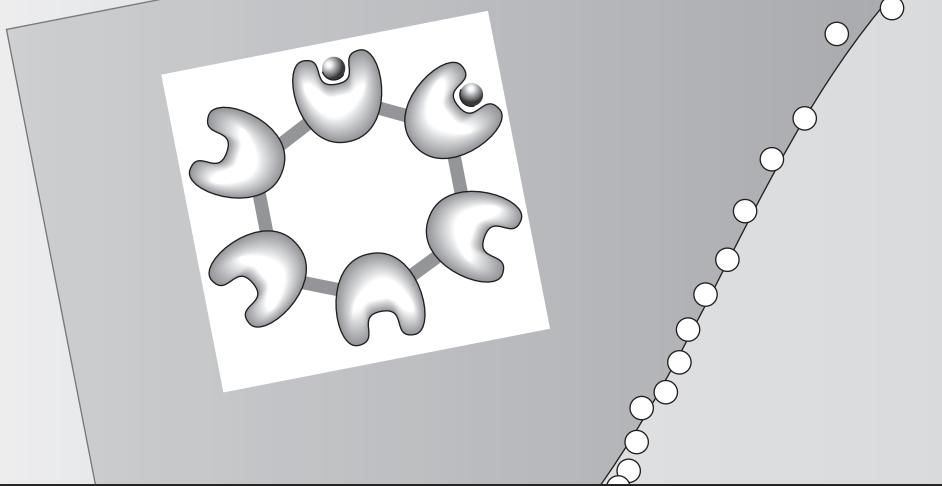
Suggested Reading

- Excellent treatments of surfaces, adsorption, and catalysis are:
- AW Adamson and AP Gast, *Physical Chemistry of Surfaces*, 6th edition, Wiley, New York, 1997.
 - RPH Gasser, *An Introduction to Chemisorption and Catalysis by Metals*, Oxford University Press, New York, 1985.
 - MJ Jaycock and GD Parfitt, *Chemistry of Interfaces*, Halstead Press, New York, 1981.
 - RI Masel, *Principles of Adsorption and Reaction on Solid Surfaces*, Wiley, New York, 1996.
 - H Maskill, *The Physical Basis of Organic Chemistry*, Oxford University Press, New York, 1985.
 - GD Parfitt and CH Rochester, eds, *Adsorption from Solution at the Solid/Liquid Interface*, Academic Press, New York, 1983.

- DM Ruthven, *Principles of Adsorption and Adsorption Processes*, Wiley, New York, 1984.
- RA van Santen and JW Niemantsverdriet, *Chemical Kinetics and Catalysis*, Plenum Press, New York, 1995.
- GA Somorjai and Y Li, *Introduction to Surface Chemistry and Catalysis*, 2nd edition, Wiley, New York, 2010.
- A Zangwill, *Physics at Surfaces*, Cambridge University Press, New York, 1988.
- An exceptionally detailed and careful analysis of membrane transport is:
- TF Weiss, *Cellular Biophysics*, Volume 1: *Transport*, MIT Press, Cambridge, MA, 1996.
- Concise introductions to ligand-protein binding include:
- IM Klotz, *Ligand-Receptor Energetics: A Guide for the Perplexed*, Wiley, New York, 1997.
- DJ Winzor and WH Sawyer, *Quantitative Characterization of Ligand Binding*, Wiley-Liss, New York, 1995.
- J Wyman and SJ Gill, *Binding and Linkage: Functional Chemistry of Biological Macromolecules*, University Science Books, Mill Valley, CA, 1990.
- Lattice model of adsorption:
- TL Hill, *Introduction to Statistical Thermodynamics*, Addison-Wesley, Reading, MA, 1960 (reprinted by Dover Publications, New York, 1986).
- For catalysis in biology see:
- A Fersht, *Enzyme Structure and Mechanism*, 2nd edition, WH Freeman, San Francisco, 1985.
- WP Jencks, *Catalysis in Chemistry and Enzymology*, McGraw-Hill, New York, 1969 (reprinted, with a new appendix, by Dover Publications, New York, 1987).
- J Kyte, *Mechanism in Protein Chemistry*, Garland, New York, 1995.

This page is intentionally left blank.

28 Multi-site & Cooperative Ligand Binding

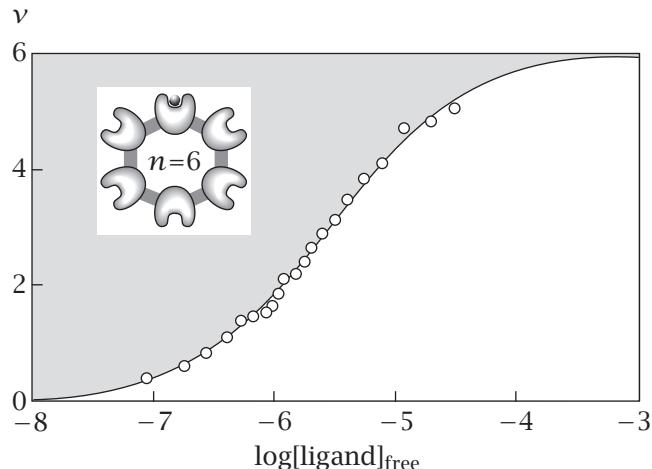


Some binding processes are *cooperative*, rather than independent. One molecule binds to a target; then a second molecule binds with a different affinity. In such cases, the independent-site model of Chapter 27 is too simple. Coupled binding is the core operating principle of molecular machines in biology, described in the next chapter. Coupled binding also explains self-assembly processes, such as the formation of soap micelles and lipid bilayers. Here, we introduce the methodology for modeling cooperative binding, called *binding polynomials*.

In experiments, you can measure *binding curves*—for example, how many ligand molecules bind to a protein for a given concentration of ligand in solution. To interpret binding experiments, you make a model of the binding process and derive its *binding polynomial*. You start with the simplest models (models with the fewest parameters) that are consistent with the data and any known symmetries. From the binding polynomial, you calculate the number of bound ligand molecules as a function of the bulk ligand concentration, adjusting the model parameters to best fit the data. If that model does not fit the data to within acceptable error, you try a different model. Here, we show how to make such models.

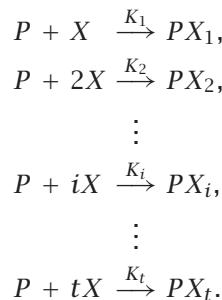
Figure 28.1 shows data for the binding of a ligand (TNP-ATP, trinitrophenyl adenosine triphosphate) to DnaB helicase protein, a hexamer of six identical subunits. This protein is involved in the replication of DNA. The figure shows that the average number of ligands bound per hexamer approaches six, or one ligand bound per protein monomer, at high ligand concentrations.

Figure 28.1 A binding curve. Nucleotide ligands bind to DnaB, a hexameric protein. The average number v of ligands bound per hexamer protein is a function of the ligand concentration in solution. Source: W Bujalowski and MM Klonowska, *Biochemistry* **32**, 5888–5900 (1993).



Binding Polynomials Are Used to Compute Binding Curves

Consider the *multiple binding equilibria* of a ligand X with $i = 1, 2, 3, \dots, t$ different binding sites on a protein or polymer P :



Each equilibrium constant K_i is defined by

$$K_i = \frac{[PX_i]}{[P]x^i},$$

where $x = [X]$ is the concentration of free ligand, $[P]$ is the concentration of free P molecules, and $[PX_i]$ is the concentration of those P molecules that have i ligands bound. There can be many different ligation states of P molecules in solution at the same time. Some P molecules contain no ligand; their concentration is $[P]$. Some P molecules contain one ligand molecule; their concentration is $[PX_1]$. Some P molecules contain i ligands; their concentration is $[PX_i]$.

What are the fractions of all the P molecules in solution that are i -liganded? To compute fractions, you need a denominator. So, first, sum the concentrations of all of the species of P :

$$\begin{aligned} Q[P] &= [P] + [PX_1] + [PX_2] + \dots + [PX_t] \\ &= [P](1 + K_1x + K_2x^2 + K_3x^3 + \dots + K_tx^t) \end{aligned}$$

$$\Rightarrow Q = 1 + K_1x + K_2x^2 + K_3x^3 + \dots + K_t x^t \\ = \sum_{i=0}^t K_i x^i. \quad (28.1)$$

This sum is called the *binding polynomial* Q . We defined Q in this way (factoring out $[P]$) for convenience, to be dimensionless. The fraction of P molecules that are in the i -liganded state is $[PX_i]/(Q[P])$.

A key quantity for comparing models with experiments is the average number $v = \langle i \rangle$ of ligands bound per P molecule, as a function of the ligand concentration x . To compute this average, use $\langle i \rangle = \sum i p(i)$, where $p(i) = [PX_i]/(Q[P])$ is the probability that a P molecule is in state i :

$$v = \langle i \rangle = \frac{[PX_1] + 2[DX_2] + 3[DX_3] + \dots + t[DX_t]}{[P] + [PX_1] + [PX_2] + [PX_3] + \dots + [PX_t]} \\ = \frac{[P](K_1x + 2K_2x^2 + 3K_3x^3 + \dots + tK_tx^t)}{[P](1 + K_1x + K_2x^2 + K_3x^3 + \dots + K_tx^t)} \\ = Q^{-1} \sum_{i=0}^t i K_i x^i \\ = \frac{x}{Q} \frac{dQ}{dx} = \frac{d \ln Q}{d \ln x}. \quad (28.2)$$

See Equations (C.4) and (C.5) in Appendix C.

The binding polynomial Q is the sum over all possible ligation states. Q resembles a partition function. Whereas a partition function is a sum over all possible energy levels, the binding polynomial Q is the sum over all possible ligation states. Equation (28.2) shows that the derivative of Q with respect to x gives an average number of ligands bound per P molecule. It is similar to the temperature derivative you would take of a partition function to get the average energy per molecule, Equation (10.33).

Other properties can also be computed from the binding polynomial. For example, you can find the *fluctuations* in the number of ligands bound from the second moment $\langle i^2 \rangle$. Take the second derivative of the binding polynomial:

$$x \frac{d}{dx} \left(x \frac{dQ}{dx} \right) = \sum_{i=0}^t i^2 K_i x^i. \quad (28.3)$$

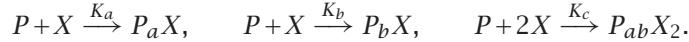
Then the second moment, which is used to determine the variance in the number of ligands bound, is (see Equation (C.6) in Appendix C)

$$\langle i^2 \rangle = \frac{x}{Q} \frac{d}{dx} \left(x \frac{dQ}{dx} \right) = \frac{\sum_{i=0}^t i^2 K_i x^i}{\sum_{i=0}^t K_i x^i}. \quad (28.4)$$

Binding polynomials are useful for describing the cooperativity of ligand binding. The simplest case of cooperative binding is a molecule P that has two binding sites that are not independent of each other.

The Simplest Model of Binding Cooperativity Involves Two Binding Sites

Suppose that a ligand molecule of type X can bind to P either at site a with an affinity described by the equilibrium binding constant K_a or at site b with affinity K_b , or that two ligand molecules X can bind at the two sites at the same time with affinity K_c :



The equilibrium constants are

$$K_a = \frac{[P_aX]}{[P]x}, \quad K_b = \frac{[P_bX]}{[P]x}, \quad K_c = \frac{[P_{ab}X_2]}{[P]x^2},$$

where $x = [X]$ is the concentration of free ligand, $[P]$ is the concentration of P molecules that have no ligands bound, $[P_aX]$ and $[P_bX]$ are the concentrations of the two types of singly liganded P molecules, and $[P_{ab}X_2]$ is the concentration of doubly liganded P .

We want to find v , the fraction of binding sites on P that are filled, as a function of the ligand concentration x . The sum over all the ligation states of P is

$$\begin{aligned} [P]Q &= [P] + [P_aX] + [P_bX] + [P_{ab}X_2] \\ &= [P](1 + K_a x + K_b x + K_c x^2), \end{aligned} \tag{28.5}$$

where

$$Q = 1 + K_a x + K_b x + K_c x^2 \tag{28.6}$$

is the binding polynomial for this system. The fraction of P molecules that have a single ligand bound *at site a* is $[P_aX]/Q[P]$. The fraction of P molecules that have a single ligand bound *at either site* is $([P_aX] + [P_bX])/Q[P]$. The fraction of P molecules that have two ligands bound is $[P_{ab}X_2]/Q[P]$.

The average number of ligands bound per P molecule, $0 \leq v \leq 2$, is

$$\begin{aligned} v(x) &= \frac{[P_aX] + [P_bX] + 2[P_{ab}X_2]}{[P]Q} \\ &= \frac{K_a x + K_b x + 2K_c x^2}{1 + K_a x + K_b x + K_c x^2} = \frac{d \ln Q}{d \ln x}, \end{aligned} \tag{28.7}$$

because $d \ln Q / d \ln x = (x/Q)(dQ/dx)$ and $dQ/dx = K_a + K_b + 2K_c x$.

When the two sites are independent, $K_c = K_a K_b$. Then there is no cooperativity, so Equation (28.6) reduces to

$$Q = (1 + K_a x + K_b x + K_a K_b x^2) = (1 + K_a x)(1 + K_b x), \tag{28.8}$$

and the average number of ligands bound is given by

$$v = \frac{x(K_a + K_b + 2K_a K_b x)}{(1 + K_a x)(1 + K_b x)} = \frac{K_a x}{1 + K_a x} + \frac{K_b x}{1 + K_b x}, \tag{28.9}$$

which corresponds to two Langmuir binding events, one at each site.

Binding *cooperativity* occurs when $K_c \neq K_a K_b$. *Positive cooperativity* occurs when $K_c > K_a K_b$, that is, when two ligands bind with higher affinity than would be expected from the two individual binding affinities alone. *Negative cooperativity* or *anti-cooperativity* occurs when $K_c < K_a K_b$. Then the two ligands bind with an overall affinity that is lower than would have been expected from the individual affinities.

The treatment above is *site-based*. It is useful when experiments provide all three quantities K_a , K_b , and K_c . To use it, you need a way to distinguish when your ligand is binding at site *a* versus site *b*. But this amount of information is often not available. Notice that the terms K_a and K_b always appear as the sum $K_a + K_b$ in Equation (28.7). If you use the binding constants as parameters to fit experimental data of v versus x , and if you have no additional information, you won't be able to choose K_a and K_b independently. You can only determine their sum, so you are only justified in using two independent parameters to fit the data, not three. The *stoichiometry-based* model in the next section uses only two parameters to describe the same two-site binding process.

The Stoichiometric Approach to Two Binding Sites

Suppose that one ligand X can bind to P with equilibrium constant K_1 and the second ligand X can bind with equilibrium constant K_2 :



We are no longer assuming that we can distinguish sites *a* and *b*. We only know the number of ligands bound to P . When the reactions are written in this form, the equilibrium constants are

$$K_1 = \frac{[PX_1]}{[P]x} \quad \text{and} \quad K_2 = \frac{[PX_2]}{[PX_1]x} = \frac{[PX_2]}{K_1[P]x^2},$$

where $x = [X]$ is the concentration of free ligand, $[P]$ is the concentration of free P molecules, $[PX_1]$ is the concentration of all P molecules with one ligand bound and $[PX_2]$ is the concentration of P molecules with two ligands bound. Now the binding polynomial is

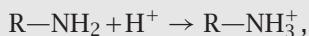
$$Q = 1 + K_1x + K_1K_2x^2. \tag{28.10}$$

In this stoichiometry-based approach, the binding curve is

$$v(x) = \frac{d \ln Q}{d \ln x} = \frac{K_1x + 2K_1K_2x^2}{1 + K_1x + K_1K_2x^2}. \tag{28.11}$$

If you don't independently know the individual site affinities, then this stoichiometry-based model is preferable to the site-based model because it requires one less parameter. Example 28.1 demonstrates both approaches.

EXAMPLE 28.1 Two binding sites for titratable protons on glycine. The amino acid glycine has two sites that bind a proton. At the amino end,



and at the carboxyl end,

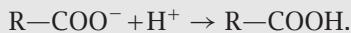


Figure 28.2 shows the binding curve $v(x)$, the average number of protons bound, versus x . (In the figure, $x = [\text{H}^+]$ is plotted as pH = $-\log[\text{H}^+]$.) At high pH (low $[\text{H}^+]$), no ligands are bound. At intermediate pH, one ligand is bound. And at low pH (high $[\text{H}^+]$), both binding sites are filled with ligand.

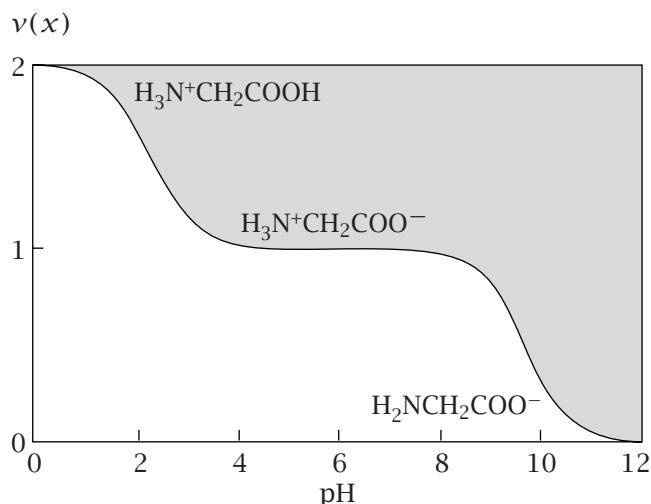
The two separate transitions in Figure 28.2 show that the two sites have very different affinities, $K_1 \gg K_2$. In that case, for low to intermediate ligand concentrations ($K_2x \ll 1$), Equation (28.11) gives $v(x) \approx K_1x/(1+K_1x)$. You can find K_1 from the midpoint ($v = 1/2$) of the transition on the right side of Figure 28.2, where $x = K_1^{-1} = 2 \times 10^{-10} \text{ M}$. So $K_1 = 5 \times 10^9 \text{ M}^{-1}$. To get K_2 , focus on the second ligand (proton), from $v = 1$ to $v = 2$. For this ligand, it is most convenient to look at the quantity $v-1$, which ranges from 0 to 1. Equations (28.10) and (28.11) give

$$v-1 = \frac{K_1x + 2K_1K_2x^2 - Q}{Q}. \quad (28.12)$$

Because the second titration takes place for intermediate ligand concentrations x , where $K_1x \gg 1$, $K_2x \approx 1$, and $Q \approx K_1x + K_1K_2x^2$, Equation (28.12) gives

$$\begin{aligned} v-1 &\approx \frac{K_1K_2x^2}{K_1x + K_1K_2x^2} \\ &= \frac{K_2x}{1 + K_2x}. \end{aligned}$$

Figure 28.2 Glycine has two titratable binding sites for protons. When the concentration of ligand $[\text{H}^+]$ is low (high pH), the number of protons bound is small ($v \rightarrow 0$). As the ligand concentration increases (toward low pH), the glycines become saturated and have $v \rightarrow 2$ protons each. Source: J Wyman and SJ Gill, *Binding and Linkage*, University Science Books, Mill Valley, CA, 1990.



The midpoint of this second titration is given from Figure 28.2 by $K_2^{-1} = 5 \times 10^{-3} \text{ M}$. So $K_2 = 2 \times 10^2 \text{ M}^{-1}$. K_1 and K_2 are the two affinities from the stoichiometric model. To obtain the same binding curve from the site-based model, equate the coefficients of the binding polynomials:

$$\begin{aligned} Q_{\text{site}} &= 1 + (K_a + K_b)x + (K_a K_b c)x^2 \\ &= Q_{\text{stoichiometric}} = 1 + K_1 x + K_1 K_2 x^2, \end{aligned} \quad (28.13)$$

where $c = K_c / (K_a K_b)$ defines the degree of cooperativity. Suppose one site has higher affinity than the other, and suppose we call that site a . Comparison of the first terms of Equation (28.13) gives

$$K_a \approx K_1 = 5 \times 10^9 \text{ M}^{-1}.$$

Comparing the second terms of Equation (28.13) gives

$$K_b c = K_2 = 2 \times 10^2 \text{ M}^{-1}.$$

Independent experiments show that $K_b = 2 \times 10^4 \text{ M}^{-1}$ [1], leading to the conclusion that $c = 10^{-2}$. This negative cooperativity has been attributed to electrostatic repulsion between the two sites [1].

'Cooperativity' Describes the Depletion of Intermediate States

In Chapter 26, we distinguished between two types of transition (see Figure 28.3): two-state and one-state. At the midpoint of a two-state transition, two states are populated, and the states between them have small populations. At the midpoint of a one-state transition, there is a single broad population distribution, so intermediates are highly populated.

You can apply these same classifications to binding cooperativity. Using the binding polynomial, Equation (28.10), you can express the relative populations of the three states as: no ligands bound, $1/Q$; one ligand, $K_1 x / Q$; and two ligands, $2K_1 K_2 x^2 / Q$. Figure 28.4(a) shows a one-state transition as a function of ligand concentration. At the left (small x), most P molecules are unliganded. In the middle (intermediate x), three states are equally populated: zero-liganded, one-liganded, and two-liganded. At the right (high x), most P molecules have two ligands. Figure 28.4(b) shows two-state cooperativity: at the midpoint concentration in x , most P molecules are either zero-liganded or two-liganded. There is little population of the one-liganded intermediate state. More quantitative details are given in Example 28.2.

EXAMPLE 28.2 Binding intermediate states and cooperativity. Figure 28.5 shows three different cases of systems that have zero, one, or two ligands bound. In each case, suppose that the midpoint of the binding curve is at $x_{\text{mid}} = 0.01$ (see Figure 28.3).

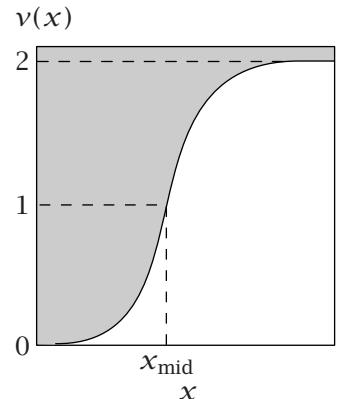
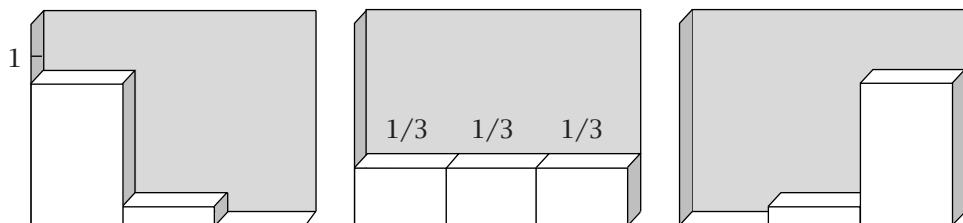


Figure 28.3 A binding curve. To determine the type of cooperativity, look at the ligation distribution at the transition midpoint, x_{mid} where $v = 1$. Examples of distributions are given in Figures 28.4 and 28.5.

(a) One-State Transition

Population



(b) Two-State Transition

Population

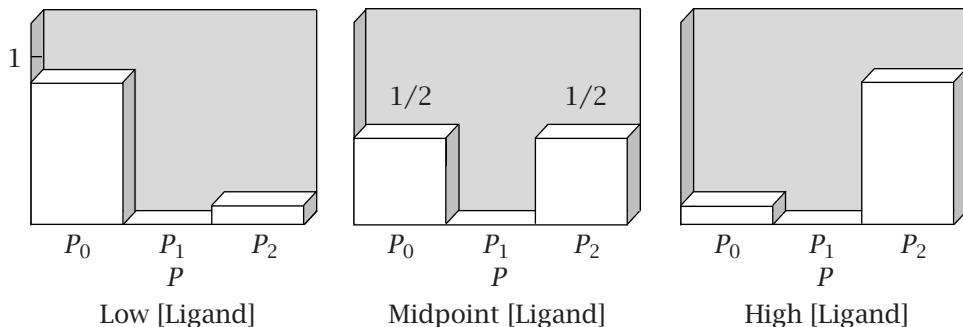


Figure 28.4 Ligation states for P molecules with two binding sites: P_0 is the fraction of P molecules with no ligands, P_1 is the fraction with one ligand bound, and P_2 is the fraction having two ligands bound. (a) One-state transition. (b) Two-state transition. From left to right: at low ligand concentrations, P molecules are mostly unliganded. At the midpoint of binding, a one-state transition has a substantial intermediate population while a two-state transition involves no intermediates. At high ligand, most P molecules are doubly ligated.

Suppose $K_1 = 1$ and $K_2 = 10^4$ (Figure 28.5(a)). Then Equation (28.10) gives $Q = 2.01$. The relative populations of the different ligation states are (0 ligands, 1 ligand, 2 ligands) = $(1, K_1 x_{\text{mid}}, K_1 K_2 x_{\text{mid}}^2) = (1, 0.01, 1)$. This gives two-state binding (see Figure 28.5). At low ligand concentrations, no ligands are bound. At high ligand concentrations, two ligands are bound. At the midpoint of the transition, there are no intermediates; P is mostly unligated or doubly ligated.

Now suppose instead that $K_1 = 1000$ and $K_2 = 10$ (Figure 28.5(b)). Then $Q = 12$ at $x_{\text{mid}} = 0.01$. This is an example of one-state binding. Starting at low ligand concentrations $x \ll 1$, the P molecules have no ligands. Increasing the ligand concentration leads to predominantly singly ligated P molecules, then ultimately to doubly ligated P molecules at high ligand concentrations. At the midpoint of the transition, intermediate states are populated. The relative populations are 1 : 10 : 1.

Finally, suppose instead that $K_1 = 100$ and $K_2 = 100$ (Figure 28.5(c)). Then $Q = 3$ at $x_{\text{mid}} = 0.01$. This system has equal populations of the three states, 1 : 1 : 1, at the midpoint of the binding curve. This binding is not cooperative.

Relative Populations of Ligation States

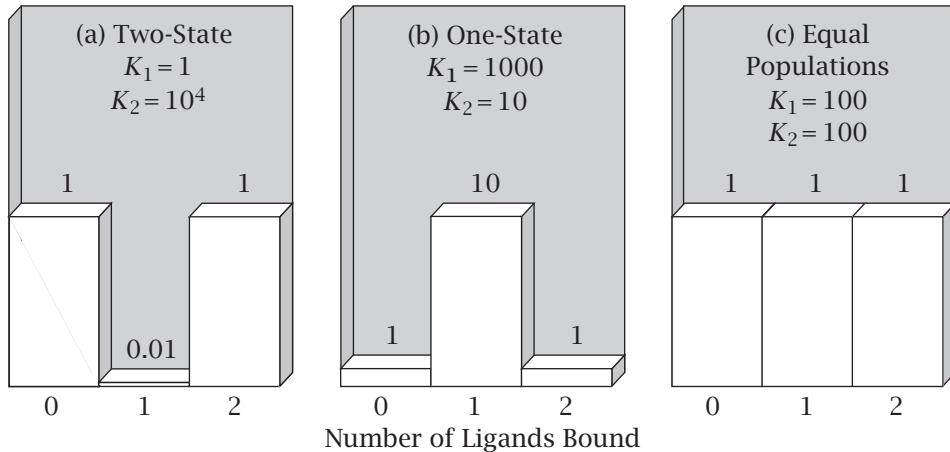


Figure 28.5 Relative populations of P molecules having zero, one, or two ligands bound at the midpoint of the binding transition described in Example 28.2: (a) two-state cooperativity; (b) one-state cooperativity; (c) independent, equal populations.

The next section shows how to construct binding polynomials for more complicated systems.

Binding Polynomials Can Be Constructed Using the Addition and Multiplication Rules of Probability

To predict binding equilibria, the first step is to devise an appropriate model and compute the binding polynomial. In this section, we show a short cut for writing the binding polynomial directly from the model, skipping the step of writing out the equilibria. We first show simple cases, but the power of the method is in describing more complex cases. The key to this approach is to use the addition and multiplication rules of probability (see Chapter 1).

EXAMPLE 28.3 The Langmuir model illustrates the binding polynomial. The binding polynomial is the sum of concentrations over all the ligation states of P . For the Langmuir model, this is

$$[P]Q = [P] + [PX] = [P](1 + Kx) \implies Q = 1 + Kx. \quad (28.14)$$

Using $dQ/dx = K$ and Equation (28.2), you can compute the fraction θ of sites filled as

$$\theta = \frac{d \ln Q}{d \ln x} = \frac{x}{Q} \frac{dQ}{dx} = \frac{Kx}{1 + Kx}. \quad (28.15)$$

You can also get the binding polynomial in Equation (28.14) by using the addition rule of probabilities described in Chapter 1. According to the addition rule, if two states are mutually exclusive (bound and unbound, for example), then you can sum their statistical weights, in the same way that terms are summed in partition functions. Use 1 as the statistical weight for the empty site and Kx as the statistical weight for the filled site, and add them to get Q .

Similarly, according to the multiplication rule (also described in Chapter 1), if two ligands bind independently, the contribution to the binding polynomial is the product of the two statistical weights. For example, for two independent sites, Equation (28.8) gives $Q = (1+K_a x)(1+K_b x)$. The a site can be empty OR filled, and, because these are mutually exclusive options, the addition rule dictates that their statistical weights are summed to get the factor $(1+K_a x)$. Likewise, the b site can be empty OR filled, and these are also mutually exclusive, giving the factor $(1+K_b x)$. Because a and b sites are independent of each other, the factors $(1+K_a x)$ and $(1+K_b x)$ are multiplied to get the full binding polynomial. The main value of this logic is for more complex equilibria, such as those in the next section.

***n* Independent Sites with Identical Affinities**

Suppose there are n independent binding sites on P , all having the same affinity, K . Using the addition rule, the statistical weight for each site is $(1+Kx)$. Because there are n independent sites, the binding polynomial is the product

$$Q = (1+Kx)^n. \quad (28.16)$$

Taking the derivative according to Equation (28.2) (see also Equation (C.4)) gives v , the *occupancy*, the average number of ligands bound per P molecule:

$$v(x) = \frac{x}{Q} \left(\frac{dQ}{dx} \right) = \frac{nxK(1+Kx)^{n-1}}{(1+Kx)^n} = \frac{nKx}{1+Kx}. \quad (28.17)$$

Equation (28.17) can also be derived from the Langmuir adsorption equation, because θ is the *fractional occupancy*, the fraction of sites filled:

$$\begin{aligned} \theta &= \frac{v}{n} = \frac{Kx}{1+Kx} \\ &= \frac{\text{average number of ligands bound per } P \text{ molecule}}{\text{total number of sites on each } P \text{ molecule}}. \end{aligned} \quad (28.18)$$

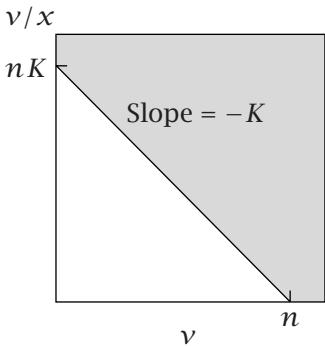


Figure 28.6 Scatchard plot of $v(x)/x$ versus v , where v is the average number of ligands bound per P molecule and x is the ligand concentration. When there are n independent binding sites with affinity K , the slope is $-K$ and the intercepts are nK and n .

The Scatchard Plot for n Independent Binding Sites

In the same way that a linear formulation of the Langmuir and Michaelis-Menten equations has practical advantages for fitting experimental data, the *Scatchard plot* (Figure 28.6) is a useful linearized form of Equation (28.17). Rearranging Equation (28.17) gives

$$\frac{v}{x} = \frac{nK}{1+Kx} \implies \frac{v}{x} + vK = nK \implies \frac{v}{x} = nK - vK. \quad (28.19)$$

A plot of v/x versus v will be linear with a slope of $-K$, and it will have an intercept (at $v = 0$) of nK , when a binding process involves n independent sites

of binding constant K . The limiting values at $v = 0$ and $v/x = 0$ are found by extrapolation, because they cannot be measured. When such a plot is not linear, however, you should not try to fit it with multiple straight lines: instead you should try a different model [2].

EXAMPLE 28.4 A Scatchard plot. Figure 28.7 shows a Scatchard plot for the binding of tryptophan to the *Escherichia coli* tryptophan aporepressor protein. The horizontal-axis intercept indicates that the number of sites is $n = 2$. The binding constant is found to be about $K = 2.0 \times 10^4 \text{ M}^{-1}$, taken either from the slope of this figure, or from the vertical-axis intercept, which is nK .

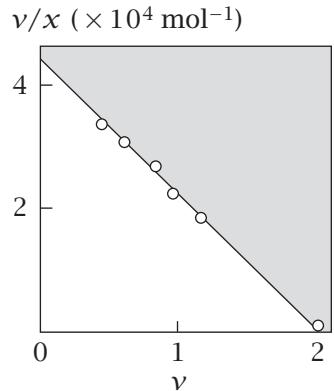
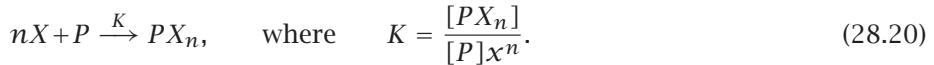


Figure 28.7 Scatchard plot for binding tryptophan to Trp aporepressor. Source: DN Arvidson, C Bruce, and RP Gunsalus, *J Biol Chem* **261**, 238–243 (1986).

The Hill Plot for Cooperative Binding

Ligand binding is sometimes highly cooperative: the number of ligands bound may be either zero or many, but nothing in between. If a P molecule binds either zero molecules or exactly n ligand molecules at a time, then the Hill model is useful:



The sum over the states of P is given by

$$[P]Q = [P] + [PX_n] = [P](1 + Kx^n), \quad (28.21)$$

so the binding polynomial is $Q = 1 + Kx^n$. Then $dQ/dx = nKx^{n-1}$, and the average number of ligands bound per P molecule is

$$v(x) = \left(\frac{d \ln Q}{d \ln x} \right) = \left(\frac{x}{Q} \right) \left(\frac{dQ}{dx} \right) = \frac{nKx^n}{1 + Kx^n}, \quad (28.22)$$

according to Equation (28.2).

For a Hill process, a plot of v versus x will be sigmoidal (see Figure 28.8(a)). The steepness of the transition region increases with n . To put Equation (28.22) into a linearized form, use the average number of ligands bound per site $\theta = v/n$:

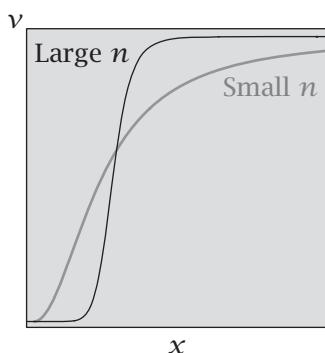
$$\theta = \frac{Kx^n}{1 + Kx^n}. \quad (28.23)$$

Because $1 - \theta = 1/(1 + Kx^n)$, you have

$$\frac{\theta}{1 - \theta} = Kx^n. \quad (28.24)$$

A *Hill plot* is $\ln[\theta/(1 - \theta)] = \ln K + n \ln x$ versus $\ln x$. The slope gives the Hill exponent (or Hill coefficient) n and the intercept is $\ln K$.

(a) Cooperative Binding



(b) Hill Plot

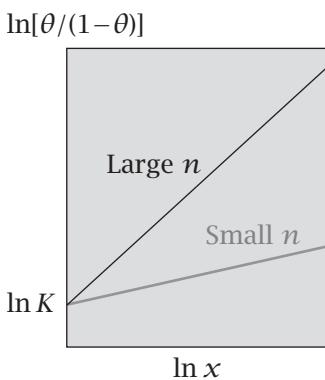


Figure 28.8 Cooperativity described by the Hill model. (a) When the fraction v of sites filled is plotted versus ligand concentration x , the steepness of the curve increases with n , the Hill coefficient. (b) The Hill plot linearizes the data. The Hill coefficient n is found from the slope and the binding constant K is found from the intercept (where $\ln x = 0$; the figure shows the range $x > 1$).

EXAMPLE 28.5 Biochemical switching is a cooperative process. Some biological processes are all-or-none, like electrical switches. For example, a dose of the hormone progesterone that is above a threshold concentration can trigger a frog's eggs to mature. Below the threshold concentration, the progesterone has no effect. Such behavior can be expressed in terms of large Hill coefficients, and may result from cooperative binding processes. Figure 28.9 shows a step in the progesterone process in which the concentration of a protein called MOS leads to a sharp increase in the phosphorylation of a protein called MAPK, with a Hill exponent $n \geq 5.1$. This is the Hill coefficient observed for groups of egg cells; within individual cells, the Hill coefficient is found to be much larger, $n \approx 35$ [3]. Such ultrasensitivity to ligand concentrations is often important in biological regulation.

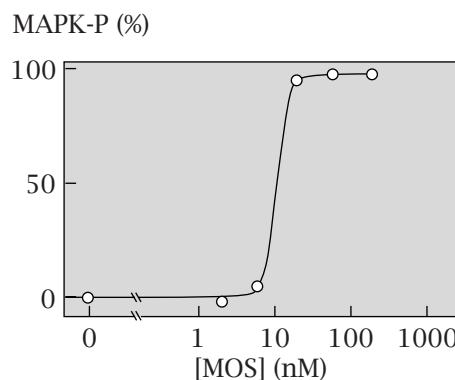


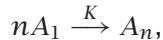
Figure 28.9 For the maturation of frog eggs, the production of phosphorylated mitogen-activated protein kinase (MAPK-P) is governed by small changes in the concentration of the MOS protein. This relationship can be described by the Hill model. Source: JE Ferrell and EM Machleder, *Science* **280**, 895–898 (1998).

Another type of cooperative process is the formation of soap micelles, lipid bilayer membranes, and other molecular assemblies.

Aggregation and Micellization Are Cooperative Processes

Surfactant molecules can assemble into noncovalent aggregates or clusters called *micelles* (see Figure 28.10). Assembly is cooperative. Such clusters are often spherical, cylindrical, or planar. A typical spherical micelle may contain 60 surfactant molecules and will usually have a narrow size distribution, say 55–65 surfactant molecules.

Here's a model for the cooperative assembly of surfactant molecules into micelles and bilayers. As a first approximation, suppose that every micelle contains exactly n surfactant molecules. The concentration of micelles is $[A_n]$. The concentration of free surfactant molecules in solution (which we call monomers) is $[A_1]$. The equilibrium is



where the equilibrium constant is

$$K = \frac{[A_n]}{[A_1]^n}.$$

The total number of *objects* (monomers plus micelles) in the solution (per unit volume) is $[A_1] + [A_n]$. The total number of *surfactant molecules*

is $[A_1] + n[A_n]$. So the number of molecules per object is

$$v = \frac{[A_1] + n[A_n]}{[A_1] + [A_n]}. \quad (28.25)$$

Let $x = [A_1]$, and use the definition of K to express Equation (28.25) as

$$v(x) = \frac{[A_1](1+nKx^{n-1})}{[A_1](1+Kx^{n-1})} = \frac{1+nKx^{n-1}}{1+Kx^{n-1}}. \quad (28.26)$$

This model predicts a sharp transition at a surfactant concentration called the *critical micelle concentration* (cmc); see Figure 28.11. Using Equation (28.26), you can see that the midpoint of the transition from monomers to micellar assemblies occurs at a concentration x that satisfies $x^{n-1} = 1/K$. At surfactant concentrations below the cmc, the solution contains mostly monomers. For small x , $x^{n-1} \ll 1/K$, and the number of surfactant molecules per object is $v \approx 1$. At low surfactant concentrations, the surfactant molecules cannot overcome the large translational entropy needed to come together in solution. However, at high surfactant concentrations, above the cmc, the surfactant molecules form micelles because the free energy of association of the hydrocarbon tails is more favorable than the loss in translational entropy. For large x , you have $x^{n-1} \gg 1/K$, so the number of surfactant molecules per object approaches n : $v \rightarrow (nKx^{n-1})/(Kx^{n-1}) = n$.

Why Do Micelles Have Such Well-Defined Finite Sizes?

Why aren't micelles either very small or very large? Why don't surfactant molecules grow into a macroscopic phase separation, for example? Let's generalize the model above. Rather than assuming that a micelle has a single fixed number of surfactant molecules, let's use the binding-polynomial method to allow for multiple equilibria and multiple sizes. Replace Equation (28.26) with

$$v(x) = \frac{1+2K_2x+3K_3x^2+\dots+nK_nx^{n-1}+\dots}{1+K_2x+K_3x^2+\dots+K_nx^{n-1}+\dots}, \quad (28.27)$$

to give $v(x)$, the number of surfactant molecules per micelle, as a function of the bulk concentration x of surfactant molecules. This binding-polynomial generalization allows you to explore why one particular micellar size is more favorable than another. K_n is the equilibrium constant for assembling a micelle starting from n monomeric surfactant molecules dissolved in water. You can express each K_n in terms of its chemical potential:

$$K_n = e^{-\Delta\mu_n^\circ/kT}, \quad (28.28)$$

where $\Delta\mu_n^\circ$ is the chemical potential change for forming an n -mer micelle from n monomers. In this generalized model, we are expressing that the partitioning, K_n and $\Delta\mu_n^\circ$, of a surfactant molecule depends on the size of the micelle, through n , the number of molecules in the micelle. We are allowing for the possibility that a surfactant molecule partitions more strongly into micelles of some sizes than of other sizes. What drives surfactant molecules to preferentially partition into micelles of different sizes? According to the model of C Tanford and J Israelachvili [4, 5], micelle formation is driven by the tendency

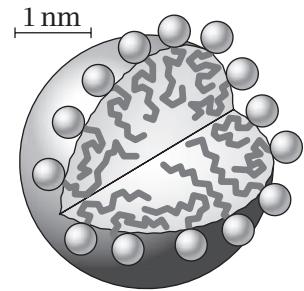


Figure 28.10 Micelles have polar or charged head groups on the outside and hydrocarbon tails on the inside.

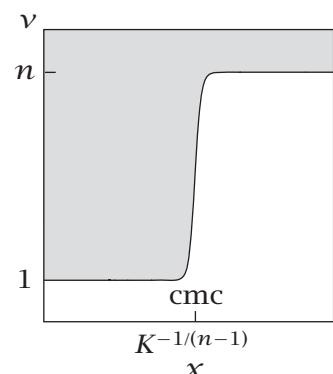


Figure 28.11 Micelle formation is cooperative. When the surfactant concentration is below the critical micelle concentration (cmc), most surfactant molecules are monomers. When the surfactant concentration is above the cmc, most molecules are in n -mer micelles.

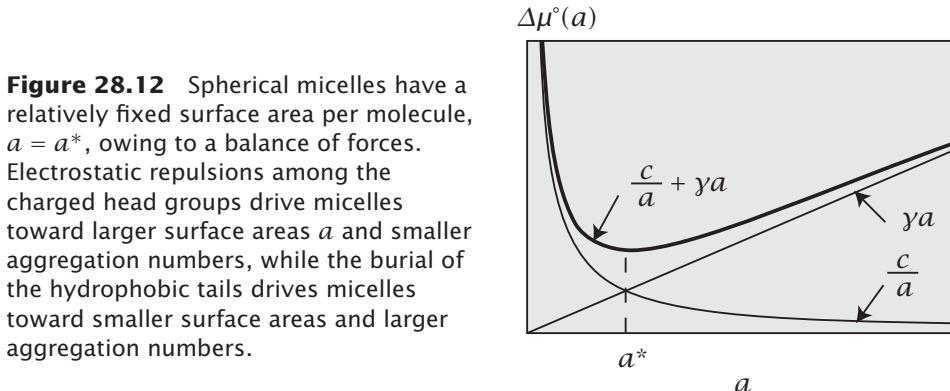


Figure 28.12 Spherical micelles have a relatively fixed surface area per molecule, $a = a^*$, owing to a balance of forces. Electrostatic repulsions among the charged head groups drive micelles toward larger surface areas a and smaller aggregation numbers, while the burial of the hydrophobic tails drives micelles toward smaller surface areas and larger aggregation numbers.

of soap molecules to come together to bury their hydrocarbon tails away from water into an oil-like core, but micellar growth is limited by the repulsions of the electrostatic charges among the surfactant head groups. A micelle should be big enough to have a hydrophobic core (to reduce the hydrocarbon surface exposure to water), but small enough that the head groups are well separated from each other by sufficient curvature of the surface. Consider spherical micelles. Instead of expressing $\Delta\mu_n^\circ$ as a function of the number n of monomers in the micelle, let's equivalently express it as $\Delta\mu^\circ(a)$, the chemical potential as a function of the surface area a of a micelle. Tanford and Israelachvili describe the preferential partitioning as a sum of two tendencies:

$$\Delta\mu_n^\circ = \Delta\mu^\circ(a) = \gamma a + \frac{c}{a}, \quad (28.29)$$

where γ is the interfacial tension between the oil-like spherical micelle with its surrounding water and c is a constant (see Figure 28.12). γa describes the hydrophobic effect; increasing the surface area a of the micelle is unfavorable. The hydrophobic free energy per molecule increases with a . c/a describes the head-group repulsions; increasing the surface area a is favorable. The electrostatic free energy decreases with a . The reason that micelles have a particular size is because the balance of these two tendencies leads to a stable state. Take the derivative $(d\Delta\mu^\circ(a)/da)_{a=a^*} = 0$ of Equation (28.29) to get the stable-state head-group area $a^* = \sqrt{c/\gamma}$.

To see why micelles have such a narrow distribution of sizes, substitute $c = \gamma(a^*)^2$ into Equation (28.29) to get $\mu^\circ(a) = (\gamma/a)[a^2 + (a^*)^2]$. Now use $a^2 + (a^*)^2 = (a - a^*)^2 + 2aa^*$ to get

$$\Delta\mu^\circ(a) = 2\gamma a^* + \frac{\gamma}{a}(a - a^*)^2 \quad (28.30)$$

Because $(a - a^*)^2$ is exponentiated in Equation (28.28), it predicts a Gaussian distribution of micellar sizes, centered around $a = a^*$, as shown in Figure 28.13. The width of the micelle size distribution (which we don't calculate here) is given by the variance of this Gaussian function.

What Determines the Shapes of Micelles?

Not all surfactant aggregates are spherical. Surfactants and lipids sometimes form cylindrical micelles or planar lipid bilayers or vesicles instead. The shape

Population of n -mer

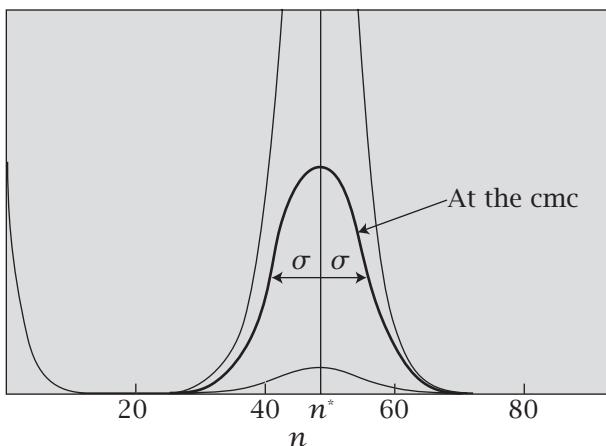


Figure 28.13 The populations of micelles having different numbers of surfactant molecules. σ is the standard deviation around the mean value, $n = n^*$.

of a surfactant aggregate is determined by the chemical structure of the surfactant molecules, by their solution concentration, and by variables such as temperature and pH. Surfactant molecules will assemble into spheres if the individual surfactant molecule has a chemical structure such that its head-group area is large enough relative to its tail, on average, to form a cone shape; see Figure 28.10. For a sphere, the volume is $v = (4/3)\pi R^3$ and the equilibrium area is $a^* = 4\pi R^2$. So, the radius of a spherical micelle will be $R = 3v/a^*$. To avoid holes inside, the micellar radius $R \leq \ell_c$ cannot exceed the extended length ℓ_c of the surfactant chain. These constraints mean that if your surfactant molecules have a large enough head-group area ($a^* \geq 3v/\ell_c$), they will assemble into micelles having a spherical shape. Correspondingly, surfactant molecules having intermediate head-group areas of $3v/\ell_c \geq a^* \geq 2v/\ell_c$ will form cylindrical micelles. And molecules having smaller head-group areas, $2v/\ell_c \geq a^* \geq v/\ell_c$, will form planar structures.

EXAMPLE 28.6 Relating head-group sizes to micellar shapes. For surfactants having a single hydrocarbon tail, use the approximation $v/\ell_c \approx 21 \text{ \AA}$ to estimate that surfactants with head groups bigger than $a^* = 63 \text{ \AA}^2$ will form spherical micelles, surfactant molecules with $a^* = 42 \text{ \AA}^2$ to 63 \AA^2 will form cylindrical micelles, and those with $a^* = 21 \text{ \AA}^2$ to 42 \AA^2 will form planar bilayers or vesicles. This shows how surfactant molecules having different molecular structures will lead to micelles having different shapes.

Example 28.7 describes the cooperative formation of multilayers of molecules on surfaces.

EXAMPLE 28.7 The BET adsorption model of layers of ligands. Sometimes ligands can ‘pile up’ on a site, layer upon layer, so a surface does not become saturated. This type of adsorption behavior can often be described by a model developed by S Brunauer, PH Emmett, and E Teller in 1938, called the BET model [6, 7]. The first molecule that adsorbs directly to a site on the surface binds with affinity K_1 . Second and subsequent ligands need not bind directly to the surface; they can bind on top of a molecule that has already bound (see Figure 28.14). Suppose that such ‘stacked’ ligands bind with affinity K_2 . The binding

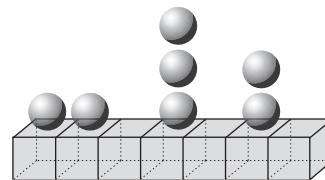


Figure 28.14 Ligands can bind on top of each other on surfaces in the BET model.

polynomial for a given site is

$$Q = 1 + K_1x + K_1K_2x^2 + K_1K_2^2x^3 + K_1K_2^3x^4 + \dots, \quad (28.31)$$

because a site can be unbound (with statistical weight 1) OR have one ligand stuck to the surface (with statistical weight K_1x) OR have two ligands piled up (with statistical weight $K_1K_2x^2$), etc. You can simplify Equation (28.31) to

$$\begin{aligned} Q &= 1 + K_1x \left[1 + K_2x + (K_2x)^2 + (K_2x)^3 + \dots \right] \\ &= 1 + \frac{K_1x}{1 - K_2x} = \frac{1 - (K_2 - K_1)x}{1 - K_2x}, \end{aligned} \quad (28.32)$$

by using Equation (C.6) in Appendix C: $1 + K_2x + (K_2x)^2 + (K_2x)^3 + \dots = 1/(1 - K_2x)$, if $K_2x < 1$. To find ν , the average number of ligands bound per surface site, as a function of the ligand concentration, take the derivative of Equation (28.32)

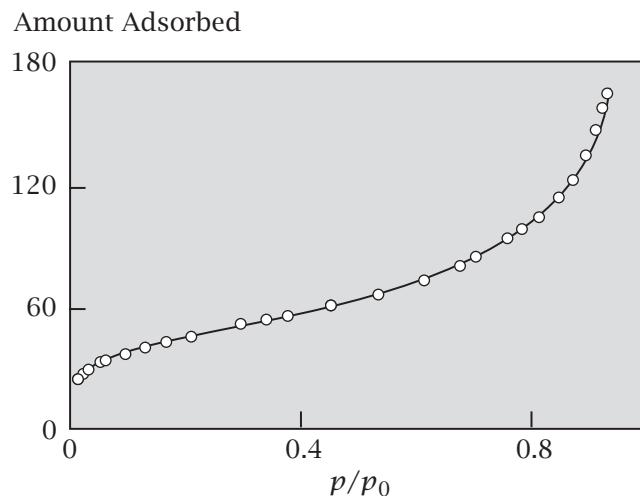
$$\frac{dQ}{dx} = \frac{K_1}{(1 - K_2x)^2}, \quad (28.33)$$

to get, according to Equation (C.4),

$$\nu(x) = \frac{x}{Q} \left(\frac{dQ}{dx} \right) = \frac{K_1x}{(1 - K_2x)[1 - (K_2 - K_1)x]}. \quad (28.34)$$

In this model, the first layer of ligands seeds the formation of subsequent layers if $K_2 > K_1$. The amount of ligand adsorbed to the surface increases sharply with increasing ligand concentration as $(K_2 - K_1)x \rightarrow 1$. An example of the BET binding curve is shown in Figure 28.15.

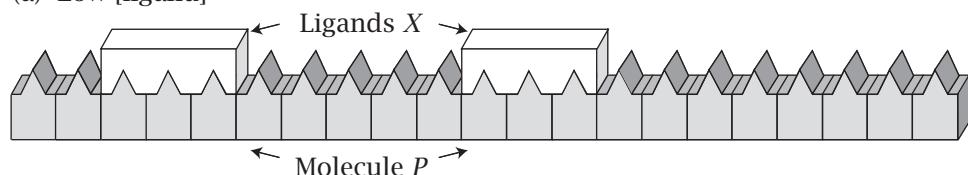
Figure 28.15 Adsorption of nitrogen on silica as a function of nitrogen pressure (relative to the nitrogen saturation pressure p_0) at 77 K follows the BET model. Source: PC Hiemenz, *Principles of Colloid and Surface Chemistry*, Marcel Dekker, New York, 1977. Data are from DH Everett, GD Parfitt, KSW Sing, and R Wilson, *J Appl Chem Biotechnol* **24**, 199–217 (1974).



The Model of McGhee and von Hippel Treats Ligands that Crowd Out Other Ligands

Some ligand binding processes involve a ‘parking problem.’ Imagine two cars parked parallel to a curbside. If there is empty space between them, but not enough space to park another car, you will not be able to park cars at maximum density. The empty spaces will be unfillable. Figure 28.16(b) shows the similar situation for ligand binding—how one ligand molecule X can crowd out another ligand from binding to a molecule P . The ligand X occupies more than a single binding site on P . At low ligand concentrations, each ligand molecule can readily find space available on P to bind (Figure 28.16(a)). But at high concentrations, ligands will have a parking problem or ‘excluded-volume problem.’

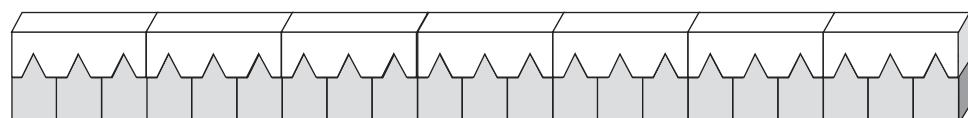
(a) Low [ligand]



(b) Intermediate [ligand]



(c) Saturating [ligand]



(d) McGhee-von Hippel Model

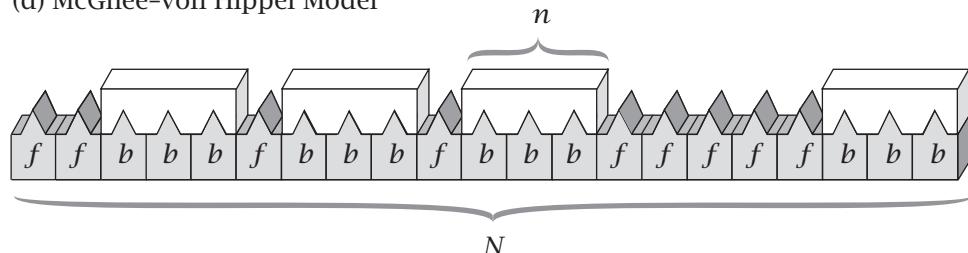


Figure 28.16 Multisite ligands can bind to a linear molecule such as DNA, excluding other ligands from binding. Molecule P is represented as a linear lattice having N sites. Each ligand occupies n sites ($n = 3$ shown here). (a) Low ligand concentrations. (b) At ligand concentrations that are less than full coverage, further binding is impossible because the remaining free sites are not distributed in stretches long enough for additional ligands to bind. (c) Full saturation of P by ligand molecules. (d) In the model of McGhee and von Hippel, lattice sites are either blocked (b) by a bound ligand, or free (f).

Figure 28.16(c) shows P saturated by X , the true binding equilibrium that will be reached at very high concentrations of X , given enough time. Figure 28.16(b) shows the problem at intermediate concentrations: P is not yet fully saturated by X , but nevertheless is unable to accept any new ligands. Sites remain available, but they are *not distributed appropriately* to allow additional ligands to bind. We have not faced this excluded-volume problem before, because we have considered only ligands that occupy a single site. However, with multisite ligands, a ligand that binds at one site can prevent another ligand from binding at a different site.

An apparent negative cooperativity can arise from *ligand shape and packing*, rather than from the energetic factors that we have considered so far. Also, the *kinetics* of multisite binding can be very slow, because new ligands can bind only if ligands that are already bound dissociate and rebind. JD McGhee and PH von Hippel developed a model to predict the binding curves for multisite ligands [8].

In the model of McGhee and von Hippel, each P molecule is a linear lattice of N sites. Each bound ligand occupies n sites on P . The binding affinity is

$$K = \frac{[PX]}{[P]x}, \quad (28.35)$$

where x is the concentration of free ligand (the number of molecules per unit volume) and $[PX]$ is the number of ligand molecules that are bound (per unit volume). $[P]$ is the number of ligand-sized stretches of sites on P molecules that are empty, per unit volume. In our previous modeling, $[P]$ would have been the concentration of 'receptor' molecules, the number of free P molecules per unit volume, because this gives the number of binding sites that are accessible to a ligand. But for multisite binding, determining $[P]$, the number of binding sites that are accessible to a ligand, is a little more challenging.

Figure 28.16(d) shows one particular distribution of ligands bound to P , expressed as a one-dimensional string of two characters, f and b . A site is free (f) if it is not occupied by a ligand. A site is blocked (b) if a bound ligand covers it. You can find $[P]$, the number of locations that are available for an n -mer ligand, by counting the number of stretches of the linear lattice that have n free sites in a row. You can estimate the number of such stretches by using conditional probabilities:

$$\begin{aligned} [P] &= (\text{number of individual sites } (i) \text{ that are free}) \\ &\quad \times \text{probability(site } i+1 \text{ is free, given that site } i \text{ is free)} \\ &\quad \times \text{probability(site } i+2 \text{ is free, given that site } i+1 \text{ is free)} \\ &\quad \vdots \\ &\quad \times \text{probability(site } i+n-1 \text{ is free, given that site } i+n-2 \text{ is free)}. \end{aligned} \quad (28.36)$$

If $p(f)$ is the *fraction* of sites that are free, then $Np(f)$ is the *number* of sites that are free. Let $p(f|f)$ be the *conditional probability* of finding a site $i+1$ free, given that site i is free. Equation (28.36) becomes

$$[P] = Np(f)p(f|f)^{n-1}. \quad (28.37)$$

First we compute $p(f)$. The number of ligands bound (per unit volume) is $B = [PX]$. Each ligand molecule blocks n sites at a time, so the total number

of blocked sites is nB . The rest of the $N - nB$ sites are free. So the fraction of sites that are free is

$$p(f) = \frac{N - nB}{N}. \quad (28.38)$$

To determine the conditional probability $p(f|f)$, first determine how to satisfy the condition. The horse race problem, Example 1.12, shows how to compute conditional probabilities by deleting the options that do not satisfy the given conditions. In this case, the condition that must be satisfied to identify an appropriate site is that an f must occupy the site immediately to its left. The b 's always come in runs: $b_1 b_2 b_3 \dots b_n$, where b_1 is the first site blocked by a ligand, b_2 is the second site blocked, etc. The only b sites that have an f to their left are b_1 positions, because positions b_2, b_3, \dots, b_n all have b sites to their left. The number of sites that do not satisfy our condition is $(n-1)B$, because this is the number of b_2, b_3, \dots, b_n sites. All the other

$$N - (n-1)B$$

sites satisfy the condition of having an f to the left.

The number of f sites is $N - nB$. So the conditional probability that a site is f , given that the site to the left of it is f , is

$$\begin{aligned} p(f|f) &= \frac{\text{number of } (f) \text{ sites}}{\text{number of } (ff) \text{ pairs} + \text{number of } (fb_1) \text{ pairs}} \\ &= \frac{N - nB}{N - (n-1)B}. \end{aligned} \quad (28.39)$$

We have made an approximation by neglecting cases of two ligands with no f sites separating them, i.e., cases of a b_1 following a b_n . Substituting Equation (28.39) into Equation (28.37) gives

$$[P] = (N - nB) \left[\frac{N - nB}{N - (n-1)B} \right]^{n-1}. \quad (28.40)$$

To put the binding equilibrium in terms of a Scatchard plot, let $\nu = B/N$ represent the number of ligands bound per P site. This gives $[P]$ in terms of ν :

$$[P] = N(1 - n\nu) \left[\frac{1 - n\nu}{1 - (n-1)\nu} \right]^{n-1}. \quad (28.41)$$

Using $[PX] = B = \nu N$ in Equation (28.35) gives $K = \nu N / ([P]\chi)$. Rearranging this expression and substituting Equation (28.41) gives the binding curve:

$$\frac{\nu}{\chi} = K(1 - n\nu) \left[\frac{1 - n\nu}{1 - (n-1)\nu} \right]^{n-1}. \quad (28.42)$$

Figure 28.17(a) shows the prediction of Equation (28.42) that excluded volume causes curvature in Scatchard plots. Figure 28.17(b) shows experimental data that fit the model. McGhee and von Hippel also generalized this model to allow for next-neighbor interactions among ligands to explain, for example, how gene-32 protein from bacteriophage T4 can line up within milliseconds and saturate single-strand DNA molecules to assist in DNA replication.

In this chapter, we have worked with binding polynomials Q . This approach has its foundations in the grand canonical ensemble, described below.

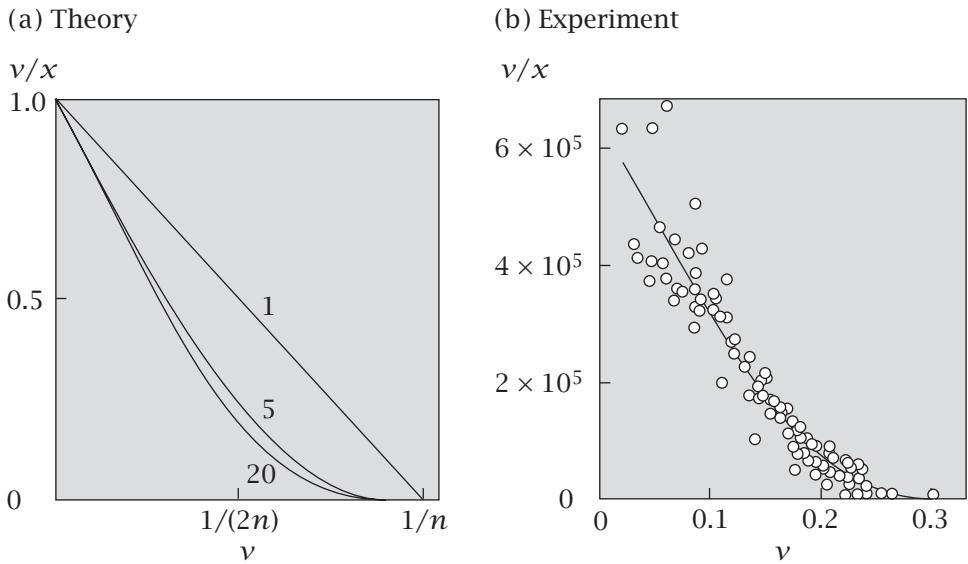


Figure 28.17 (a) Predicted binding curves from the McGhee–von Hippel model for block sizes $n = 1, 5, 20$. Source: JD McGhee and PH von Hippel, *J Mol Biol* **86**, 469–489 (1974). (b) Experiments on daunomycin binding to calf thymus DNA. The data fit the McGhee–von Hippel model, with parameters $n = 3$ and $K = 5.9 \times 10^5 \text{ M}^{-1}$. Source: JB Chaires, *Meth Enzymol* **340**, 3–22 (2001).

The Grand Canonical Ensemble Gives the Basis for Binding Polynomials

Recall from Chapter 8 that when you switch from a system having fixed energy, (U, V, N) , to a system having fixed temperature, (T, V, N) , you must also switch from maximizing the entropy S to minimizing the free energy F to predict the state of equilibrium. When T is constant, energy fluctuates and exchanges between the system and its surrounding *bath*. Now we make an additional switch. The present chapter focuses on systems that exchange both energy *and particles* with a surrounding bath. We switch from the *canonical ensemble* (T, V, N) to the *grand canonical ensemble* (T, V, μ) . Previously, we took as our system a whole beaker having a fixed number N of ligand molecules. But now, our system of interest is a single P molecule; ligands can jump back and forth between the P molecule and the surrounding solution in the beaker. The reason for choosing the P molecule as your system in this chapter is because the property you measure is the average number of ligand molecules bound to it. Different P molecules have different ligation states. Any given P molecule may have $0, 1, 2, \dots$ ligand molecules bound to it. When your system of interest is the P molecule (say a protein or DNA molecule), then what is fixed is μ , the chemical potential (i.e., the concentration) of the ligand molecules in the bath. And, what varies (i.e., fluctuates) is the ligation state of the P molecule in solution. We show below that for the grand canonical ensemble, the extremum quantity that predicts equilibrium is $\tilde{F}(T, V, \mu)$, not the canonical free energy $F(T, V, N)$. We also show that the companion partition function for this thermodynamic function \tilde{F} is just the binding polynomial.

We consider here only a single chemical component: a ligand species that has chemical potential μ in the solution. The methods of Chapter 8 show that the extremum function, for fixed μ , for predicting equilibrium is

$$\tilde{F} = F - \mu N = U - TS - \mu N. \quad (28.43)$$

At equilibrium, \tilde{F} is at a minimum, and the differential expression is $d\tilde{F} = -S dT - p dV - N d\mu = 0$.

We show below how minimizing $\tilde{F}(T, V, \mu)$ to predict the state of equilibrium leads naturally to the binding-polynomial mathematics that we have developed in this chapter. Express Equation (28.43) in terms of averages $U = \langle E \rangle$ and $N = \langle M \rangle$ over the energy and the numbers of particles $M = 0, 1, 2, \dots, M_0$ that are bound to the system molecule P , averaged over all the ligation states of the system:

$$d\tilde{F} = d\langle E \rangle - T dS - \mu d\langle M \rangle. \quad (28.44)$$

Let p_{jM} represent the probability that the system is in energy level j and has M particles. We want to find the minimum, $d\tilde{F} = 0$, subject to three constraints: (i) the probabilities must sum to one; (ii) the energies sum to the appropriate average; and (iii) the particle numbers sum to the appropriate average. The constraints are

$$\sum_{j=1}^t \sum_{M=0}^{M_0} p_{jM} = 1, \quad (28.45)$$

$$\sum_{j=1}^t \sum_{M=0}^{M_0} E_{jM} p_{jM} = \langle E \rangle, \quad (28.46)$$

$$\sum_{j=1}^t \sum_{M=0}^{M_0} M p_{jM} = \langle M \rangle. \quad (28.47)$$

To solve $d\tilde{F} = 0$ (Equation (28.44)), subject to the constraint equations (28.45)-(28.47), use the method of Lagrange multipliers (see Chapter 4), to get

$$\sum_j \sum_M \left[E_{jM} + kT(1 + \ln p_{jM}^*) + \alpha - \mu M \right] dp_{jM}^* = 0, \quad (28.48)$$

where α is the Lagrange multiplier for the constraint equation (28.45). Because the term in the square brackets in Equation (28.48) equals zero according to the Lagrange method, you have

$$\begin{aligned} \ln p_{jM}^* &= -\frac{E_{jM}}{kT} - \frac{\alpha}{kT} - 1 + \frac{M\mu}{kT} \\ \Rightarrow p_{jM}^* &= \frac{e^{-E_{jM}/kT} e^{M\mu/kT}}{\tilde{Q}}, \end{aligned} \quad (28.49)$$

where \tilde{Q} is the grand partition function. \tilde{Q} is a sum over both energy levels and particle numbers:

$$\begin{aligned}\tilde{Q}(T, V, \mu) &= \sum_{j=1}^t \sum_{M=0}^{M_0} e^{-E_{jM}/kT} e^{M\mu/kT} \\ &= \sum_{M=0}^{M_0} e^{M\mu/kT} \sum_{j=1}^t e^{-E_{jM}/kT} \\ &= \sum_{M=0}^{M_0} Q(T, V, M) e^{M\mu/kT},\end{aligned}\tag{28.50}$$

where $Q(T, V, M)$ is the canonical partition function of Equation (10.10). Although we have used a rather formal notation in this section in order to connect as clearly as possible to the statistical thermodynamics of the grand canonical ensemble, nevertheless, you can see that \tilde{Q} is exactly the binding polynomial that we have used throughout this chapter. Substitute the concentration of ligand in the bath, $x = e^{\mu/kT}$ (since $\mu = kT \ln x$), and substitute $K_M = \sum_{j=1}^t e^{-E_{jM}/kT}$, and replace the symbol that we use in this section, \tilde{Q} , with the simpler symbol Q that we have used throughout this chapter, and you will see that Equation (28.50) is nothing more than an expression for the binding polynomials, such as $Q = 1 + K_1x + K_2x^2 + \dots$, that we have described in this chapter.

Substituting Equation (28.49) into Equation (28.47) gives an expression for the average particle number:

$$\langle M \rangle = \frac{\sum_M \sum_j M e^{-E_{jM}/kT} e^{M\mu/kT}}{\tilde{Q}} = kT \left(\frac{\partial \ln \tilde{Q}}{\partial \mu} \right)_{V,T}.\tag{28.51}$$

Use Equation (28.44) to get $\langle M \rangle = -(\partial \tilde{F} / \partial \mu)$ and combine it with Equation (28.51) to get

$$\tilde{F} = -kT \ln \tilde{Q}.\tag{28.52}$$

This shows that the partition function for the grand canonical ensemble (T, V, μ) is the binding polynomial \tilde{Q} and that the extremum quantity that predicts equilibrium is $\tilde{F}(T, V, \mu)$.

EXAMPLE 28.8 Here's how you use the grand canonical free energy \tilde{F} . Use $\tilde{F}(x)$ in the same way for ligand binding processes that you used $F(T)$ for thermal processes (see Example 8.1 and Figure 8.5). Imagine that ligand molecules, at concentration x in solution, either bind to a single site on a protein P , with grand free energy $\tilde{F}_1(x)$, or bind to 10 sites, with $\tilde{F}_{10}(x)$. Compute the two quantities, \tilde{F}_1 and \tilde{F}_{10} . Whichever is the more negative at a given ligand concentration x identifies the most stable binding mode. Do this for a series of concentrations x ; this may predict that increasing x causes increased ligand bound to the protein, for example.

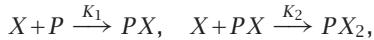
Summary

Binding processes can be described by binding polynomials. A binding polynomial is a sum over all possible ligation states, similar to a partition function, which is a sum of Boltzmann factors over energy levels. A binding polynomial embodies a model that can be used to calculate quantities such as the binding curve, the average number of ligands bound versus ligand concentration. This function can be used to interpret experiments. In particular, some binding processes are cooperative, meaning that when a ligand binds to one site, the affinity of another site may change.

Problems

1. The entropy in the grand ensemble. Write an expression for the entropy in the grand canonical ensemble (T, V, μ).

2. Three-site binding. A ligand X can bind to a macromolecule P at three different binding sites with binding constants K_1, K_2 , and K_3 :



(a) Write the binding polynomial, Q .

(b) Write an expression for the number of ligands v bound per P molecule.

(c) Compute v for $x = [X] = 0.05$, assuming $K_1 = 1$, $K_2 = 1$, and $K_3 = 1000$.

(d) Assume the same K values as in (c). Below ligand concentration $x = x_0$, most of the macromolecular P molecules have zero ligands bound. Above $x = x_0$, most of the P molecules have three ligands bound. Compute x_0 .

(e) For $x = x_0$ in (d), show the relative populations of the ligation states with zero, one, two, and three ligands bound.

3. Drug binding to a protein. A drug D binds to a protein P with two different equilibrium binding constants, K_1 and K_2 :



where $K_1 = 1000 \text{ M}^{-1}$, and $K_2 = 2000 \text{ M}^{-1}$ at $T = 300 \text{ K}$.

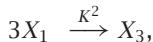
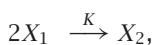
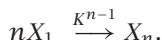
(a) Write an algebraic expression for the fraction of protein molecules that have two drug molecules bound as a function of drug concentration $x = [D]$.

(b) If the drug concentration is $x = 10^{-3} \text{ M}$, what is the fraction of protein molecules that have one drug molecule bound?

(c) If the drug concentration is $x = 10^{-3} \text{ M}$, what is the fraction of protein molecules that have zero drug molecules bound?

(d) If the drug concentration is $x = 10^{-3} \text{ M}$, what is the fraction of protein molecules that have two drug molecules bound?

4. Polymerization equilibrium. Consider the process of polymerization and assume the principle of equal reactivity, that each monomer adds with the same equilibrium constant as the previous one:


 \vdots


(a) Write the binding polynomial Q .

(b) Write the average chain length $\langle n \rangle$ in terms of Q .

(c) Plot $\langle n \rangle$ versus x for $K = 1$ and for $Kx < 1$.

5. Scatchard plot for an antibody. The plot in Figure 28.18 shows the result of binding dinitrophenol (DNP) ligand to an antibody (Ab). Compute the binding constant K and the number of binding sites per Ab molecule n .

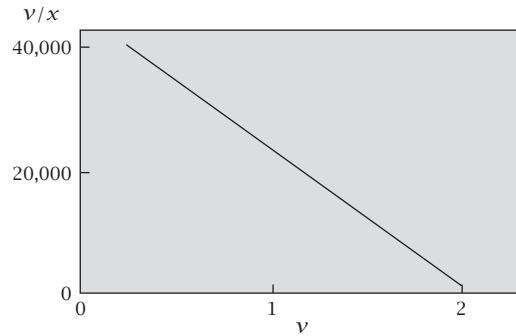
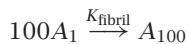
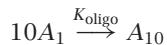


Figure 28.18 v is the number of moles of DNP bound per mole of Ab and x is the bulk concentration of DNP in solution in moles per liter. Source: J Darnell, H Lodish, and D Baltimore, *Molecular Cell Biology*, 2nd edition, WH Freeman, San Francisco, 1990.

6. Relating stoichiometric to site constants. Derive the relationship between stoichiometric-model binding constants K_1 and K_2 and site-model binding constants K_a and K_b . Suppose the sites are independent, $K_c = K_a K_b$.

7. Scatchard plots don't work for multiple types of binding site. If a binding process involves n_1 sites with affinity K_1 , a Scatchard plot of v/x versus v gives a straight line with K_1 as slope and n_1 as horizontal-axis intercept (see Figure 28.6). Show that if you have two types of sites having (number, affinity) = (n_1, K_1) and (n_2, K_2) , a Scatchard plot doesn't give two straight lines from which you can get (n_2, K_2) . To fit multiple types of sites, you need a different model.

8. A model for Alzheimer's fibrillization. Consider a solution of protein molecules in three possible states of aggregation: in state A_1 , the protein is a monomer in solution; in state A_{10} , the protein is hydrophobically clustered into oligomers having 10 molecules each; and in state A_{100} , 100 protein molecules are all aggregated into long fibrils. The equilibrium is given by



(a) Express the binding polynomial Q for this association equilibrium in terms of K_{oligo} , K_{fibril} , and x , where x is the concentration of protein A .

- (b) Write an expression for the average number v of protein molecules per object (monomer, oligomer, or fibril).

9. Alzheimer's again: chain-length dependence. In a variant of the problem above, suppose the oligomer and fibril can be formed from any number of monomers, M . Suppose the oligomer is driven by hydrophobic clustering:

$$K_{\text{oligo}} = K_1^M$$

where K_1 represents the equilibrium constant for adding a monomer to an oligomer.

Suppose the fibril is held together by hydrogen bonds: $K_{\text{fibril}} = K_2^M$, where K_2 is the equilibrium constant for adding a monomer to a fibril.

- Now, express v as a function of x , K_1 , K_2 , and M .
- If $x = 1$, $K_1 = 2$, $K_2 = 1.1$, and $M = 40$, what is the value of v ?
- If $x = 1$, $K_1 = 2$, $K_2 = 0.9$, and $M = 40$, what is the value of v ?

10. Cooperative assembly of membrane proteins. You have constructed an alpha-helical peptide. It partitions into lipid bilayer membranes and assembles into oligomeric complexes within the membrane. You measure the concentration x of the peptide in the membrane and the fraction θ of helical peptide molecules that form oligomeric complexes in the membrane. At temperature $T = 300\text{ K}$, the peptide concentrations x and the ratios of oligomeric complexes $\theta/(1-\theta)$ are as given in Table 28.1.

Table 28.1

x (μM)	$\theta/(1-\theta)$
1	10
2	160
3	810
4	2560
5	6250

- What are the association constant K and the Hill coefficient n ?
- Make a Hill plot.
- Plot θ versus x .
- At temperature $T = 330\text{ K}$, you find that all the values of $\theta/(1-\theta)$ in the table are doubled. Compute the ΔH and ΔS of binding.

11. Misusing the Scatchard plot. Consider a binding problem in which you have a ligand x and a protein P that have N_1 independent sites with binding affinity K_1 , and N_2 additional independent sites with binding affinity K_2 .

- Write the binding polynomial for this system.

- (b) Write an expression for v/x , where x is the ligand concentration in solution and v is the average number of ligands bound per P molecule.

(c) Now, consider the Scatchard plot shown in Figure 28.19, which has been used in an attempt to find the parameters N_1 , K_1 , N_2 , and K_2 from the model above:

$$v/x = a_0 - b_0 v + a_1 - b_1 v$$

Show that the parameters a_0 , b_0 , a_1 , and b_1 that you obtain from this curve fitting are not able to give you the quantities N_1 , K_1 , N_2 , and K_2 .

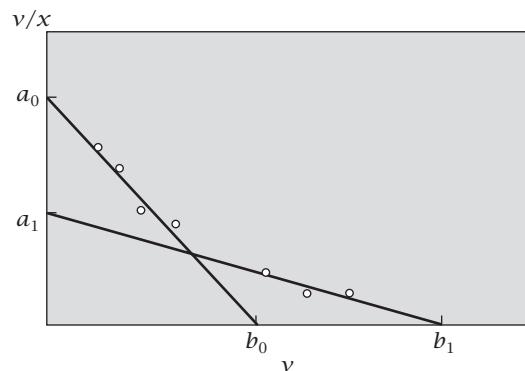


Figure 28.19 Using a Scatchard plot to try to fit non-Scatchard binding.

References

- J Wyman and SJ Gill, *Binding and Linkage: Functional Chemistry of Biological Macromolecules*, University Science Books, Mill Valley, CA, 1990.
- IM Klotz, *Ligand-Receptor Energetics: A Guide for the Perplexed*, Wiley, New York, 1997.
- JE Ferrell Jr and EM Machleider, *Science* **280**, 895–898 (1998).
- JN Israelachvili, *Intermolecular and Surface Forces*, 2nd edition, Academic Press, London, 1992.
- C Tanford, *The Hydrophobic Effect: Formation of Micelles and Biological Membranes*, 2nd edition, Wiley-Interscience, New York, 1980.
- DM Ruthven, *Principles of Adsorption and Adsorption Processes*, Wiley, New York, 1984.
- PC Hiemenz and R Rajagopalan, *Principles of Colloid and Surface Chemistry*, 3rd edition, Marcel Dekker, New York, 1997.
- JD McGhee and PH von Hippel, *J Mol Biol* **86**, 469–489 (1974).

Suggested Reading

Treatments of the principles of binding polynomials:

CR Cantor and PR Schimmel, *Biophysical Chemistry, Part III: The Behavior of Biological Macromolecules*, WH Freeman, San Francisco, 1980.

E Di Cera, *Thermodynamic Theory of Site-Specific Binding Processes in Biological Macromolecules*, Cambridge University Press, New York, 1995.

JT Edsall and H Gutfreund, *Biothermodynamics: the Study of Biochemical Processes at Equilibrium*, Wiley, New York, 1983.

IM Klotz, *Ligand-Receptor Energetics: A Guide for the Perplexed*, Wiley, New York, 1997.

JA Schellman, *Biopolymers* **14**, 999–1008 (1975).

DJ Winzor and WH Sawyer, *Quantitative Characterization of Ligand Binding*, Wiley-Liss, New York, 1995.

J Wyman and SJ Gill, *Binding and Linkage: Functional Chemistry of Biological Macromolecules*, University Science Books, Mill Valley, CA, 1990.

Treatments of the grand canonical ensemble:

TL Hill, *An Introduction to Statistical Thermodynamics*. Addison-Wesley, Reading, MA, 1960 (reprinted by Dover Publications, New York, 1986).

DA McQuarrie, *Statistical Mechanics*, 2nd edition, University Science Books, Sausalito, CA, 2000.

The BET binding process is treated in:

PC Hiemenz and R Rajagopalan, *Principles of Colloid and Surface Chemistry*, 3rd edition, Marcel Dekker, New York, 1997.

DM Ruthven, *Principles of Adsorption and Adsorption Processes*, Wiley, New York, 1984.

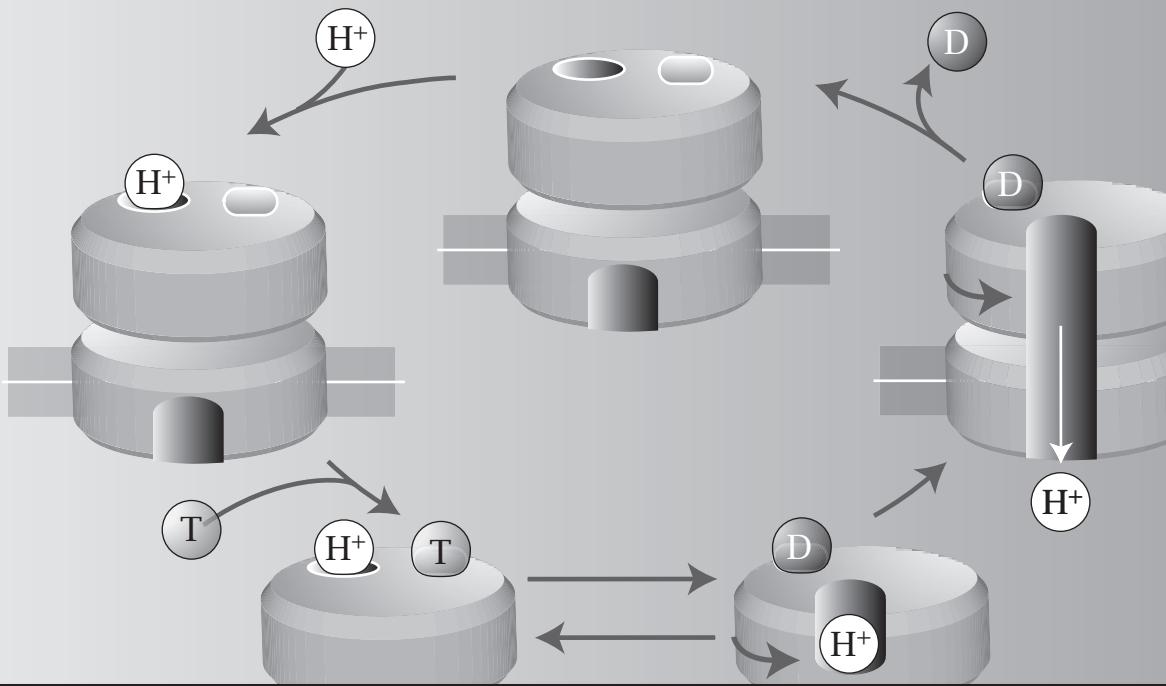
For extensive discussions of micelles, bilayers, and other colloidal properties, see:

JN Israelachvili, *Intermolecular and Surface Forces*, 2nd edition, Academic Press, London, 1992.

RJ Hunter, *Foundations of Colloid Science*, 2nd edition, Oxford University Press, Oxford, 2000.

C Tanford, *The Hydrophobic Effect: Formation of Micelles and Biological Membranes*, 2nd edition, Wiley-Interscience, New York, 1980.

29 Bio & Nano Machines



Biochemical Machines Are Powered by Coupled Binding Processes

Here, we explore how nanoscale machines and engines work. In biology, molecular machines perform energy conversion, signal transduction, regulation, transcription, ion pumping, molecular transport, and more. A molecular machine can be as simple as a single protein or nucleic acid molecule. How can a single protein or DNA molecule *transduce* one type of energetic action into another, in repeated cycles? A cyclic transducer is another name for an engine. The key to transduction is coupled binding events. For example, a molecular motor, such as myosin, can create motion because the motor's binding to a molecular 'track protein,' such as actin, is coupled to the motor's binding to ATP. We explore the principles of biomolecular machines using binding polynomials. We begin with oxygen binding to globin proteins, probably the best understood process of coupled binding and cooperativity in biology. Hemoglobin's affinity to bind one oxygen molecule depends on how many other oxygens are already bound. Because of this cooperativity, hemoglobin can pick up oxygen in the lungs and drop it off in the tissues. And, oxygen transport can be regulated by pH because hemoglobin's binding to oxygen is coupled to hemoglobin's binding to protons.

Oxygen Binding to Hemoglobin Is a Cooperative Process

Hemoglobin is a protein that binds four oxygen molecules. Hemoglobin has four subunits. So, to model binding, it would be simplest to first assume four identical independent sites on hemoglobin, each with affinity K , giving a binding polynomial

$$Q = (1+Kx)^4 = 1 + 4Kx + 6(Kx)^2 + 4(Kx)^3 + (Kx)^4,$$

where x represents the oxygen concentration (pressure). The binomial coefficients in this expression reflect the numbers of ways of arranging a given number of ligands on the four sites: $4!/4!0! = 1$ is the number of arrangements in which a hemoglobin tetramer has either no ligands or four ligands bound. There are $4!/3!1! = 4$ different ways to arrange one or three ligands on the four sites. There are $4!/2!2! = 6$ ways two ligands can bind to four sites. This is just the Langmuir model for four independent identical sites, $v = 4Kx/(1+Kx)$. But even the earliest experiments showed that oxygen binding to hemoglobin is cooperative and does not fit the independent-site model. For example, the fourth oxygen molecule binds with 500-fold more affinity than the first oxygen molecule.

One of the earliest models of hemoglobin cooperativity was *the Adair equation*, developed by G Adair in 1925 [1], which can be expressed as

$$Q = 1 + 4K_1x + 6K_2x^2 + 4K_3x^3 + K_4x^4. \quad (29.1)$$

K_1, K_2, K_3 , and K_4 are adjustable parameters used to fit binding curves of oxygen to hemoglobin. An extensive literature has developed since 1925 to improve upon the Adair model: to reduce the number of parameters to fewer than four, to interpret the parameters in terms of the protein structure, and to account for how ligands other than oxygen influence hemoglobin-oxygen binding. Two key models, originally developed for hemoglobin-oxygen binding, and later also applied to many other proteins, involve a mechanism called *allosteric regulation*. In allostery, a ligand molecule binds to a biomolecule on a site that is not its active site, causing the biomolecule to change its shape and often its activity.

The Pauling Model

In 1935, Linus Pauling assumed that the four hemoglobin subunits are identical to each other and that each subunit binds to one O_2 molecule with affinity K [2]. He proposed that there is an additional interaction free energy ε_0 whenever two oxygen molecules are bound to subunits that are *adjacent* to each other. He described the additional energy in terms of a Boltzmann weight f :

$$f = e^{-\varepsilon_0/kT}.$$

Long before the structure of the protein was known, Pauling postulated that the four subunits of hemoglobin might have different possible geometric relationships with each other. Two geometries that he considered were the square and the tetrahedron. Let's consider the tetrahedron (see Table 29.1).

Table 29.1 In the Pauling model, the four subunits of hemoglobin are assumed to have a tetrahedral arrangement. Each sphere represents one bound ligand molecule. Nearest-neighbor ligand interactions are indicated by continuous lines. The table shows the count of nearest-neighbor interactions.

					
Number of Ligands Bound	0	1	2	3	4
Number of Pairwise Interactions	0	0	1	3	6

Every vertex of a tetrahedron is adjacent to every other. For every pair of oxygen molecules that are bound at adjacent sites, Pauling introduced the interaction factor f . Oxygen bound to one subunit might propagate some conformational change to the adjacent subunit that would affect its affinity for its oxygen ligand. When only one oxygen binds, there is no interaction energy so f is not included. For any of the six possible arrangements in which two oxygens are adjacent to each other, the factor f appears once. When three ligands are bound, there will be three pairwise nearest-neighbor interactions, so the interaction factor is included as f^3 . When four ligands are bound, there will be six pairwise nearest-neighbor interactions, so the interaction factor appears as f^6 . Combining the intrinsic binding energies with these extra interaction energies gives the Pauling model for the binding polynomial:

$$Q = 1 + 4Kx + 6(Kx)^2f + 4(Kx)^3f^3 + (Kx)^4f^6.$$

The Pauling model treats oxygen binding to hemoglobin using only two parameters, and attributes the binding cooperativity to pairwise interactions between adjacent subunits.

An important idea here is that you can multiply terms in the binding polynomial by Boltzmann factors such as $f = \exp(-\varepsilon_0/kT)$. This is justified by the form of Equation (28.50). In general, energetic contributions such as this cooperative energy, or pairs of conjugate variables such as (pressure, volume) or (force, length) can be included as multiplicative Boltzmann factors within binding polynomials, as we will see later in this chapter.

A different model of oxygen-hemoglobin cooperativity, due to J Monod, J Wyman, and P Changeux [3], arose after M Perutz et al. showed in the 1960s that hemoglobin has two different structures [4, 5].

The Monod-Wyman-Changeux (MWC) Allosteric Model

Monod, Wyman, and Changeux proposed that hemoglobin has two different conformations: *T* and *R* (originally *tense* and *relaxed*, but these are just labels).

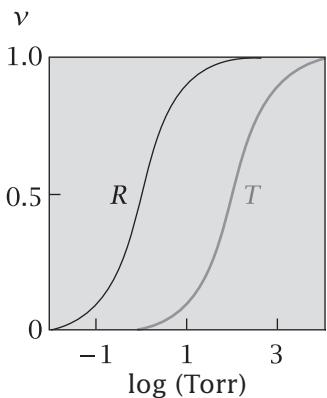


Figure 29.1 In the MWC model of hemoglobin cooperativity, the ligand oxygen has greater affinity for the *R* conformation than for *T*. State *R* saturates at a lower ligand concentration than state *T* does. Source: J Wyman and SJ Gill, *Binding and Linkage: Functional Chemistry of Biological Macromolecules*, University Science Books, Mill Valley, CA, 1990.

T and *R* are in equilibrium, with equilibrium constant *L*:

$$R \xrightarrow{L} T, \quad \text{where} \quad L = \frac{[T]}{[R]}.$$

The ligand has different affinities for the two protein conformations (see Figure 29.1):



In the absence of ligand, *T* is more stable than *R*. That is *L* > 1. Cooperativity results if the ligand binds to *R* more tightly than to *T*. In that case, adding ligand shifts the system from *T* states (liganded plus unliganded) toward the *R* states. To illustrate the principle, consider the simplest case, the binding of one ligand. There are four possible states of the system: *R*, *T*, *RX*, and *TX*, where *R* and *T* have no ligand and *RX* and *TX* each have one ligand bound. The binding polynomial *Q* is the sum over all ligation states of both *T* and *R* species:

$$[P]Q = [R] + [RX] + [T] + [TX], \quad (29.2)$$

where *[P]* is the total concentration of hemoglobin, *[R]* is the concentration of unligated *R* conformers, *[T]* is the concentration of unligated *T* conformers, and *[RX]* and *[TX]* are the concentrations of ligated *R* and *T* molecules, respectively. In terms of the equilibrium constants, *[T] = L[R]*, *[RX] = K_R[R]x*, and *[TX] = K_T[T]x = K_T L[R]x*, Equation (29.2) becomes

$$Q = \left(\frac{1}{1+L} \right) [(1+K_Rx) + L(1+K_Tx)]. \quad (29.3)$$

The fraction *f_R* of protein molecules that are in the *R* state is

$$f_R = \frac{[R] + [RX]}{[R] + [RX] + [T] + [TX]} = \frac{1 + K_Rx}{1 + K_Rx + L(1 + K_Tx)} \quad (29.4)$$

and the fraction *f_T* of protein molecules in the *T* state is

$$f_T = \frac{[T] + [TX]}{[R] + [RX] + [T] + [TX]} = \frac{L(1 + K_Tx)}{1 + K_Rx + L(1 + K_Tx)}. \quad (29.5)$$

Dividing Equation (29.5) by Equation (29.4) gives

$$\frac{f_T}{f_R} = \frac{L(1 + K_Tx)}{(1 + K_Rx)}.$$

Consider the two limits

$$x \rightarrow 0 \quad \Rightarrow \quad \frac{f_T}{f_R} \rightarrow L,$$

$$x \rightarrow \infty \quad \Rightarrow \quad \frac{f_T}{f_R} \rightarrow \frac{LK_T}{K_R}.$$

This shows that if *K_T/K_{R < 1, then increasing the oxygen concentration shifts the equilibrium from *T* to *R*; see Figure 29.2.}*

Fraction Populated

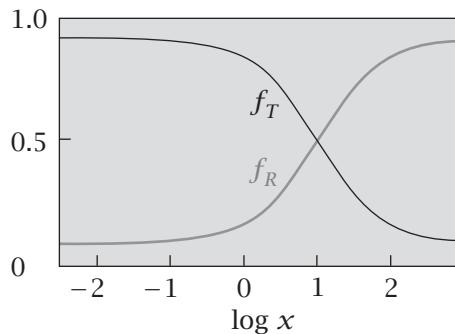


Figure 29.2 Increasing ligand concentration x in the MWC model of allostery causes a conformational change from T to R (Equations (29.4) and (29.5)), for $K_R = 1$, $K_T = 0.01$, and $L = 10$. The fraction f_T of molecules in the T state falls off as the fraction of molecules in the R state increases to 1.

Generalized to handle four binding sites, the MWC binding polynomial for hemoglobin is $Q = (1+L)^{-1}[(1+K_Rx)^4 + L(1+K_Tx)^4]$. In that case, the average number v of ligands bound is

$$v(x) = \frac{d \ln Q}{d \ln x} = [Q(L+1)]^{-1}[4K_Rx(1+K_Rx)^3 + 4LK_Tx(1+K_Tx)^3]. \quad (29.6)$$

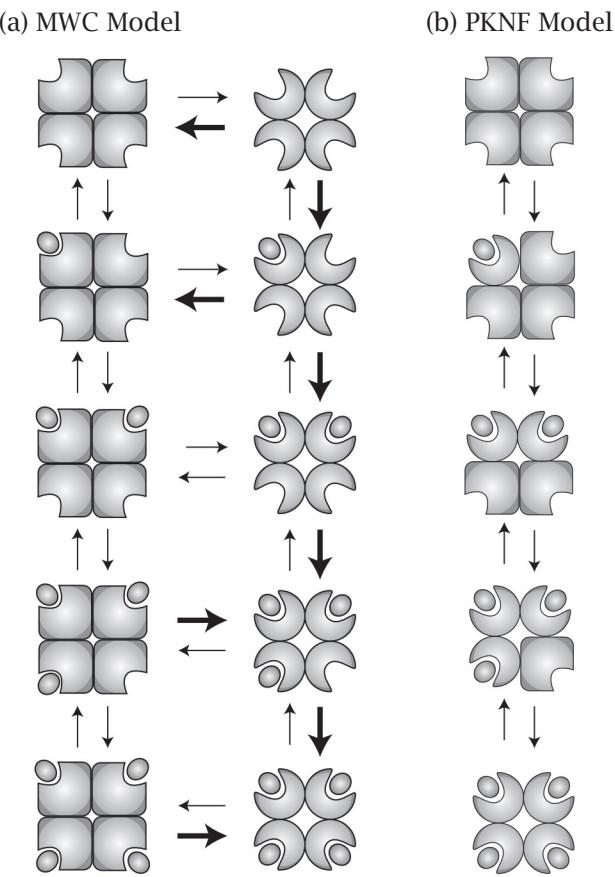
The MWC model describes oxygen binding to hemoglobin in terms of three parameters: K_R , K_T , and L . The key assumption is that the R and T monomer subunit structures of the hemoglobin protein have different affinities for the ligand. Equation (29.6) is also useful for understanding how ligands can drive the opening and closing of membrane channels [6].

Figure 29.3 compares the MWC model (also known as the *concerted-allostery model*) with the model of Pauling, later modified by Koshland, Nemethy, and Filmer (PKNF) (also known as the *sequential-allostery model*). In MWC, ligand binding does not perturb the R or T protein conformations. In MWC, if a ligand binds to an R conformation, the protein remains in R ; if a ligand binds to T , the protein remains in T . One ligand simply binds with a different affinity to a protein molecule that is in the R conformation than to a protein molecule in the T conformation. In contrast, in PKNF, ligand binding *induces* the protein into a new conformation. In PKNF, when a ligand molecule binds to R , the protein sub-unit switches to the T conformation. In PKNF, as the number of ligands bound increases from 0 to 4, the protein structure sequentially shifts, one subunit at a time, from T towards R . The PKNF process is called *induced fit*, because the act of binding the ligand tips each hemoglobin subunit conformation from T to R .

These models were among the first and simplest to capture the principles of hemoglobin-binding cooperativity. More extensive data are now available, and give a more refined perspective on hemoglobin–oxygen interactions [4, 5, 7–10].

So far, we have only considered a single type of ligand: one oxygen molecule binds, changing hemoglobin's affinity to bind another oxygen. But, binding processes can also involve combinations of different types of ligands. Hemoglobin can also bind to protons, for example. Changing the pH of your blood changes hemoglobin's affinity for oxygen. Carbon monoxide can also bind to hemoglobin, displacing oxygen, leading to toxicity and asphyxiation. Coupled binding is the basis for signaling, transduction, inhibition, activation, and regulation in biology.

Figure 29.3 Two different models of allostery: MWC and PKNF. The *T* state of each hemoglobin subunit is square; the *R* state is round. (a) MWC model. All four subunits are in state *T* or all four are in state *R*. Each protein molecule in solution is either in the fully-*T* state or in the fully-*R* state. A ligand binds to *T* with different affinity than it binds to *R*. Ligand binding does not shift a protein monomer from *T* to *R*. (b) PKNF model. Binding a ligand molecule induces one *T* monomer to shift to its *R* conformation, called *induced fit*.

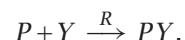
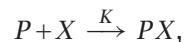


Binding Polynomials Can Treat Multiple Types of Ligand

Some molecules can inhibit, activate, or regulate the binding of others (see Figure 29.4). Let's apply binding polynomials again.

Competitive Inhibition

Suppose a ligand of type *X* binds with affinity *K* to a site on molecule *P*. A different type of ligand molecule, *Y*, can bind to the same site on *P* with affinity *R*:



Assume the solution concentrations of the two ligands are $x = [X]$ and $y = [Y]$. The binding polynomial will be

$$Q = 1 + Kx + Ry, \quad (29.7)$$

because the P binding site can either be free (with a statistical weight of 1), OR it can bind X (with weight Kx) OR it can bind Y (with weight Ry). The average fraction of P sites that are filled by X is $\nu_X = (\partial \ln Q / \partial \ln x)$:

$$\nu_X = \frac{Kx}{Q} = \frac{Kx}{1+Kx+Ry}. \quad (29.8)$$

Similarly, the fraction of P sites filled by Y is $\nu_Y = (\partial \ln Q / \partial \ln y) = Ry/Q$. Y is called a *competitive inhibitor*, because Y competes with X for the binding site. Figure 29.5(a) shows this inhibition: the amount of X bound is reduced as Y increases in concentration or affinity (ν_X decreases as Ry increases). For example, methotrexate is a small-molecule competitive inhibitor of the enzyme dihydrofolate reductase, important in cancer and autoimmune diseases. Equation (29.8) also applies to impurities that can ‘poison’ a surface or catalyst by displacing its ligand.

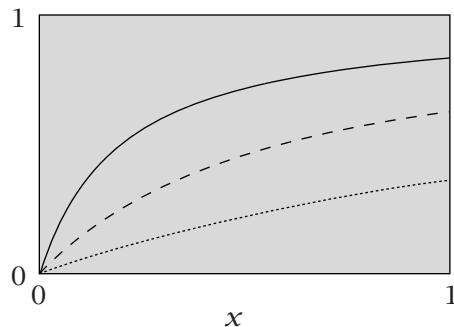
Binding polynomials tell you populations of all the possible states, bound and unbound. But they don’t tell you which binding states are active. Consider two different scenarios. First, in the model above, suppose the biological action of P happens only when P binds to X but not to Y . Then, biological action will correlate with ν_X . And Y will inhibit that action. But another possibility is that the biological action happens when P binds to either ligand X or Y . In that case, the biological action will correlate instead with

$$\text{action} \propto \frac{Kx + Ry}{1+Kx+Ry}. \quad (29.9)$$

In this second mechanism, Y is now an *activator*, not an inhibitor. In order to compute the properties of molecular machines, you need to know which terms in a binding polynomial are responsible for the action of interest. The binding polynomial itself does not tell you this—you need additional knowledge of the mechanism of action. Here’s another type of coupled ligand binding.

(a) Competitive

$$\nu(x)$$



(b) Prebinding

$$\nu(x)$$

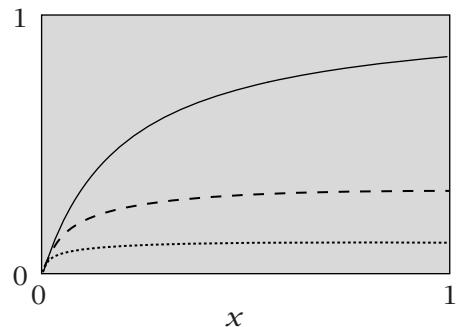


Figure 29.5 Binding isotherms for a ligand X : (a) in the presence of a competitive inhibitor Y ; (b) in the presence of a molecule Y that can bind only if X also binds (Equation (29.14)) (—), $Ry = 0$; (---), $Ry = 2$; (····), $Ry = 8$.

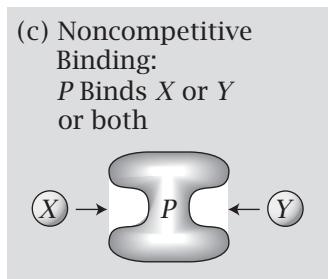
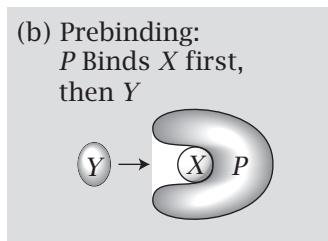
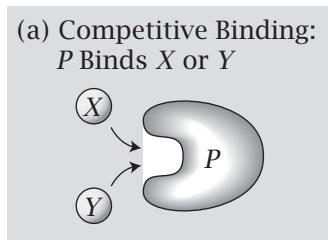
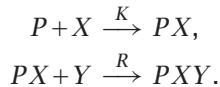


Figure 29.4 Types of inhibitors. (a) Competitive inhibitor Y competes for the site where X binds to P . (b) Uncompetitive inhibitor Y does not bind unless a ligand X also binds. (c) Noncompetitive inhibitor Y has no effect on the binding of X to P , for example, because it binds at an independent site. Y affects P in some other way than through the process of binding X .

Prebinding: Ligand Y Binds Only if X Also Binds

Suppose that molecule Y can bind to P only if X also binds to P :



Now the binding polynomial is

$$Q = 1 + Kx + KRxy, \quad (29.10)$$

because the binding site can be empty (statistical weight of 1) OR bind to X (statistical weight Kx) OR bind to (X AND Y) (statistical weight $KRxy$). Ligand Y cannot bind alone. The average number of X molecules bound per P is

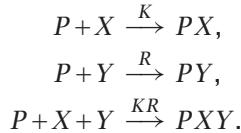
$$\nu_X = \frac{Kx + KRxy}{1 + Kx + KRxy}. \quad (29.11)$$

Figure 29.5(b) illustrates prebinding behavior. The BET adsorption in Example 28.7 is a type of prebinding: a second monolayer cannot form until a first layer has formed.

Related models describe *noncompetitive* and *uncompetitive* inhibitors.

Noncompetitive Inhibition

A noncompetitive inhibitor can be described by the process



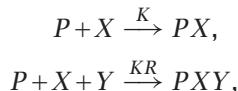
For this process, the binding polynomial is

$$Q = 1 + Kx + Ry + KRxy = (1 + Kx)(1 + Ry). \quad (29.12)$$

In the noncompetitive inhibitor model, the binding of X to P is independent of the binding of Y to P , sometimes because the two ligands bind to independent sites. For example, nifedipine is a calcium-channel blocker drug that is given for high blood pressure; it is a noncompetitive inhibitor of a cytochrome P450 protein called CYP2C9.

Uncompetitive Inhibition and Activators

An uncompetitive inhibitor corresponds to the process



and has the binding polynomial

$$Q = 1 + Kx + KRxy. \quad (29.13)$$

Uncompetitive inhibition is the same as prebinding and can be interpreted in terms of a site that binds either X alone or both X and Y at the same time. Now,

if state PX is active and PXY is inactive, the binding quantity of interest is

$$\nu = \frac{Kx}{1+Kx+Kxy}, \quad (29.14)$$

and Y will be an inhibitor.

Below are some useful thermodynamic Maxwell relationships that apply to coupled binding processes.

The Wyman–Gill Linkage Theory

Linkage is a term that describes how the binding of a ligand of type X affects the binding of a ligand of type Y [7]. Linkage relationships are helpful for understanding processes such as ATP-driven energy transduction, or how oxygen binding to hemoglobin is controlled by salts, pH, and carbon monoxide, or molecular regulation in biology.

For a system containing two types of ligand, X and Y , the free energy is $G(T, p, N_X, N_Y)$. But, we are not interested here in fixed N_X and N_Y . We are interested in fixed values of the chemical potentials. Why? In this chapter, our thermodynamic system is just the P molecule, not the whole beaker of molecules. X and Y molecules jump onto and off of the P molecule. The binding polynomials in this chapter express the full distributions of how many X or Y molecules are bound to P . The quantities that are fixed in this case are the concentrations x and y of the ligands in the surrounding medium. (This is just another way of saying that the chemical potentials of the ligand in solution are fixed, because $\mu_X(x) = \mu_X^\circ + kT \ln x$ and $\mu_Y(y) = \mu_Y^\circ + kT \ln y$.) We know the *average numbers*, $\nu_X = \langle N_X \rangle$ and $\nu_Y = \langle N_Y \rangle$, of ligands on a P molecule. So, we switch to the *semi-grand* ensemble, for which the predictor of equilibrium is the quantity

$$\tilde{G}(T, p, \mu_X, \mu_Y) = G(T, p, N_X N_Y) - \mu_X N_X - \mu_Y N_Y. \quad (29.15)$$

In Chapter 28, we focused on the grand canonical ensemble and a single type, X , of ligand molecule. Here, the prefix *semi-* just means that you have more than one type of ligand. In differential form, Equation (29.15) becomes

$$d\tilde{G} = -S dT + V dp - \mu_X \nu_X - \mu_Y \nu_Y. \quad (29.16)$$

Now, differentiate $\mu_X(x)$ and $\mu_Y(y)$ to get $d\mu_X = d \ln x$ and $d\mu_Y = d \ln y$, and use the Maxwell strategy of taking the relevant cross-derivatives in Equation (29.16) to get the Wyman–Gill linkage relationship:

$$\left(\frac{\partial \nu_X}{\partial \ln y} \right) = \left(\frac{\partial \nu_Y}{\partial \ln x} \right). \quad (29.17)$$

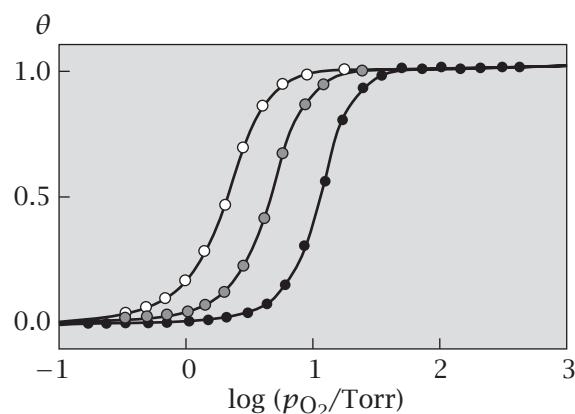
Linkage relations are simply Maxwell relationships (Chapter 9) applied to coupled binding.

EXAMPLE 29.1 Linkage of pH and oxygen binding to hemoglobin. Your breathing helps to buffer the pH of your blood. When your blood is too acidic, your hemoglobin doesn't bind oxygen very tightly (see Figure 29.6). This triggers neural circuits in your brain to cause you to breathe faster. In this case, a useful linkage equation (29.17) is

$$\frac{\partial \nu_{O_2}}{\partial \text{pH}} = -\frac{\partial \nu_{H^+}}{\partial \log p_{O_2}}. \quad (29.18)$$

That is, the average number ν_{O_2} of oxygens bound to hemoglobin varies with $\text{pH} = -\log[H^+]$ in inverse proportion to how the average number ν_{H_2} of bound protons depends on the logarithm of the oxygen partial pressure p_{O_2} .

Figure 29.6 Fractional binding of oxygen to hemoglobin. Different pH values shift the oxygen binding, indicating a coupling of the binding of oxygen and protons to hemoglobin. The pH values are 6.95 (●), 7.71 (○), and 7.91 (○). Source: J Wyman and SJ Gill, *Binding and Linkage: Functional Chemistry of Biological Macromolecules*, University Science Books, Mill Valley, CA, 1990.



In the next section, we note that binding polynomials are just as applicable to rate processes as to binding thermodynamics.

Rates Can Often Be Treated by Using Binding Polynomials

In this chapter, we have used binding polynomials to make models for the equilibrium number of ligands bound to P as a function of the concentration x of ligand in solution. Such equilibrium expressions are also useful for predicting how *rates* depend on the concentrations of ligands in solution. This was noted before in Equation (27.28), which shows that Langmuir-like saturation is the basis for Michaelis–Menten enzyme kinetics.

When enzymes, carriers, and transport processes involve binding steps that are more complex than simple Langmuir binding, binding polynomials can help. Consider two different enzymes that act on the same substrate. They have binding constants K_1 and K_2 and maximum velocities $v_{\max 1}$ and $v_{\max 2}$. Generalize Equation (27.28) to give the velocity v as the sum of the velocities from the two different enzyme reactions:

$$v = \frac{K_1 x v_{\max 1}}{1 + K_1 x} + \frac{K_2 x v_{\max 2}}{1 + K_2 x}. \quad (29.19)$$

This corresponds to a two-site binding process.

Inhibitors can also slow down enzyme reactions in ways that are predicted by binding polynomials. For example, if ligand Y is a competitive inhibitor of substrate X , Equation (29.8) can be used to predict the rate:

$$\frac{v}{v_{\max}} = \frac{Kx}{1+Kx+Ry}. \quad (29.20)$$

If an enzyme requires n identical substrate molecules to bind at the same time before it can catalyze a reaction, and if each molecule binds with affinity K , the Hill model can be used to predict the rate v of product formation:

$$\frac{v}{v_{\max}} = \frac{Kx^n}{1+Kx^n}, \quad (29.21)$$

where n is the Hill coefficient (see Equation (28.23)). In short, many rate processes can be expressed in terms of binding polynomials because they have underlying equilibrium binding steps.

The Molecular Logic of Biology Is Encoded in Coupled-Binding Actions

Biological machines are varied and powerful, like computers. The power of computers comes from the multiplicity of useful ways that you can link together a few elementary logic units, such as AND gates, OR gates, and bistable flip-flops. In a simple logic device, you have two inputs x and y . Each input is given to be in one of two states, call them 0 or 1. A *truth table* (Figure 29.7) describes the outputs for the different possible inputs.

Biology uses this principle too. Biomolecules can act as logic elements. For example, some types of proteins are activated only when two noncompetitive ligands, X and Y , are both bound at the same time. Such proteins act like AND

AND gate			OR gate		
x	y	Output	x	y	Output
0	0	0	0	0	0
0	1	0	0	1	1
1	0	0	1	0	1
1	1	1	1	1	1

Figure 29.7 A logical AND gate turns ON (1) only when both inputs are ON: $x = 1$ and $y = 1$. A logical OR gate turns OFF (0) only when both inputs are OFF: $x = 0$ and $y = 0$. Gates of these types are common in computers.

gates, with an action proportional to $f_{\text{AND}} = KRxy/Q$, where Q is given by Equation (29.12). In this case, the role of ‘input signals,’ ON or OFF, is played by whether each type of ligand is BOUND or UNBOUND. The role of an ‘output signal’ is played by a protein’s conformation or its biological activity. Another logical device is an OR gate. You can realize an OR gate with a protein that is active when either X OR Y OR both ligands are bound: $f_{\text{OR}} = (Kx + Ry + KRxy)/Q$. Or, you can realize a NOR gate with a protein that is active when no ligand is bound: $f_{\text{NOR}} = 1/Q$. In addition, you can make flip-flop memory elements from bistable switches. In a protein, bistable switching simply results from high binding cooperativity, i.e., a high Hill coefficient. Like logic elements in computers, biochemical logic elements can be chained together into higher-level combinations. Below is an example of the binding logic of an important biological machine, the *lac* operon.

EXAMPLE 29.2 Transcriptional regulation and the *lac* operon. Figure 29.8 shows the *lac* operon, one of the best understood regulatory systems in biochemistry. For elucidating this paradigm of prokaryotic gene regulation, J Monod and F Jacob shared the 1965 Nobel Prize in Physiology or Medicine. Normally, a bacterial cell eats glucose. But if necessary, the cell can also eat a more complex sugar, a disaccharide called lactose. To eat lactose, however, the bacterium must go to some extra trouble. The cell must first manufacture an enzyme, called β -galactosidase, that breaks lactose down into two monosaccharides: galactose and glucose. Then it eats that glucose.

Consider the biochemical logic. The cell could face four possible states: (+glucose, +lactose), (+glucose, -lactose), (-glucose, -lactose), or (-glucose, +lactose), where + means the sugar is present in the medium around the cell and – means the sugar is absent. In only one state, (-glucose, +lactose), does the bacterium need to produce the enzyme to digest lactose. When glucose is present, (+glucose, +lactose) or (+glucose, -lactose), the cell does not need to go to the trouble of breaking down the lactose. And, when neither sugar is present, (-glucose, -lactose), the bacterium has no food at all, so again there is no point in making the enzyme.

Figure 29.8 shows how this logic is implemented in molecular binding events. To make the β -galactosidase (*BG*) enzyme, the corresponding gene must first be transcribed from DNA to RNA. To do this, a protein called RNA polymerase (*P*) binds to a region of the DNA called the *promoter*. But the polymerase protein doesn’t bind to DNA very tightly. The polymerase protein only jumps onto the DNA when the DNA is *activated*. DNA is activated when an activator protein molecule (*A*), called CAP (catabolite activator protein), binds to a small molecule called cyclic AMP (cAMP, *C*), and both bind together to the DNA.

There are two additional facts you need to know. (a) When the cell’s glucose level is high, the concentration of *C* is low. (b) Another protein is also involved. A *repressor protein R* can bind to DNA, blocking transcription. Lactose can bind to the repressor protein, preventing *R* from binding to DNA, thus facilitating transcription. So, here’s how the system works.

State: (+glucose, +lactose). Glucose is high \Rightarrow *C* is low \Rightarrow DNA is not activated \Rightarrow **no *BG* enzyme.**

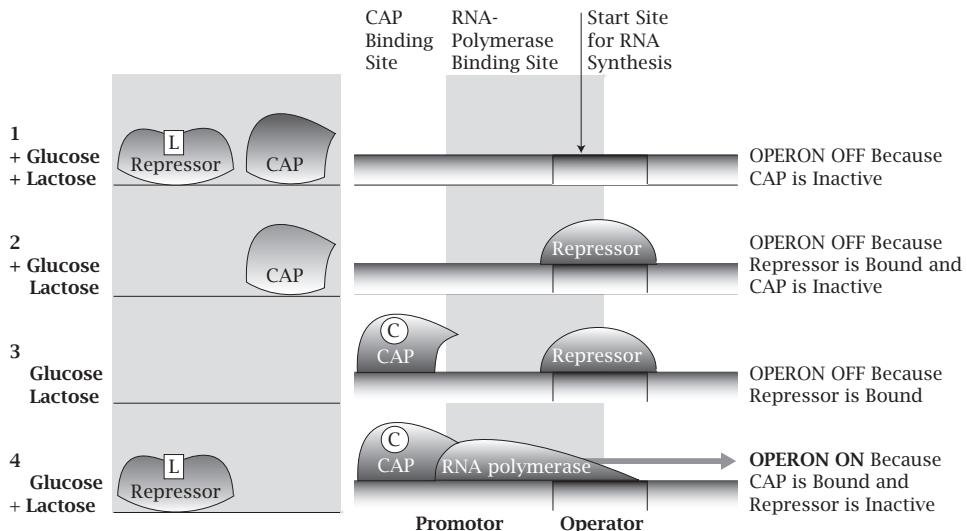


Figure 29.8 Logic of the *lac* operon. (1) (+G, +L) does not activate the DNA, so no *BG* enzyme. (2) (+G, -L) represses the DNA, so no *BG* enzyme. (3) (-G, -L) activates, but also represses the DNA, so no *BG* enzyme. (4) (-G, +L) activates and no repressor binds, so the *BG* enzyme is made.

State: (+glucose, -lactose). Again, because glucose is high, DNA is not activated. In addition, lactose is low $\Rightarrow R$ represses DNA \Rightarrow **no *BG* enzyme**.

State: (-glucose, -lactose). Glucose is low $\Rightarrow C$ is high \Rightarrow DNA is activated, but lactose is low $\Rightarrow R$ represses DNA \Rightarrow **no *BG* enzyme**.

State: (-glucose, +lactose). Glucose is low $\Rightarrow C$ is high \Rightarrow DNA is activated. In addition, lactose is high $\Rightarrow R$ does not repress DNA \Rightarrow ***BG* enzyme is produced**.

Few molecular machines are yet fully understood. So, our modest aim here, and in other examples below, is simply to develop a minimal plausible model. We will use binding polynomials. Here is a binding-polynomial model for the four different states of the DNA:

$$Q = 1 + K_1 r + K_2 r a c + K_3 p a c, \quad (29.22)$$

where p , r , a , and c are the concentrations of RNA polymerase protein, repressor protein, CAP activator protein (A), and cAMP (C), respectively, and the K 's are the corresponding equilibrium binding coefficients. The four states of the DNA are: (1) unligated, (2) repressor alone is bound, (3) repressor is bound to activated DNA (R and A and C are all bound), and (4) polymerase is bound to activated DNA (P and A and C are all bound). The logic above suggests that the machine's action, f_{action} , which corresponds to the concentration of the *BG* enzyme, will be proportional to the $p a c$ term:

$$f_{\text{action}} = \frac{K_3 p a c}{Q}. \quad (29.23)$$

This model predicts how the biological action $f_{\text{action}}(p, r, a, c)$, describing the level of BG enzyme, depends on the concentrations of DNA's binding partners. To make quantitative predictions, you need to know the quantities K_1 , K_2 , and K_3 . You can see that f_{action} increases with p , a , and c and decreases as the concentration of repressor R increases. In a fancier model, you might also want to include explicitly $c(\text{glucose})$, how the cAMP concentration depends on glucose, and $r(\text{lactose})$, how the repressor level depends on lactose. As your experimental knowledge of the system grows, you can refine your model to obtain additional insights. This example illustrates how binding polynomials can capture the biochemical logic of activation and inhibition of binding partners that act on a molecular machine—the *lac* operon region of DNA in this case.

Biochemical Engines Harness Downhill Processes to Drive Uphill Processes

Now, we explore energy transduction. In the macroscopic world, energy-utilizing (uphill) processes can be driven by coupling them to downhill chemical reactions, for example in batteries. Similarly, in the microscopic world of a biological cell, uphill processes—including chemical reactions, forces and motions, hauling cargo, and the pumping of ions from regions of low to high concentration—are driven by coupling to the hydrolysis of ATP to ADP. How is this energy transduced? Biological machines use coupled binding. While a macroscopic engine might function by burning fuel in a chamber, transduction in a biochemical engine can simply involve an individual protein or DNA molecule that binds to multiple ligands with coupled affinities.

Figure 29.9(a) shows a general thermodynamic cycle that explains how the conversion of ATP to ADP (a downhill process) can drive an uphill process, indicated here in general as a conversion of X to Y . Figure 29.9(b) illustrates this with a particular example, a protein called *F1-ATPase*. This protein pumps protons from a region of low concentration X to a region of high proton concentration Y through the following steps: (1) a proton from region X binds to the empty protein site; (2) ATP (call it T) also binds to the protein; (3) ATP is hydrolyzed to ADP (call it D), causing a small rotation of the protein's rotor; (4) pumping the proton uphill; then (5) ADP is released. Then the cycle begins again.

Let's express these states in terms of a binding polynomial. Let x be the concentration of a proton X in the low-concentration region. Let y be the concentration of a proton Y in the high-concentration region. Let t be the concentration of ATP. Let d be the concentration of ADP. The binding polynomial is the sum of all the binding states of the ATPase protein P :

$$[P]Q = [P] + [PX] + [PTX] + [PDY] + [PD] \quad (29.24)$$

$$\Rightarrow Q = 1 + K_1x + K_1R_1tx + K_2R_2dy + R_2y, \quad (29.25)$$

where $[P]$ is the concentration of free protein, and the K 's are defined by the following equilibrium relationships: $[PX] = [P]K_1x$ is the concentration

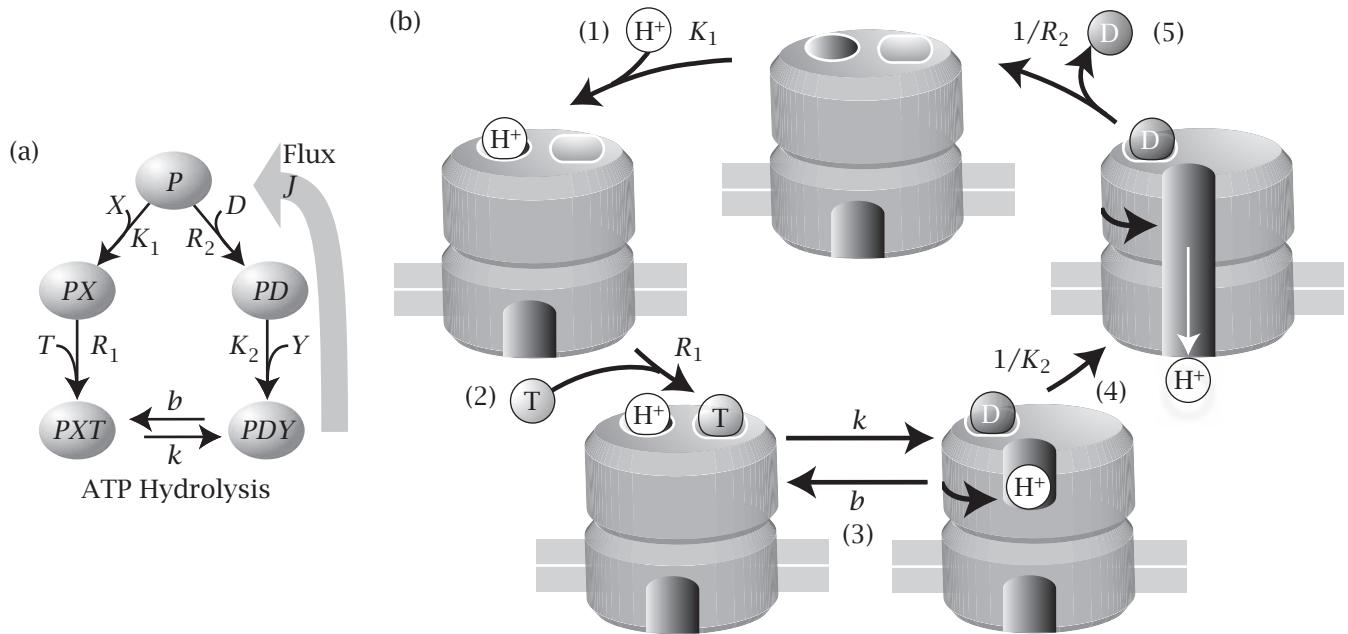


Figure 29.9 (a) Protein P converts X to Y , or pumps material up a concentration gradient from X to Y in a cyclic process driven by the hydrolysis of ATP. (b) A structure-based mechanism for this process in the F1-ATPase motor protein. (1) Protein takes up a proton H^+ from above the membrane. (2) Protein then also binds ATP. (3) ATP hydrolyzes to ADP, rotating the motor by a partial turn, positioning the proton to flow from inside (above) to outside (below) the membrane. (4) Proton crosses the membrane through the motor. (5) ADP is released. Ready to start again.

of protein bound only to X , $[PTX] = [P]K_1R_1tx$ is the concentration of protein bound to both X and ATP, $[PD] = [P]R_2y$ is the concentration of protein bound to both ADP, and $[PDY] = [P]K_2R_2dy$ is the concentration of protein bound to ADP and Y . Note that in Figure 29.9(b), all the arrows point counter-clockwise around the cycle, in the direction of the forward flux, because this is most intuitive. But in Figure 29.9(a), some arrows point in the opposite direction instead, so that all the constants refer to a binding process. You can draw the arrows either way, depending on your preference, but if a forward arrow has equilibrium constant K , the reverse arrow has equilibrium constant $1/K$.

This model illustrates how a downhill step, ATP → ADP, can drive an uphill process, $X \rightarrow Y$. This ATPase engine will spin counterclockwise if the ratio of concentrations ATP/ADP is high enough. To see this, define the flux J as the number of conversions of T to D per second per engine (i.e., per P molecule). Assume that $T \rightarrow D$ is the rate-limiting step and that all other steps happen much faster. Then, at steady state, the flux will be determined by the relative forward and backward rates on the bottom step of Figure 29.9(a):

$$\begin{aligned}
 \frac{J}{[P]} &= \frac{k[PTX]}{[P]Q} - \frac{b[PDY]}{[P]Q} \\
 &= \frac{kK_1R_1tx}{Q} - \frac{bK_2R_2dy}{Q} \\
 &= \frac{kK_1R_1tx - bK_2R_2dy}{1 + K_1x + K_1R_1tx + K_2R_2dy + R_2y}.
 \end{aligned} \tag{29.26}$$

Here are the principles of energy transduction that are illustrated by this model:

'Burning energy' drives the pumping of X uphill: This engine will spin forward ($J > 0$), pumping protons uphill, provided that $kK_1R_1tx > bK_2R_2dy$. You can express this 'spin-forward' condition as

$$\left(\frac{kK_1R_1}{bK_2R_2}\right) \left(\frac{t}{d}\right) > \frac{y}{x} \quad (29.27)$$

Equation (29.27) tells you how large a ratio of ATP/ADP you need in order to pump protons up a concentration gradient of y/x , for a given 'machine constant' $kK_1R_1/(bK_2R_2)$. Increasing the ATP concentration beyond that value will increase the spin rate of the cycle.

Running X downhill reverses the process and 'recharges the battery': If such a molecular machine is reversible, it can run backwards instead: protons flowing down their concentration gradient can convert ADP to ATP. If $y/x \gg 1$ is sufficiently large, the machine will run backwards to convert D to T . The F_1 -ATPase protein is an example of a reversible molecular machine.

At equilibrium, the engine doesn't spin: The engine will halt ($J = 0$) when the ATP 'battery,' i.e., the ratio of concentrations t/d , is sufficiently small to just balance y/x in Equation (29.27).

We have described $X \rightarrow Y$ as a pumping of protons up a concentration gradient. But $X \rightarrow Y$ could, instead, represent any other uphill transformation from low chemical potential μ_X to higher chemical potential μ_Y .

This model illuminates the *chemiosmotic hypothesis*, proposed in 1961, for which P Mitchell won the 1978 Nobel Prize in Chemistry. At that time, the question was how biological systems convert the energy in glucose to ATP. The view was that in the process of synthesizing ATP, energy was stored in high-energy chemical bonds in some hypothetical intermediate compound. However, no such intermediate was ever found. Instead, the chemiosmotic hypothesis predicted that it was the electrochemical gradient of protons that drove ATP synthesis, by the kind of coupling described above. Proof of chemiosmosis came with careful stoichiometric studies of the electrons, the protons and the phosphate balance, and the later discovery of ATPase proteins.

EXAMPLE 29.3 Operating characteristics of a typical biochemical engine. What are typical values for the 'machine constant'? In the *Escherichia coli* bacterium, the relative concentration of ATP and ADP is approximately $t/d = 10$. Because the pH is approximately 8.0 inside a mitochondrion and pH = 7 outside, it means that pumping protons outside the mitochondrion leads to $y/x = 10$. So, pumping protons uphill in this case can be achieved with the biochemical machine above if $kK_1R_1 \approx bK_2R_2$.

EXAMPLE 29.4 Biological signaling and phosphorylation. Here, we make a simple model of *signal transduction*, in which a protein molecule is triggered to perform some action by the concentration of some signal effector ligand molecule X . Figure 29.10 shows that binding the ligand X to a protein P activates or phosphorylates another protein A , converting it to a state B .

How does the binding of signal X drive the protein machine in Figure 29.10 to run counterclockwise? To model signal transduction, you need only to make small changes to our model ATPase energy-transduction machine above. To convert notation from Equations (29.24) and (29.26), let $a = [A]$, the concentration of A , replace t ; let $b = [B]$, the concentration of B , replace d ; let $K = K_1 = K_2$; and, let X replace Y , since the signal molecule X is not changing state. With these changes, Equation (29.26) now gives the machine cycle rate per protein molecule as

$$\frac{J}{[P]Q} = \frac{Kx(kR_1a - bR_2b)}{1 + 2Kx + KR_1ax + KR_2bx} \quad (29.28)$$

Equation (29.28) predicts the following signal-transduction behavior. First, if $kR_1a > bR_2b$, increasing the signal concentration x drives the cycle to spin, leading to activation of A to B . Second, the signaling saturates: the machine ultimately reaches a maximum speed with increasing signal concentration x , for a fixed value of $kR_1a - bR_2b$. Third, supplying more protein in the inactivated state A (increasing a) increases J , leading to more activated protein B .

Another type of biochemical machine creates molecular motion.

EXAMPLE 29.5 Molecular motors convert chemical energy to directed force and motion. *Molecular motors* are machines that convert energy to motion. Motors provide the motility for cell division and they haul cargo along the axons of neurons. How do molecular motors convert energy to motion in a particular direction? We model a protein called *kinesin*. Kinesin has two feet, interestingly called heads. When supplied with ATP, kinesin ‘walks’ along another

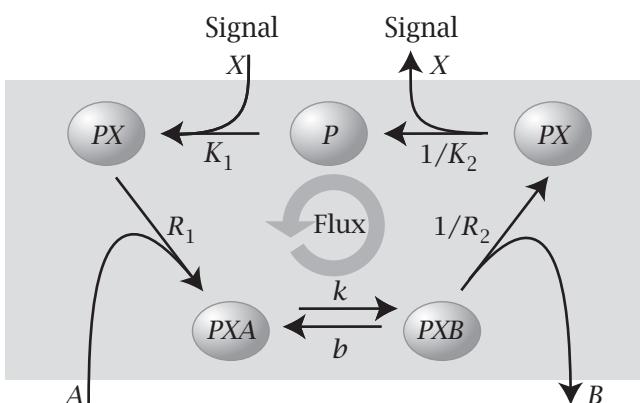


Figure 29.10 A model of signal transduction. As the concentration of signal molecule X increases, it drives protein P to convert a protein from state A to B , for example by phosphorylating or activating it.

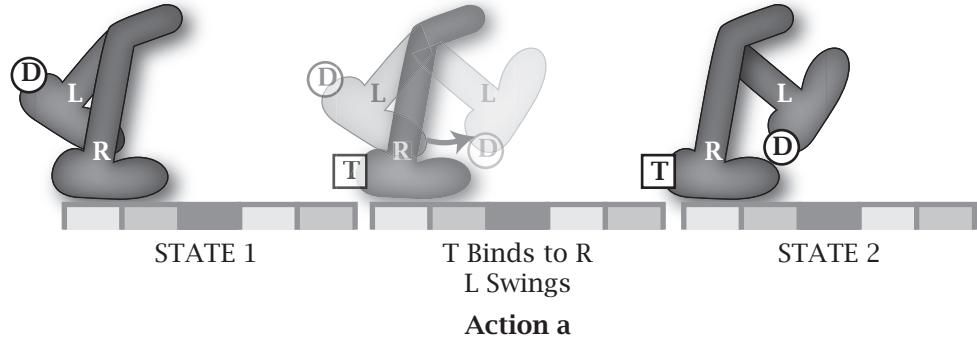


Figure 29.11 Kinesin walking is driven by the hydrolysis of ATP.

biomolecule, called a *microtubule*, which is like a train track. Motion results from a series of coupled binding events; see Figure 29.11. Let's label the feet: right (R) and left (L). 'Down' means the foot is bound to the track. 'Up' means the foot is not bound to the track. T means an ATP is bound to a foot, D means ADP is bound, and 0 means no ligand is bound. For one step of kinesin walking along the microtubule track, Figure 29.11 shows the three states and the three intervening transitions:

State 1, Front Foot Down (FFD): Right foot down in front (0). Left foot up in back (D). **Action a,** Left leg swings forward. ATP binds to the back right foot on the track.

State 2, Back Foot Down (BFD): Left foot, now in front, is still up (D). Right foot, now in back, is still down (T). **Action b,** Front foot now binds to the track and releases ADP.

State 3, Both Feet Down (2FD): Left foot down in front (0). Right foot down in back (T). **Action c,** ATP hydrolyzes to ADP, releasing the back foot from the track.

State 1 again, but now with the feet switched. Left foot is now down in front (0). Right foot is up in back (D).

Kinesin is *processive*, meaning that at least one foot is usually on the track; fully unbound states are rare. Let's model this in terms of three states of kinesin. Here is a binding polynomial that captures these basic features:

$$Q = K_{\text{FFD}} d + K_{\text{BFD}} dt + f K_{\text{2FD}} t, \quad (29.29)$$

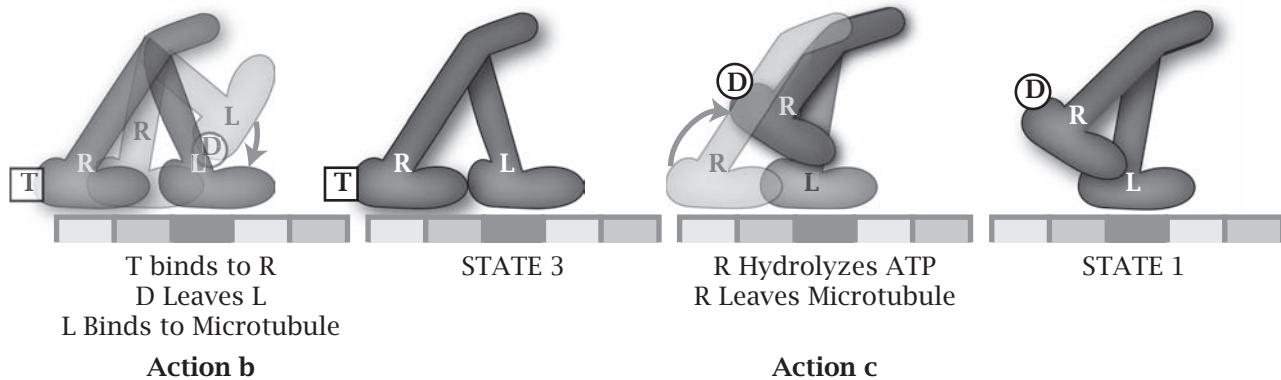


Figure 29.11 Continued.

where t is the ATP concentration, d is the ADP concentration, and the K 's are the equilibrium constants for the corresponding states. The quantity

$$f = e^{-F_0 \ell / kT} \quad (29.30)$$

is a Boltzmann factor that we introduce here to account for how the binding equilibrium in one step of the cycle is affected by the force F_0 acting against the walker molecule. ℓ is the distance moved per step. For kinesin, experiments show that $\ell = 8.3$ nm. The applied force serves as a sort of ‘headwind’ against the walker, reducing the population of the state 2FD relative to the other two states.

Because the rate-limiting step is the hydrolysis of ATP to ADP (Action c) [11], the velocity v of walking, per kinesin molecule, is the product of (step length, $\ell = 8.3$ nm) \times (number of ATP molecules hydrolyzed per second, k_2) \times (population of the rate-limiting state, 2FD):

$$v = k_2 \ell \frac{f K_{2FD} t}{K_{FFD} d + K_{BFD} t d + f K_{2FD} t} \quad (29.31)$$

Figure 29.12 shows the predictions of this simple model compared with experiments [12]. First, the model predicts that more ATP leads to faster walking (larger v). Second, the walking speed reaches a maximum at high ATP concentrations. Third, as the opposing force increases, walking slows down, and ultimately stalls. You can explain the directionality of the walking motion as arising either from ATP’s preferential binding and release of the back foot from the track or from ADP’s preferential release from the front foot. An improved model might aim to better understand why the best-fit value of the step length $\ell = 2.5$ nm is smaller than the value $\ell = 8.3$ nm measured in independent structural studies.

Our last example of a molecular transducer is an ion channel that senses temperature and converts that information to a flux of ions across a membrane.

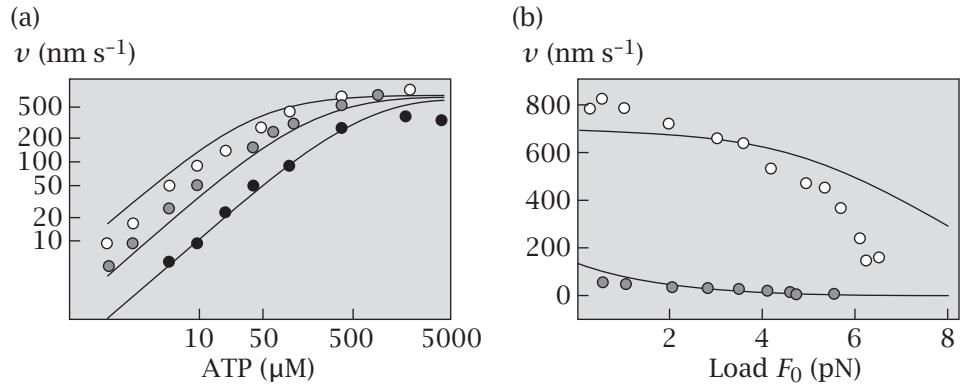


Figure 29.12 (a) Kinesin walking velocity along microtubules, v , increases with ATP concentration (\circ , 1.05 pN loading force; \bullet , 3.59 pN; \blacksquare , 5.63 pN). (b) Walking slows down under increasing loading force (\circ , 2 μM ATP; \bullet , 5 μM). Predictions are from Equation (29.31); experiments are from [12]. For reference, $kT = 4.1$ pN nm at $T = 300$ K. The following parameters were used to fit the data: $\ell = 2.5$ nm; $k_2 = 280$ s $^{-1}$; $K_{\text{BFD}} = 0$; and $(K_{\text{FFD}}d/K_{\text{2FD}}) = 21.1$ μM.

EXAMPLE 29.6 How do you sense hot and cold? Biological cells sense temperature. How? And, why do we use the same words—hot and cold—to refer to foods that taste spicy (like chili peppers) or cool (like menthol) that we also apply to high and low temperatures? These actions are explained by voltage-gated ion-channel membrane proteins. Changing the temperature or binding a ligand to an ion channel can shift the voltage dependence of its ion currents. A protein called TRPV1 is a membrane ion channel that opens up to ion flow either when the temperature is increased or when a spicy molecule, like *capsaicin*, binds to it [13]. Another membrane ion channel, called TRPM8, opens up either when the temperature is lowered or when menthol binds to the channel.

Here is a model for how a voltage-gated ion channel can sense temperature or ligand binding. The channel protein opens or closes depending on an applied voltage. Assume that the binding polynomial is

$$Q = 1 + f\delta(1 + Kx) \quad (29.32)$$

for the three states of such a system: (1) The closed state of the protein channel has a statistical weight of 1. (2) The channel can be opened by voltage alone, and modulated by temperature; this state has statistical weight $f\delta$, described below. (3) Channel opening can also be influenced by ligand binding; this state has a statistical weight $f\delta Kx$, where K is the binding affinity of the ligand for the protein and x is the concentration of ligand in solution.

Here is the basis for these factors. $f(T)$ is a Boltzmann factor that expresses the degree of channel opening and its dependence on temperature T :

$$f = e^{-\Delta G_0/RT}. \quad (29.33)$$

R is the gas constant, $\Delta G_0 = \Delta H_0 - T \Delta S_0$ is the free energy required to open the channel, and ΔH_0 and ΔS_0 are the enthalpic and entropic components (since we are interested in the temperature dependence here). $\delta(T, V)$ is a Boltzmann

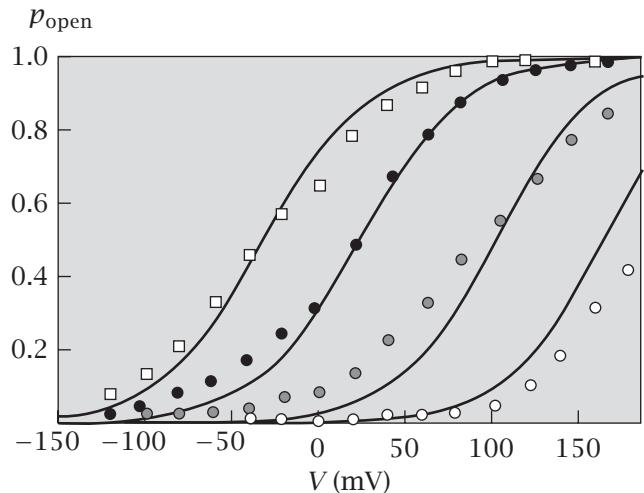


Figure 29.13 Capsaicin receptor (hot) channel opening probability at 17°C (○), 25°C (●), 35°C (●), 42°C (□). Lines are given by Equation (29.35) using the fit parameters $\Delta H_0 = +50.5 \text{ kcal mol}^{-1}$, $\Delta S_0 = +162.3 \text{ cal K}^{-1} \text{ mol}^{-1}$, and $z = 0.9$. This shows that: (a) for fixed temperature, applied voltage opens the channel, and (b) for fixed voltage, heating opens the channel. The channel is highly sensitive. In some cases, the channel changes from almost fully closed, $p_{\text{open}} \approx 0$, to fully open, $p_{\text{open}} \approx 1$, with only a temperature increase of 25°C. Data are from Figure 3d of T Voets, G Droogmans, U Wissenbach, et al., *Nature* **430**, 748–754 (2004).

factor that expresses how channel opening is modulated by an applied voltage V :

$$\delta = e^{zFV/RT} \quad (29.34)$$

where zF is the effective charge per mole on the protein's dipole and $F = 23,060 \text{ cal mol}^{-1} \text{ V}^{-1}$. Notice that the combined factor $f\delta$ is exactly what we used in Equation (21.20) to treat a voltage-gated ion channel.

Because states (2) and (3) are both open, the fraction of channels that are open is given by

$$p_{\text{open}} = \frac{f\delta(1+Kx)}{1+f\delta(1+Kx)} \quad (29.35)$$

How does the channel behave when no ligand is present, $x = 0$? Applying a voltage opens the channel. In this case, Equation (29.35) reduces to the voltage-gated ion-channel model, Equation (21.20). Figure 29.13 compares this model with experiments on the capsaicin channel [13] and shows the strong temperature dependence of this voltage gating.

You get interesting insights from the values of the model parameters that best fit the data: $\Delta H_0 = +50.5 \text{ kcal mol}^{-1}$, $\Delta S_0 = +162.3 \text{ cal K}^{-1} \text{ mol}^{-1}$ (so $T \Delta S_0 = 48.7 \text{ kcal mol}^{-1}$ at $T = 300 \text{ K}$), and $z = 0.9$. First, you see that the enthalpy and entropy of channel opening are both large positive values. So, at room temperature, this hot receptor is held closed by the enthalpy. Heating any system increases its entropy, so the hot channel's open state has higher entropy than its closed state. Second, the high sensitivity to temperature is because of enthalpy-entropy compensation. The enthalpy and entropy are both large and nearly equal, so their difference, the free energy, is small and highly sensitive to temperature. Third, our quantity z (which is called z_{eff} in Equation (21.22)) gives a useful structural prediction. z is a product of two factors: (1) the valency

of the charge that moves up or down the bilayer when the gate is opened and (2) the dimensionless ratio of the distance of that charge motion up or down the bilayer, divided by the bilayer thickness. Because a lipid bilayer is approximately 20 Å thick, and because our best-fit parameter value is $z = 0.9$, our model predicts that the opening of the gate shifts one charge $20 \times 0.9 = 18$ Å up or down the bilayer, or it shifts six charges 3 Å each within the bilayer, or some other such combination of charges \times distance. Such predictions could be tested by structural experiments.

How does the cold channel work? Figure 29.14 shows the same model applied to the menthol channel, with no ligand. In this case, the experimental data are best fit by parameters $\Delta H_0 = -30.7 \text{ kcal mol}^{-1}$, $\Delta S_0 = -110 \text{ cal K}^{-1} \text{ mol}^{-1}$ (so $T \Delta S_0 = -33 \text{ kcal mol}^{-1}$ at $T = 300 \text{ K}$), and $z = 0.7$. For this cold channel, like the hot channel, you see that the high temperature sensitivity again comes from enthalpy–entropy compensation. However, while the hot channel has *positive* values of ΔH_0 and ΔS_0 , the cold channel has large *negative* values of ΔH_0 and ΔS_0 . The cold channel is closed by heating. For the cold channel, the closed state has the higher entropy.

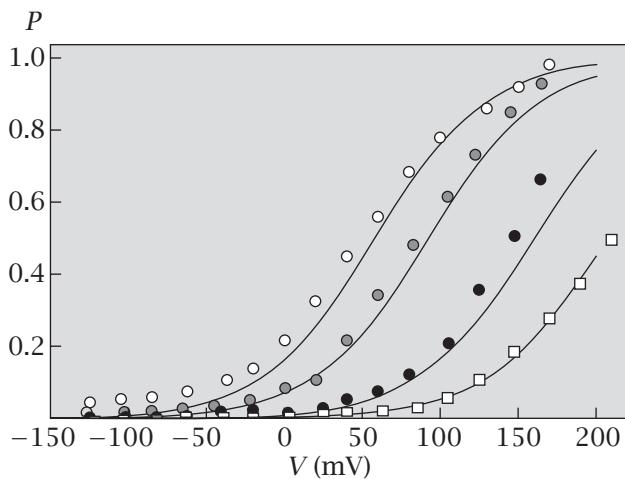
Now, consider the effect of ligand binding of capsaicin or menthol. Experimentalists measure how the binding of a ligand (in solution, at concentration x) shifts the midpoint voltage V_{mx} , the voltage at which $p_{\text{open}} = 1/2$. To compute the midpoint voltage, substitute $p_{\text{open}} = 1/2$ into Equation (29.35) to get $f \delta(V_{mx})(1+Kx) = 1$, i.e.,

$$f(1+Kx)e^{zFV_{mx}/RT} = 1. \quad (29.36)$$

Take the logarithm and rearrange Equation (29.36) to get

$$V_{mx} = -\frac{RT}{zF} \ln[f(1+Kx)], \quad (29.37)$$

Figure 29.14 The opening probability for the menthol (cold) channel at 15°C (○), 20°C (◐), 30°C (●), and 37°C (□). Lines are given by Equation (29.35) with the fit parameters $\Delta H_0 = -30.7 \text{ kcal mol}^{-1}$, $\Delta S_0 = -110 \text{ cal K}^{-1} \text{ mol}^{-1}$, and $z = 0.7$. This channel is opened by cooling. Data points are from Figure 2d of T Voets, G Droogmans, U Wissenbach, et al., *Nature* **430**, 748–754 (2004).



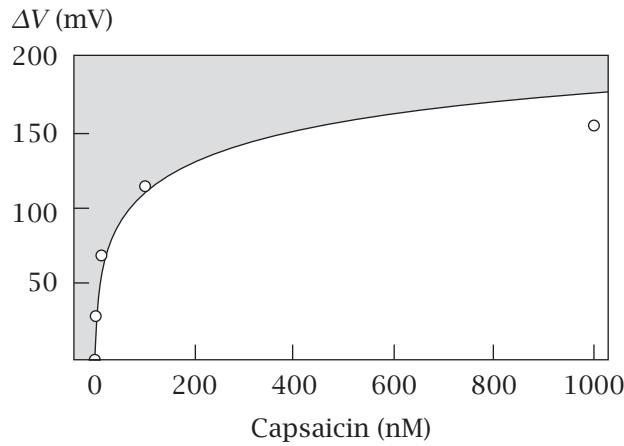


Figure 29.15 Capsaicin receptor midpoint voltage, $\Delta V(x)$ versus capsaicin concentration x , at 24°C . Capsaicin concentrations of 1000 nM, 100 nM, 10 nM, 1 nM, and 0 nM. The model curve is a fit of Equation (29.39) with $K = 0.5 \text{ nM}^{-1}$; the rest of the parameters are given in the caption of Figure 29.13. Source: T Voets, G Droogmans, U Wissenbach, et al., *Nature* **430**, 748–754 (2004).

where ‘ mx ’ refers to the midpoint for ligand concentration x . As a reference, we also want the midpoint voltage in the absence of ligand, V_{m0} . Substitute $x = 0$ into Equation (29.37) to get

$$V_{m0} = -\frac{RT}{zF} \ln f. \quad (29.38)$$

(Another useful relation is $V_{m0} = \Delta G_0/(zF)$, obtained by combining Equation (29.38) with Equation (29.33).) Combine Equations (29.37) and (29.38) to get the negative shift $\Delta V(x)$ of the midpoint voltage upon capsaicin binding to the receptor:

$$\Delta V(x) = V_{m0} - V_{mx} = \frac{RT}{zF} \ln(1+Kx). \quad (29.39)$$

Figure 29.15 compares Equation (29.39) with experiments. Like increasing the temperature, binding capsaicin to the hot receptor lowers the midpoint voltage of the channel-opening transition. This model illustrates how (1) the capsaicin receptor can convert changes in temperature to changes in voltage-gated ion currents and (2) effects of the binding of a ligand can mimic the effects of temperature. This is why capsaicin tastes hot and menthol tastes cool.

Our approach to modeling has been to hypothesize a binding-polynomial model and to include relevant Boltzmann factors for the physical factors that are important: temperature, voltage gating, and ligand binding. You then compute the consequences in terms of binding curves. You find model parameters that fit the data to the desired accuracy. Then, you use those curve-fit parameters to provide insights into the mechanism of the machine. If the model were not sufficient, you would then hypothesize a different model and try again.

Summary

Biological function depends on biochemical and genetic circuits. The elementary components of those circuits are proteins and nucleic acids. The principles of operation derive from coupled and cooperative binding and conformational changes of ligands with the biomolecules. When one ligand binds to a protein, it can change how other ligands bind, leading to the pumping of ions and protons, the synthesis of ATP from ADP, the transport of oxygen, the sensing of temperature, the conversion of chemical energy to force and motion, the control of catalytic processes, the winding and unwinding of DNA, and the networks of regulatory logic that control and regulate and monitor and drive biological systems.

Problems

1. A molecular machine. You have a complex binding process, a ‘molecular machine,’ in which three ligands X , Y , and Z bind to a molecule P . The free concentrations of the ligands in the solution are x , y , and z , respectively. Molecule X can bind with either ‘foot’ on P , with binding constant K_1 , or with both feet on P , with binding constant K_1^2 . Molecule Y can bind only if X is bound with both feet on P . Molecule Z can bind to P only if X is doubly bound and Y is bound. The binding equilibria are shown in Figure 29.16.

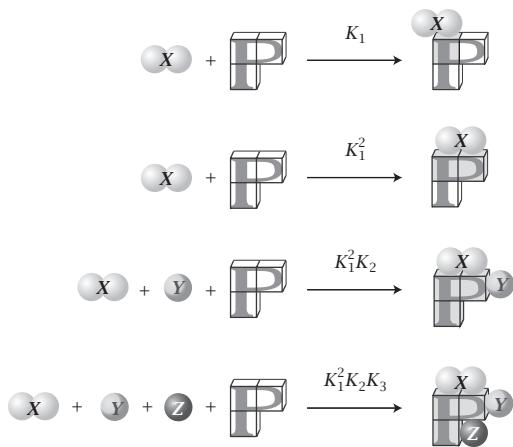


Figure 29.16 Different possible states of binding.

- Write the binding polynomial for this molecular machine.
 - What is the fraction of P molecules that have Y and Z and both feet of X bound?
 - What species dominate(s) at small x ?
- 2. Saturation of myoglobin.** Suppose that O_2 molecules bind to myoglobin with association constant $K = 2 \text{ Torr}^{-1}$ at 25°C and pH 7.4.
- Show a table of the fractional saturation of myoglobin for pressures of 1, 2, 4, 8, and 16 Torr O_2 .
 - Does the fractional saturation double for each doubling of the pressure?
- 3. Oxygen shifts the dimer/tetramer equilibrium in hemoglobin.** Hemoglobin dissociates in two steps: tetramer \rightarrow dimers \rightarrow monomers. Oxygen binding alters the equilibrium between hemoglobin dimers and tetramers. The free energies shown in Figure 29.17 have been measured.
- What is ΔG_4 ?
 - Does oxygen binding shift the equilibrium toward the dimers or toward the tetramer?

4. Helicases unwind DNA. Suppose DNA has two states: wound (helical) and unwound. In chromosomal DNA, the wound state is more stable than the unwound form of DNA by $-4.1 \text{ kcal mol}^{-1}$ (per unit length).

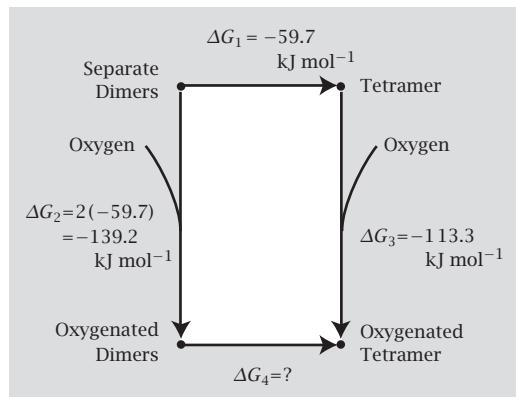


Figure 29.17 Thermodynamic cycle for O_2 binding and tetramer formation of hemoglobin. Source: FC Mills and GK Ackers, *J Biol Chem* **259**, 2881–2887 (1979); GK Ackers, *Biophys J* **32**, 331–346 (1980).

- Compute the equilibrium constant K at $T = 300 \text{ K}$ for unwound DNA \xrightarrow{K} wound DNA.
- Helicase is a protein that binds with equilibrium binding constant $B_h = 10$ to wound DNA (see Figure 29.18). Using the thermodynamic cycle in Figure 29.18, compute the binding constant B_u for binding helicase to unwound DNA that would be required to unwind the DNA (half the DNA is unwound and half is helical when helicase is unbound and $K_0 = 1$).

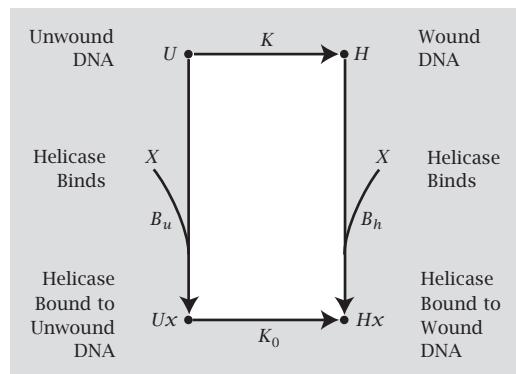


Figure 29.18 Thermodynamic cycle for binding helicase to DNA. B_u and B_h are the equilibrium binding constants for binding helicase protein to unwound and wound DNA, respectively. $K = [\text{helical DNA}]/[\text{unwound DNA}]$ is the equilibrium constant for helix formation in DNA in the absence of helicase and K_0 is that equilibrium constant in the presence of ligand.

- (c) Write an expression for the fraction of DNA molecules f_u that are unwound (including both u and ux) as a function of the concentration x of helicase in solution and of K , B_h , and B_u .
- (d) For the values of K and B_u that you found, and for $B_h = 0.1$, find the value of x that gives $f_u \approx 1/2$.

5. Oxygen binding to hemoglobin and myoglobin.

Figure 29.19 compares the single-site binding of oxygen to myoglobin with the four-site cooperative binding to hemoglobin. What biological advantage of hemoglobin is evident from these curves?

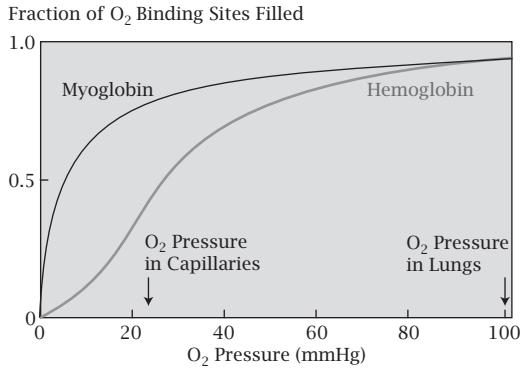


Figure 29.19 The binding of O_2 to hemoglobin and myoglobin. Source: J Darnell, H Lodish, and D Baltimore, *Molecular Cell Biology*, 2nd edition, WH Freeman, San Francisco, 1990.

6. Inhibitors of enzyme kinetics.

The Michaelis-Menten model of enzyme kinetics gives the reaction velocity v in terms of a maximum rate v_{\max} as

$$v = v_{\max} \left(\frac{Kx}{1+Kx} \right), \quad (29.40)$$

where K is the binding constant ($K = 1/K_M$, where K_M is the Michaelis constant) and x is the substrate concentration. A Lineweaver-Burk plot is the linearized version of Equation (29.40): a plot of $1/v$ versus $1/x$.

- (a) Give the quantities (i), (ii), and (iii) shown in Figure 29.20 in terms of K and v_{\max} .
 (b) Write the linearized form: $1/v$ versus $1/x$.

Competitive inhibitors obey the expression

$$v = v_{\max} \left(\frac{K_x x}{1+K_x x + K_y y} \right).$$

Noncompetitive inhibitors obey the expression

$$v = v_{\max} \left(\frac{K_x x}{1+K_x x + K_y y + K_x K_y x y} \right).$$

Uncompetitive inhibitors obey the expression

$$v = v_{\max} \left(\frac{K_x x}{1+K_x x + K_x K_y x y} \right).$$

- (c) Sketch the plots of $1/v$ versus $1/x$ for competitive, noncompetitive, and uncompetitive inhibition. In each case, include the curve for the uninhibited rates.

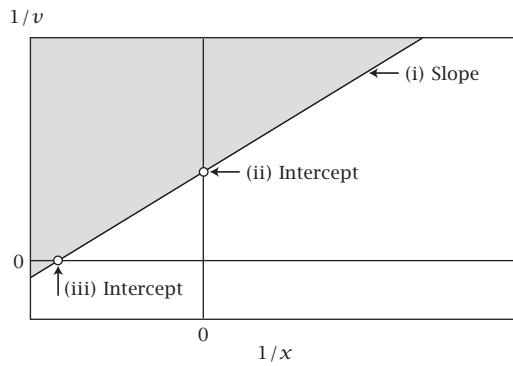


Figure 29.20 A Lineweaver-Burk plot has a slope and two intercepts that can help determine the nature of an inhibitor.

7. DNA-binding protein cooperativity. Single-strand binding protein (SSB) is a tetrameric protein that binds single-stranded DNA oligomers. One DNA oligomer can bind to each protein monomer. A maximum of four oligomers can bind one tetrameric protein. The binding constant for one oligonucleotide is K . x is the concentration of oligonucleotide. Binding is observed to be cooperative. Bujalowski and Lohman [14] have proposed a Pauling-like model of cooperativity. Think of the four protein subunits as being arranged in a square. A multiplicative ‘cooperativity’ factor f applies whenever two ligands occupy *adjacent* sites on the protein square, but not when two ligands are *diagonally across* from each other.

- (a) Write the binding partition function in terms of K , f , and x .
 (b) Compute $v(x)$, the average number of DNA ligands bound per protein tetramer.

8. Helicase nucleotide binding site. DNA B helicase is a hexameric protein that binds nucleotide ligands (see Figure 28.1). The binding fits a Pauling-like model involving six sites with binding constant K , arranged in a hexamer, and with interaction parameter f whenever two ligands are on adjacent sites.

- (a) Write the binding polynomial for this model.
 (b) Write an expression for the average number of sites filled as a function of ligand concentration x .

9. Running a molecular transporter backwards. A molecular transporter undergoes the thermodynamic cycle shown in Figure 29.21. Equilibrium and rate constants are given in the diagram. Assume that the concentration of the ligand on the left (C_L) is smaller than that

on the right ($C_L/C_R < 1$). Certain sets of values T, K_L, K_R, k_f, k_r will allow for the transport of the ligand up a concentration gradient from a region of lower concentration (C_L) to a higher concentration (C_R).

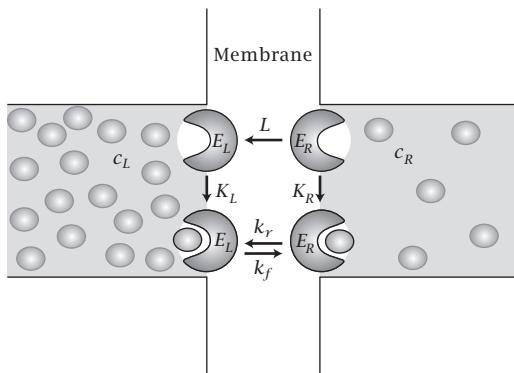


Figure 29.21 The four states of a membrane transporter protein.

- (a) Show that the flux around the cycle is

$$\frac{k_f T K_L C_L - k_r K_R C_R}{(1 + K_L C_L) + T(1 + K_R C_R)}$$

- (b) If you want to pump uphill from left to right, for a given value of $C_L/C_R < 1$, what ratio of the quantities T, K_L, K_R, k_f, k_r is required?
(c) For $C_L = 0$, write an expression for the cycle flux J versus C_R that gives a linearized plot. What is the slope and what is the intercept?

10. A model transcriptional regulatory system. Consider the genetic regulatory system shown in Figure 29.22. The action of the system depends on the concentrations $[DNA]$, $[P]$, $[R]$, and $[A]$, where DNA is the operon DNA, P is the RNA promoter protein, R is the repressor protein, and A is the activator protein.

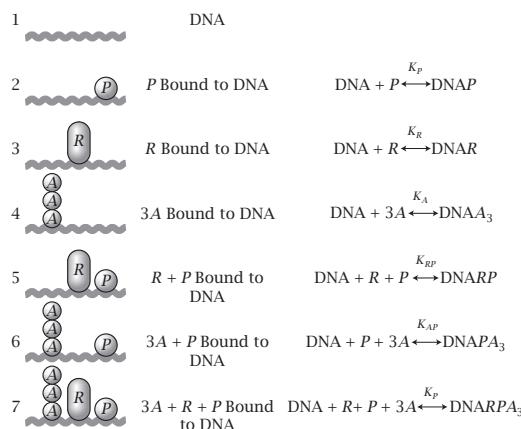


Figure 29.22 Different binding sites of a transcriptional regulator.

- Write the binding polynomial Q for this system.
- It is found that only state 6 leads to transcription of the *lac* genes. Write an expression for f_6 , the population of molecules in state 6, as a function of the species concentrations.
- Draw a qualitative plot of f_6 versus $[A]$, with all else constant.
- Draw a qualitative plot of f_6 versus $[R]$, all else constant.
- Draw a qualitative plot of f_6 versus $[P]$, all else constant.

11. Affinities of protein inhibitors. A receptor protein P binds a metabolite molecule M with affinity $K_{\text{assoc}} = 10^6 \text{ M}^{-1}$.

- What ligand concentration in solution is required so that 75% of the receptor sites are occupied?
- A competitive inhibitor drug binds to the same protein with nanomolar affinity, i.e., $R_{\text{assoc}} = 10^9 \text{ M}^{-1}$. If the inhibitor has concentration $\gamma = 10^{-7} \text{ M}$ in solution, in the situation described above, what is the fractional occupancy of P now?

12. Cooperative protein binding to a DNA plasmid. Many different classes of proteins can bind to DNA. By binding to a promoter, these proteins can act to either increase or decrease the expression of a gene. The portion of the DNA sequence that a protein binds to is called a box.

There is a protein HU that binds to sites on a DNA plasmid, which helps the DNA to maintain an unkinked state. A second protein IHF binds to specific boxes and forces the plasmid to adopt a 90° angle. The resulting kinked structure can cause a cascade of stress response events. For example, this kinked structure is observed in bacteria when activating a virulence plasmid.

Consider a simplified model of the process. There are only three IHF-binding sites on the plasmid. Each IHF-binding site can bind either one IHF protein or up to n HU proteins. When IHF binds, it excludes all n binding sites in a mutually exclusive manner. An idealized schematic is shown in Figure 29.23.

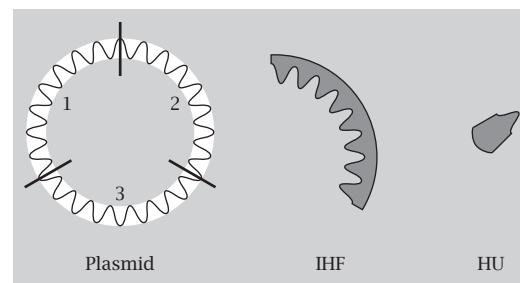


Figure 29.23 Two proteins, HU and IHF, bind to a DNA plasmid.

For purposes of illustration, a possible binding configuration is shown in Figure 29.24.

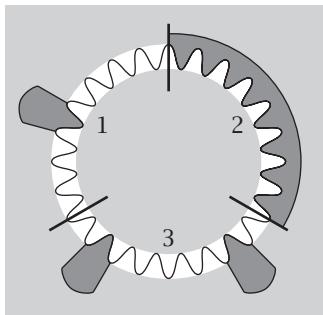


Figure 29.24 The bound-state complex.

In this case, there is an IHF bound to site 2 (excluding the possibility of an HU binding there), two HU's bound to site 3, and one HU bound to site 1.

- Express the binding polynomial for HU, assuming there is no IHF present, in terms of the binding constant K_H . Also express the binding polynomial for IHF, assuming there is no HU present, in terms of the binding constant K_I .
- Now, let's consider the effect of cooperativity. Using the Pauling model for cooperativity, express the binding polynomial for IHF, assuming there is no HU present. Use ϵ as the interaction free energy for each pair of IHF molecules that are bound.
- For the model in (b), what is the average number of IHF ligands bound per plasmid?
- Express the binding polynomial, assuming both HU and IHF are present. Assume there is no cooperativity among any of the sites.
- According to the model in (d), what is the average number of HU ligands bound to DNA? What happens in the limit of large HU? What happens in the limit of large IHF?

13. Activator or inhibitor? From binding experiments, you determine that a binding polynomial is

$$Q = 1 + Kx + RKxy + TRKxyz^4,$$

where x , y , and z represent the concentrations of three different ligands, and K , R , and T represent their binding affinities. Explain how ligand z binds and whether it can be an activator and inhibitor.

14. Binding polynomials. Here are two binding polynomials:

- $Q = 1 + K_3z + K_3K_2zy + K_3K_2K_1xyz;$
- $Q = 1 + K_3z + K_2zy + K_1xyz.$

- Which describes the following process: x can bind to P (with affinity K_1) only if y and z are bound, y can bind to P (with affinity K_2) only if z is bound, and z binds to P with affinity K_3 .
- For the Q you chose, write an expression for the fraction of molecules v_y that have y bound.

15. Cooperative binding to an enzyme. An enzyme has three binding sites: A , B , and C . A ligand can bind to the enzyme at all three sites; however, site C will not take up ligand until both sites A and B have been bound. And, sites A and B are independent of each other. This system is a single domain on the protein.

- Express the binding polynomial for this system.
- Write an expression for the average number of bound ligands.
- Express a different binding polynomial, assuming that the enzyme is uncompetitively inhibited by a different inhibitor at site C and that the overall system consists of three independent domains.
- What is the average number of bound ligands in this latter case?

References

- [1] GS Adair, *J Biol Chem* **63**, 529–545 (1925).
- [2] L Pauling, *Proc Natl Acad Sci USA* **21**, 186–191 (1935).
- [3] J Monod, J Wyman, and JP Changeaux, *J Mol Biol* **12**, 88–118 (1965).
- [4] MF Perutz, AJ Wilkinson, M Paoli, and GG Dodson, *Annu Rev Biophys Biomol Struc* **27**, 1–34 (1998).
- [5] WA Eaton, *Dahlem Workshop on Simplicity and Complexity in Proteins and Nucleic Acids*, Berlin, 1998.
- [6] J Li, WN Zagotta, and HA Lester, *Q Rev Biophys* **30**, 177–193 (1997).
- [7] J Wyman and SJ Gill, *Binding and Linkage: Functional Chemistry of Biological Macromolecules*, University Science Books, Mill Valley, CA, 1990.
- [8] A Szabo and M Karplus, *J Mol Biol* **72**, 163–197 (1972).
- [9] GK Ackers, *Adv Prot Chem* **51**, 185–253 (1998).
- [10] Q Cui and M Karplus, *Protein Sci* **17**, 1295–1307 (2008).
- [11] T Mori, RD Vale, and M Tomishige, *Nature* **450**, 750–754 (2007).
- [12] MJ Schnitzer, K Visscher, and SM Block, *Nat Cell Biol* **2**, 718–723 (2000).
- [13] T Voets, G Drogomas, U Wissenbach, et al., *Nature* **430**, 748–754 (2004).
- [14] W Bujalowski and T Lohman, *J Mol Biol* **207**, 249–268 (1989).

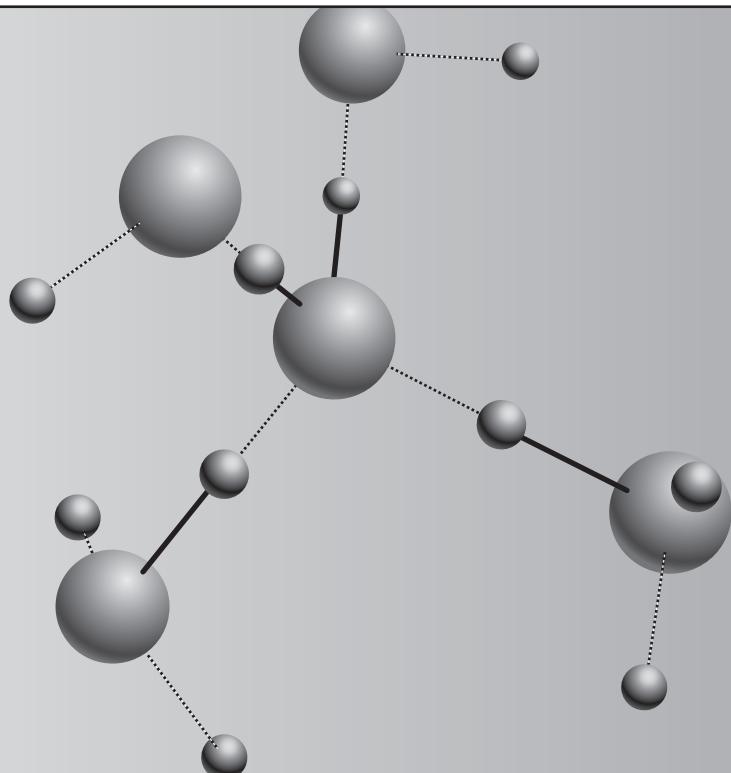
Suggested Reading

The basics of binding polynomials, emphasizing hemoglobin:

- U Alon, *An Introduction to Systems Biology: Design Principles of Biological Circuits*, Chapman and Hall, London, 2007.
- L Bintu, NE Buchler, HG Garcia, et al., *Curr Opin Genet Dev* **15**, 116–124 (2005).
- E Di Cera, *Thermodynamic Theory of Site-Specific Binding Processes in Biological Macromolecules*. Cambridge University Press, New York, 1995.
- R Phillips, J Kondev, and J Theriot, *Physical Biology of the Cell*, Garland Science, New York, 2009.
- J Wyman and SJ Gill, *Binding and Linkage: Functional Chemistry of Biological Macromolecules*, University Science Books, Mill Valley, CA, 1990.
- A detailed discussion of inhibitors is given in:
- IH Segel, *Biochemical Calculations: How to Solve Mathematical Problems in General Biochemistry*, 2nd edition, Wiley, New York, 1976.
- Excellent reviews of hemoglobin cooperativity:
- GK Ackers, *Adv Prot Chem* **51**, 185–253 (1998).
- WA Eaton, *Dahlem Workshop on Simplicity and Complexity in Proteins and Nucleic Acids*, Berlin, 1998.
- MF Perutz, AJ Wilkinson, M Paoli, and GG Dodson, *Annu Rev Biophys Biomol Struc* **27**, 1–34 (1998).
- A Szabo and M Karplus, *J Mol Biol* **72**, 163–197 (1972).
- Outstanding treatments of biochemical machines and circuits:
- U Alon, *An Introduction to Systems Biology: Design Principles of Biological Circuits*, Chapman and Hall, London, 2007.
- DA Beard and H Qian, *Chemical Biophysics: Quantitative Analysis of Cellular Systems*, Cambridge University Press, Cambridge, 2008.
- TL Hill, *Cooperativity Theory in Biochemistry: Steady-State and Equilibrium Systems*, Springer-Verlag, New York, 1985.
- TL Hill, *Free Energy Transduction and Biochemical Cycle Kinetics*, Springer-Verlag, New York, 1989 (reprinted by Dover Publications, New York, 2004).
- J Howard, *Mechanics of Motor Proteins and the Cytoskeleton*, Sinauer Associates, Sunderland, MA, 2001.
- R Phillips, J Kondev, and J Theriot, *Physical Biology of the Cell*, Garland Science, New York, 2009.
- Definitive works on the biology of gene regulation:
- B Alberts, A Johnson, J Lewis, et al., *Molecular Biology of the Cell*, 5th edition Garland Science, New York, 2008.
- M Ptashne, *A Genetic Switch: Phage Lambda Revisited*, 3rd edition, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, 2004.

This page is intentionally left blank.

30 Water



Water Has Anomalous Properties

Water is the most abundant liquid on Earth. Yet it is often considered *anomalous* because it behaves somewhat differently from simpler and better understood liquids such as argon. Argon atoms can be modeled as Lennard-Jones spheres. Water molecules attract and repel each other like argon atoms do, but water can also form hydrogen bonds. Hydrogen bonding introduces orientation dependence into the intermolecular interactions. So, water has some properties of a normal liquid and some properties of a fluctuating tetrahedral network of hydrogen bonds. In this chapter, we look at the properties of pure water. In Chapter 31, we look at water as a solvent for polar and nonpolar molecules.

Hydrogen Bonds Are a Defining Feature of Water Molecules

Composed of only three atoms, water is small and compact. To a first approximation, a water molecule is spherical. Water's two hydrogen atoms and two lone pairs of electrons are arranged in near-tetrahedral symmetry around an oxygen atom. Water has no net charge, but it has a permanent dipole moment, which can be described as a partial negative charge on the oxygen and two partial positive charges, one near each of the two hydrogens. The permanent dipole moment is the vector sum of these two O–H vectors.

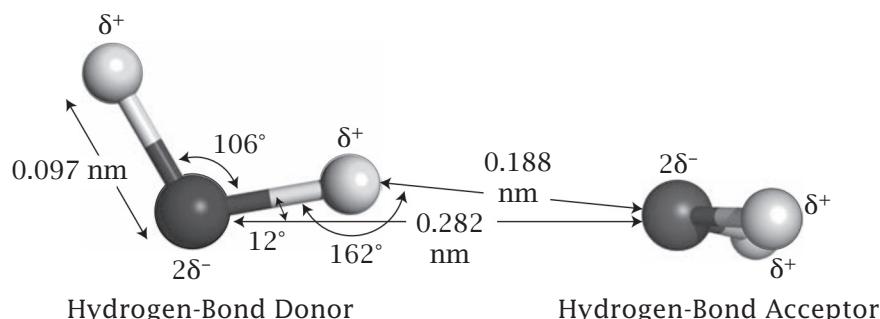


Figure 30.1 The angles and distances that define an average hydrogen bond between two water molecules. Source: M Chaplin, Water's hydrogen bond strength, arXiv: 0706.1355 [cond-mat.soft].

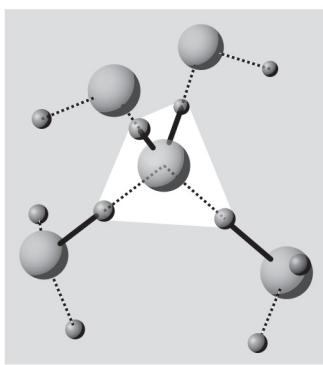


Figure 30.2 Water forms a tetrahedral bonding arrangement with neighboring water molecules.

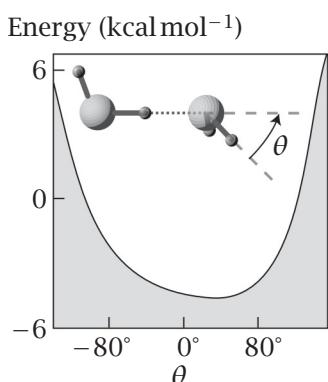


Figure 30.3 Energy versus angle θ of a hydrogen bond between two water molecules in the gas phase, from molecular simulations. Source: FH Stillinger, *Science* **209**, 451–457 (1980).

Two water molecules can form hydrogen bonds with each other. Hydrogen bonds are formed when a hydrogen *donor* atom is near an *acceptor* atom in the appropriate orientation. Some types of atom are good at donating hydrogens, some are good at accepting, and some do neither. Water molecules can act as both donors and acceptors for hydrogen bonds. A water molecule *A* can donate a hydrogen bond to a water molecule *B* and accept a hydrogen bond from another water *C*. The OH bond is a donor: it can stretch to share its hydrogen with an acceptor. Each lone pair of electrons on the oxygen of a water molecule can be an acceptor, and can bond to a donor's hydrogen. Figure 30.1 shows the average structure of a water-water hydrogen bond. A water molecule can participate in any number of hydrogen bonds from zero to four. Figure 30.2 shows a water molecule in the tetrahedral hydrogen-bonding arrangement that is common in ice and in liquid water.

How strong is a water-water hydrogen bond? Figure 30.3 shows a plot of energy versus hydrogen-bond angle for a water dimer in the vapor phase, based on molecular orbital calculations reviewed by Stillinger [1]. The most favorable orientation between two water molecules occurs when the hydrogen bond between them is approximately straight. The energy of an optimal water-water hydrogen bond in the gas phase is estimated to be about $5.5 \text{ kcal mol}^{-1}$.

Hydrogen Bonding in Water Is Cooperative

The hydrogen bonds in water are *cooperative*, meaning that the strength of a hydrogen bond depends on its relationship with other neighboring water molecules. We know this from studies of small polygonal clusters of water molecules in the gas phase. Such clusters serve as useful models for the types of hydrogen-bonded polygons that are components of bulk water and of solvation shells around solutes (see Chapter 31). For example, Figure 30.4 shows how the oxygen-oxygen distance between water molecules and the average dipole moment per water molecule change with the number of water molecules in the cluster. The dipole moment is 1.855 debyes (D) for a single isolated water molecule; 2.7 D for some of the polyhedral clusters having five to eight water molecules; and 2.4–2.6 D for the average water molecule in bulk liquid water at 0°C .

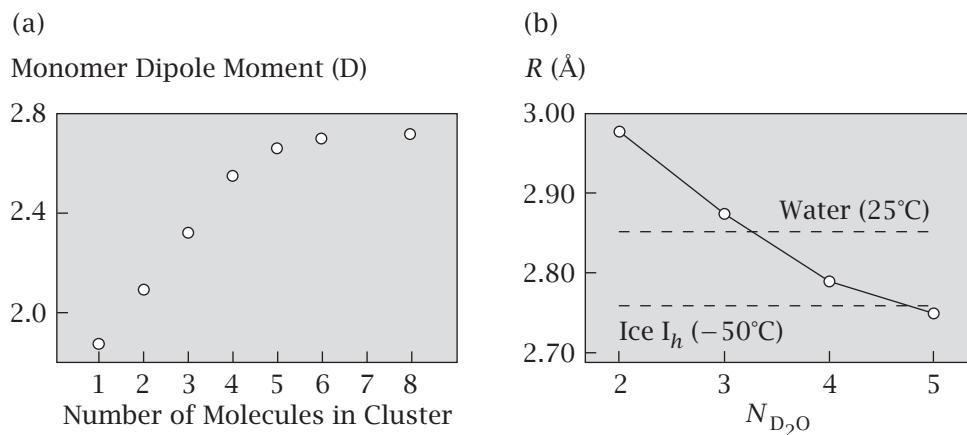


Figure 30.4 Two measures of hydrogen bond cooperativity. (a) Quantum-mechanical calculations show that the dipole moment per water increases with the size of a water cluster. (b) The average oxygen–oxygen distance R decreases with increasing cluster size N_{D_2O} . Source: JK Gregory, DC Clary, K Liu, et al., *Science* **275**, 814–817 (1997); JD Corzan, LB Braly, K Liu, et al., *Science* **271**, 59–62 (1996).

Hydrogen-bonding cooperativity arises from the shifting of electron density among water molecules. Suppose water molecule A donates a hydrogen from its OH group to the lone pair of electrons on the acceptor water molecule B. When a hydrogen bond forms, the hydrogen atom shifts toward molecule B, while electron density shifts toward water molecule A. The loss of electron density on molecule B results in more positive charge on its hydrogens. So molecule B can more readily donate its hydrogens to water molecule C. The shifting of electron density accounts for the higher dipole moments of water molecules in cluster and bulk states than in the gas phase. Similar cooperativities appear in other highly polar hydrogen-bonded liquids, such as HF, H₂O₂, and HCN.

Water Has Tetrahedral Coordination

The tetrahedral symmetry of a water molecule defines the structural framework for the solid, ice. Crystal structures show that the water molecules in ice I_h (the form that is stable at 0°C at 1 atm pressure) have tetrahedral symmetry (see Figure 30.5). Hydrogen bonding appears to be a dominant component of the energetics of water in its solid phase.

Liquid water also has some tetrahedral structure because melting does not fully disrupt the icelike cage structuring. This is known from the radial distribution functions for liquid water that are observed in x-ray and neutron-diffraction experiments. Figure 30.6 shows the radial distribution functions of two liquids: water and argon.

Figures 30.6 and 30.7 indicate that liquid water is tetrahedrally coordinated. Integrating under the first peak of the radial distribution function in Figure 30.6 gives the average number of nearest neighbors surrounding each molecule (see the discussion of Figure 24.10). The number of nearest neighbors around a water molecule is 4.4 at 4°C, and remains about 4 over a broad range of

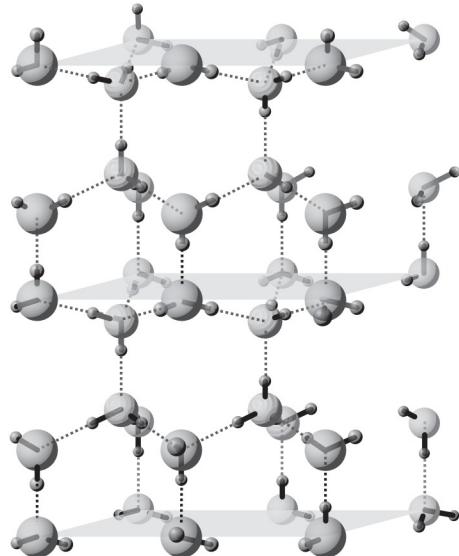


Figure 30.5 The structure of ice I_h .

temperatures from 4°C to 200°C (under high pressures at 200°C , to keep water liquid). In contrast, the number of nearest neighbors in liquid argon is 10 at low temperatures. That number diminishes at higher temperatures. Argon packing is comparable to the closest packing of hard spheres, in which each sphere has 12 nearest neighbors. The peak of the radial distribution function shows

Figure 30.6 The oxygen–oxygen pair correlation function $g(R^*)$ for liquid water at 4°C (—) compared with the pair correlation function for argon at $T = 84.25\text{ K}$ and 0.71 atm (—). $R^* = r/\sigma$ is the distance r between two oxygens, normalized by the bond length, where σ is 2.82 \AA for water and 3.4 \AA for argon.

The integral under the first peak shows that water molecules have fewer first-shell neighbors (about 4.4) than argon (about 10). Source: F Franks, *Water*, Royal Society of Chemistry, London, 1983.

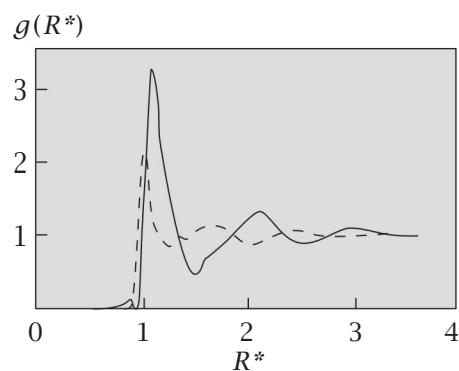
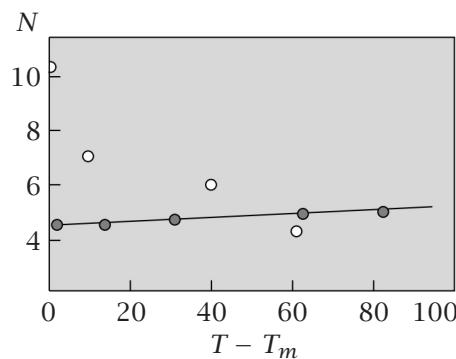


Figure 30.7 The number N of closest neighbors around water (●), and around argon (○) as a function of the temperature difference from the melting point, $T - T_m$. Water has 4–5 nearest neighbors through the full liquid range. Argon has 10 neighbors in its cold liquid state. Source: F Franks, *Water*, Royal Society of Chemistry, London, 1983.



that liquid water has oxygen–oxygen distances from 2.82 Å at 4°C to 2.92 Å at 200°C. Water dimers have longer bonds in the vapor phase (2.98 Å), and shorter ones in ice (2.74 Å) near 0 K.

Liquid water is more ordered than argon. This is indicated in Figure 30.6. The width of the first peak of the correlation function provides a measure of disorder in the first-neighbor shell. It indicates the variability in the numbers of first neighbors. Also, the entropy of vaporization is larger for water ($109\text{ J K}^{-1}\text{ mol}^{-1}$) than for other simple liquids ($70\text{--}90\text{ J K}^{-1}\text{ mol}^{-1}$), indicating that more structure is broken by boiling water than by boiling argon.

The oscillations in the oxygen–oxygen pair correlation function become negligible after about the third-neighbor shell in water, as in argon. While this gives some evidence that the structuring influence of each water molecule on other waters is relatively localized, the orientational correlations—which may be a better measure of water’s structural organization—extend out to more distant neighbors.

Hydrogen Bonding Weakens with Temperature in Liquid Water

What is the strength of a hydrogen bond in liquid water, and how does it change with temperature? Estimates of the numbers, strengths, and angles of the hydrogen bonds in liquid water come from vibrational spectroscopy. Raman spectroscopy has been used to study the stretch vibration of O–D bonds when a small amount of deuterated water D₂O is dissolved in a solvent of H₂O. The spectral shift from the D₂O peak at 2500 cm⁻¹ (at 20°C) toward 2650 cm⁻¹ (at 400°C, liquid at high pressure) is generally taken to indicate that hydrogen bonds bend or break with increasing temperature.

To say that a hydrogen bond has a ‘strength’ requires that there be two distinguishable states, made and broken, for which a free energy difference can be determined. For liquid water, it has long been debated whether hydrogen bonds fall into two such classes, or whether they are stretched or bent through a continuum of states. The first ‘two-state’ model was due to WK Röntgen in 1892 [2], who was better known for his discovery of x-rays, for which he won the first Nobel Prize in Physics in 1901. He proposed that liquid water has two components: bulk and dense. A model by Nemethy and Scheraga [3] postulates five species: waters having zero to four hydrogen bonds. While much evidence supports the view that there is a continuum of hydrogen bonding in water, the two-state model is often found to be a useful approximation. Support for the two-state model comes partly from Raman experiments that show an *isosbestic point* (see Example 26.2), in which spectra at a series of different temperatures all intersect at a single point (see Figure 30.8).

Computer simulations of liquid water at $T = 298\text{ K}$ also show that the continuum of hydrogen bonding in liquid water falls into two classes. Simulations show that breaking a hydrogen bond increases the free energy by $\Delta G = 480\text{ cal mol}^{-1}$, with corresponding enthalpy and entropy increases of $\Delta H = 1.9\text{ kcal mol}^{-1}$ and $\Delta S = 4.77\text{ cal K}^{-1}\text{ mol}^{-1}$ [4]. Figure 30.9 shows the prediction for how the amount of hydrogen bonding in liquid water decreases with temperature from the melting point to the boiling point. Most models predict

Fraction Hydrogen-Bonded

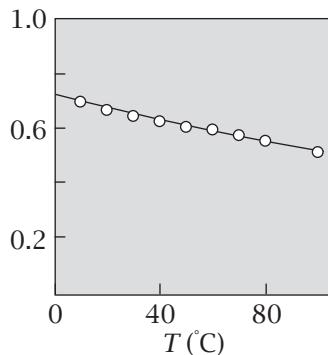
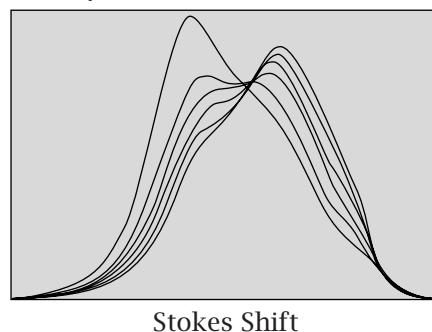


Figure 30.9 Fraction of hydrogen bonds made in liquid water over the liquid temperature range. Sources: DE Hare and CM Sorensen, *J Chem Phys* **93**, 6954–6960 (1990); KAT Silverstein, ADJ Haymet, and KA Dill, *J Am Chem* **122**, 8037–8041 (2000).

Figure 30.8 Raman spectra of water at different temperatures. The isosbestic point supports a model involving two states. Source: G D'Arrigo, G Maisana, F Mallamace, et al., *J Chem Phys* **75**, 4264–4270 (1981).

Intensity



Fraction of Molecules

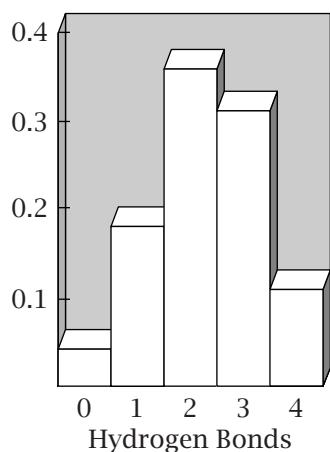


Figure 30.10 Simulated distribution of the number of hydrogen bonds per molecule in liquid water at 25°C. Source: FH Stillinger, *Science* **209**, 451–457 (1980).

that not all the hydrogen bonds are made at the melting point and not all the hydrogen bonds are broken at the boiling point.

Computer simulations suggest that liquid water has the structure of a fluctuating network of hydrogen bonds. Figure 30.10 shows a model distribution of hydrogen bonds in cold water, around room temperature. The distribution shifts with increasing temperature toward fewer average hydrogen bonds per water molecule. *Random network models* and *continuum models* predict that hydrogen bonds are stretched and bent through a continuum of angles. Figure 30.11 shows how hydrogen bonds loosen (stretch and bend) with temperature and how they compress and bend with pressure, according to a model of Henn and Kauzmann [5] that is based partly on measured spectral changes with temperature.

Pure Water Has Anomalous Properties

Increasing the temperature of water causes the solid phase (ice) to melt to a liquid, and then causes the liquid phase to boil to become a gas. In this regard, water is normal. But water is unusual in more subtle ways. Table 30.1 compares the physical properties of water with those of other liquids and solids. Water has a high dielectric constant, in part because hydrogen bonds are polarizable. So water is a better solvent for ions than many other liquids. Water is more cohesive than liquids of comparable molecular size because of water-water hydrogen bonding. This cohesion is indicated in Figure 30.12, which shows the high boiling and melting temperatures of water, and in Table 30.1, which shows water's relatively high enthalpy of vaporization. Water also has high melting and critical temperatures, and a high surface tension (about 70 dyn cm⁻¹) compared with the surface tensions of alkanes (about 30 dyn cm⁻¹).

Liquid water has a relatively large heat capacity for its size (see Table 30.1 and Figure 30.13(a)). The heat capacity $C_p = (\partial H / \partial T)$ describes the storage of energy (or enthalpy) in bonds that break or weaken with increasing temperature. A statistical mechanical model of Dahl and Andersen [6] shows that water is able to store energy in hydrogen bonds that can weaken, as well as van der Waals interactions that can break or weaken.

An interesting property of water is that a localized change in pH propagates through the liquid unusually fast—faster than is predicted by the diffusion constant for the hydrogen ions. Why? A model of Stanley and Teixeira [7] shows

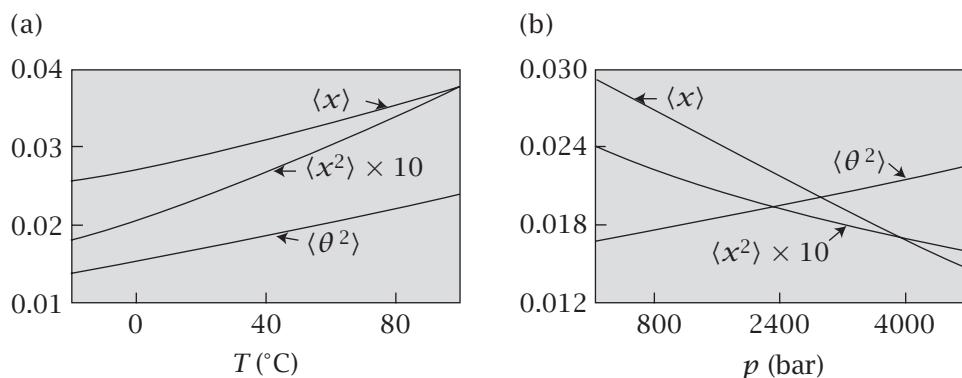


Figure 30.11 The model of Henn and Kauzmann indicates that (a) increasing the temperature T stretches and bends the water–water hydrogen bonds, while (b) increasing the pressure p shortens and bends those bonds. $x = (R - R_0)/R_0$ is the dimensionless O–O bond length, representing the hydrogen bonding length, where $R_0 = 2.75 \text{ \AA}$ is the O–O separation in ice, and θ is the hydrogen-bond angle (O–H–O, between the donor O–H and the O–O axis, in radians). Source: AR Henn and W Kauzmann, *J Phys Chem* **93**, 3776–3783 (1989).

Table 30.1 Physical properties of some typical liquids.

	Argon		Benzene		Water	
	Solid	Liquid	Solid	Liquid	Solid	Liquid
Density (kg m^{-3})	1636	1407	1000	899	920	997
Heat capacity ($\text{J K}^{-1} \text{ mol}^{-1}$)	30.9	41.9	11.3	132	37.6	75.2
Isothermal compressibility ($\text{N}^{-1} \text{ m}^{-2}$)	1	20	3	8.7	2	4.9
Self-diffusion coefficient ($\text{m}^2 \text{ s}^{-1}$)	10^{-13}	1.6×10^{-9}	10^{-13}	1.7×10^{-9}	10^{-14}	2.2×10^{-9}
Thermal conductivity ($\text{J s}^{-1} \text{ m}^{-1} \text{ K}^{-1}$)	0.3	0.12	0.27	0.15	2.1	0.58
Liquid range (K)		3.5		75		100
Surface tension (mJ m^{-2})		13		28.9		72
Viscosity (P)		0.003		0.009		0.01
Latent heat of fusion (kJ mol^{-1})		1.18		10		5.98
Latent heat of evaporation (kJ mol^{-1})		6.69		35		40.5
Melting point (K)		84.1		278.8		273.2

Source: F Franks, *Water*, Royal Society of Chemistry, London, 1983.

that liquid water is above a ‘critical percolation threshold,’ which means that uninterrupted pathways in hydrogen-bonding networks extend over macroscopic distances in water. The H^+ and OH^- ions can be transported rapidly via a ‘bucket brigade’ mechanism: one water donates a hydrogen to the next water, which passes a different hydrogen to water number three, and so on.

Another view of the cohesion of water comes from comparing H_2O with its deuterated and tritiated isotopes D_2O and T_2O . Figure 30.14 shows the entropy of transferring a single water molecule from the vapor phase into the pure liquid of each of these materials. The heavier isotopes have stronger bonds, presumably due to lower zero-point energies. The entropies measured in these

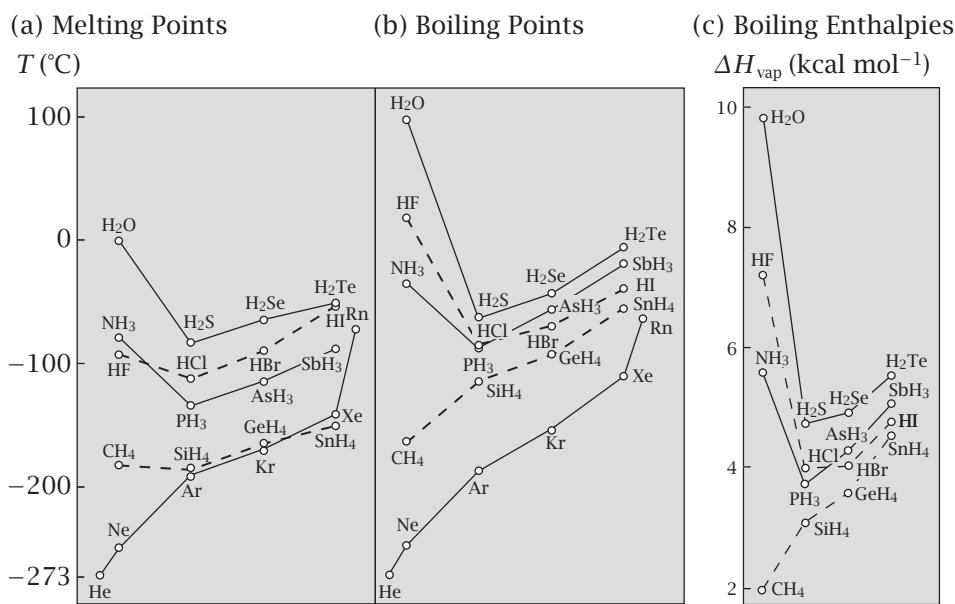


Figure 30.12 Water has higher (a) melting temperature, (b) boiling temperature, and (c) enthalpy of vaporization than comparable molecules. Each connected line represents a series of isoelectronic molecules, stepping downward in the periodic table. Alternate lines are broken to make them easier to follow. Source: L Pauling, *General Chemistry*, Dover Publications, New York, 1970.

experiments show that water composed of the heavier isotopes is more ordered than H₂O.

The Volumetric Anomalies of Water

Perhaps the most recognizable anomaly of water is its volume as a function of temperature. Ice floats. Water's solid is less dense than its liquid. In contrast, the solids of most other materials are denser than their liquids, so they sink. Melting most materials causes them to loosen up and expand but Figure 30.13(b) shows that melting ice increases water's density. Ice has low density because of its open tetrahedral structure.

This volumetric anomaly is also manifested in the $p(T)$ phase diagram (see Figure 30.15). For simple materials, the slope of the $p(T)$ phase boundary between the liquid and the solid is positive. Applying pressure to simple liquids squeezes the molecules together and drives them to freeze. But for water, the opposite happens: squeezing ice by applying pressure causes it to melt. For water, the slope of $p(T)$, of the liquid-solid phase boundary, is negative.

The slope of the phase boundary gives useful information about structural changes. The Clapeyron equation (14.17) gives $dp/dT = \Delta s/\Delta v$, where Δs and Δv are the changes in entropy and volume between the two phases. Fundamentally, the entropy of a liquid is always higher than its corresponding solid: $\Delta s = s_{\text{liquid}} - s_{\text{solid}} > 0$ (provided the liquid is the stable phase at the higher temperature). If dp/dT is positive, as it is for simple materials, the solid must be the denser phase: $\Delta v = v_{\text{liquid}} - v_{\text{solid}} > 0$. But if dp/dT is negative, as it is for

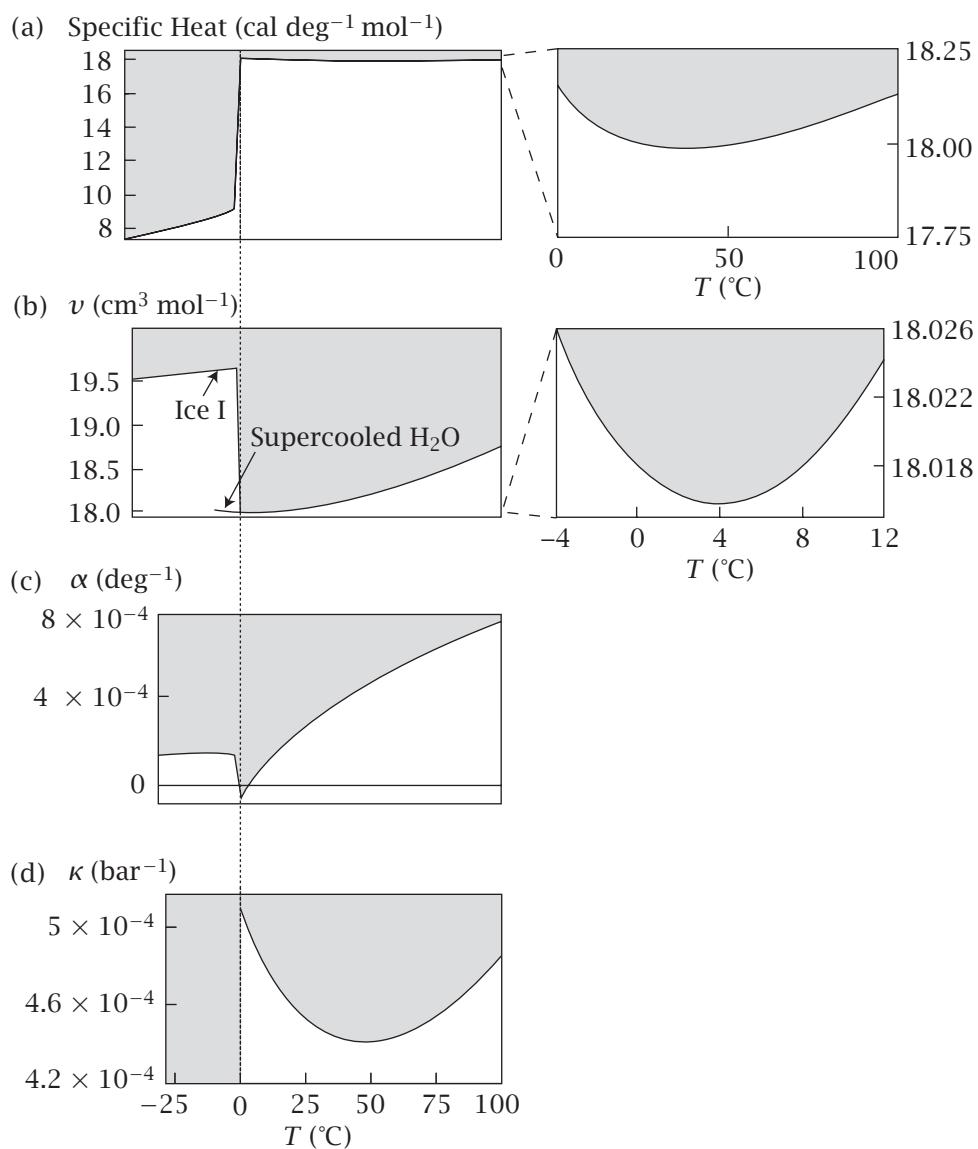


Figure 30.13 (a) The heat capacity of water versus temperature T , from ice at below -25°C to the boiling point. Comparison with Figure 7.2 shows that melting ice leads to a bigger increase in heat capacity $C_p = dH/dT$ than in simpler liquids. An expanded scale of the specific heat of water versus temperature is shown over the liquid range. Source: D Eisenberg and W Kauzmann, *The Structure and Properties of Water*, Oxford University Press, New York, 1969. (b) The molar volume ν of water versus temperature T . Ice is less dense than liquid water. The inset shows an expanded scale around the temperature of maximum density, 3.984°C . (c) The thermal expansion coefficient $\alpha = (1/V)(\partial V/\partial T)_p$ of water versus temperature T . (d) The isothermal compressibility $\kappa = (1/V)(\partial V/\partial p)_T$ of water. The minimum is around 46°C . Source of (b)-(d): GS Kell, *J Chem Eng Data* **12**, 66–69 (1967).

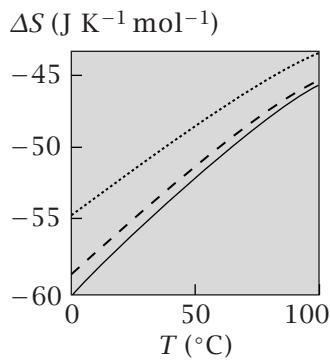


Figure 30.14 Entropies of transferring (.....) H_2O , (---) D_2O , or (—) T_2O from vapor to liquid water, versus temperature. The heavier isotopes have more ordered liquid phases. Source: A Ben-Naim, *Hydrophobic Interactions*, Plenum Press, New York, 1980.

water, the liquid must be denser: $v_{\text{liquid}} < v_{\text{solid}}$. Example 30.1 shows how Δs is computed from the slope of the phase boundary.

EXAMPLE 30.1 Computing Δs from the phase diagram. The slope of the liquid-solid boundary for the water-to-ice transition (Figure 30.15(a)) is found experimentally to be

$$\frac{dp}{dT} = -133 \text{ atm K}^{-1}.$$

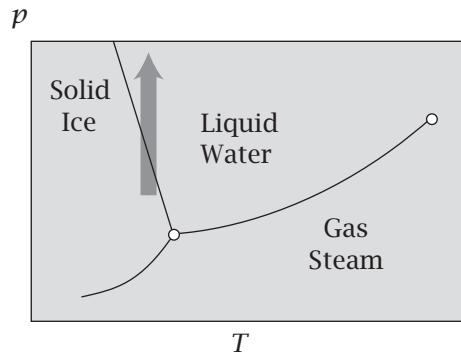
This means that you can lower the melting temperature of water by 1 K for every 133 atm of pressure that you apply. To compute Δs , use $v_{\text{liquid}} - v_{\text{solid}} = 18.02 \text{ cm}^3 \text{ mol}^{-1} - 19.66 \text{ cm}^3 \text{ mol}^{-1} = 1.64 \text{ cm}^3 \text{ mol}^{-1}$, and rearrange Equation (14.19) to get

$$\begin{aligned}\Delta s &= \Delta v \frac{dp}{dT} = \frac{(1.64 \text{ cm}^3 \text{ mol}^{-1})(-133 \text{ atm K}^{-1})(1.987 \text{ cal K}^{-1} \text{ mol}^{-1})}{(82.058 \text{ cm}^3 \text{ atm K}^{-1} \text{ mol}^{-1})} \\ &= 5.28 \text{ cal K}^{-1} \text{ mol}^{-1}.\end{aligned}$$

For the units conversion, we multiplied and divided by the gas constant R in the numerator and denominator. As a check, you can also get Δs from Δh , for the fusion process at equilibrium:

$$\Delta s_{\text{fusion}} = \frac{\Delta h_{\text{fusion}}}{T_{\text{fusion}}} = \frac{1.44 \text{ kcal mol}^{-1}}{273.15 \text{ K}} = 5.28 \text{ cal K}^{-1} \text{ mol}^{-1}.$$

(a) Phase Diagram for Water



(b) Phase Diagram for CO_2

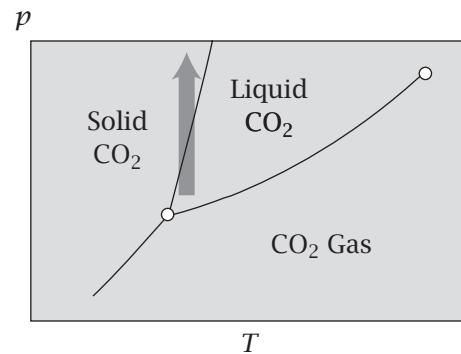


Figure 30.15 Phase diagrams, pressure p versus temperature T for (a) water and (b) CO_2 show that $dp/dT < 0$ for water and $dp/dT > 0$ for CO_2 . Applied pressure melts ice, but freezes CO_2 . Source: P Atkins, *Physical Chemistry*, 6th edition, WH Freeman, New York, 1977.

The reason that you can skate on ice has been explained as the pressure-induced melting of the ice beneath the blade of an ice skate. You can't skate on other solids. But if you use the phase diagram to calculate the reduction in melting temperature that would be caused by the pressure of an ice skater

(Problem 1 at the end of this chapter), you will find that skaters are melting ice with less pressure than the phase diagram predicts. Surface effects might also play a role in pressure-induced melting from ice skates. Less pressure is required to melt the first few molecular surface layers than is predicted by the phase diagram, because surface molecules are more fluid, because they have fewer intermolecular bonds.

Figure 30.13(b) shows a more subtle volumetric property of water called the *density anomaly*. Heating simple liquids expands them because the greater thermal motions push the molecules apart. In this regard, water is normal above 3.984°C. But below that temperature (and above the freezing point 0°C), heating *increases* water's density. 3.984°C is called the *temperature of maximum density*. The maximum density is subtle (see the exploded view in Figure 30.13(b)): the density increase from 0°C to 4°C is less than 1% of the density increase upon melting.

The volume anomalies are also reflected in Figure 30.13(c), which shows the thermal expansion coefficient $\alpha = (1/V)(\partial V / \partial T)_p$, the temperature derivative of the molar volume (Figure 30.13(b)). Water is anomalous where $\alpha < 0$, which occurs in the cold liquid state. For most materials, α is positive at all temperatures. For example, Figures 9.2 and 9.3 show that the molar volumes of polyethylene, benzene, and other liquids increase monotonically with temperature.

A few other systems, including silicon and zirconium tungstate, (ZrW_2O_8), are also orientationally structured and, like water, have negative thermal expansion coefficients. For these unusual materials, applying pressure melts out the structure and drives the system toward denser amorphous states.

The compressibility $\kappa = (1/V)(\partial V / \partial p)_T$ specifies how the volume of a material changes with pressure. In simple materials, the liquid state is more compressible than the solid state. Less pressure is needed to squeeze down the volume of a liquid than a solid. Temperature also affects the compressibility. Hot materials are usually more compressible than colder materials because hot materials are expanded and have broken bonds. In this regard, hot water (above 46°C) is normal; heating increases the compressibility of hot water. But cold water behaves differently. Figure 30.13(d) shows that water has an anomalous minimum in its isothermal compressibility. Heating cold water (below 46°C) makes it less compressible, presumably by breaking the hydrogen bonding structure, leading to denser water, which is less compressible than cage-like water. Interestingly, the minimum in the compressibility is at 46°C, which is different than the temperature of maximum density. Figures 9.3 and 9.5 show that the thermal expansion coefficient and compressibility are both substantially smaller for water than for small organic molecules, indicating that water's underlying framework is more solid-like and less deformable than that of other liquids.

Another anomalous property of water is the pressure dependence of its viscosity. The viscosity of a liquid is a measure of the friction that hinders the flow of molecules past each other. So, you can increase the viscosities of materials by changing the temperature and pressure in ways that strengthen their intermolecular attractions. Normally, heating a liquid loosens the noncovalent bonds, lowering the viscosity. And normally, applying pressure drives molecules together into stronger interactions, increasing the viscosity. Water

Figure 30.16 The viscosity of hot water increases with pressure, resembling simple liquids. But at low pressures, the viscosity of cold water (below about 20°C) decreases with applied pressure. Source: T DeFries and J Jones, *J Chem Phys* **66**, 896–901 (1977).

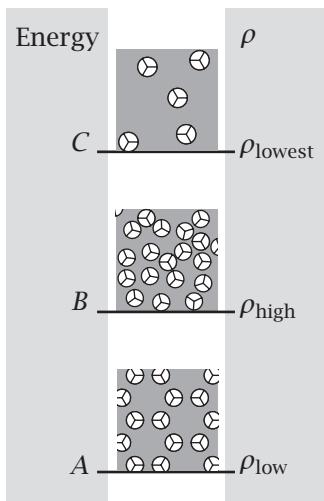
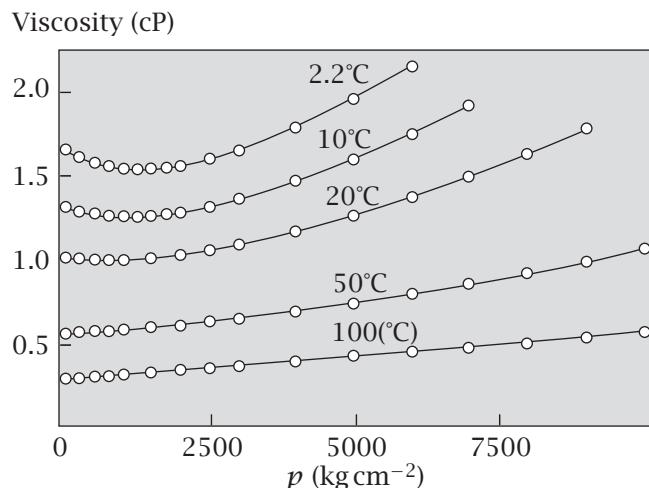


Figure 30.17 An energy diagram for three states of liquid water molecules. At energy level *A*, water has hydrogen-bonded tetrahedral ice-like structures of low density. At energy level *B*, higher-energy structures have weaker or broken hydrogen bonding, more van der Waals interactions, and higher density than *A*. At energy level *C*, bonds (van der Waals and hydrogen bonds) are weaker and looser than in *B*, so the density is lower. Increasing temperature leads from *A* to *B* to *C*, first increasing, then decreasing, the density.

follows this behavior at high pressures. But Figure 30.16 shows that applying pressure to cold water (at pressures of less than 1000 kg cm⁻²) decreases water's viscosity, presumably by crunching, bending, and breaking the hydrogen-bonding structure, weakening those interactions.

What causes the unusual volumetric behaviors of water? An explanation is given by the energy ladder in Figure 30.17. The state of lowest energy, *A*, arises from water's strongest bonds, its hydrogen bonds. Water's most stable structures are its cage-like open tetrahedral structures. So, cold water and ice have both low energy and low density. The next higher energy level is *B*, a state in which some hydrogen bonds are broken but additional (weaker) van der Waals interactions are made. So, warmer liquid water has some denser water clusters. The highest energy level is *C*, where both hydrogen bonds and van der Waals bonds are broken. Hot water has the lowest densities, ultimately boiling to become steam.

The latter process, of breaking bonds and reducing density from *B* to *C*, is common to most materials. The former process, of melting the hydrogen bonds and increasing the density from *A* to *B*, is relatively unique to water. Water's high heat capacity can be explained as the thermal energy storage that comes from breaking two types of bonds—hydrogen bonds and van der Waals interactions—rather than just one type.

This model explains the effects of pressure. By Le Chatelier's principle, pressure forces systems into high-density states. So, squeezing on hot water, shifts to denser populations from *C* to *B*, increasing water's structure, density, and viscosity, similar to the effects seen in simpler liquids. Squeezing on cold water also increases water's density, but it shifts water populations from *A* to *B* in this case, disrupting water's cages and decreasing water's bonding, structure and viscosity.

Supercooled and Stretched Water

Water can be supercooled to -39.5°C at a pressure of one atmosphere, if ice nucleation and impurities are prevented. Water can also remain liquid if its pressure is lowered below its freezing point, at fixed temperature: this is called

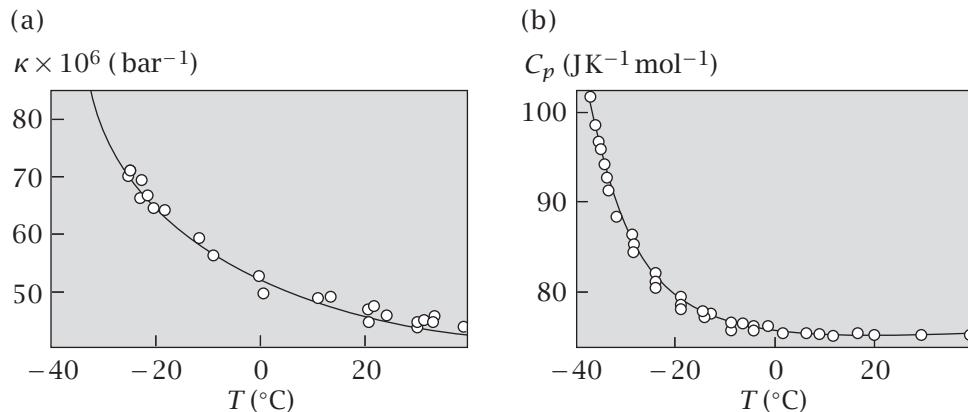


Figure 30.18 (a) The isothermal compressibility κ and (b) the heat capacity C_p of water diverge upon supercooling to -45°C . Source: PG Debenedetti, *Metastable Liquids, Concepts and Principles*, Princeton University Press, Princeton, 1966.

stretched water. Compared with other supercooled and stretched materials, water has two anomalous properties: the heat capacity and the isothermal compressibility grow large as the temperature approaches -45°C (see Figure 30.18). Such divergences often indicate a phase transition. The microscopic origins of this behavior are not yet understood [8].

Summary

Water molecules have tetrahedral symmetry and the ability to form strong orientation-dependent hydrogen bonds. The tetrahedral symmetry appears in the structure of ice and in the underlying fluctuating hydrogen-bonded network structure of liquid water. Hot water acts like a normal liquid. It expands when heated. However, cold liquid water has anomalous volumetric properties because of the competition between the hydrogen bonds, which favor open tetrahedral structures, and the van der Waals interactions, which favor denser disordered structures.

Problems

1. Does pressure-induced melting explain ice skating?

Suppose an ice skater weighs $m = 70\text{ kg}$. An ice skate exerts a pressure $\Delta p = mg/A$ on ice, where $A = 1.0\text{ cm}^2$ is the surface area of a skate and g is the gravitational acceleration constant.

- Compute the decrease in freezing temperature of water caused by this applied pressure.
- Is this pressure sufficient to melt ice?

2. Comparing water with other small molecules.

Table 14.1 shows the melting and boiling temperatures and enthalpies of several substances.

- Compute the melting and boiling entropies for these substances.
- Plot the enthalpies, entropies, and boiling and freezing temperatures versus molecular size, and explain the anomalies.

3. The residual entropy of ice.

The entropy of ice at $T = 0\text{ K}$ is $-3.43\text{ JK}^{-1}\text{ mol}^{-1}$. This can be explained in two steps. If there are N water molecules in an ice crystal, there are $2N$ hydrogens. Each hydrogen is situated between two oxygens (see Figure 30.2). Imagine that the oxygens define a three-dimensional grid and that the hydrogens are located between oxygens.

- How many configurations of the hydrogens would there be in the absence of any other constraints?
- How many configurations of the hydrogens would there be if every oxygen must have exactly two hydrogens assigned to it?

4. Global warming of the oceans.

The oceans on Earth weigh $1.37 \times 10^{21}\text{ kg}$ [9]. If global warming caused the average ocean temperature to increase by 1°C , what would be the enthalpy increase ΔH ?

5. The tip of the iceberg.

When ice floats on water, only a fraction h of the ice is above the water. Compute h .

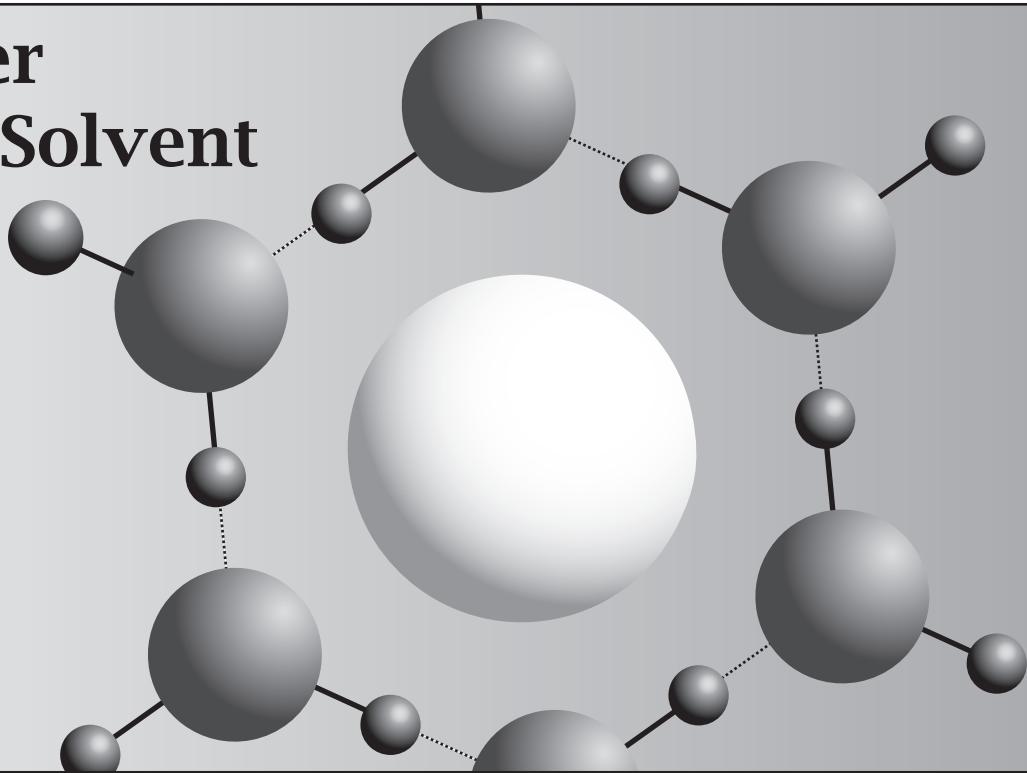
References

- [1] FH Stillinger, *Science* **209**, 451–457 (1980).
- [2] WK Röntgen, *Ann Phys Chem* **45**, 91 (1892).
- [3] G Nemethy and HA Scheraga, *J Chem Phys* **36**, 3382 (1962).
- [4] KAT Silverstein, ADJ Haymet, and KA Dill, *J Am Chem Soc* **122**, 8037–8041 (2000).
- [5] AR Henn and W Kauzmann, *J Phys Chem* **93**, 3770–3783 (1989).
- [6] LW Dahl and HC Andersen, *J Chem Phys* **78**, 1980–1993 (1983).
- [7] HE Stanley and J Teixeira, *J Chem Phys* **73**, 3404–3422 (1980).
- [8] PG Debenedetti, *J Phys Cond Matt* **15**, R1669–R1726 (2003).
- [9] W Stumm and JJ Morgan, *Aquatic Chemistry: Chemical Equilibria and Rates in Natural Waters*, 3rd edition, Wiley, New York, 1996.

Suggested Reading

- General references on the properties of water:
- M Chaplin, Water's hydrogen bond strength, arXiv: 0706.1355 [cond-mat.soft].
- PG Debenedetti, *Metastable Liquids: Concepts and Principles*, Princeton University Press, Princeton, 1996.
- PG Debenedetti, *J Phys Cond Matt* **15**, R1669–R1726 (2003).
- KA Dill, T Truskett, V Vlachy, and B Hribar-Lee, *Annu Rev Biophys Biomol Struct* **34**, 173–199 (2005).
- D Eisenberg and W Kauzmann, *The Structure and Properties of Water*, Oxford University Press, New York, 1969.
- F Franks, ed., *Water: A Comprehensive Treatise*, 7 volumes. Plenum Press, New York, 1972–1982.
- F Franks, *Water: A Matrix of Life*, 2nd edition, Royal Society of Chemistry, London, 2000.

31 Water as a Solvent



We now explore how water solvates nonpolar, polar, and ionic solutes. On the one hand, water is called a *universal solvent*. It can dissolve a wide range of polar and ionic molecules. On the other hand, water has little ability to dissolve inert gases or hydrocarbons. The low solubility of oil in water is called the *hydrophobic effect*. We begin with nonpolar solvation.

Oil and Water Don't Mix: The Hydrophobic Effect

Oil and water don't mix. But insolubility, *per se*, is not remarkable. Many compounds are insoluble in others. Benzene won't dissolve in water, but benzene also won't dissolve in formamide or glycerol. What is so special about the disaffinity between oil and water? Three properties set the mixing of oil and water apart from other solvation processes: (1) The free energy of mixing oil with water is typically much more positive than for other pairs of compounds. (2) The opposition to mixing is mainly entropic at 25°C. And, (3) the transfer of oil into water is accompanied by a large positive change in the heat capacity.

To see how water is special, first consider 'normal' solvation. Based on Figure 16.3, you expect the vapor pressure of a solute *B* dissolved in a solvent *A* to be (Equation (16.5))

$$\frac{p_B}{p_{B,\text{int}}^\circ} = \chi_B e^{\Delta\mu^\circ/RT}, \quad \text{where} \quad \Delta\mu^\circ = z \left(w_{AB} - \frac{w_{AA}}{2} \right), \quad (31.1)$$

where $\Delta\mu^\circ$ is the chemical potential per mole and the molar energies $w_{AA}/2$ for opening a cavity in the solvent and w_{AB} for the interaction of the solute with the cavity are both assumed to be independent of temperature. Likewise, we expect the liquid partition coefficient for the transfer of solute B from its pure liquid phase into a solvent A (see Figure 16.11) to be (see Equations (16.42) and (16.43))

$$K_B^A = e^{-\Delta\mu^\circ/RT}, \quad (31.2)$$

where

$$\Delta\mu^\circ = z \left(w_{SB} - w_{SA} + \frac{w_{AA}}{2} - \frac{w_{BB}}{2} \right), \quad (31.3)$$

where again the energy quantities w are assumed to be independent of temperature.

In Chapters 14–16, we developed a lattice model of simple solutions. We took the pair attraction energies w_{AA} , w_{BB} , and w_{AB} to be independent of temperature, so $\Delta\mu^\circ$ for transferring the solute into the solvent would also be independent of temperature. In that model, the disaffinity of A for B has no entropic component, and arises because the AA and BB attractive energies are stronger than the AB attractions. The next section describes how oil–water immiscibility differs from this.

The Signature of Hydrophobicity Is Its Temperature Dependence

The unusual aspect of mixing oil with water is its temperature dependence. One way to find the temperature dependence of $\Delta\mu^\circ$ is to measure the concentration c_{SW} of the solute in the water phase and the concentration c_{SO} of the solute in the oil phase, at equilibrium. The ratio is the partition coefficient $K(T) = c_{SW}/c_{SO}$, which can depend on temperature T . Such measurements give $\Delta\mu^\circ(T) = -RT \ln K(T)$. You can express $\Delta\mu^\circ(T)$ in terms of its enthalpic and entropic components (see Equation (9.32)):

$$\Delta\mu^\circ = \Delta h^\circ - T \Delta s^\circ, \quad (31.4)$$

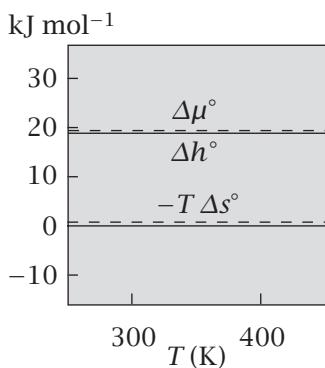
where Δh° is the molar enthalpy and Δs° is the molar entropy of solvation. Because Δh° and Δs° in Equation (31.4) could themselves depend on temperature, measuring $\Delta\mu^\circ(T)$ alone is not sufficient to tell you the temperature dependences of $\Delta h^\circ(T)$ or $\Delta s^\circ(T)$ individually. However, calorimetry can help. Calorimetry gives the molar heat capacity change upon transfer, Δc_p . According to Equations (8.25) and (8.26), Δc_p can be used to uniquely determine both the enthalpy and entropy:

$$\Delta h^\circ(T) = \Delta h^\circ(T_h) + \int_{T_h}^T \Delta c_p \, dT' = \Delta c_p (T - T_h) \quad (31.5)$$

and

$$\Delta s^\circ(T) = \Delta s^\circ(T_s) + \int_{T_s}^T \frac{\Delta c_p}{T'} \, dT' = \Delta c_p \ln \left(\frac{T}{T_s} \right). \quad (31.6)$$

(a) Simple Solution



(b) Aqueous Solution

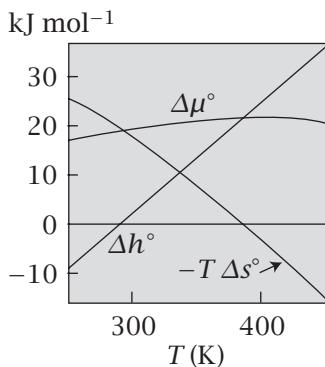


Figure 31.1 Comparison of the molar enthalpy Δh° , entropy Δs° , and free energy $\Delta\mu^\circ$ for transferring (a) a solute from its pure liquid into a simple solution and (b) benzene from its pure liquid into water. The entropy at high temperature is extrapolated assuming constant heat capacity. Sources: (a) KA Dill, *Biochemistry* **29**, 7133–7155 (1990); (b) PL Privalov and SJ Gill, *Adv Prot Chem* **39**, 191–234 (1988).

The last equality in Equations (31.5) and (31.6) holds if Δc_p is a constant, which is usually a good approximation for solvation. T_h is defined as the temperature at which $\Delta h^\circ = 0$. T_s is the temperature at which $\Delta s^\circ = 0$.

So, here's how you can interpret experiments. Using a calorimeter, measure Δc_p for solute transfer, say from its pure phase B into the solvent A . For *simple solutions*, you would find that Δc_p is small, so Δh° and Δs° are independent of temperature (see Figure 31.1(a)), and then you could readily get Δh° and Δs° from $-RT \ln K(T)$. However, experiments show that $\Delta c_p \gg 0$ for dissolving oils in water. In that case, substituting Equation (31.5) into Equation (31.4) gives

$$\Delta\mu^\circ(T) = \Delta c_p \left[(T - T_h) - T \ln\left(\frac{T}{T_s}\right) \right]. \quad (31.7)$$

For dissolving oils in water, you will find that $\Delta\mu^\circ(T)$ is a curved function of temperature T . To fit such data, you use three parameters: T_h , T_s , and Δc_p . First find the temperature T_h at which the enthalpy of solvation is zero. To do this, plot the function $\Delta\mu^\circ(T)/RT$. The temperature at which this function reaches a maximum is T_h . You can readily prove that that is the point at which $\Delta h^\circ = 0$, but we won't do so here. Then, find the temperature T_s at which the function $\Delta\mu^\circ(T)$ has a maximum or minimum; that is the temperature at which $\Delta s^\circ = 0$. Finally, given those values of T_h and T_s , find the value of Δc_p that causes Equation (31.7) to best fit your data.

Table 31.1 gives experimental values of $\Delta\mu^\circ$ and its components Δh° , $T \Delta s^\circ$, and Δc_p for transferring nonpolar solutes from their pure liquid phase into liquid water. The table shows the defining features of the hydrophobic effect: $\Delta\mu^\circ$ is positive and large, $\Delta c_p \gg 0$, $\Delta h^\circ \approx 0$, and $\Delta s^\circ \ll 0$ around room temperature. The aversion of oil for water is entropic at 25°C. This aversion becomes enthalpic at higher temperatures. The large heat capacity implies curvature in the free energy function (see Figure 31.1(b)).

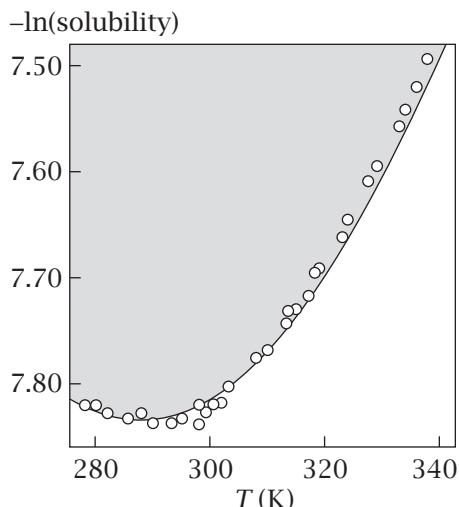
Figure 31.2 shows this curvature in the logarithm of the solubility ($= \Delta\mu^\circ / RT$) for benzene in water. There is a minimum in the solubility as a function of temperature, around 18°C. The same general behavior is observed with alkanes, alcohols, or inert gases dissolved in water. In the next section, we explore the microscopic basis for the hydrophobic effect.

Table 31.1 Thermodynamic properties for the transfer of small hydrocarbons from their pure liquid state to water. The temperature T_h (where solubility is minimum and $\Delta h^\circ = 0$) has an average value of $T_h = 22^\circ\text{C}$ among the six compounds shown. The temperature T_s (where $\Delta s^\circ = 0$) has an average value of $T_s = 113^\circ\text{C}$. $\Delta\mu^\circ$ is computed using Equation (31.7) with $T = 298\text{ K}$.

Compound	Solubility (mol frac $\times 10^{-4}$)	$\Delta\mu^\circ$ (kJ mol $^{-1}$)	Δh° (kJ mol $^{-1}$)	Δs° (J deg $^{-1}$ mol $^{-1}$)	ΔC_p° (J deg $^{-1}$ mol $^{-1}$)
Benzene	4.01	19.38	2.08	-58.06	225
Toluene	1.01	22.80	1.73	-70.7	263
Ethyl benzene	0.258	26.19	2.02	-81.0	318
Cyclohexane	0.117	28.12	-0.1	-94.8	360
Pentane	0.095	28.50	-2.0	-102.8	400
Hexane	0.020	32.53	0.0	-109.1	440

Source: RL Baldwin, *Proc Natl Acad Sci USA* **83**, 8069–8072 (1986).

Figure 31.2 Solubility of benzene in water versus temperature T . Benzene is least soluble around 18°C. Source: C Tanford, *The Hydrophobic Effect: Formation of Micelles and Biological Membranes*, 2nd edition, Wiley, New York, 1980. Data are from SJ Gill, NF Nichols, I Wadsö, *J Chem Thermodyn* **8**, 445–452 (1976).



The Hydrophobic Entropy Is Due to the Ordering of Waters Around the Solute

What is the basis for the excess entropy $\Delta s^\circ \ll 0$ that opposes the mixing of oil and water? It cannot be due to solute-solute interactions because the experiments are performed at sufficiently low solute concentrations.

Inserting a nonpolar solute molecule induces ordering of surrounding water molecules, mainly in the first solution shell. As evidence that the main effect is on the first-shell waters, the quantities $\Delta\mu^\circ$, Δh° , Δs° , and Δc_p are proportional to the solute surface area, which corresponds to the numbers of water molecules in the first solvation shell (see, e.g. [1]). First-shell water structuring has been studied extensively in *natural gas clathrate hydrates* [2]. A clathrate is a polyhedral cage of water molecules that contains a hydrocarbon (often a single molecule) such as CH₄. Much is known about clathrate structures because they are commercially important. When natural gas flows from offshore drilling rigs through pipelines, clathrate-encaged gas molecules clog the pipelines, reducing the flows.

The water structuring around nonpolar solutes is illustrated in Figure 31.3. Computer simulations show that at 25°C, the first-shell water molecules orient so that they can make hydrogen bonds with neighboring waters in the shell and avoid ‘wasting’ hydrogen bonds by pointing those hydrogen-bonding arms towards the solute.

Why is the disaffinity between oil and water so strong? Why is the hydrophobic entropy so large at $T = 300\text{ K}$? Here’s a qualitative argument. A water molecule that sits in a first solvation shell around a solute molecule has orientational restrictions. To maximize its hydrogen bonding, a first-shell water will point its hydrogen bonds towards other nearby waters and not towards the solute. Suppose that half of a water’s possible orientations would ‘waste’ a hydrogen bond pointed at the solute. Then, if a small solute molecule is surrounded by 15 first-shell water molecules (Table 31.2), the hydrophobic entropy due to the orientational restrictions of ordering those 15 waters would be

$$\Delta S = k \ln \left(\frac{W_{\text{shell waters}}}{W_{\text{bulk waters}}} \right) = k \ln (1/2)^{15} = -86 \text{ JK}^{-1}\text{mol}^{-1}, \quad (31.8)$$

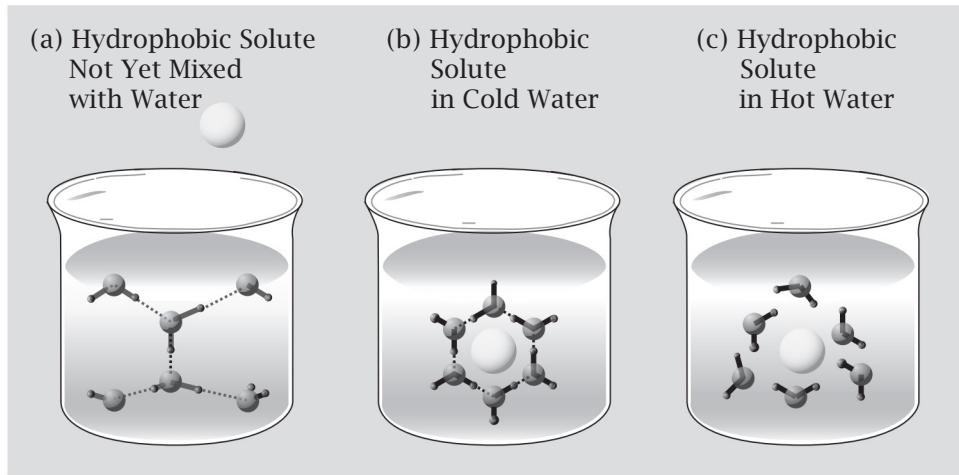


Figure 31.3 The ‘iceberg’ model of the heat capacity of (a) the transfer of nonpolar solutes into water [3, 4]. (This picture is only a two-dimensional cartoon, showing fewer first-shell waters than in real three-dimensional water.) (b) In cold water (near room temperature), the water molecules surrounding the nonpolar solute form good hydrogen bonds (low energy), in structured cages (low entropy) that avoid ‘wasting’ hydrogen bonds. (c) In hot water, more first-shell water conformations are accessible (higher entropy), but some of them have weaker or unformed hydrogen bonds and/or van der Waals interactions (higher energy). This contributes to the heat capacity because the system enthalpy increases with temperature.

corresponding to a free energy of about $6.1 \text{ kcal mol}^{-1}$ at $T = 300 \text{ K}$ opposing dissolution of oil in water.

Figure 31.4 shows why the restriction factor of $1/2$ per water molecule is a reasonable estimate [5, 6]. Figure 31.4(a) shows a water molecule at the center of a tetrahedron. The figure shows the six different ways the central water can point its two hydrogen-bonding arms toward the vertices of the tetrahedron, hydrogen bonding to neighboring waters in bulk water. Now, suppose the back gray dot water molecule in Figure 31.4(a) is replaced by a solute molecule. How many ways can the central water now orient if it must avoid pointing either of its hydrogen-bonding arms towards the solute vertex? Figure 31.4(b) shows that there are now only three ‘good’ hydrogen-bonding arrangements, out of the six possible arrangements, accounting for our estimate of $1/2$ above. The reason the hydrophobic solvation entropy (per solute molecule) is so large is because it is not the solute molecule itself that is being ordered; rather, the ordering involves the *many water molecules* in the first shell of the solute.

Why Is the Hydrophobic Enthalpy Small at 25°C ?

In simple systems, the mixing of molecules of A with molecules of B is opposed by an unfavorable enthalpy, $\Delta h^\circ > 0$. With oil and water, however, the enthalpies of mixing are not unfavorable, $\Delta h^\circ \approx 0$; see Table 31.1 [7, 8]. Modeling suggests that this arises from a combination of a traditional unfavorable contribution plus a favorable contribution (a negative component of Δh° of solvation), because inserting a solute helps open up water cages, strengthening water–water hydrogen bonding in the first shell. Small solutes help water

Table 31.2 Surface areas A of small nonpolar solutes, and numbers N_w of first-neighbor waters.

Molecule	$A (\text{\AA}^2)$	N_w
CH_4	152.36	17
C_2H_6	191.52	21
C_3H_8	223.35	25
N_2	142.41	16
CO	145.31	16
O_2	135.74	15
He	105.68	12
Ne	116.13	13
Ar	143.56	16
Kr	155.70	17
Xe	168.33	19

Source: SJ Gill, SF Dec, G Olofsson, and I Wadsö, *J Phys Chem* **89**, 3758–3761 (1985).

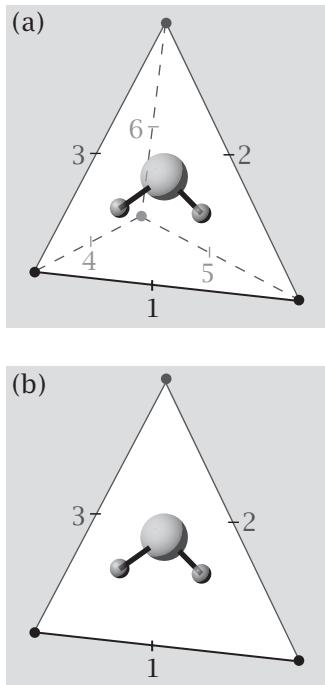


Figure 31.4 (a) A water molecule has six ways it can form two hydrogen bonds with its four tetrahedral neighboring waters. Neighboring water molecules are shown as vertices of a tetrahedron. The six edges of the tetrahedron indicate the pairs of acceptors for the two hydrogen-bond donors of the central water. (b) When a nonpolar solute replaces the water at the back vertex, the central water now has only three fully hydrogen-bonded configurations with neighboring water molecules. Source: G Ravishankar and DL Beveridge, *Theoretical Chemistry of Biological Systems*, G Náray-Szabó, ed. Elsevier, Amsterdam, 1986.

to favor the polygons and clathrate cages that are slightly larger and better hydrogen-bonded than the tetrahedra that water normally forms in the pure liquid.

What Is the Basis for the Large Hydrophobic Heat Capacity?

Inserting oil into water is distinguished by a large positive heat capacity of solvation, $\Delta c_p \gg 0$. This implies a mechanism of thermal energy storage (because $\Delta c_p = \partial h / \partial T$; see Chapter 12) beyond that of pure water alone. Figure 31.5 shows that the hydrophobic heat capacity change Δc_p grows with the solute size, in proportion to the number of water molecules in the first shell, implying that the first-shell waters play a key role in the energy storage mechanism.

Energy can be stored by bond breaking or bending. Figure 31.3 shows the ‘iceberg’ model of heat capacity, from Frank and Evans [4]. At low temperatures (25°C), first-shell water molecules are ordered (low entropy) and form good hydrogen bonds with other water molecules (low enthalpy). Heating melts these icebergs. Any melting process involves an increase of entropy and enthalpy. Warming up cold water increases the enthalpy by bending, breaking, or loosening water–water hydrogen bonds in the first solvation shell around the solute. Water also gains entropy from this bending and loosening. But the term ‘iceberg’ should not be taken to imply crystal-like ordering in cold liquid water. The degree of ordering of liquid water around nonpolar solutes is substantially less than the ordering in ice.

In summary, nonpolar solvation in water at 25°C is strongly opposed by the entropy, while the solvation enthalpy is small. However, the large Δc_p means that the mechanism changes at higher temperatures. Approaching the boiling temperature, 100°C, nonpolar solvation becomes strongly opposed by the enthalpy, while the solvation entropy is small. This is because high temperatures cause (1) first-shell water molecules to orient more randomly to gain entropy, but also (2) a corresponding breakage of hydrogen bonds, leading to a large positive solvation enthalpy (see Figure 31.3(c)).

Water Is Structured Near Cavities

Computer simulations by Postma et al. [9] show that water molecules around small empty cavities behave like water molecules around nonpolar solutes: the first-shell water molecules orient to avoid pointing their hydrogen-bonding groups toward the cavity, which would waste their hydrogen bonds. Also, ‘nature abhors a vacuum’: the energy of opening a cavity increases substantially with radius d (see Figure 31.6). So most cavities in water are relatively small. An exception is water near its critical point.

Hydrophobic Solvation Near the Critical Temperature of Water

Figure 31.7 shows the thermal fingerprints of hydrophobic interactions of neon dissolved in water near the critical temperature of water, 647.3 K (at $p = 218$ atm). The hydrophobic enthalpy and the partial molar volume of a nonpolar solute in water grow large and diverge as the critical temperature is approached. Why? Fluctuations become large near critical points (see Chapters 25 and 26). In some regions of space, water is clustered as in the liquid, while in other

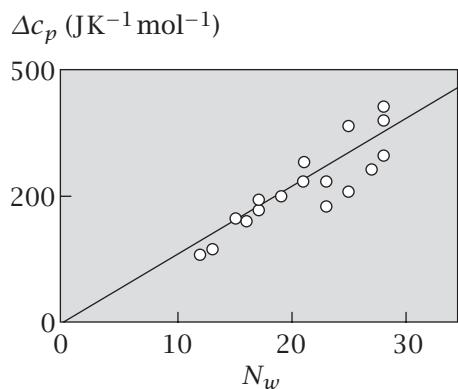


Figure 31.5 Experimental heat capacity of transfer Δc_p versus the number N_w of water molecules in the first solvation shell for hydrocarbon gases. Source: SJ Gill, SP Dec, G Olofsson, and I Wadsö, *J Phys Chem* **89**, 3758–3761 (1985).

regions, there are large holes, corresponding to many broken intermolecular interactions, a state of high enthalpy. Near the critical point, nonpolar solutes create large holes in the structure of water.

Near Large or Planar Surfaces, Water Loses Hydrogen Bonds

Water molecules adjacent to planar nonpolar surfaces behave differently from waters adjacent to small spherical solutes [10, 11]. Around a small solute at room temperature, water has an ‘entropy problem’ because it is restricted in its ability to form hydrogen bonds. Around a large solute (or next to a planar nonpolar surface), water has an ‘enthalpy problem’ because it is geometrically prevented from forming a maximum number of hydrogen bonds. A planar geometry forces each first-shell water molecule to waste one hydrogen bond, pointing it directly towards the surface. Sacrificing one hydrogen bond is the best the water can do near a plane. Because water has a high chemical potential near planar surfaces, it will evacuate the intervening space between two planar surfaces that are brought close enough together [12].

So far, we have considered the insertion of a single nonpolar molecule into water. This is sometimes called *hydrophobic hydration* or *hydrophobic solvation*. Now consider bringing two nonpolar solutes together in water. This is called *hydrophobic interaction*.

In water, two nonpolar solute molecules will be attracted to each other because pairing up reduces the surface area of nonpolar contact with water (see Figure 31.8). This reduces the unfavorable entropy at 25°C.

When some nonpolar solutes pair together in water, they come into direct contact, squeezing out intervening waters. When other nonpolar solutes associate in water, they do not come into direct contact. Rather, the two solute molecules are separated by an intervening water layer. Modeling by Pratt and Chandler [13], Zichi and Rossky [8, 14], Watanabe and Andersen [15], and Ravishanker et al. [6, 16] shows that this *solvent-separated state* is sometimes stable. For example, Figure 31.9 shows the krypton-krypton pair correlation function in water computed by Watanabe and Andersen [15]; the second small peak represents a population of solvent-separated pairs. This state is populated because of its compatibility with stable water cages around solutes. Solvent-separated states are more likely to occur when the solutes are small [18].

Solute molecules that are polar, such as alcohols, which are highly soluble, can alter water structure differently at different solute concentrations.

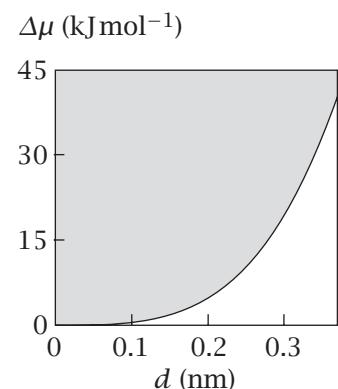


Figure 31.6 The free energy of creating a cavity of radius d in water. Source: G Hummer, S Garde, AE Garcia, et al., *Proc Natl Acad Sci USA* **93**, 8951–8955 (1996).

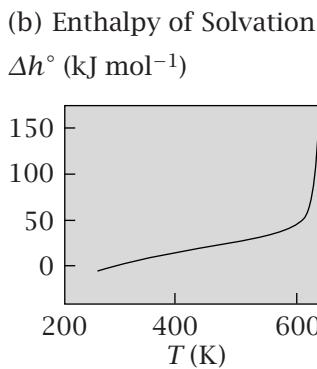
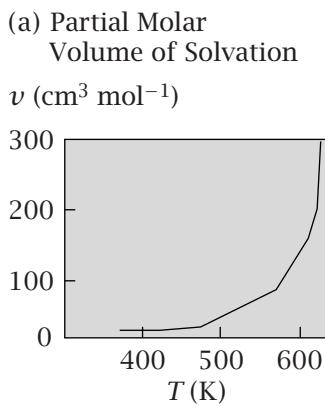
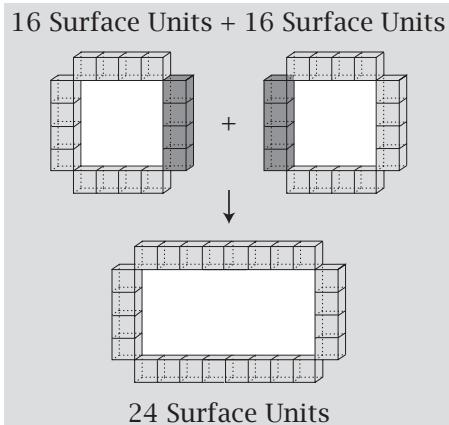


Figure 31.7 (a) The partial molar volume v and (b) the enthalpy Δh° of solvation of neon in water diverge as the temperature approaches the critical temperature of water. Source: R Crovetto, R Fernandez-Prini, and ML Japas, *J Chem Phys* **76**, 1077–1086 (1982).

Figure 31.8 Why do nonpolar solutes associate in water? Association reduces the total surface area of contact between solute molecules and the surrounding water. When the two solutes are separated, they have, in total, 32 units of surface contact with the solvent. But when the solutes contact each other, they have only 24 units of surface contact with water.



Alcohols Constrict the Volumes in Mixtures with Water

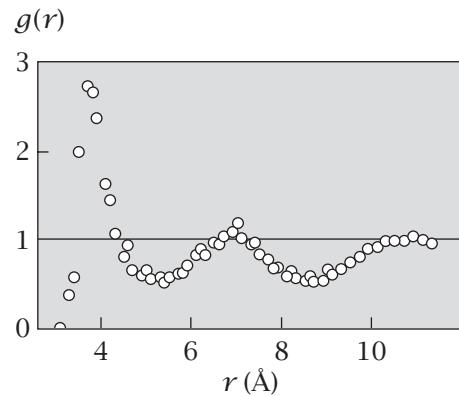
Solutes at high concentrations can disrupt the structure of water. Figure 31.10 shows the partial molar volumes of alcohols in mixtures of alcohol and water, as a function of alcohol concentration. Adding a small amount of alcohol to water leads to a constriction. The volume of the mixture is smaller than the sum of volumes of the pure alcohol and pure water components. The water molecules pack better around the alcohol molecules than around other water molecules. However, at higher concentrations, alcohol molecules pervade the system so extensively that they disrupt the fluctuating network of hydrogen-bonded water polygons.

Ions Can Make or Break Water Structure

Kosmotropic and Chaotropic Ions Cause Water Ordering or Disordering

Insert an ion, such as Na^+ or Cl^- , into water. In the first solvation shells around the ion, water will be structured. But the water structuring around ionic solutes is different than the structuring around nonpolar solutes.

Figure 31.9 Computed krypton–krypton pair correlation function in water. $g(r)$ describes the enhancement (or deficit) of krypton atoms at a given distance from another krypton, relative to the value at large separations, $g(r) = 1$ as $r \rightarrow \infty$. The second peak indicates that some solute pairs are separated by water. Source: K Watanabe and HC Andersen, *J Phys Chem* **90**, 795–802 (1986).



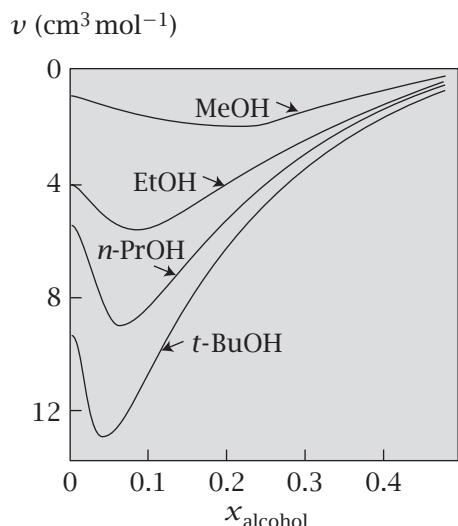


Figure 31.10 Partial molar volumes v of alcohol–water mixtures, versus alcohol concentration, for methanol (MeOH), ethanol (EtOH), *n*-propanol (*n*-PrOH), and *t*-butanol (*t*-BuOH).
Source: F Franks, *Water*, Royal Society of Chemistry, London, 1983.

Some ions, called *kosmotropes* (structure makers) increase the local structuring of water. Other ions, called *chaotropes* (structure breakers) decrease the local structuring of water (see Figure 31.13 below). Ion-induced water structuring has been explored in various experiments, such as how salts affect the viscosity of water and the entropies of solvation of salts in water.

Among the earliest studies were the experiments of Jones and Dole in 1929 on the viscosities of ions in dilute aqueous solutions [19]. Jones and Dole found that the viscosity η of an aqueous solution depends on the ion concentration c according to the relation

$$\eta = \eta_0(1 + A\sqrt{c} + Bc), \quad (31.9)$$

where η_0 is the viscosity of pure water and A and B are constants determined by the ionic solute. The term $A\sqrt{c}$ dominates over the Bc term at low concentrations (below about 0.01 M salt). It is positive for all ions, and originates in the same type of electrostatic interactions that give rise to the \sqrt{c} term in the Debye–Hückel theory of activity coefficients (see page 464). Positive and negative ions are attracted to one another, contributing extra cohesion to the liquid, making it more viscous.

The Jones–Dole B term becomes important at higher ion concentrations and defines so-called *specific ion* effects. The B term can be either negative or positive. It carries information about how much structure the ions impose on water. Ions with $B > 0$ are considered structure makers because increasing their concentrations increases the viscosity of an aqueous solution. Each flowing ion drags along some water. Ions with $B < 0$ are considered to be structure breakers because increasing their concentrations decreases the viscosity.

Table 31.3 lists some B coefficients. Comparison with Figure 31.11 shows that cations that have high charge density (either small size or high charge) will increase water structure ($B > 0$). Ions that have low charge density (either large size or small charge) will decrease the structure of water ($B < 0$). As temperature increases, B values grow more positive.

The same conclusions follow from experiments on the entropies of dissolving salts in water. Figure 31.12 shows that ions having high charge densities have negative entropies of solvation (water ordering), while large or weakly

Table 31.3 Jones–Dole B coefficients at $T = 25^\circ\text{C}$ (see Equation (31.9)).

	Cl^-	Br^-	I^-	NO_3^-
Li^+	0.139	0.106	0.081	0.101
Na^+	0.07866			
K^+	-0.0140	-0.0480		-0.0531
Rb^+	-0.037	-0.061	-0.11	
Cs^+	-0.050		-0.1184	-0.092

Source: RW Gurney, *Ionic Processes in Solution*, McGraw-Hill, New York, 1953. Data are from G Jones and SK Talley, *J Am Chem Soc* 55, 624–642 (1933); G Jones and RE Stauffer, *J Am Chem Soc* 62, 335–337 (1940); and VD Laurence and JH Wolfenden, *J Chem Soc* 1144–1147 (1934).

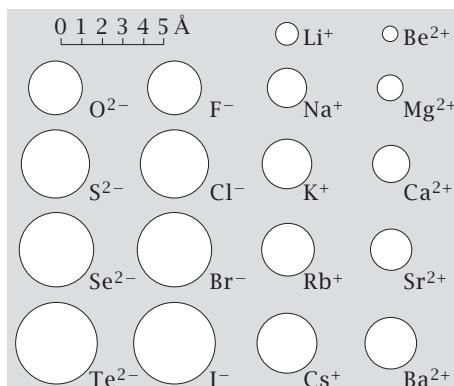


Figure 31.11 Relative radii of ions, from crystal structures. Source: *CRC Handbook of Chemistry and Physics*, 81st edition, DR Lide, ed., CRC Press, Boca Raton, FL, 2000.

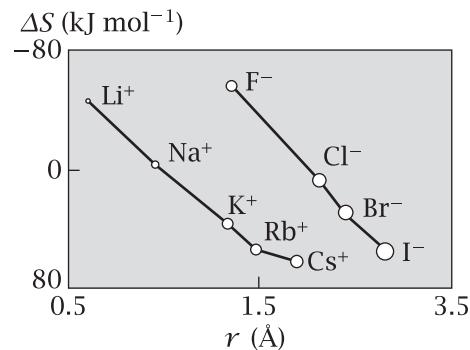


Figure 31.12 Entropies of ion solvation, versus ion radius. More negative values are upwards on the vertical axis. By this measure, small ions order water. Large ions disorder water. There is a charge asymmetry: it takes a larger anion to cause the same degree of ordering as a smaller cation.

charged ions have positive entropies of solvation (water disordering) [20]. (For free energies of solvation, see Table 31.4, and for enthalpies of solvation, see Table 22.3.) Whereas nonpolar solutes affect water structure mainly through water's hydrogen bonding, ionic solutes affect water structure mainly through their electrostatic interactions with water's dipole.

Ions Order Water Molecules by an Electrostatic Mechanism

Water has an electrostatic dipole. Its oxygen is partially negative and its hydrogens are partially positive. Molecular simulations indicate that water's dipole will point away from Na^+ or other cations (see Figure 31.13). Its dipole will

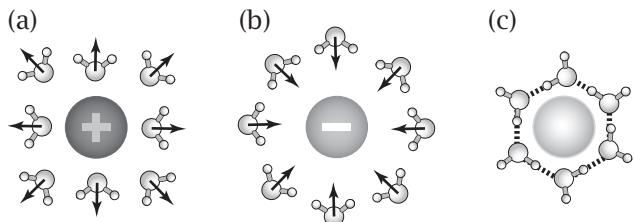


Figure 31.13 Different model mechanisms of water ordering: (a) water's dipole points away from a cation; (b) water's dipole points towards an anion; (c) water's hydrogen bonding is the dominant structuring mechanism of waters around nonpolar solutes.

point toward Cl^- or other anions. If an ion has a small radius, it will have strong electrostatic interactions with water's dipole. This is why smaller ions order waters more strongly than larger ions do. Water's dipole interacts more weakly with a large ion. For example, to a water molecule, a large ion, like Cs^+ , 'looks' much like a nonpolar solute because the electrostatic interactions are so weak. Around a Cs^+ ion, waters will form water-water hydrogen bonds, much like they do around nonpolar solutes.

So, different solutes affect water structuring by different mechanisms (see Figure 31.13): (a) water's dipole points away from small (or highly charged) cations; (b) water's dipole points toward small (or highly charged) anions; and (c) water forms hydrogen-bonding cages around nonpolar solutes or around ions that are large or weakly charged. The strong interactions of water with small ions are reflected in their very negative enthalpies of solvation (see Figure 22.17): water molecules are bonded more tightly to small ions such as Li^+ than to large ions such as K^+ . These strong ion-water attractions also cause *electrostriction*, the negative partial molar volume that results when ions are inserted into water. According to computer simulations [21], all three mechanisms of water structuring contribute to solvation entropies and heat capacities: nonpolar solutes increase the heat capacity of water, while ions can sometimes decrease the heat capacity.

Some salts are more soluble in water than others. Qualitatively, this can be explained by the same water structuring mechanisms as above.

Ion Pairing Preferences in Water Depend on Charge Densities

Compounds containing Na^+ , K^+ , NH_4^+ , Cl^- , and SO_4^{2-} tend to be soluble in water, while most phosphates PO_4^{3-} , sulfides S^{2-} , and carbonates CO_3^{2-} are insoluble. Table 31.5 shows the solubilities of various salts in water. Figure 31.14 shows an explanation, due to Collins [20]. In a salt having both a small cation and anion (like LiF), strong electrostatic interactions hold them together, so they are not very soluble in water. A salt having both a large cation and anion (like

Table 31.5 Molar solubilities of salts.

	F^-	Cl^-	Br^-	I^-
Li^+	0.1	19.6	20.4	8.8
Na^+	1.0	6.2	8.8	11.9
K^+	15.9	4.8	7.6	8.7
Rb^+	12.5	7.5	6.7	7.2
Cs^+	24.2	11.0	5.1	3.0

Source: KD Collins, *Biophys J* 72, 65–76 (1997).

Table 31.4 Free energies of ion solvation.

Ion	Free Energy of Solvation (kcal mol ⁻¹)
H^+	-265.9
Li^+	-128.4
Na^+	-103.2
K^+	-86.0
NH_4^+	-85.2
Rb^+	-80.6
Cs^+	-75.1
OH^-	-104.7
F^-	-104.4
HCO_3^-	-76.2
Cl^-	-74.5
Br^-	-68.3
I^-	-59.9

Source: CP Kelly, CJ Cramer, and DG Truhlar, *J Phys Chem B* 110, 16066–16081 (2006).

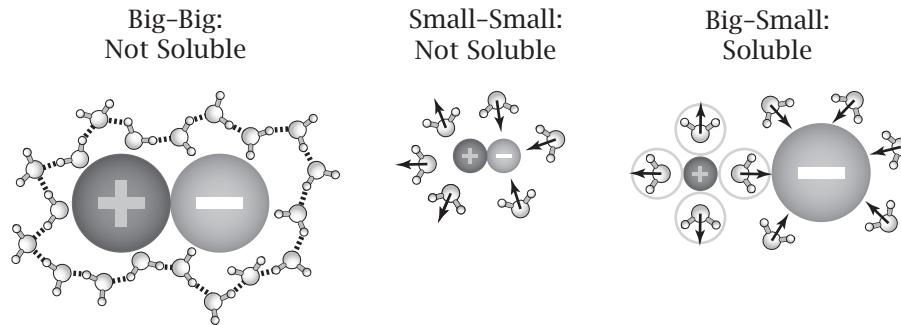


Figure 31.14 Salt solubilities in water. When the anion and cation are both big, they act like hydrophobes; such salts are relatively insoluble in water. When an anion and cation are both small, they have strong electrostatic attraction for each other; such salts are also insoluble in water. But when one ion is big and the other is small, the small ion attracts a shell of water, causing the insertion of water between the two ions, resulting in high solubilities of such salts in water.

CsI) is held together by hydrophobic-like attractions, surrounded by structured water, so it is also not very soluble in water. However, a salt having one large and one small ion (like CsF) is highly soluble because the small ion prefers to solvate a water dipole than to interact with the large ion.

When nonpolar solutes are dissolved in salty water, there are *Hofmeister* effects.

The Hofmeister Series: Nonpolar Solvation in Salt Solutions

Inserting oil into pure water is unfavorable. Adding salt to the water can either increase or decrease the free energy of inserting the oil. In 1888, F Hofmeister [22] discovered that different salts have different propensities to precipitate or dissolve proteins in solution. This series is of practical importance for ‘salting-in’ (dissolving) or ‘salting-out’ (precipitating) proteins. The Hofmeister effect is observed only if the salt is in sufficiently high concentration (typically greater than about 0.1 M). The *Hofmeister series* is a rank ordering of salt types that affect the partitioning of oils into aqueous salt solutions. For example, the solubilities of benzene and acetyl tetraglycyl ethyl ester increase in the following order: for anions, SO_4^{2-} , CH_3COO^- , Cl^- , Br^- , ClO_4^- , CNS^- ; and for cations, NH_4^+ , K^+ , Na^+ , Li^+ , Ca^{2+} . Salts can strengthen or weaken the hydrophobic interaction of oil with water. Salts that strengthen hydrophobic interactions cause proteins to associate with each other, leading to protein precipitation. However, Hofmeister effects are subtle and not yet fully understood [23].

Summary

The hydrophobic effect describes the unusual thermodynamics of the disaffinity of oil for water. In cold water, nonpolar solutes induce structuring in first-shell waters. Increasing the temperature melts out this structure. Ions, too, can cause the ordering of waters, by an electrostatic mechanism. The charge densities on ions have a role in determining their heats of dissolution and the viscosities of aqueous solutions, and can modulate the solubilities of oil in salt water.

Problems

1. Compute the extrema of $\Delta\mu^\circ(T)$ and $\Delta\mu^\circ(T)/RT$.

- Use the Gibbs-Helmholtz equations (13.40) and (13.41) to show that $\Delta s^\circ = 0$ at the temperature at which $\mu^\circ(T)$ is a maximum or minimum.
- Show that $\Delta h^\circ = 0$ at the temperature at which $\Delta\mu^\circ/RT$ is a maximum or minimum.

2. The heat capacity of protein unfolding. Figure 31.15 shows the enthalpy of folding a protein.

- Determine ΔC_p for folding from this graph.
- If $\Delta S_{\text{fold}} \approx 0$ at $T = 28^\circ\text{C}$, compute ΔS_{fold} at $T = 100^\circ\text{C}$.
- Compute ΔG_{fold} at $T = 25^\circ\text{C}$.

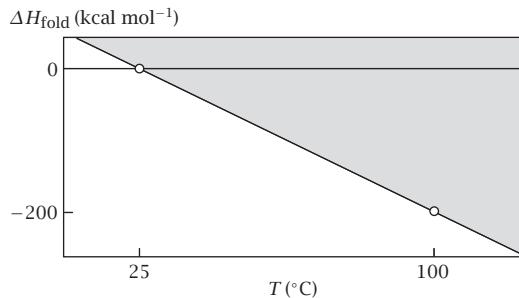


Figure 31.15

3. The forces of micelle formation. A micelle contains 80 sodium dodecyl sulfate molecules, each of which has a headgroup with one negative charge at the surface. The micellar radius is 15 Å.

- Compute the electrostatic energy required to charge up the head groups at the surface of the spherical micelle in water at $T = 300\text{K}$.
- The free energy of transfer of a CH_2 group from water to oil is $-0.8\text{ kcal mol}^{-1}$. Dodecyl chains have 12 CH_2 groups. Is hydrophobic association sufficient to overcome the electrostatic repulsions to form micelles?

4. ‘Icebergs’ are less ordered than ice. Compare the entropy per molecule of inserting a nonpolar solute into water with the entropy of freezing water.

5. The electrostriction of ions in water. Formic acid (HCOOH) at infinite dilution in water at $T = 298.15\text{ K}$ has partial molar volume $v = 34.69\text{ cm}^3\text{ mol}^{-1}$. The change in volume upon ionization is $\Delta v = -8.44\text{ cm}^3\text{ mol}^{-1}$. If this Δv is distributed uniformly among the first-shell water molecules, how much volume reduction can be attributed to each first-shell water molecule?

6. Ion solvation enthalpies scale inversely with ion radius. Ionic radii increase from Li^+ to Na^+ to K^+ , yet the enthalpies of solution decrease (see Figure 22.17). Write an electrostatic expression to rationalize this trend.

7. Estimating the cavity size distribution in water.

Suppose that the energy cost of creating a spherical cavity of radius r in water is

$$\varepsilon(r) = 4\pi r^2 \gamma,$$

where $\gamma = 7.2 \times 10^{-2}\text{ N m}^{-1}$ is the surface tension of water at $T = 300\text{K}$.

- Write an expression for the size distribution $p(r)$ of cavities in water.
- Compute the average radius $\langle r \rangle$ of a cavity at $T = 300\text{K}$.

References

- WL Jorgensen, J Gao, and C Ravimohan, *J Phys Chem* **89**, 3470–3473 (1985).
- AM Buswell and WH Rodebush, *Sci Am* **194**, 76–89 (1959).
- KA Dill, *Biochemistry* **29**, 7133–7155 (1990).
- HS Frank and MW Evans, *J Chem Phys* **13**, 507–532 (1945).
- SJ Gill, SF Dec, G Olofsson, and I Wadsö, *J Phys Chem* **89**, 3758–3761 (1985).
- G Ravishankar and DL Beveridge, *Theoretical Chemistry of Biological Systems*, G Náray-Szabó, ed., Elsevier, Amsterdam, 1986.
- A Geiger, A Rahman, and FH Stillinger, *J Chem Phys* **70**, 263–276 (1979).
- DA Zichi and PJ Rossky, *J Chem Phys*, **83**, 797–808 (1985).
- JPM Postma, HJC Berendsen, and JR Haak, *Faraday Symp Chem Soc* **17**, 55–67 (1982).
- CY Lee, JA McCammon, and PJ Rossky, *J Chem Phys* **80**, 4448–4455 (1984).
- NT Southall and KA Dill, *J Phys Chem B* **104**, 1326–1331 (2000).
- K Lum, D Chandler, and JD Weeks, *J Phys Chem B* **103**, 4570–4577 (1999).
- LR Pratt and D Chandler, *J Chem Phys* **67**, 3683–3704 (1977).
- DA Zichi and PJ Rossky, *J Chem Phys* **84**, 2814–2822 (1986).
- K Watanabe and HC Andersen, *J Phys Chem* **90**, 795–802 (1986).
- G Ravishankar, M Mezei, and DL Beveridge, *Faraday Symp Chem Soc* **17**, 79–91 (1982).
- WH New and BJ Berne, *J Am Chem Soc* **117**, 7172–7179 (1995).
- RH Wood and PT Thompson, *Proc Natl Acad Sci USA* **87**, 946–949 (1990).

- [19] RW Gurney, *Ionic Processes in Solution*, McGraw-Hill, New York, 1953.
- [20] KD Collins, *Biophys J* **72**, 65–76 (1997).
- [21] BM Pettitt and PJ Rossky, *J Chem Phys* **84**, 5836–5844 (1986).
- [22] PH von Hippel and T Schleich, *Acc Chem Res* **2**, 257–265 (1968).
- [23] RL Baldwin, *Biophys J* **71**, 2056–2063 (1996).

Suggested Reading

Solvation thermodynamics is treated extensively in:

- A Ben-Naim, *Hydrophobic Interactions*, Plenum Press, New York, 1980.
- A Ben-Naim, *Solvation Thermodynamics*, Plenum Press, New York, 1987.
- J Herzfeld and DJ Olbris, *Hydrophobic effect*, *Encyclopedia of Life Sciences*, Wiley, Chichester, 2002, <http://www.els.net/> [doi: 10.1038/npg.els.0002974].

PL Privalov and SJ Gill, *Adv Protein Chem* **39**, 191–234 (1988).

For excellent reviews of the Hofmeister effect, see:

- RL Baldwin, *Biophys J* **71**, 2056–2063 (1996).
- W Kunz, P Lo Nostro, and BW Ninham, *Curr Opin Coll Interf Sci* **9**, 1–18 (2004).

- MT Record, W Zhang, and CF Anderson, *Adv Protein Chem* **51**, 281–353 (1998).

Biological applications of hydrophobicity are discussed in:

- NT Southall, KA Dill, and ADJ Haymet, *J Phys Chem B* **106**, 521–533 (2002).

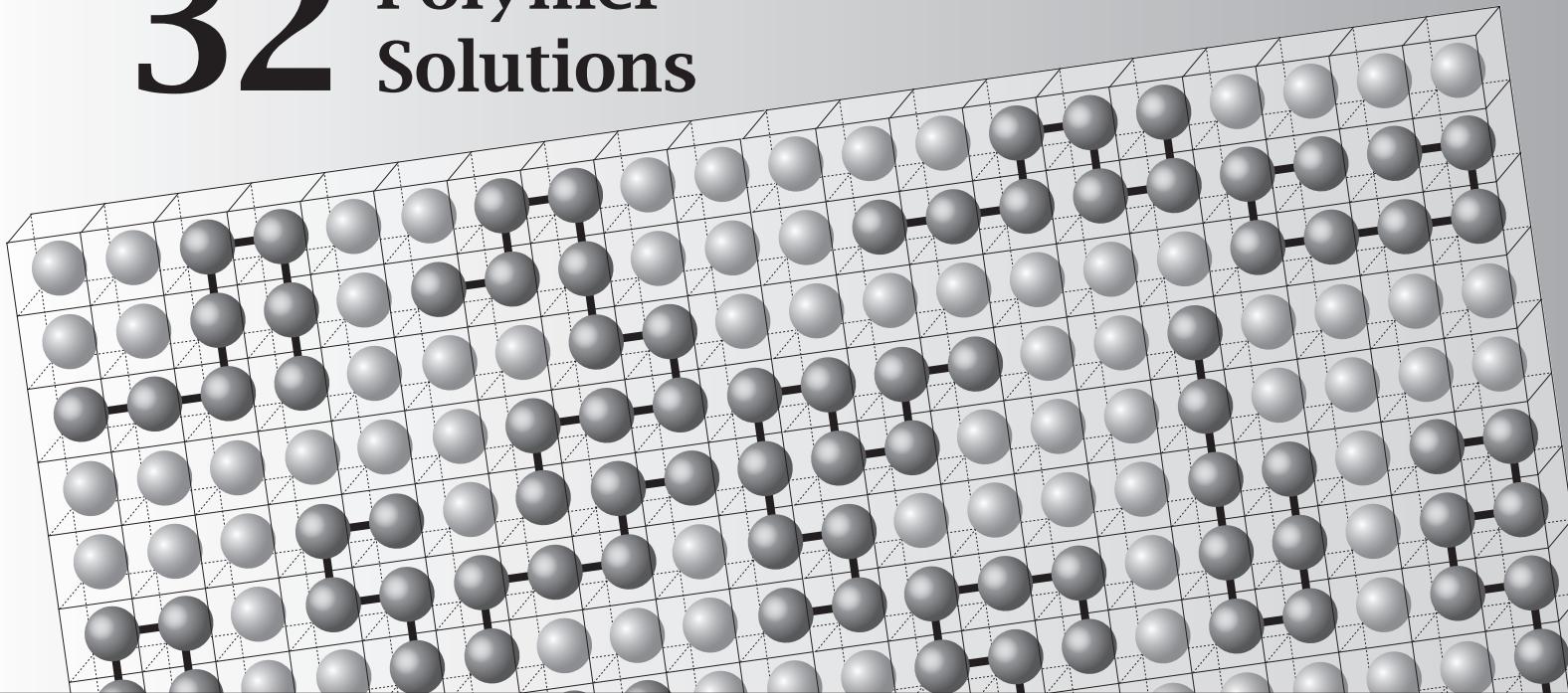
C Tanford, *The Hydrophobic Effect: Formation of Micelles and Biological Membranes*, 2nd edition, Wiley, New York, 1980.

Potentials of mean force are treated in statistical mechanics texts such as:

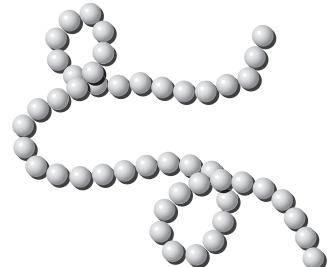
- HL Friedman, *A Course in Statistical Mechanics*, Prentice-Hall, Englewood Cliffs, NJ, 1985.

- DA McQuarrie, *Statistical Mechanics*, 2nd edition, University Science Books, Sausalito, CA, 2000.

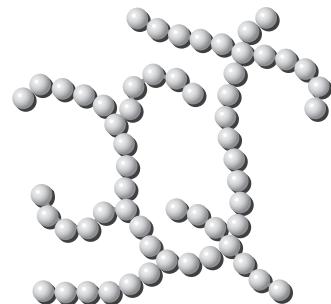
32 Polymer Solutions



(a) Linear Polymer



(b) Branched Polymer



Polymer Properties Are Governed by Distribution Functions

Polymer molecules are chains of covalently bonded monomer units. Some polymers are *linear*, like beads on a necklace; others are *branched* (see Figure 32.1). For predicting the properties of all types of polymers, statistical mechanics plays a key role because of two intrinsic polymer distribution functions. First, synthetic polymers have *distributions of chain lengths*. Second, polymer chains have *distributions of conformations*. Many of the unique properties of polymeric solutions and solids, such as the elasticity of rubber and the viscoelasticity of slimy liquids, are consequences of the entropies that arise from the conformational freedom of chain molecules. We begin with the properties of polymer mixtures.

Polymers Have Distributions of Conformations

Chain molecules have conformational degrees of freedom. Different conformations arise when bonds in the chains have rotational isomers of nearly equal energies. Multiple conformers of polymers are often populated at room temperature.

The energies of rotation around the C–C single bonds in butane and ethane are shown in Figure 32.2. At low temperatures, the molecules are found mainly

Figure 32.1 Polymers may be (a) linear or (b) branched chains of monomer units.

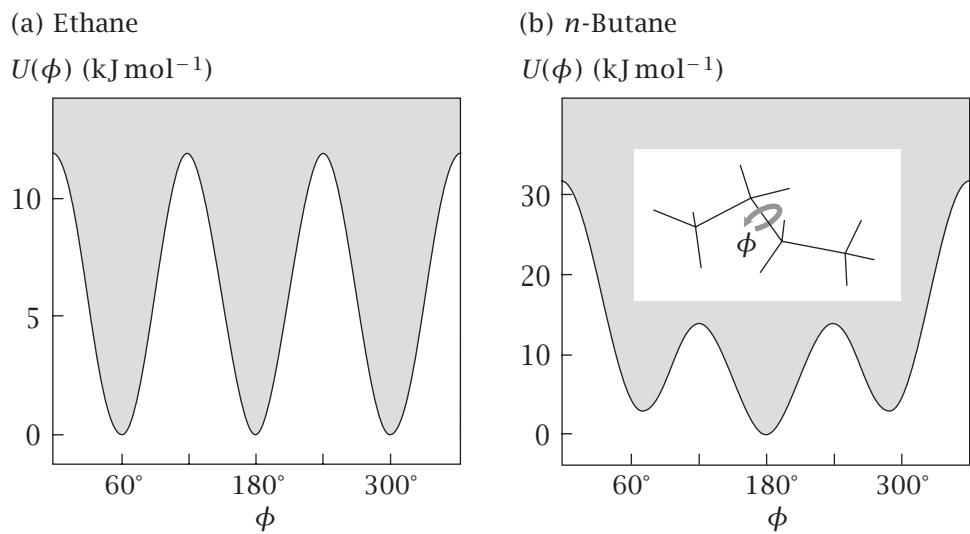


Figure 32.2 Bonds in polymers have torsional freedom. In saturated hydrocarbons, three torsional angles ϕ , called *trans*, *gauche⁺*, and *gauche⁻*, are favored by the minima in $U(\phi)$. (a) For ethane, the three *rotational isomers* have equal energy. (b) For butane, one minimum has lower energy than the other two. Source: WL Mattice and UW Suter, *Conformational Theory of Large Molecules: The Rotational Isomeric State Molecule in Macromolecular Systems*, Wiley, New York, 1994.

in the *trans* conformation, following the Boltzmann distribution. Molecular motions at low temperature are confined to vibrations within the lowest potential well. When the temperature is high enough, however, some molecules escape over the potential barrier (about 2.9 kcal mol⁻¹ in butane) to populate two states called *gauche⁺* and *gauche⁻*. Polyethylene, a chain of CH₂ groups, is one of the most common synthetic polymers. At high temperatures a polyethylene chain of N bonds has approximately 3^N conformations, a large number if N is large. Such configurational freedom underlies the properties of polymers.

Polymer Solutions Differ from Small-Molecule Solutions

Solutions of polymer molecules differ from solutions of small molecules in several ways. First, the colligative properties differ from those of the simple mixtures that we considered in Chapters 15 and 16 because polymer molecules are typically much larger than solvent molecules. Second, small-molecule solutions have a strong entropic tendency to mix, but polymer solutions do not. Polymers rarely mix with other polymers. This is unfortunate, because it would be useful to blend polymers in the same way that metals are alloyed, to gain the advantageous properties from each of the components. Third, polymeric liquids, solids, and solutions are rubbery, owing to chain conformational freedom.

In this chapter, we describe the *Flory-Huggins theory*, the simplest model that explains the first two of these thermodynamic properties of polymer solutions. We consider the elasticity and viscoelasticity of polymeric materials in Chapters 33 and 34.

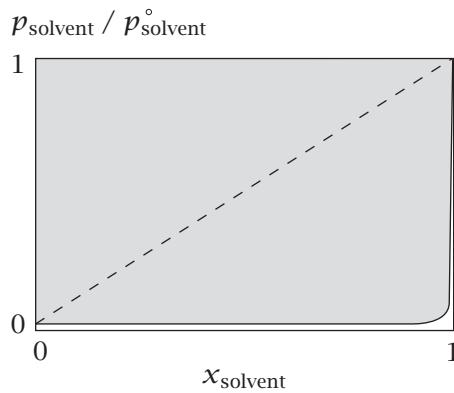
A Polymer Chain Is Typically Much Larger Than a Solvent Molecule

A polymer molecule can be composed of up to tens of thousands of monomers. Because each chain monomer is about the size of a solvent molecule, a polymer molecule occupies much more volume than a solvent molecule. That makes the volume fraction of a polymer in solution very different from its mole fraction. If a polymer chain has 1000 monomers, and if each monomer is the size of a solvent molecule, and if there is one polymer chain in the solution for every 1000 solvent molecules, then the volume fraction concentration of the polymer in the solution is 0.5, while the mole fraction is only 0.001 (see Problem 1 at the end of this chapter). The solvent volume fraction is 0.5 and its mole fraction is 0.999.

Colligative properties reflect this difference. Figure 32.3 shows the vapor pressure of the solvent benzene over a solution containing rubber, which is a polymer: (a) as a function of mole fraction and (b) as a function of volume fraction. The vapor pressure of a small-molecule solvent over a polymer solution shows nonideal behavior when plotted versus mole fraction x . The Flory-Huggins theory described in the next section shows that a better measure of concentration in polymer solutions is the volume fraction ϕ .

Figure 32.3(a) shows that vapor pressures over polymer solutions are different than vapor pressures over small-molecule solutions. First, the polymer molecules themselves have low vapor pressures because each chain is so large that it has many points of attraction to the neighboring solvent molecules. It cannot readily escape. However, the figure also shows that the solvent has a much smaller vapor pressure over a polymer solution than it would have over the corresponding small-molecule solution at the same mole fraction concentration, say $x = 0.5$. Why should the *solvent molecules* be so strongly attracted to the polymers in the solution? In earlier chapters, we considered solution nonidealities that are due to interaction energies, expressed

(a) Mole Fraction



(b) Volume Fraction

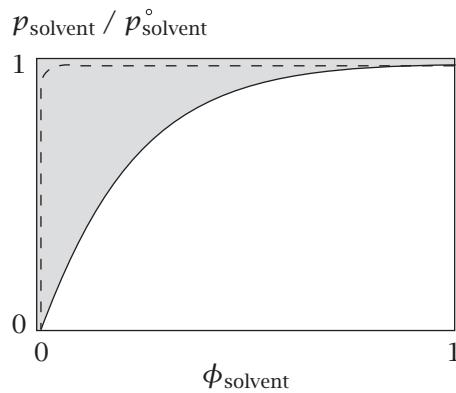


Figure 32.3 (a) The vapor pressure over an ideal solution (---) and the vapor pressure of a small-molecule solvent over a polymer solution (—) plotted as a function of mole fraction x . (b) The vapor pressure of a small-molecule solvent over a polymer solution (—) may have only a small deviation from the diagonal, if plotted versus *volume fraction* ϕ rather than against *mole fraction*. Source: PJ Flory, *Principles of Polymer Chemistry*, Cornell University Press, Ithaca NY, 1953.

as χ_{AB} . In this chapter, we explore another basis for solution nonidealities. Independent work by PJ Flory and ML Huggins in the early 1940s showed how the nonidealities of polymer solutions can arise from the very different sizes of a polymer chain and its small-molecule solvent. In short, to compute the entropies of polymer solutions, you don't treat each *polymer molecule* on an equal footing with each solvent molecule. Rather, you treat each *solvent-sized chain segment* of the polymer molecule as equivalent to a solvent molecule. Here's the idea in more detail.

The Flory–Huggins Model Describes Polymer Solution Thermodynamics

To compute the thermodynamics of polymer solutions, consider a lattice of M sites (see Figure 32.4). Each site has z nearest neighbors. Suppose that there are n_s solvent molecules, each of which occupies a single lattice site, and that there are n_p polymer molecules. Each polymer molecule has N *chain segments*, representing the monomers. Each chain segment occupies one lattice site. If the polymer and solvent molecules completely fill the lattice, then

$$M = Nn_p + n_s. \quad (32.1)$$

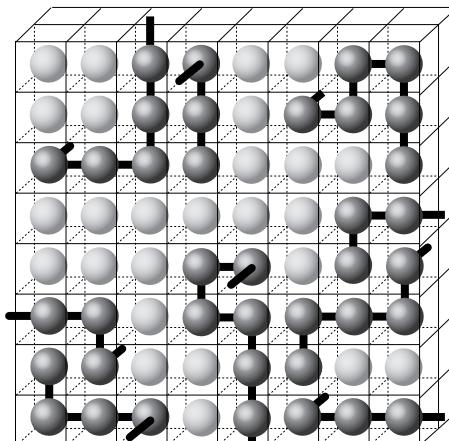
The mole fractions of polymer and solvent are $n_p/(n_p+n_s)$ and $n_s/(n_p+n_s)$, respectively. The volume fractions, ϕ_s for the solvent and ϕ_p for the polymer, are

$$\phi_s = \frac{n_s}{M} \quad \text{and} \quad \phi_p = \frac{Nn_p}{M}. \quad (32.2)$$

The Entropy of Mixing

To compute the entropy of mixing, you can count the number of arrangements of the n_p polymers and the n_s solvent molecules on the lattice. Let's first count the number of configurations of a single chain molecule by using a 'chain growth' method. The first monomer can be located on any of the M lattice sites. If you were counting the arrangements of the molecules of a gas or a simple

Figure 32.4 Polymer lattice model used for the Flory–Huggins theory. The dark beads show the polymer chain(s). The light beads represent the solvent molecules.



lattice solution, the second monomer could be located on any of the remaining $M - 1$ sites. But since the second monomer is in a chain and is connected to the first monomer, there are only z sites available to monomer 2 that are adjacent to monomer 1. Likewise, the third monomer must be connected to the second, so it has $(z - 1)$ options (since it cannot be in the same site as monomer 1). The fourth connects to the third, etc. The total number v_1 of conformations of one chain on the lattice is

$$v_1 = Mz(z-1)^{N-2} \approx M(z-1)^{N-1}, \quad (32.3)$$

which we have simplified by approximating the lone factor of z by $z - 1$.

Equation (32.3) is incomplete because it does not account for *excluded volume* among distant segments—the possibility that one monomer may land on a lattice site that has already been occupied by a more distant monomer. Equation (32.3) overestimates the number of conformations v_1 . To account for excluded volume, PJ Flory (1910–1985), who won the 1974 Nobel Prize in Chemistry for his contributions to polymer science, made the approximation that the volume excluded to each segment is proportional to the amount of space filled by the chain segments that have already been placed on the lattice, as if they were randomly dispersed in space. According to this approximation, $(M - 1)/M$ is the fraction of sites available for the second monomer, $(M - 2)/M$ is the fraction available for the third monomer, and so on. Therefore a better estimate of the number of conformations available for one chain is

$$\begin{aligned} v_1 &= M \left[(z) \left(\frac{M-1}{M} \right) \right] \left[(z-1) \left(\frac{M-2}{M} \right) \right] \dots \left[(z-N+1) \left(\frac{M-N+1}{M} \right) \right] \\ &\approx \left(\frac{z-1}{M} \right)^{N-1} \frac{M!}{(M-N)!}. \end{aligned} \quad (32.4)$$

Now use the same logic for putting all of the n_p chains onto the lattice. But, to simplify the derivation, instead of laying down the whole first chain, then the whole second chain, just lay down the first bead of chain 1, then the first bead of chain 2, ... until you have laid down each first bead of all n_p chains. Count the number of arrangements of those first beads and call it v_{first} . Then lay down bead 2 of chain 1, bead 3 of chain 1, ..., bead N of chain 1, then bead 2 of chain 2, bead 3 of chain 2, etc. Count all these arrangements of all the remaining chain segments and call it $v_{\text{subsequent}}$. The number of arrangements v_{first} for placing the first monomer segments for all of the n_p chains is

$$v_{\text{first}} = M(M-1)(M-2) \dots (M-n_p+1) = \frac{M!}{(M-n_p)!}. \quad (32.5)$$

Next, we count the number of arrangements of the $n_p(N-1)$ remaining segments of the n_p chains. So far, n_p monomers (the first segment of each chain) have been placed on the lattice. This means that $M - n_p$ sites are available for the first of the subsequent segments. And $M - Nn_p - 1$ sites are available for the last segment. So, following the logic of Equation (32.4), the number of arrangements of all the subsequent segments, $v_{\text{subsequent}}$, is

$$v_{\text{subsequent}} = \left(\frac{z-1}{M} \right)^{n_p(N-1)} \frac{(M-n_p)!}{(M-Nn_p)!}. \quad (32.6)$$

To compute the total number of arrangements $W(n_p, n_s)$ of the system of all n_p polymer molecules and n_s solvent molecules, multiply the two factors together from Equations (32.5) and (32.6) to get

$$\begin{aligned} W(n_p, n_s) &= \frac{V_{\text{first}} V_{\text{subsequent}}}{n_p!} \\ &= \left(\frac{z-1}{M}\right)^{n_p(N-1)} \frac{M!}{(M-Nn_p)!n_p!}. \end{aligned} \quad (32.7)$$

The factor $n_p!$ in the denominator on the right-hand side of Equation (32.7) accounts for the indistinguishability of one polymer chain from another. The factor $(M - Nn_p)!$ accounts for the indistinguishability of the solvent molecules.

To mix the polymer with the solvent, start with a container of pure solvent (n_s solvent molecules and no polymer) and a container of pure polymer (n_p polymer molecules and no solvent), and combine them to get a solution of n_s solvent molecules and n_p polymer molecules (see Figure 32.5). The Boltzmann expression gives the entropy change upon mixing in terms of the multiplicities W :

$$\frac{\Delta S_{\text{mix}}}{k} = \ln \left[\frac{W(n_p, n_s)}{W(0, n_s)W(n_p, 0)} \right]. \quad (32.8)$$

For a lattice of pure solvent molecules, $W(0, n_s) = 1$. You can compute $W(n_p, 0)$ for the pure polymer by substituting Nn_p for M into Equation (32.7) to get

$$W(n_p, 0) = \left(\frac{z-1}{Nn_p}\right)^{n_p(N-1)} \frac{(Nn_p)!}{n_p!}. \quad (32.9)$$

Use Equation (32.9), substitute $n_s = M - Nn_p$ into Equation (32.7) for the mixture $W(n_p, n_s)$, and use Stirling's approximation (Equation (B.3) in Appendix B) to get

$$\begin{aligned} \frac{W(n_p, n_s)}{W(0, n_s)W(n_p, 0)} &= \left(\frac{M!}{n_p!n_s!}\right) \left(\frac{z-1}{M}\right)^{n_p(N-1)} \left(\frac{n_p!}{(Nn_p)!}\right) \left(\frac{Nn_p}{z-1}\right)^{n_p(N-1)} \\ &= \left[\frac{M!}{(Nn_p)!n_s!}\right] \left(\frac{Nn_p}{M}\right)^{n_p(N-1)} \end{aligned}$$

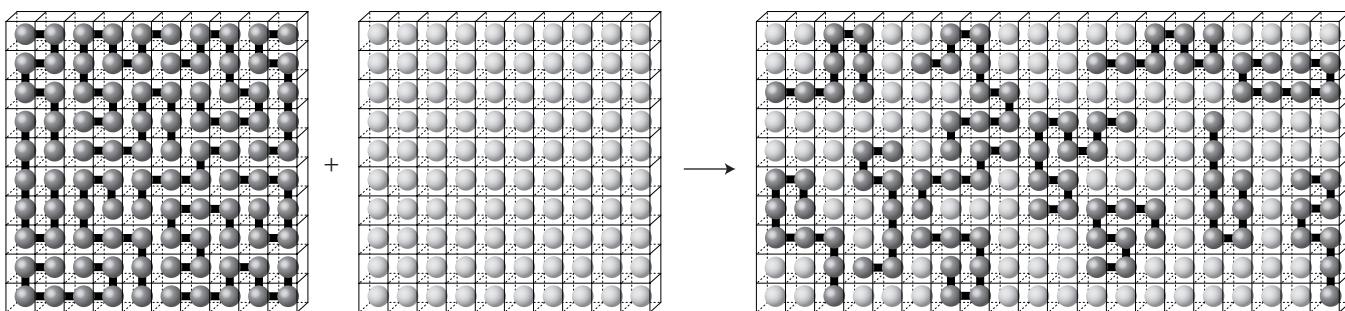


Figure 32.5 A lattice model for mixing n_p polymer molecules, and n_s solvent molecules, to get a solution of n_p polymer molecules and n_s solvent molecules.

$$\begin{aligned}
&= \left[\frac{M^M}{(Nn_p)^{Nn_p} n_s^{n_s}} \right] \left(\frac{Nn_p}{M} \right)^{n_p N} \left(\frac{M}{Nn_p} \right)^{n_p} \\
&= \left(\frac{M}{n_s} \right)^{n_s} \left(\frac{M}{Nn_p} \right)^{n_p}.
\end{aligned} \tag{32.10}$$

Substitute Equation (32.10) into Equation (32.8) to get the mixing entropy

$$\frac{\Delta S_{\text{mix}}}{k} = -n_s \ln \left(\frac{n_s}{M} \right) - n_p \ln \left(\frac{Nn_p}{M} \right). \tag{32.11}$$

Use the definitions of volume fractions, Equations (32.2), to get the entropy per lattice site:

$$\frac{\Delta S_{\text{mix}}}{Mk} = -\phi_s \ln \phi_s - \left(\frac{\phi_p}{N} \right) \ln \phi_p. \tag{32.12}$$

When $N = 1$, the Flory-Huggins mixing entropy reduces to that of the simple-solution lattice model, Equation (15.4).

Now we compute the energy of mixing.

The Energy of Mixing

Now follow the same procedure that we used for simple lattice solutions (see Equation (15.5)). If m_{ss} , m_{pp} , and m_{sp} are the numbers of contacts between solvent molecules (s) and monomeric segments (p) of the polymer chain, and if w_{ss} , w_{pp} , and w_{sp} are the corresponding contact energies for each type of pair, then the total contact energy is

$$U = m_{ss}w_{ss} + m_{pp}w_{pp} + m_{sp}w_{sp}. \tag{32.13}$$

The conservation relations between the numbers of contacts and the numbers of lattice sites are

$$zn_s = 2m_{ss} + m_{sp}$$

and

$$zNn_p = 2m_{pp} + m_{sp}. \tag{32.14}$$

We have made a simplification here. The quantity zNn_p in the second of these equations could better be approximated as $(z-2)Nn_p$. The factor of $z-2$ would be more appropriate than z because each p monomer is not fully accessible for making noncovalent contacts, since it is covalently connected to two other p monomers in the chain (or one if it is at the end of the chain). But at the level of simplicity of the model, this refinement is neglected. Substitute Equations (32.14) into Equation (32.13) to replace the unknown m 's by the known n 's. Now use the Bragg-Williams approximation (Equation (15.11)), which estimates the number of sp contacts as the product (n_s , the number of s monomers) \times (the number z of neighboring sites per s monomer) \times (the fraction Nn_p/M of such lattice sites that are occupied by p monomers). That

is, $m_{sp} \approx (zn_s N n_p)/M$, so

$$\frac{U}{kT} = \left(\frac{zw_{ss}}{2kT} \right) n_s + \left(\frac{zw_{pp}}{2kT} \right) N n_p + \chi_{sp} \frac{n_s n_p N}{M}, \quad (32.15)$$

where

$$\chi_{sp} = \frac{z}{kT} \left(w_{sp} - \frac{w_{ss} + w_{pp}}{2} \right) \quad (32.16)$$

is the dimensionless energy quantity introduced in Chapters 15 and 16 that describes the preference for polymer-solvent interactions relative to polymer-polymer and solvent-solvent interactions.

The Free Energy and Chemical Potential

To get the free energy, combine the entropy equation (32.11) with the energy equation (32.15):

$$\begin{aligned} \frac{F}{kT} &= \frac{U}{kT} - \frac{S}{k} = n_s \ln \phi_s + n_p \ln \phi_p + \left(\frac{zw_{ss}}{2kT} \right) n_s \\ &\quad + \left(\frac{zw_{pp}}{2kT} \right) N n_p + \chi_{sp} \frac{n_s n_p N}{M}. \end{aligned} \quad (32.17)$$

To generalize, suppose that you have a mixture of n_A molecules of a polymer A containing N_A monomer units, and n_B molecules of a polymer B having N_B monomers per chain. Then Equation (32.17) gives

$$\begin{aligned} \frac{F_{\text{mix}}}{kT} &= n_A \ln \phi_A + n_B \ln \phi_B + \left(\frac{zw_{AA}}{2kT} \right) N_A n_A \\ &\quad + \left(\frac{zw_{BB}}{2kT} \right) N_B n_B + \chi_{AB} \frac{n_A n_B N_A N_B}{M}. \end{aligned} \quad (32.18)$$

The volume fractions are $\phi_A = (n_A N_A)/M$ and $\phi_B = (n_B N_B)/M$, where $M = n_A N_A + n_B N_B$. As a check, notice that if $N_B = 1$, Equation (32.18) reduces to Equation (32.17).

Now we compute the chemical potential of the molecules of B (which can be either a polymer or small molecule, depending on N_B). Hold n_A (not M) constant and take the derivative of Equation (32.18) (see Equations (15.17) and (15.18)):

$$\begin{aligned} \frac{\mu_B}{kT} &= \left(\frac{\partial}{\partial n_B} \left(\frac{F_{\text{mix}}}{kT} \right) \right)_{n_A, T} \\ &= \ln \phi_B + 1 - \frac{n_B N_B}{M} - \frac{n_A N_B}{M} + \chi_{AB} N_B (1 - \phi_B)^2 + \left(\frac{zw_{BB}}{2kT} \right) N_B. \end{aligned} \quad (32.19)$$

Use the expression

$$\frac{n_A N_B}{M} = \frac{n_A N_A}{M} \left(\frac{N_B}{N_A} \right) = \phi_A \left(\frac{N_B}{N_A} \right) = (1 - \phi_B) \left(\frac{N_B}{N_A} \right)$$

and the definition $\phi_B = n_B N_B/M$ to reduce Equation (32.19) to

$$\frac{\mu_B}{kT} = \ln \phi_B + (1 - \phi_B) \left(1 - \frac{N_B}{N_A} \right) + \chi_{AB} N_B (1 - \phi_B)^2 + \left(\frac{zw_{BB}}{2kT} \right) N_B. \quad (32.20)$$

EXAMPLE 32.1 Mixing a polymer with a solvent. Mix a small-molecule solvent B , $N_B = 1$, with a polymer A of chain length $N_A = N$ to reach a polymer concentration $\phi_p = \phi_A$. The change in chemical potential upon mixing is $\Delta\mu_{\text{mix}} = \mu_B(\phi_B) - \mu_B(\phi_B = 1)$. So, substitute $\phi_B = 1 - \phi_p$ into Equation (32.20), then substitute $\phi_B = 1$ into Equation (32.20), and subtract, to get

$$\frac{\Delta\mu_{\text{mix}}(\phi_p)}{kT} = \ln(1 - \phi_p) + \phi_p \left(1 - \frac{1}{N}\right) + \chi_{AB} \phi_p^2. \quad (32.21)$$

We will find Equation (32.21) useful in the following chapters for computing the osmotic pressures of polymer solutions.

Now that you have the chemical potential, you can follow the same procedures you use for the lattice model of simple solutions to predict the colligative properties and phase separations of polymer solutions.

Flory–Huggins Theory Predicts Nonideal Colligative Properties for Polymer Solutions

Let's compute the vapor pressure of a solvent over a polymer solution by using the Flory–Huggins theory. Let component A represent the polymer and B represent the small molecule. To describe the equilibrium, follow the strategy of Equation (16.2): use Equation (32.19), and set the chemical potential of B in the vapor phase equal to the chemical potential of B in the polymer solution. The vapor pressure of B over a polymer solution is

$$\frac{p_B}{p_{B,\text{ref}}} = \phi_B \exp \left[(1 - \phi_B) \left(1 - \frac{N_B}{N_A}\right) + N_B \chi_{AB} (1 - \phi_B)^2 + \left(\frac{z w_{BB}}{2kT}\right) N_B \right]. \quad (32.22)$$

Equation (32.22) resembles Equation (16.2) for small-molecule solutions, except that the volume fraction ϕ_B replaces the mole fraction x_B , and a term $(1 - \phi_B)(1 - N_B/N_A)$ accounts for nonideality due to the difference in molecular sizes. The Flory–Huggins theory predicts that it is the number of solvent-sized monomer units in polymer chains that determine the colligative properties, not the number of polymer chains. This is the reason that volume fractions rather than mole fractions are more useful for polymer solutions.

The Phase Behavior of Polymers Differs from that of Small Molecules

For processing polymers and creating plastic materials, it is crucial to know when polymers will dissolve in solvents and when they will not. Also, some common materials are polymer blends, which are mixtures of different types of polymers. We want to know when one type of polymer is miscible in another. To explore the phase behavior of polymer solutions, we want the free energy as a function of volume fraction ϕ . Compute ΔF_{mix} as the difference in free energy resulting from mixing the two pure phases together. Using Equation (32.18) divided by M and taking this difference gives

$$\frac{\Delta F_{\text{mix}}}{MkT} = \left(\frac{\phi_A}{N_A}\right) \ln \phi_A + \left(\frac{\phi_B}{N_B}\right) \ln \phi_B + \chi_{AB} \phi_A \phi_B,$$

or, in terms of the volume fraction of A, $\phi = \phi_A$,

$$\frac{\Delta F_{\text{mix}}}{MkT} = \left(\frac{\phi}{N_A} \right) \ln \phi + \left(\frac{1-\phi}{N_B} \right) \ln(1-\phi) + \chi_{AB} \phi(1-\phi). \quad (32.23)$$

When the two types of molecules in a solution have different sizes, the mixing free energy will be an asymmetrical function of ϕ as in Figure 32.6. Systems are predicted to be miscible for compositions in which the free energy is concave upward and immiscible where that function is concave downward (see Figure 25.8). In contrast, Figures 25.10 and 25.14 show that lattice-model phase diagrams for small-molecule solutions are symmetrical around $\phi_A = \phi_B = 1/2$.

We now show the basis for the asymmetry in polymer solutions. The coexistence curve is defined by the common tangent (see Equation (25.7)),

$$\left(\frac{\partial \Delta F_{\text{mix}}}{\partial \phi} \right)' = \left(\frac{\partial \Delta F_{\text{mix}}}{\partial \phi} \right)'', \quad (32.24)$$

where the ' and '' indicate the two phases that are in equilibrium. In general, calculation of the coexistence curve involves the solution of two simultaneous nonlinear equations. You can compute the coexistence curve by substituting Equation (32.17) or (32.18) into Equation (32.24) and finding the points of common tangency. We won't do this here. Our goal here is just to show the asymmetry in polymer phase diagrams and polymer-polymer immiscibility. You can do this by focusing on the critical point.

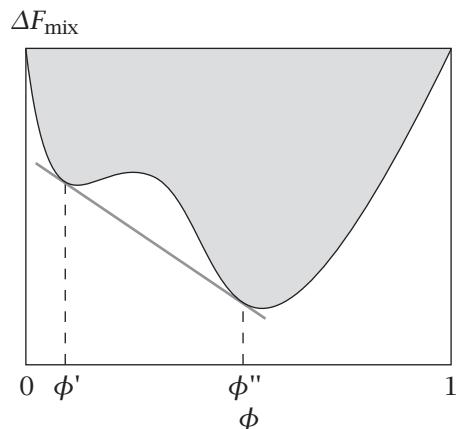
The spinodal decomposition curve is given by the values of ϕ that cause the second derivative of Equation (32.20) to equal zero:

$$\frac{\partial^2}{\partial \phi^2} \left(\frac{\Delta F_{\text{mix}}}{MkT} \right) = \frac{1}{N_A \phi} + \frac{1}{N_B (1-\phi)} - 2\chi_{AB} = 0. \quad (32.25)$$

The critical point ϕ_c is where the spinodal curve and its derivative are zero (see Chapter 25, page 504):

$$\frac{\partial^2}{\partial \phi^2} \left(\frac{\Delta F_{\text{mix}}}{MkT} \right) = \frac{\partial^3}{\partial \phi^3} \left(\frac{\Delta F_{\text{mix}}}{MkT} \right) = 0. \quad (32.26)$$

Figure 32.6 The free energy of mixing a polymer solution versus volume fraction ϕ . Find the phase equilibria by using the common tangent method (Figure 25.9). The two phases in equilibrium have volume fractions ϕ' and ϕ'' .



Taking the derivative of Equation (32.25) gives

$$\begin{aligned} \frac{\partial^3}{\partial \phi^3} \left(\frac{\Delta F_{\text{mix}}}{MkT} \right) &= -\frac{1}{N_A \phi_c^2} + \frac{1}{N_B (1-\phi_c)^2} = 0 \\ \Rightarrow \frac{\phi_c}{1-\phi_c} &= \left(\frac{N_B}{N_A} \right)^{1/2} \\ \Rightarrow \phi_c &= \frac{N_B^{1/2}}{N_A^{1/2} + N_B^{1/2}}. \end{aligned} \quad (32.27)$$

For small-molecule mixtures, the critical point depends on the energy of interaction χ_{AB} . Here, you see that the polymer-polymer critical point also depends on the chain lengths of the two polymers. Now let's use this expression to interpret phase equilibria in polymer solutions.

Polymer Phase Diagrams Are Asymmetrical Owing to Size Differences between Polymer and Solvent Molecules

Two cases are of special interest. First, suppose that component *A* is a long-chain polymer and *B* is a small molecule, $N = N_A \gg 1$ and $N_B = 1$. Then Equation (32.27) gives the critical volume fraction of the polymer (*A*) as

$$\phi_c = \frac{1}{1+N^{1/2}}. \quad (32.28)$$

Substituting Equation (32.28) into Equation (32.25) gives the value of the interaction parameter χ_c at the critical temperature:

$$\chi_c = \frac{1+N^{1/2}}{2N^{1/2}}. \quad (32.29)$$

As $N \rightarrow \infty$, $\phi_c \rightarrow 0$, which means that the peak of the mixing phase diagram shifts strongly to the left in Figure 32.7 as the chain length increases. The phase diagram becomes highly asymmetrical because of the difference in size between the polymer and the solvent molecules. Figure 32.8 shows the asymmetry of the liquid-liquid phase diagram of polystyrene in cyclohexane. In the next section, we treat a second case of interest, in which both the solute and the solvent are polymers.

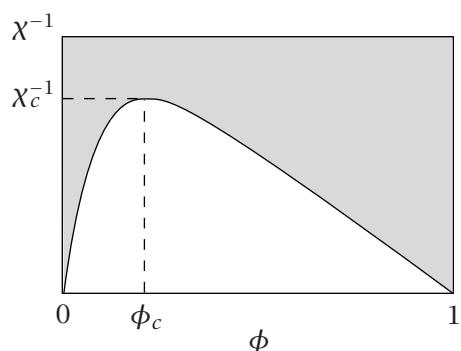
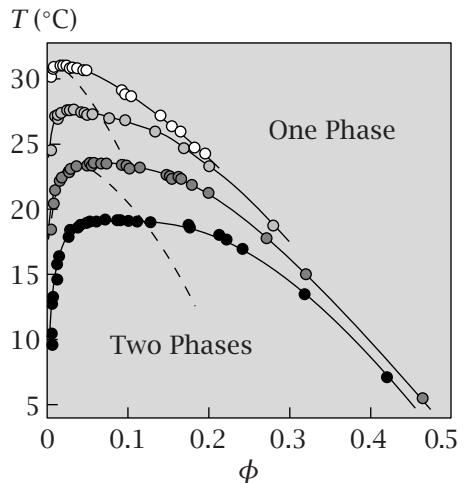


Figure 32.7 A polymer phase diagram. χ_c^{-1} on the vertical axis is proportional to temperature. Mixtures of polymers and small molecules have asymmetric phase diagrams, in which the critical volume fraction of polymer $\phi_c < 1/2$ (see Equation (32.28)). Equation (32.29) shows that as $N \rightarrow \infty$, the critical exchange parameter $\chi_c \rightarrow 1/2$.

Figure 32.8 Experimental phase diagrams of polystyrene in cyclohexane, showing the strong asymmetry indicated in Figure 32.7. Molecular weights: (○) 1.27×10^6 ; (○) 2.5×10^5 ; (●) 8.9×10^4 ; (●) 4.3×10^4 . The dashed lines indicate the Flory–Huggins theory predictions for the first and third curves from the top. Source: LH Sperling, *Introduction to Physical Polymer Science*, 2nd edition, Wiley, New York, 1992.



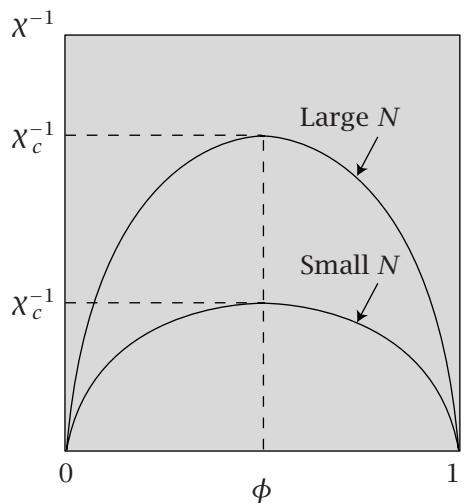
Polymers Are Immiscible with Other Polymers Because the Mixing Entropy Is so Small

Polymers rarely dissolve in other polymers. Why not? Now consider the case of two long-chain polymers of the same length $N = N_A = N_B$. Equations (32.27) and (32.25) give

$$\phi_c = \frac{1}{2} \quad \text{and} \quad \chi_c = \frac{2}{N}. \quad (32.30)$$

As $N \rightarrow \infty$, $\chi_c \rightarrow 0$ and the chains become immiscible at all temperatures (see Figure 32.9). The entropy of mixing polymers is so small that even an extremely small unfavorable enthalpy of mixing can prevent mixing. For example, even deuterated polybutadiene will not mix with protonated polybutadiene below a critical temperature of 61.5°C , if $N = 2300$. Rather than mix, polymer blends often separate into domains of each of the component polymers.

Figure 32.9 Miscibility phase diagram for two polymers having the same length N . Long polymers don't readily mix with each other: the two-phase region in the middle becomes large as $N \rightarrow \infty$.



EXAMPLE 32.2 Polymer mixing entropy. The mixing entropy for small molecules having mole fractions $x_A = x_B = 0.5$ is

$$\frac{\Delta S_{\text{mix}}}{Mk} = -0.5 \ln 0.5 - 0.5 \ln 0.5 = 0.69,$$

for a lattice having M sites. The mixing entropy for two polymers $N_A = N_B = 10,000$ having mole fractions $x_A = x_B = \phi_A = \phi_B = 0.5$ is

$$\begin{aligned}\frac{\Delta S_{\text{mix}}}{Mk} &= -\left(\frac{\phi_A}{N_A}\right) \ln \phi_A - \left(\frac{\phi_B}{N_B}\right) \ln \phi_B \\ &= -\left(\frac{0.5}{10,000}\right) \ln 0.5 - \left(\frac{0.5}{10,000}\right) \ln 0.5 \\ &= 6.9 \times 10^{-5}.\end{aligned}$$

(Take the entropic part of Equation (32.18).)

Because this entropy is so small, polymers have little tendency to mix.

Now we look at a factorization assumption of the Flory–Huggins model, and its implication for how polymer concentration affects chain conformations.

The Flory Theorem Says that the Intrinsic Conformations of a Chain Molecule Are Not Perturbed by Surrounding Chains

A fundamental assumption that underlies the polymer solution theory in this chapter is the factorization of the chain partition function in Equations (32.4) and (32.7) into a product of two terms. One term, which depends on factors $(z-1)$, describes *local* interactions that determine the number of bond rotational isomers between connected neighbors. The other term, which depends on factors $(M-i)/M$, accounts approximately for *nonlocal* interactions. Non-local interactions are due to excluded volume among monomers of different chains or among monomers that are not neighbors within a single chain. The assumption that local factors are independent of nonlocal factors is such an important premise that it has been given a name: the *Flory theorem*. According to this premise, increasing the concentration of chain molecules (or the compactness of a single-chain molecule) should diminish the partition function uniformly, not biasing the distribution of conformations in any one particular way or another, since the factors $(M-i)/M$ do not affect the factors $(z-1)$. This implies that the *distribution of the $(z-1)^{N-1}$ conformations* of a chain are exactly the same whether or not the chain is in a concentrated polymer solution (provided that it is homogeneous). The remarkable prediction is that chains will have exactly the same conformations when they are in highly tangled bulk polymer solutions as they would have if they were free in a simple solvent. What does change strongly with polymer concentration, however, is the excluded-volume entropy, which disfavors the dense states. In short, increasing the chain density diminishes the total *number* of conformations, but, to a first approximation, it does not affect their *conformational distribution*.

High polymer concentrations are predicted to decrease the populations of all the conformations uniformly. Is this remarkable theorem correct? To a very good approximation, neutron-scattering experiments first performed in the 1970s showed that it is [1]. A labeled chain in a bulk polymer medium has the same average radius as that expected for an ideal chain, which we describe in Chapter 33.

Summary

A polymer molecule is much larger than a typical small solvent molecule, often by more than a thousandfold. This difference in size has important consequences for the thermodynamics of polymer solutions. First, the colligative attraction of a polymer solution for small solvent molecules is much higher than would be expected on the basis of the mole fractions. The attraction is proportional to the volume fraction. Second, polymers are typically not miscible with other polymers because of small mixing entropies. Third, the size asymmetry between a polymer and a small-molecule solvent translates into an asymmetry in miscibility phase diagrams.

Problems

1. Concentrations: mole fractions x versus volume fractions ϕ . Consider a solution of polymers P and small molecules s , with $\phi_s = \phi_P = 0.5$. The polymer chain length is $N = 1000$.

What are the mole fractions x_s and x_P ?

2. Excluded volume in protein folding.

- Use the Flory-Huggins theory to estimate the number $v_1(c)$ of conformations of a polymer chain that are confined to be maximally compact, $M = N$. Simplify the expression by using Stirling's approximation.
- If the number of conformations of the unfolded chain is $v_1(u) = M(z-1)^{N-1}$ (since $M \gg N$), then compute the entropy of folding,

$$\Delta S_{\text{fold}} = k \ln \left[\frac{v_1(c)}{v_1(u)} \right].$$

3. Flory-Huggins mixing free energy. A polymer A with chain length $N_A = 1000$ is mixed with a small molecule B . The volume fractions are $\phi_A = 0.3$ and $\phi_B = 0.7$, and $\chi_{AB} = 0.1$. The temperature is 300 K. Using the Flory-Huggins theory,

- Compute $\Delta S_{\text{mix}}/Mk$.
- Compute $\Delta U_{\text{mix}}/MkT$.
- Compute $\Delta F_{\text{mix}}/M$.

4. The critical point of a Flory-Huggins solution. For the mixing of a polymer A ($N_A = 10^4$) with a small molecule B ($N_B = 1$), compute the critical volume fraction ϕ_c and the critical exchange parameter χ_c .

5. Volume fractions versus mole fractions. For what class of model mixing process are volume fractions identical to mole fractions?

6. Solvent chain lengths affect partition coefficients. What is the maximum effect of the solvent chain length on the partition coefficient according to the Flory-Huggins model?

7. The colligative properties of polymers. Show how a small-molecule diluent can reduce the freezing point of a polymer solution.

References

[1] PJ Flory, *Faraday Discuss Chem Soc* **68**, 14–25 (1979).

Suggested Reading

Excellent texts on polymer solutions are:

PG de Gennes, *Scaling Concepts in Polymer Physics*, Cornell University Press, Ithaca, NY, 1979.

M Doi, *Introduction to Polymer Physics*, Oxford University Press, Oxford, 1996.

PJ Flory, *Principles of Polymer Chemistry*, Cornell University Press, Ithaca, NY, 1953.

UW Gedde, *Polymer Physics*, Chapman & Hall, London, 1995.

WW Graessley, *Polymeric Liquids and Networks: Structure and Properties*, Garland Science, New York, 2004.

TL Hill, *An Introduction to Statistical Thermodynamics*, Addison-Wesley, Reading, MA, 1960 (reprinted by Dover Publications, New York, 1986).

SF Sun, *Physical Chemistry of Macromolecules: Basic Principles and Issues*, 2nd edition, Wiley, New York, 2004.

This page is intentionally left blank.

33 Polymer Elasticity & Collapse



Polymeric Materials Are Often Elastic

Rubber is elastic. Gooey liquids such as raw eggs are viscoelastic. Polymeric materials have these rubbery behaviors because chain molecules have many conformations of nearly equal energy and because stretching chains lowers their conformational entropy. The chains retract to regain entropy. Polymer chain entropies and energies also play key roles when proteins fold, when DNA becomes encapsulated within virus heads, when DNA wraps around histone proteins, and when polymer gels expand or contract. The simplest molecular description of polymer elasticity and entropy is the *random-flight model*.

Polymer Chains Can Be Modeled as Random Flights

We consider a polymer chain to be a connected sequence of N rigid vectors, each of length b . For now, each vector represents a chemical bond, but starting on page 662, we consider other situations in which each vector can represent a *virtual bond*, a collection of more than one chemical bond.

There are different ways to characterize the size of a polymer chain. The *contour length* L is the total stretched-out length of the chain,

$$L = Nb \quad (33.1)$$

(see Figure 33.1(a)). The contour length has a fixed value, no matter what the chain conformation.

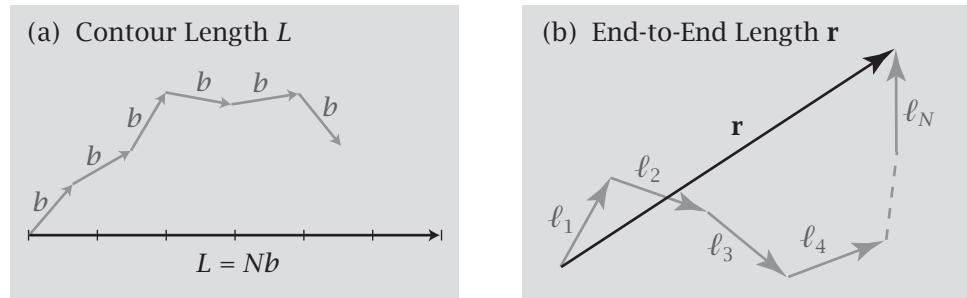


Figure 33.1 Two measures of the size of a polymer molecule. If there are N bonds and each bond has length b , the *contour length* (a) is the sum of the lengths of all bonds, $L = Nb$. The *end-to-end length* (b) is the *vector sum*, $\mathbf{r} = \sum_{i=1}^N \ell_i$, where ℓ_i is the vector representing bond i .

Another measure of polymer chain size is the *end-to-end length*, the length of the vector pointing through space from one end of the chain to the other, when the chain is in a given conformation (see Figure 33.1(b)). The end-to-end length varies from one chain conformation to the next. The end-to-end vector \mathbf{r} is the vector sum over the bonds ℓ_i for bonds $i = 1, 2, 3, \dots, N$:

$$\mathbf{r} = \sum_{i=1}^N \ell_i. \quad (33.2)$$

The end-to-end length cannot be greater than the contour length. For any vector, you can always consider the x , y , and z components individually. The x component of the end-to-end vector is

$$r_x = \sum_{i=1}^N x_i, \quad (33.3)$$

where $x_i = b \cos \theta_i$ is the projection on the x axis of the i th vector and θ_i is the angle of bond i relative to the x axis. If all the bond vectors have the same length b , you get

$$r_x = b \sum_{i=1}^N \cos \theta_i. \quad (33.4)$$

In solution, there are many molecules, each with a different conformation. Averaging over chain conformations gives experimentally measurable properties, such as the radius of the molecule as measured by light scattering, the viscosities of polymer solutions, and dynamical properties, such as the rates of polymer motion.

To get these averages, the simplest model is the *freely jointed chain*, or *random-flight* model ('flight' in three dimensions or 'walk' in two dimensions). In this model the angle of each bond is assumed to be independent of every other, including the nearest-neighboring bonds. Assuming that all conformations have equal probabilities, averaging the x component of the end-to-end

vector over all conformations gives

$$\langle r_x \rangle = \left\langle b \sum_{i=1}^N \cos \theta_i \right\rangle = b \sum_{i=1}^N \langle \cos \theta_i \rangle = 0, \quad (33.5)$$

because the average over randomly oriented vectors is $\langle \cos \theta_i \rangle = 0$ (see Example 1.26). This means that monomer N of a random-flight chain is in approximately the same location as the first monomer that defines the origin of the chain, because there are as many positive steps as negative steps, on average. Because the x , y , and z components are independent of each other, the y and z components can be treated identically and $\langle r_y \rangle = \langle r_z \rangle = 0$.

Because the mean value of the end-to-end vector is zero ($\langle r_x \rangle = 0$), $\langle r_x \rangle$ doesn't contain useful information about the size of the molecule. The *mean-square* end-to-end length $\langle r^2 \rangle$ or $\langle r_x^2 \rangle$ is a more useful measure of molecular size.

The square of the magnitude of the end-to-end vector can be expressed as a matrix of terms:

$$\mathbf{r} \cdot \mathbf{r} = \left(\sum_{i=1}^N \boldsymbol{\ell}_i \right)^2 = \begin{matrix} \boldsymbol{\ell}_1 \cdot \boldsymbol{\ell}_1 + \boldsymbol{\ell}_1 \cdot \boldsymbol{\ell}_2 + \boldsymbol{\ell}_1 \cdot \boldsymbol{\ell}_3 + \cdots + \boldsymbol{\ell}_1 \cdot \boldsymbol{\ell}_N + \\ \boldsymbol{\ell}_2 \cdot \boldsymbol{\ell}_1 + \boldsymbol{\ell}_2 \cdot \boldsymbol{\ell}_2 + \boldsymbol{\ell}_2 \cdot \boldsymbol{\ell}_3 + \cdots + \boldsymbol{\ell}_2 \cdot \boldsymbol{\ell}_N + \\ \vdots \\ \boldsymbol{\ell}_N \cdot \boldsymbol{\ell}_1 + \boldsymbol{\ell}_N \cdot \boldsymbol{\ell}_2 + \boldsymbol{\ell}_N \cdot \boldsymbol{\ell}_3 + \cdots + \boldsymbol{\ell}_N \cdot \boldsymbol{\ell}_N. \end{matrix} \quad (33.6)$$

Each product in this sum, $\boldsymbol{\ell}_i \cdot \boldsymbol{\ell}_j$, is a vector dot product, which is a product of the lengths of the two vectors multiplied by the cosine of the angle between the vectors. This sum involves two kinds of terms, 'self-terms' $\boldsymbol{\ell}_i \cdot \boldsymbol{\ell}_i$ (along the main diagonal) and 'cross-terms' $\boldsymbol{\ell}_i \cdot \boldsymbol{\ell}_j, i \neq j$. The self-terms give $\langle \boldsymbol{\ell}_i \cdot \boldsymbol{\ell}_i \rangle = b^2$, because the product of a vector with itself is the square of its length. The cross-terms give $\langle \boldsymbol{\ell}_i \cdot \boldsymbol{\ell}_j \rangle = b^2 \langle \cos \theta \rangle = 0$, because there is no correlation between the angles of any two bond vectors in the random-flight model. Therefore, the only terms that contribute to the sum are the self-terms along the diagonal. There are N of those, so

$$\langle r^2 \rangle = Nb^2. \quad (33.7)$$

The *root-mean-square* (RMS) end-to-end distance is $\langle r^2 \rangle^{1/2} = N^{1/2}b$.

A related quantity is the *radius of gyration* $R_g = [1/(N+1)] \langle \sum_{j=0}^N s_j^2 \rangle$, where $N+1$ is the number of beads connected by the N bonds, s_j is the vector from the center of mass of the chain to bead j , and the brackets $\langle \rangle$ indicate the average over all the chain configurations. The radius of gyration is used compute to moments of inertia in mechanics. For polymers, the radius of gyration, which is always positive, is a measure of the radial dispersion of the monomers. It can be shown that $R_g^2 = \langle r^2 \rangle / 6$ for a long linear chain [1]. We use $\langle r^2 \rangle$ instead of R_g because the math is simpler. Because the molecular weight of a polymer is proportional to the number of monomers it contains, one important prediction of the random-flight theory is that the mean radius (and the RMS end-to-end length) of a polymer chain increases in proportion to the square root of the molecular weight. This result is the centerpiece of much of polymer theory.

How do real chains differ from random flights? An important difference is chain stiffness.

Real Chain Molecules Have Bending Correlations and Stiffness

In real polymer chains, the orientation between each backbone bond and its neighboring bond along the chain is not random, as we assumed above. Rather, there are angular correlations among neighboring bond vectors. Contrast the mathematical description of a random flight with that of a straight rod, for which all vectors point in the same direction. For a rod, *all* the terms in Equation (33.6) contribute b^2 to the sum, and there are N^2 terms in the matrix, so the end-to-end length equals the contour length, and $\langle r^2 \rangle = (Nb)^2 = L^2$.

Because $\langle r^2 \rangle \sim N$ when bonds are uncorrelated and because $\langle r^2 \rangle \sim N^2$ when bonds are perfectly correlated, you might expect that partial correlation between bonds, which is a more realistic model for polymers, would lead to a dependence of $\langle r^2 \rangle$ on N with some exponent between 1 and 2. But this is not the case.

To illustrate, consider a model in which bonds have a weak angular correlation between first neighbors, $\langle \ell_i \cdot \ell_{i+1} \rangle = b^2 \gamma$, where $\gamma < 1$ is a positive constant. There are no correlations beyond first neighbors. If N is large, Equation (33.6) contains only terms along the main diagonal and the two adjacent diagonals. So for large N , you have $\langle r^2 \rangle \approx Nb^2 + 2Nb^2\gamma = Nb^2(1+2\gamma)$.

Angular correlations between neighboring bonds do not change the scaling, $\langle r^2 \rangle \sim N$. Neighbor correlations only change a multiplicative constant. This conclusion also holds if you include second neighbors and third neighbors, etc., provided only that the correlations decay to zero over a number of bonds $|i-j| \ll N$ that is much smaller than the chain length. So the random-flight model is useful even for real chains, for which bonds are not fully independent of their neighbors. *Chain stiffness* is a term sometimes used to describe the degree of angular correlation between neighboring bonds.

In general, the effects of bond correlations can be described by

$$\langle r^2 \rangle = C_N N b^2, \quad (33.8)$$

where the constant C_N is called the *characteristic ratio*. The subscript indicates that C_N can depend on the chain length N . The characteristic ratios of polymers are typically greater than one, indicating that there are correlations of orientations between near-neighbor bonds along the chain.

The *Kuhn model* allows you to treat real chains in terms of an equivalent freely jointed chain. The equivalent chain has *virtual bonds*, or *Kuhn segments*, each of which represents more than one real chemical bond. The number of virtual bonds N_K and the length of each bond b_K are determined by two requirements. First, the Kuhn model chain must have the same value of $\langle r^2 \rangle$ as the real chain, but it is freely jointed, so its characteristic ratio equals one:

$$\langle r^2 \rangle = C_N N b^2 = N_K b_K^2. \quad (33.9)$$

Second, the Kuhn chain has the same contour length as the real chain,

$$L = N b \cos \psi = N_K b_K, \quad (33.10)$$

where ψ is the angle between the real bond and the long axis of the extended real chain (see Figure 33.2), since that axis defines the direction of the virtual

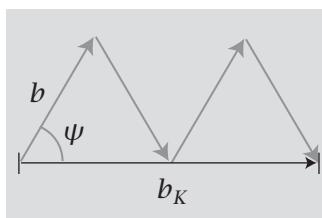


Figure 33.2 The Kuhn model of a polymer. b is the length of the chemical bond, b_K is the length of a virtual, or Kuhn, bond directed along the chain axis, and ψ is the angle between the chain axis and the chemical bond.

bonds. So, if you know ψ for a real polymer, you can solve two equations for two unknowns to determine its Kuhn parameters. Dividing Equation (33.9) by Equation (33.10) gives

$$\frac{b_K}{b} = \frac{C_N}{\cos \psi}, \quad (33.11)$$

so

$$\frac{N_K}{N} = \frac{\cos^2 \psi}{C_N}. \quad (33.12)$$

Example 33.1 gives the Kuhn length for polyethylene.

EXAMPLE 33.1 The Kuhn model of polyethylene. Polyethylene has a measured characteristic ratio of 6.7. For tetrahedral valence angles, $\psi = 70.5^\circ/2 = 35.25^\circ$ [1]. Equations (33.11) and (33.12) give the Kuhn length as $b_K/b = 6.7/\cos 35.25^\circ \approx 8$ times the chemical bond length, and there are $N/N_K = 6.7/\cos^2 35.25^\circ \approx 10$ chemical bonds per virtual bond.

The Persistence Length

Here is a more microscopic model of bond correlations. The *persistence length* ξ_p is the distance along the chain over which angular correlations persist [2]. Angular correlations lead to *stiffness*; it costs energy to bend a polymer chain. How does the persistence length depend on the bending energy of the chain? The angular correlation $g(s)$ between bond i and bond $i+s$ is defined as the average angle between two bond vectors that are s bonds apart along the chain:

$$g(s) = \frac{\langle \ell_i \cdot \ell_{i+s} \rangle}{b^2} = \langle \cos \theta(s) \rangle, \quad (33.13)$$

where $\theta(s)$ is the angle between the two bond vectors i and $i+s$ (see Figure 33.3) and $\langle \dots \rangle$ indicates the Boltzmann average over all the intervening bond angles. A reasonable approximation is that the angular correlation decays as an exponential function of s along the chain:

$$g(s) = e^{-s/\xi_p}; \quad (33.14)$$

this expression defines the persistence length ξ_p . To justify this exponential function, note that if all adjacent bonds i and $i+1$ had a fixed angle θ_0 between them, then $g(s) \propto (\cos \theta_0)^s$ diminishes by a multiplicative factor of $\cos \theta_0 < 1$ for each of the s bonds. Such repeated products are approximately exponential.

We want to compute $\langle \cos \theta \rangle$ for Equation (33.13). When there are strong correlations, $\theta(s) \approx 0$, you have $\cos \theta \approx 1 - \theta^2/2 + \dots$. So

$$\langle \cos \theta \rangle \approx \left\langle 1 - \frac{\theta^2}{2} \right\rangle = 1 - \frac{\langle \theta^2 \rangle}{2}. \quad (33.15)$$

In order to evaluate the quantity $\langle \theta^2 \rangle$, you need to know the Boltzmann probability distribution of the chain bond angles θ . For that, you need the bending energy of the chain, $\varepsilon_{\text{bend}}(\theta)$. The energy required to bend elastic-rod-like

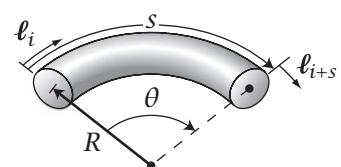


Figure 33.3 Relationship between the bond vectors ℓ_i and ℓ_{i+s} , the angle θ between them, the radius of curvature R , and the contour length s .

materials—which is often a reasonable approximation even for a single polymer chain—is a Hooke’s law spring function of angle:

$$\varepsilon_{\text{bend}} = \left(\frac{k_b}{2s} \right) \theta^2. \quad (33.16)$$

(This is called spring-like because its energy is proportional to the square of its deformation θ , just like a spring’s energy is proportional to the square of its extension x). In Equation (33.16), k_b is the spring constant for bending, and the factor of s in the denominator is because we want a bending energy per unit length s of the chain. According to the Boltzmann law, the probability of any particular angle θ is proportional to $\exp(-\varepsilon_{\text{bend}}/RT) = \exp[-k_b\theta^2/(2sRT)]$.

To compute the Boltzmann average, note that more vectors point towards larger angles θ than toward smaller angles. To capture this, use $\sin \theta d\theta$ when integrating (see Equation (1.44)) and use $\theta \approx \sin \theta$ for small θ , to get

$$\langle \theta^2 \rangle = \frac{\int_0^\pi \theta^2 e^{-a\theta^2} \sin \theta d\theta}{\int_0^\pi e^{-a\theta^2} \sin \theta d\theta} \approx \frac{\int_0^\infty \theta^3 e^{-a\theta^2} d\theta}{\int_0^\infty \theta e^{-a\theta^2} d\theta} = \frac{1}{a} = \frac{2sRT}{k_b}, \quad (33.17)$$

where $a = k_b/(2sRT)$. We evaluated the integrals using Equations (K.2) and (K.10) in Appendix K. Now, substitute Equation (33.17) into Equation (33.15), make the approximation that $\exp(-s/\xi_p) \approx 1 - s/\xi_p$ for small s , and use Equations (33.13) and (33.14) to get

$$g(s) = 1 - \frac{1}{2} \left(\frac{2sRT}{k_b} \right) = 1 - \frac{s}{\xi_p} \quad (33.18)$$

$$\implies \xi_p = \frac{k_b}{RT}. \quad (33.19)$$

Equation (33.19) shows that the persistence length ξ_p is proportional to the bending spring constant k_b . Here are two applications.

EXAMPLE 33.2 The bending of DNA. DNA is stiff because it is a double helix. At room temperature in dilute solution, the persistence length of DNA molecules is found to be approximately $\xi_p = 500 \text{ \AA}$. Use Equation (33.19) to get the bending spring constant:

$$k_b = \xi_p RT = (500 \text{ \AA})(0.6 \text{ kcal mol}^{-1}) = 300 \text{ \AA kcal mol}^{-1}. \quad (33.20)$$

EXAMPLE 33.3 The energy cost of bending DNA around histone proteins. In chromosomes, DNA molecules are wrapped tightly around protein molecules called histones. Because DNA is such a stiff molecule, the tight curvature of the DNA is energetically costly. Consider wrapping a DNA molecule around a full circle; see Figure 33.4 [2].

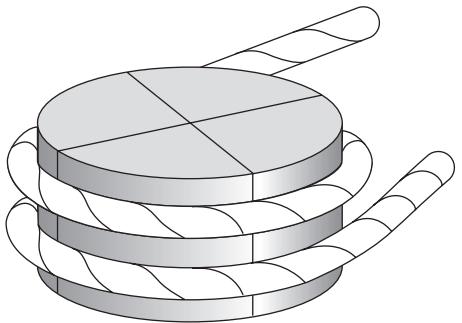


Figure 33.4 A chromosomal DNA molecule (the rope-like object) is wrapped around a histone protein (the post-like object). Because DNA is stiff, it costs energy to bend the DNA so tightly. Source: R Phillips, J Kondev, and J Theriot, *Physical Biology of the Cell*, Garland Science, New York, 2009.

The radius of a histone protein is 45 Å. Take k_b from Example 33.2. Substitute $\theta = 2\pi$ and the chain length $s = 2\pi R$ into Equation (33.16) to get

$$\varepsilon_{\text{bend}} = \frac{\pi k_b}{R} = \frac{(3.14) \times (300 \text{ Å kcal mol}^{-1})}{45 \text{ Å}} = 21 \text{ kcal mol}^{-1}. \quad (33.21)$$

We neglected the chain entropy here (see Example 33.4) because the entropic restriction due to cyclization is much smaller than the bending energy, in this case.

Other models of chain stiffness include the *wormlike chain* model [3] and the *rotational isomeric state model* [1, 4]. The latter accounts not only for the angles and lengths of chemical bonds, but also for the different statistical weights of the various possible bond conformations.

To understand why polymeric materials are elastic, you need to know more than just the first and second moments, $\langle r \rangle = 0$ and $\langle r^2 \rangle = Nb^2$; you need the full distribution function of the end-to-end lengths.

Random-Flight Chain Conformations Are Described by the Gaussian Distribution Function

To get the distribution of end-to-end lengths of a polymer chain, we use the random-flight model, also commonly used to describe the diffusion of a particle (see Chapter 18). In diffusion, a particle starts at the origin and moves a distance b in a random direction at each time step. After N time steps, the particle has moved an average of $\langle r^2 \rangle^{1/2} \propto \sqrt{N}$ steps away from where it started.

Random-flight polymers behave similarly. Figure 33.5 shows that the conformation of a random-flight polymer chain looks like the path of a diffusing particle. For a polymer chain, we first consider a one-dimensional random walk, then we generalize to three dimensions. Fix the beginning of the chain at the origin. Each bond vector, because it is oriented randomly, can have a different projection onto the x axis. Imagine laying down one bond at a time, from bond $i = 1$ to N . Each step of length b is taken in a random direction. We found above that if a polymer chain has N bonds, its terminal end will be an average distance $\langle r^2 \rangle^{1/2} = b\sqrt{N}$ from its beginning end. The probability distribution for the terminal end of the chain is (see Equation (18.9))

$$P(x, N) = \left(\frac{\beta}{\pi}\right)^{1/2} e^{-\beta x^2} = \left(\frac{3}{2\pi Nb^2}\right)^{1/2} e^{-3x^2/2Nb^2}, \quad (33.22)$$



Figure 33.5 Example of a random-walk chain conformation.

where $\beta = 3/(2Nb^2)$ (not to be confused with $\beta = (kT)^{-1}$) is standard polymer notation. Equation (33.22) gives you the relative numbers of all N -mer chain conformations that begin at the origin of the x axis and end at x . The Gaussian distribution has a peak at $x = 0$, implying that most of the chains have about as many + steps as - steps.

Relatively few chains are highly stretched. If you want to convert from the *fractions* of chain conformations to *total numbers* of chain conformations, you can multiply by the approximate total number of chain conformations, z^N , where z is the number of rotational isomers per bond, to get $z^N P(x, N)$. Now we show that the Gaussian model predicts that rubber and other polymeric materials are highly deformable and elastic.

Polymer Elasticity Follows Hooke's Law

When a polymer is stretched and then released, it retracts like a Hooke's law spring. The retractive force is proportional to the extension. The retractive force is entropic.

Suppose you pull the two ends of a polymer chain apart along the x direction. Because all conformations of random-flight chains have the same energy, the free energy is purely entropic: $F = -TS$. Substituting Equation (33.22) into the expression $S/R = \ln P(x, N)$ gives, for the entropy and free energy of stretching,

$$\frac{F}{RT} = -\frac{S}{R} = -\ln P(x, N) = \beta x^2 + \text{constant} = \frac{3x^2}{2Nb^2} + \text{constant}. \quad (33.23)$$

The retractive force f_{elastic} is defined as the derivative of the free energy:

$$f_{\text{elastic}} = -\frac{dF}{dx} = -2RT\beta x = -\frac{3RTx}{Nb^2}. \quad (33.24)$$

Equation (33.24) shows that the retractive force f_{elastic} is linear in the displacement x , like a Hooke's law spring. Polymeric materials can be stretched to many times their undeformed size. (Metal can be stretched too, but only to a much smaller extent, and by a different, enthalpic, mechanism.) Fully stretched, the size of a chain is determined by its contour length $L = Nb$. In its undeformed state, its average size is $\langle r^2 \rangle^{1/2} = N^{1/2}b$, so a polymer chain can be stretched by nearly a factor of \sqrt{N} .

Figure 33.6 shows how the retractive force depends on the stretched lengths of single molecules of DNA. You can see the linearity between force and extension for small stretching. At large deformations, the retractive forces become much stronger. Then the Gaussian distribution no longer applies.

Now we generalize to three dimensions to describe elasticity as a *vector* force that drives the chain ends together (see Figure 33.7).

Elasticity in Two and Three Dimensions

Let \mathbf{r} represent the vector from the origin, where the chain begins, to (x, y, z) , where the chain ends (see Figure 33.8(a)). The length of the end-to-end vector is r , where $r^2 = x^2 + y^2 + z^2$. Because each bond is oriented randomly, the x ,

Force (pN)

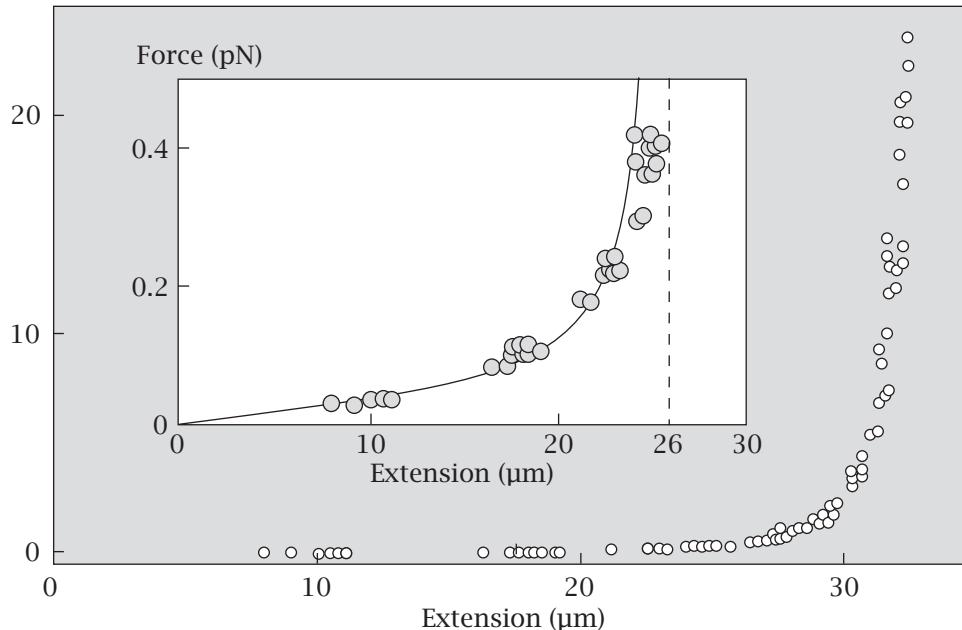


Figure 33.6 Stretching a single molecule of DNA leads to a retractive force that is linear in the extension at low extensions, but steeper at higher extensions. The inset is an enlargement of the vertical axis, for small deformations. Source: SB Smith, L Finzi, and C Bustamante, *Science* **258**, 1122–1126 (1992).

y , and z projections of a random-flight chain are independent of each other. So the probability density of finding the chain end in a small volume element between (x, y, z) and $(x+dx, y+dy, z+dz)$ is the product of independent factors:

$$\begin{aligned} P(x, y, z, N) dx dy dz &= P(x, N)P(y, N)P(z, N) dx dy dz \\ &= \left(\frac{\beta}{\pi}\right)^{3/2} e^{-\beta(x^2+y^2+z^2)} dx dy dz. \end{aligned} \quad (33.25)$$

The probability that a chain conformation has the particular end-to-end vector \mathbf{r} is

$$P(\mathbf{r}, N) = P(x, y, z, N) = \left(\frac{\beta}{\pi}\right)^{3/2} e^{-\beta r^2}. \quad (33.26)$$

Equations (33.25) and (33.26) give the probability that the chain terminus is located at a specific location \mathbf{r} between (x, y, z) and $(x+dx, y+dy, z+dz)$, if the chain originates at $(0, 0, 0)$ (see Figure 33.8(a)). However, sometimes this is not exactly the quantity you want. Instead you may want to know the probability that the chain terminates *anywhere in space at a distance r from the origin* (see Figure 33.8(b)). We denote this quantity $P(r, N)$, without the bold \mathbf{r} notation. Because the number of elements having a volume $dx dy dz$ grows with distance r from the origin as $4\pi r^2$, the probability of finding the two ends separated by a distance r in any direction is $P(r, N) = 4\pi r^2 P(\mathbf{r}, N)$. The normalization



Figure 33.7 Chain conformational elasticity is a vector force tending to cause the chain ends (labeled N and 1) to be near each other, on average.

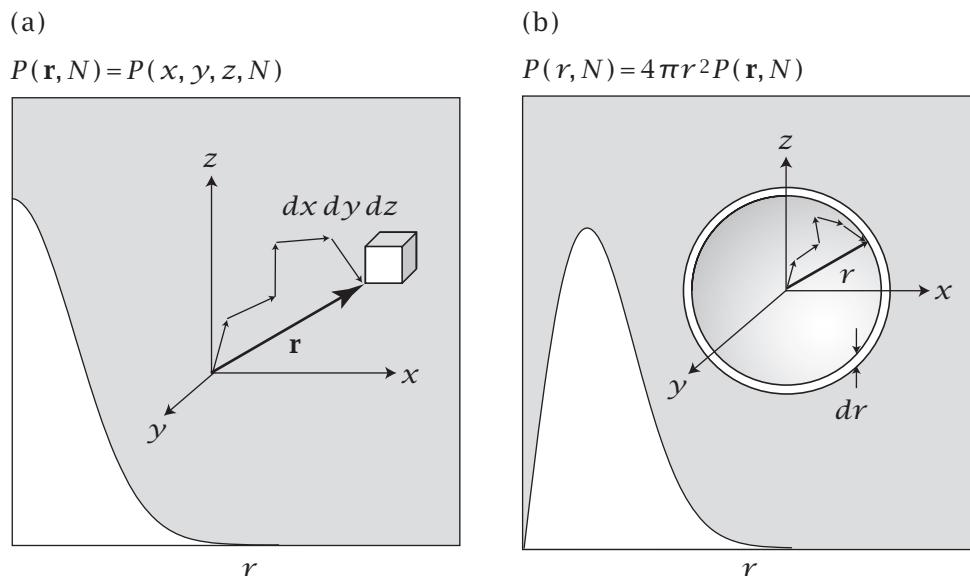


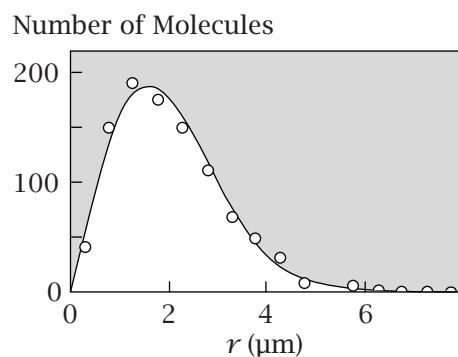
Figure 33.8 Two different probability densities for the termini of chains that begin at the origin $(0,0,0)$. (a) The probability density $P(\mathbf{r}, N) = P(x, y, z, N)$ that the chain end is *in a particular location* between (x, y, z) and $(x+dx, y+dy, z+dz)$. The most probable termination point is the origin. (b) The probability density $P(r, N) = 4\pi r^2 P(\mathbf{r}, N)$ that the chain end is *anywhere in a radial shell* between r and $r+dr$. The peak of this distribution is predicted by Equation (33.30). Source: CR Cantor and PR Schimmel, *Biophysical Chemistry, Part II: Techniques for the Study of Biological Structure and Function*, WH Freeman, San Francisco, 1980.

is given by

$$\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} P(x, y, z, N) dx dy dz = \int_0^{\infty} 4\pi r^2 P(r, N) dr = 1. \quad (33.27)$$

The function $P(r, N)$ has a peak because it is a product of two quantities: the probability $P(\mathbf{r}, N)$ decreases monotonically with distance from the origin, but the number of volume elements increases as $4\pi r^2$. Figure 33.9 shows that the two-dimensional end-to-end distance distribution of T3 DNA molecules adsorbed on surfaces and counted in electron micrographs is well predicted

Figure 33.9 Distribution of end-to-end distances r of T3 DNA adsorbed on cytochrome *c* film. The line represents the theoretical distribution for two-dimensional random walks. Source: D Lang, H Bujard, B Wolff, and D Russell, *J Mol Biol* **23**, 163–181 (1967).



by two-dimensional random-walk theory. Example 33.4 shows how the Gaussian distribution function is used for predicting polymer cyclization equilibria and kinetics.

EXAMPLE 33.4 Polymer cyclization (Jacobson–Stockmayer theory). What is the probability that the two ends of a polymer chain come close enough together for them to react with each other? This probability is useful for calculating the rate and equilibrium constants for cyclization processes in which linear chains form circles (see Figure 33.10). Here we follow the random-flight model of H Jacobson and WH Stockmayer [5].

Suppose that the polymer has N bonds. To determine the probability that the chain end is within a bond distance b of the chain beginning, integrate $P(r, N)$, the probability of finding the ends a distance r apart, from 0 to b :

$$\begin{aligned} P_{\text{cyclization}} &= \int_0^b P(r, N) dr \\ &= \left(\frac{3}{2\pi Nb^2} \right)^{3/2} \int_0^b e^{-3r^2/2Nb^2} 4\pi r^2 dr. \end{aligned} \quad (33.28)$$

When the ends are together, $r = b$ is small ($r^2/Nb^2 \ll 1$), so $e^{-3r^2/2Nb^2} \approx 1$. Then the integral in Equation (33.28) is $\int_0^b 4\pi r^2 dr = 4\pi b^3/3$, and

$$P_{\text{cyclization}} = \left(\frac{3}{2\pi Nb^2} \right)^{3/2} \left(\frac{4}{3}\pi b^3 \right) = \left(\frac{6}{\pi} \right)^{1/2} N^{-3/2}.$$

The main prediction of the Jacobson–Stockmayer theory is that the cyclization probability diminishes with chain length as $N^{-3/2}$. The longer the chain, the smaller is the probability that the two ends are close enough to react. This prediction is confirmed by experiments shown in Figures 33.11 and 33.12.

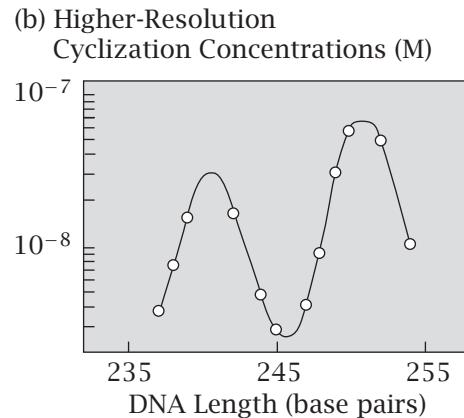
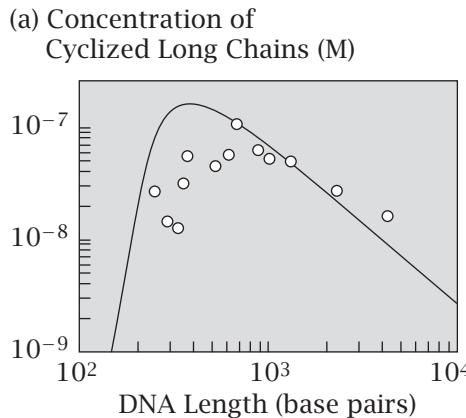


Figure 33.11 (a) For long chains the DNA cyclization probability diminishes as predicted by the Jacobson–Stockmayer theory. Shortening the chain stiffens it, reducing the cyclization probability below the Jacobson–Stockmayer value. (b) At higher resolution, the cyclization probability depends on periodicities in the polymer, neglected in the random-flight treatment. Source: D Shore and RL Baldwin, *J Mol Biol* **170**, 957–981 (1983).



Figure 33.10 For polymer cyclization, the two chain ends must be close together.

Figure 33.12 Experimental molar cyclization equilibrium constants K (in mol dm⁻³) for cyclic $[O(CH_2)_{10}OCO(CH_2)_4CO]_n$ in poly(decamethylene adipate) melts at 423 K (\circ) versus chain length n are compared with values calculated (\bullet) from the Jacobson–Stockmayer theory. Source: JA Semlyen, ed., *Cyclic Polymers*, Elsevier, London, 1986.

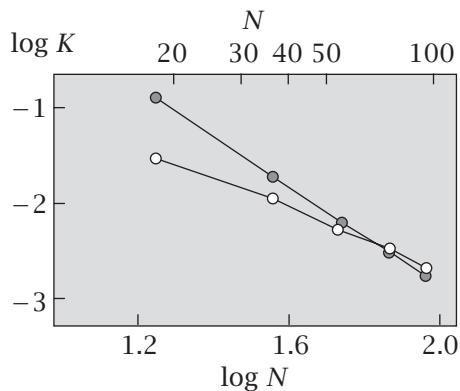


Figure 33.11 shows that the probability of DNA cyclization diminishes with chain length for long chains, as predicted by the theory, but you can also see the effect of chain stiffness, which is not treated by the theory. If chains are too short, their stiffness prevents them from bending enough to cyclize, so their cyclization probability is smaller than is predicted by the random-flight model. Also, chemical details can matter for short chains: the ends must be oriented correctly with respect to each other to cyclize.

Now we compute another useful measure of average size of a Gaussian chain, called the *most probable radius*.

EXAMPLE 33.5 Most probable radius R_0 . Begin with the entropy $S(r) = R \ln P(r, N)$. Because $F = -TS$, using Equation (33.26) gives

$$\frac{F}{RT} = -\frac{S}{R} = -\ln[4\pi r^2 P(r, N)] = \beta r^2 - 2 \ln r + \text{constant}. \quad (33.29)$$

To get the equilibrium value, or most probable value, $r = R_0$, find the minimum free energy:

$$\begin{aligned} \left[\frac{d}{dr} \left(\frac{F}{RT} \right) \right]_{r=R_0} &= 2\beta R_0 - \frac{2}{R_0} = 0 \\ \Rightarrow R_0^2 &= \frac{1}{\beta} = \frac{2Nb^2}{3}. \end{aligned} \quad (33.30)$$

The quantity $R_0^2 = 2Nb^2/3 = (2/3)\langle r^2 \rangle$ is the square of the most probable radius of the Gaussian chain. Starting on page 673, we use it for treating polymer collapse and folding processes.

Now we use the theory of polymer chains to describe elastic materials.

The Elasticity of Rubbery Materials Results from the Sum of Chain Entropies

Rubber and other polymeric materials are elastic. Polymeric elastomers are covalently cross-linked networks of polymer chains. Here we describe one of the simplest and earliest models for the retractive forces of polymeric materials: the *affine network model*.

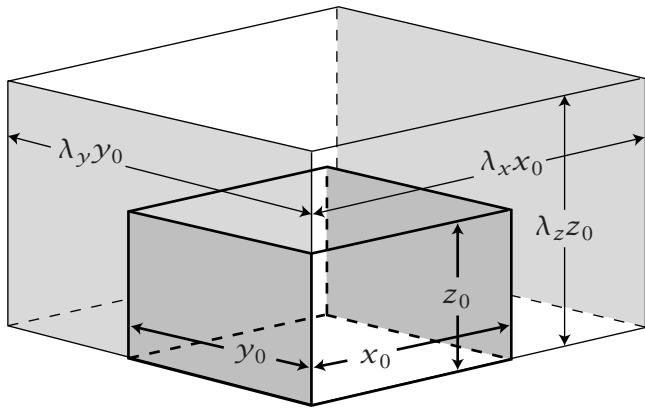


Figure 33.13 An *affine* deformation of a material from dimensions (x_0, y_0, z_0) to $(\lambda_x x_0, \lambda_y y_0, \lambda_z z_0)$.

Suppose you have m chains, all of the same length N . The bond length is b . The chain ends are covalently cross-linked at junction points. Assume that there are no intermolecular interactions, and that the total elastic free energy of the material is the sum of the elastic free energies of each of the chains. For the undeformed material, indicated by subscript 0, the end of a chain is at (x_0, y_0, z_0) and

$$r_0^2 = x_0^2 + y_0^2 + z_0^2 = Nb^2. \quad (33.31)$$

If the material is deformed by a factor λ_x in the x direction, λ_y in the y direction, and λ_z in the z direction (see Figure 33.13), so that $x = \lambda_x x_0$, $y = \lambda_y y_0$, and $z = \lambda_z z_0$, the chain end is moved to (x, y, z) , and

$$r^2 = x^2 + y^2 + z^2 = \lambda_x^2 x_0^2 + \lambda_y^2 y_0^2 + \lambda_z^2 z_0^2. \quad (33.32)$$

Using $F = -RT \ln P(x, y, z, N)$, $\beta = 3/(2Nb^2)$, and Equation (33.26), you have the free energy for deforming a single chain, F_1 :

$$\Delta F_1 = F_{\text{deformed}} - F_{\text{undeformed}} = -RT \ln \left[\frac{P(x, y, z, N)}{P(x_0, y_0, z_0, N)} \right] = RT\beta(r^2 - r_0^2).$$

Summing over m independent chains gives the total free energy F_m :

$$\Delta F_m = RT\beta \sum_m (r^2 - r_0^2). \quad (33.33)$$

Assuming that all chains are equivalent, you have $\sum_m (r^2 - r_0^2) = m(\langle r^2 \rangle - \langle r_0^2 \rangle)$, and Equation (33.33) becomes

$$\begin{aligned} \Delta F_m &= RT\beta m (\langle r^2 \rangle - \langle r_0^2 \rangle) \\ &= RT\beta m [(\lambda_x^2 - 1)\langle x_0^2 \rangle + (\lambda_y^2 - 1)\langle y_0^2 \rangle + (\lambda_z^2 - 1)\langle z_0^2 \rangle]. \end{aligned} \quad (33.34)$$

If the undeformed chains are isotropic (all the directions are equivalent), then $\langle x_0^2 \rangle = \langle y_0^2 \rangle = \langle z_0^2 \rangle = (Nb^2)/3$, and Equation (33.34) becomes

$$\frac{\Delta F_m}{RT} = \frac{m}{2} (\lambda_x^2 + \lambda_y^2 + \lambda_z^2 - 3). \quad (33.35)$$

In a macroscopic material, applying a force in one direction can cause forces and deformations in other directions. The forces per unit area in various directions are called *stresses* and the relative displacements are called *strains*.

Stresses are derivatives of the free energy with respect to strains. If you deform a material by a factor α , where the x -axis, y -axis, and z -axis deformations depend on α , the free energy is a function $F(\lambda_x(\alpha), \lambda_y(\alpha), \lambda_z(\alpha))$. The derivative can be written in terms of λ_x^2 , λ_y^2 , and λ_z^2 as

$$\left(\frac{\partial F}{\partial \alpha} \right) = \left(\frac{\partial F}{\partial \lambda_x^2} \right) \frac{d\lambda_x^2}{d\alpha} + \left(\frac{\partial F}{\partial \lambda_y^2} \right) \frac{d\lambda_y^2}{d\alpha} + \left(\frac{\partial F}{\partial \lambda_z^2} \right) \frac{d\lambda_z^2}{d\alpha}. \quad (33.36)$$

Suppose that you stretch an elastomer along the x direction by a factor $\alpha = x/x_0$. Then the x -direction force is

$$f_x = - \left(\frac{\partial \Delta F_m}{\partial \alpha} \right) = - \frac{1}{x_0} \left(\frac{\partial \Delta F_m}{\partial \alpha} \right) = - \frac{\alpha}{x} \left(\frac{\partial \Delta F_m}{\partial \alpha} \right). \quad (33.37)$$

The stress τ acting in the opposite direction equals the force given by Equation (33.37) divided by the deformed cross-sectional area yz :

$$\tau = - \frac{f_x}{yz} = \frac{\alpha}{V} \left(\frac{\partial \Delta F_m}{\partial \alpha} \right), \quad (33.38)$$

where $V = xyz$ is the final volume of the material [6, 7]. Examples 33.6 and 33.7 are applications of this elasticity model.

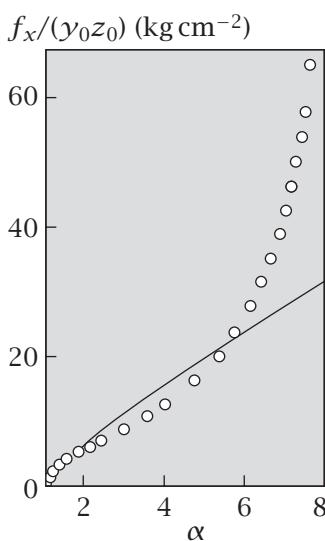


Figure 33.14 Stretching a rubber string gives stress $f_x/(y_0 z_0)$ versus elongation α : (○) experimental data; (—) curve predicted by Equation (33.41). Source: P Munk, *Introduction to Macromolecular Science*, Wiley, New York, 1989. Data are from LRG Treloar, *Trans Faraday Soc* **40**, 59–69 (1944).

EXAMPLE 33.6 Stretch a rubber band along the x axis. If you stretch rubber, its volume remains approximately constant. Therefore, stretching along the x direction by $\lambda_x = \alpha$ leads to

$$\lambda_y = \lambda_z = \frac{1}{\sqrt{\alpha}}. \quad (33.39)$$

Then Equation (33.35) becomes

$$\frac{\Delta F_m}{RT} = \frac{m}{2} \left(\alpha^2 + \frac{2}{\alpha} - 3 \right).$$

Because

$$\frac{\partial \Delta F_m}{\partial \lambda_x^2} = \frac{\partial \Delta F_m}{\partial \lambda_y^2} = \frac{\partial \Delta F_m}{\partial \lambda_z^2} = mRT/2,$$

Equation (33.35) gives the force

$$f_x = \frac{1}{x_0} \frac{\partial F_m}{\partial \alpha} = \frac{mRT}{x_0} \left(\alpha - \frac{1}{\alpha^2} \right). \quad (33.40)$$

Substituting Equation (33.40) into Equation (33.38) gives the stress, which is the force per unit area, as

$$\tau = \frac{mRT}{V} \left(\alpha^2 - \frac{1}{\alpha} \right). \quad (33.41)$$

Figure 33.14 shows that Equation (33.40) predicts experimental data adequately at extensions below about $\alpha = 3$ to 5, but that the chains become harder to stretch than the model predicts at higher extensions.

EXAMPLE 33.7 Stretch a rubber sheet biaxially. Stretch a rubber sheet along the x axis by an amount $\lambda_x = \alpha_1$ and along the y axis by an amount $\lambda_y = \alpha_2$, where α_1 is independent of α_2 . If the volume is constant, $\lambda_z = 1/(\alpha_1\alpha_2)$. Then the stress along the x axis is

$$\tau_x = \frac{mRT}{2V} \alpha_1 \left(\frac{d\lambda_x^2}{d\alpha_1} + \frac{d\lambda_z^2}{d\alpha_1} \right) = \frac{mRT}{V} \left(\alpha_1^2 - \frac{1}{\alpha_1^2 \alpha_2^2} \right). \quad (33.42)$$

The retractive stress in elastomers depends not only on the deformation but also on the cross-link density, through m/V , the density of chains. In a unit volume, there are $2m$ total chain ends. If each junction is an intersection of j chain ends, then there will be $(1 \text{ junction}/j \text{ chain ends}) \times (2m \text{ chain ends}) = (2m/j)$ junctions, so the number of junctions is proportional to m . Therefore, the retractive force increases linearly with the cross-link density of the network. Bowling balls are made of a type of rubber that has a much higher cross-link density than rubber bands.

EXAMPLE 33.8 Polymer swelling or contraction. Suppose you expand a polymeric material uniformly in all directions by an amount $\lambda = \lambda_x = \lambda_y = \lambda_z$. Equation (33.35) says that the elastic free energy will increase by an amount

$$\frac{\Delta F_m}{RT} = \frac{3m}{2} (\lambda^2 - 1). \quad (33.43)$$

The main advance embodied in theories of polymer elasticity was the recognition that the origin of the force is mainly the conformational freedom of the chains, and is entropic, not enthalpic. The conformational entropies of polymers are important not only for stretching processes. They also oppose the contraction of a polymer chain to a radius smaller than its equilibrium value.

Polymers Expand in Good Solvents, Are Random Flights in θ Solvents, and Collapse In Poor Solvents

Put a polymer molecule into a solvent. It may expand or contract. Extreme contraction is called *collapse*. Examples of polymer collapse include the folding of proteins in water or the insertion of DNA molecules into viruses. Here we describe a simple model, originated by PJ Flory [8, 9], to describe polymer collapse and expansion. We consider three different situations. (1) Random-flight behavior, in which the polymer size grows as $\langle r^2 \rangle^{1/2} = bN^{1/2}$, is dominated by local interactions. This applies only to a limited class of solvent and temperature conditions, called *θ conditions*, *θ solvents*, or *θ temperatures*. (2) In a *good solvent*, monomer-solvent interactions are more favorable than are monomer-monomer interactions. Chains expand in good solvents: the radius of a polymer molecule grows as $\langle r^2 \rangle^{1/2} = bN^{3/5}$. Such polymers are expanded relative to random flights because of the self-avoidance of the segments of the chain. (3) In *poor solvents*, the chain monomers are attracted to each other more strongly than they are attracted to the solvent. Poor solvents cause an isolated chain

to collapse into a compact globule, with radius $\langle r^2 \rangle^{1/2} = bN^{1/3}$. Poor solvents can also cause multiple chains to aggregate with each other. For example, oil chains such as polymethylene phase-separate from water because water is a poor solvent. We derive these results below.

How does the conformational free energy depend on the radius? Consider a chain with N monomers. As in Equation (33.30), our strategy is to find the radius $r = R$ that maximizes the entropy, or minimizes the free energy. However, now we include an additional contribution to the free energy. In addition to the elastic free energy F_{elastic} , we now also account for the solvation free energy $F_{\text{solvation}}$, using the Flory-Huggins theory of Chapter 32:

$$\left(\frac{d}{dr} [F_{\text{elastic}}(r) + F_{\text{solvation}}(r)] \right)_{r=R} = 0. \quad (33.44)$$

The solvation free energy depends on the radius through the mean chain segment density $\rho = N/M$. M is the number of sites of a lattice that contains the chain. We define M in terms of the chain radius below. Low density, $\rho \rightarrow 0$, means the chain is expanded, while high density, $\rho \rightarrow 1$, means the chain is compact.

Chain collapse is opposed by the conformational entropy but is driven by the gain in monomer-monomer contacts, which are favorable under poor-solvent conditions. To get the conformational entropy for $F_{\text{solvation}}$, you can begin with Equation (32.4), which gives v_1 , the number of conformations of a single chain. But we make two changes. First, we leave out the factor $(z-1)^{N-1}$ because it is a constant that doesn't change with the density ρ , and leaving it out simplifies the math. (The independence of the conformational entropy from the factor $(z-1)^{N-1}$ in the Flory model has the important implication that chain stiffness and local interactions do not contribute to entropies of collapse or expansion.) Second, our focus on a single chain means that its center-of-mass position in space is irrelevant. This is equivalent to neglecting the placement of the first monomer, so we replace the factor $M^{-(N-1)}$ with M^{-N} , and the conformational entropy of the chain is

$$\frac{S}{k} = \ln v_1 = \ln \frac{M!}{(M-N)!M^N}. \quad (33.45)$$

Multiply the numerator and denominator of Equation (33.45) by $N!$ to put part of this expression into a more familiar form:

$$\begin{aligned} \frac{S}{k} &= \ln \left(\frac{M!}{(M-N)!N!} \right) + \ln \left(\frac{N!}{M^N} \right) \\ &= -N \ln \left(\frac{N}{M} \right) - (M-N) \ln \left(1 - \frac{N}{M} \right) + N \ln \left(\frac{N}{M} \right) - N \ln e \\ &= -(M-N) \ln \left(1 - \frac{N}{M} \right) - N, \end{aligned} \quad (33.46)$$

by using Stirling's approximation $N!/M^N \approx (N/eM)^N = (N/M)^N e^{-N}$. Now divide by N to get the entropy per monomer, and express the result in terms of the average density $\rho = N/M$:

$$\frac{S}{Nk} = - \left(\frac{1-\rho}{\rho} \right) \ln(1-\rho) - 1. \quad (33.47)$$

You can check that $S \rightarrow 0$ as $\rho \rightarrow 0$ and $S \rightarrow -Nk$ as $\rho \rightarrow 1$.

EXAMPLE 33.9 Polymer collapse entropy. When a protein folds, it collapses to a nearly maximally compact state. What is the entropic component of the free energy opposing collapse at $T = 300\text{K}$? If the chain length is $N = 100$ monomers, use $\rho = 1$ and $\rho = 0$, respectively, for the compact and open states in Equation (33.47), to get

$$\begin{aligned}\Delta F_{\text{collapse}} &= -T(S_{\text{compact}} - S_{\text{open}}) = NRT \\ &\approx (100)(1.987 \text{ cal K}^{-1} \text{ mol}^{-1})(300\text{K}) \approx 60 \text{ kcal mol}^{-1}.\end{aligned}$$

Next, we determine how the contact energy depends on the chain radius, or density. Let w_{pp} , w_{ss} , and w_{sp} represent the energies of a contact between two polymer segments, between two solvent molecules, and between a solvent molecule and a polymer segment, respectively. Figure 33.15 and Equation (15.13) indicate that the free energy Δg for desolvating two polymer segments and forming a contact is

$$\Delta g = -2 \left(w_{sp} - \frac{w_{ss} + w_{pp}}{2} \right) = -\frac{2}{z} \chi RT. \quad (33.48)$$

The minus sign indicates that the process defining Δg is the reverse of the process defining χ (see Figure 15.6). The mean-field approximation gives an estimate of the number of contacts among chain monomers. The probability that a site adjacent to a polymer segment contains another polymer segment is $\rho = N/M$, and there are z sites that are neighbors of each of the N monomers, so the contact energy U relative to the fully solvated chain is

$$U = \frac{N\rho z}{2} \Delta g = -NRT\rho\chi. \quad (33.49)$$

The factor of 1/2 corrects for the double counting of interactions. Combining the entropy equation (33.47) with the energy equation (33.49) gives the conformational free energy of the chain as a function of its compactness:

$$\frac{F_{\text{solvation}}}{NRT} = \frac{U}{NRT} - \frac{S}{NR} = \left(\frac{1-\rho}{\rho} \right) \ln(1-\rho) + 1 - \rho\chi. \quad (33.50)$$

When the chain is relatively open and solvated, the density is small, $\rho \ll 1$. Then you can use the approximation $\ln(1-\rho) \approx -\rho - (1/2)\rho^2 - \dots$ to get

$$\begin{aligned}\left(\frac{1-\rho}{\rho} \right) \ln(1-\rho) + 1 &\approx \left(\frac{1-\rho}{\rho} \right) \left(-\rho - \frac{\rho^2}{2} - \dots \right) + 1 \\ &\approx \frac{1}{2}(\rho + \rho^2).\end{aligned} \quad (33.51)$$

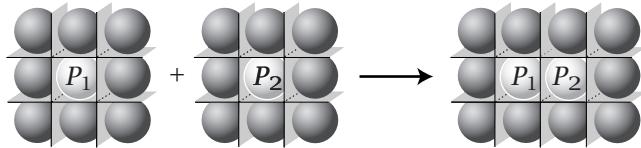


Figure 33.15 The contact free energy Δg describes desolvating two polymer chain segments P_1 and P_2 and bringing them into contact. The interaction parameter χ is defined for the process with the opposite sign (see Figure 15.6).

Substituting Equation (33.51) into Equation (33.50), and keeping only the first-order approximation, gives the solvation free energy as a function of the average segment density:

$$\frac{F_{\text{solvation}}}{RT} = N\rho \left(\frac{1}{2} - \chi \right). \quad (33.52)$$

In Flory theory, a simple approximation relates the density ρ to the size r :

$$\rho = \frac{N}{M} = \frac{Nv}{r^3}, \quad (33.53)$$

where v is the volume per chain segment. This relationship defines the value of M , the number of sites on the lattice that contains the polymer chain. Combining Equation (33.53) with Equations (33.29) and (33.52) gives

$$\frac{F_{\text{elastic}}}{RT} + \frac{F_{\text{solvation}}}{RT} = \beta r^2 - 2 \ln r + \frac{N^2 v}{r^3} \left(\frac{1}{2} - \chi \right) + \text{constant}. \quad (33.54)$$

Taking the derivative of Equation (33.54) and finding the value $r = R$ that causes the derivative to be zero (the most probable value of r ; Equation (33.44)) gives

$$2\beta R - \frac{2}{R} - \frac{3N^2 v}{R^4} \left(\frac{1}{2} - \chi \right) = 0. \quad (33.55)$$

You can express this in terms of $R_0^2 = 2Nb^2/3 = \beta^{-1}$, the most probable radius of the unperturbed chain (see Equation (33.30)). Multiplying both sides of Equation (33.55) by $R^4/(2R_0^3)$ and rearranging gives

$$\begin{aligned} \left(\frac{R}{R_0} \right)^5 - \left(\frac{R}{R_0} \right)^3 &= \left(\frac{3N^2 v}{2R_0^3} \right) \left(\frac{1}{2} - \chi \right) \\ &= \left(\frac{3}{2} \right)^{5/2} \frac{v}{b^3} \left(\frac{1}{2} - \chi \right) \sqrt{N}. \end{aligned} \quad (33.56)$$

Equation (33.56) provides the basis below for explaining how polymers expand or contract in different solvents.

In θ Solvents, Polymers Have Random-Flight Conformations

For solvent and temperature conditions that cause $\chi = 0.5$, the right-hand side of Equation (33.56) equals zero. In that case, multiplying Equation (33.56) by $(R_0/R)^3$ gives the random-flight prediction,

$$\left(\frac{R}{R_0} \right)^2 = 1 \implies R^2 = R_0^2 = \frac{2Nb^2}{3}, \quad (33.57)$$

given by Equation (33.30). That is, when the solvation free energy is zero (the monomer-monomer attraction just balances the excluded volume), the most probable radius of the chain is given simply by the elastic free energy alone.

Good Solvents Expand Polymers

For solvent and temperature conditions that cause $\chi < 0.5$ (called good solvents), the right-hand side of Equation (33.56) is positive. In this regime, excluded volume causes chain expansion. For large N , the fifth-power term is much larger than the third-power term, so

$$\begin{aligned} \left(\frac{R}{R_0}\right)^5 &\approx \left(\frac{3}{2}\right)^{5/2} \frac{\nu}{b^3} \left(\frac{1}{2}-\chi\right) \sqrt{N} \quad \Rightarrow \quad R^5 \propto R_0^5 N^{1/2} \propto N^3 \\ &\Rightarrow \quad R \propto N^{0.6}. \end{aligned} \quad (33.58)$$

In good solvents, the chain radius grows more steeply with chain length ($R \propto N^{0.6}$) than in θ solvents ($R \propto N^{0.5}$). The exponent $\nu = 3/5$ is called the *Flory exponent* for a good solvent. As an example, this law is quite an accurate predictor of the sizes of denatured states of protein molecules over a wide range of protein chain lengths. Below, you see that the same approach gives other Flory exponents for other situations.

EXAMPLE 33.10 The Flory exponent for polymers in good solvents in two dimensions, $R \propto N^{3/4}$. How does a polymer molecule's radius depend on its chain length N if the polymer molecule lies on a two-dimensional surface? Let's modify the Flory argument. Replace the three-dimensional density in Equation (33.53) with the two-dimensional density $\rho = N\nu/r^2$. Following the logic above, the factor of R^4 in the denominator of Equation (33.55) now becomes R^3 . For good solvents, Equation (33.56) becomes

$$\begin{aligned} \left(\frac{R}{R_0}\right)^4 - \left(\frac{R}{R_0}\right)^2 &= \left(\frac{3N^2\nu}{2R_0^2}\right) \left(\frac{1}{2}-\chi\right) \quad \Rightarrow \\ R^4 &= \nu b \left(\frac{1}{2}-\chi\right) N^3. \end{aligned} \quad (33.59)$$

This shows that the Flory argument predicts $R \propto N^{3/4}$ in two dimensions. Figure 33.16 shows experimental data consistent with this prediction, giving $R \propto N^{0.79 \pm 0.04}$ for negatively charged DNA on positively charged surfaces.

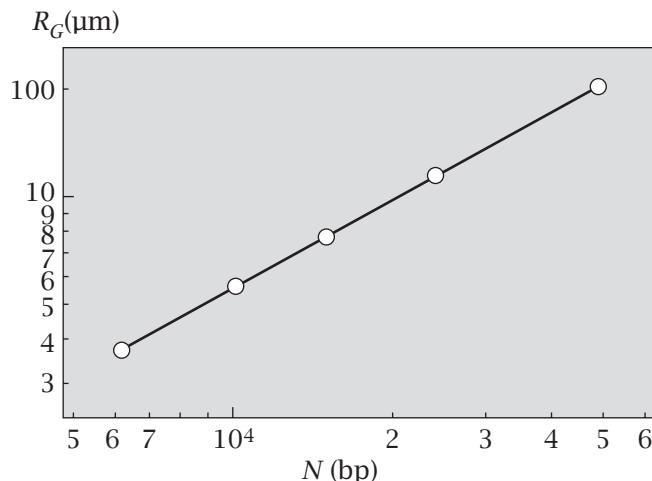
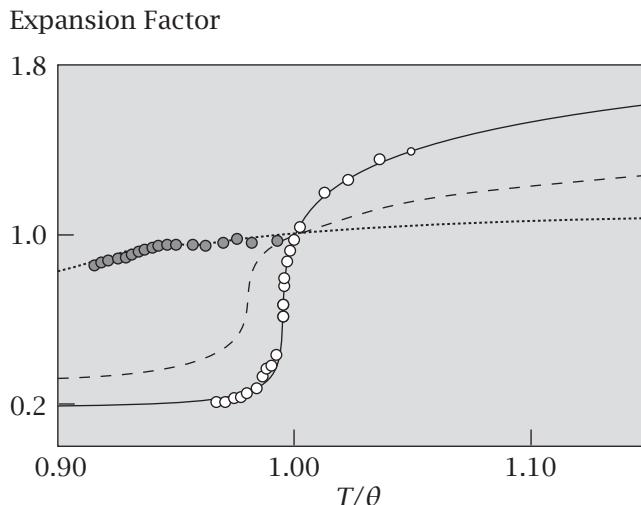


Figure 33.16 Radii of gyration R_G of DNA molecules of different lengths flattened onto positively charged surfaces, observed under microscopes. $R \propto N^{0.79 \pm 0.04}$, consistent with the predicted Flory exponent of $3/4$. Source: B Maier and JO Rädler, *Phys Rev Lett* **82**, 1911–1914 (1999).

Figure 33.17 Homopolymers collapse when the temperature T is less than the temperature θ at which $\chi = 0.5$. Chains expand at higher temperatures ($\chi < 0.5$, good solvents). The transition steepens with increasing chain lengths. (●) molecular weight $M = 2.9 \times 10^3$; (—) $M = 1 \times 10^5$; (○) $M = 2.6 \times 10^7$. Source: ST Sun, I Hishio, G Swislow, and T Tanaka, *J Chem Phys* **73**, 5971–5975 (1980).



Polymers Collapse in Poor Solvents

Solvent and temperature conditions that cause $\chi > 0.5$ are called poor solvents. For a poor solvent, the right-hand side of Equation (33.56) is negative, so $R/R_0 < 1$, and the polymer collapses into a compact conformation. For poor solvent conditions, Equation (33.56) is no longer sufficient, because of the low-density approximation we used to derive it. A collapsed chain has a high segment density. If you were to keep the next higher term in the density expansion for $\ln(1-\rho)$, however, you would find that the model predicts $R \propto N^{1/3}$, as expected for a compact chain.

Polymers can undergo very sharp transitions from the coil to compact states as the solvent and temperature are changed. These are called *coil-to-globule transitions*. The reason they are so sharp is found in Equation (33.56). For large N , the right-hand side of Equation (33.56) can change abruptly from being very positive to being very negative with only a small change in χ at about $\chi = 0.5$. Figure 33.17 shows the collapse process in homopolymers.

Two natural collapse processes are the folding of proteins into their compact native states in water, and the compaction of DNA molecules for insertion into virus heads and cell nuclei. While this homopolymer collapse model illustrates the principle of coil-to-globule transitions, neither protein folding nor DNA collapse follow it exactly, because both polymers also have electrostatic interactions and specific monomer sequences.

Polymer Solutions Exert Osmotic Pressures

Recall from Chapter 16 that if you put large molecules A , such as polymers, that cannot permeate a membrane, on the left side of a membrane, and if you put a small-molecule solvent B that can freely permeate the membrane on the right side, then B will be drawn to the left into the polymer solution.

Figure 33.18 shows a variation of that arrangement. A polymer gel is a network of polymer chains that can also contain solvent molecules and can become swollen by drawing solvent into itself, even without the need for a

membrane. Think of a sponge taking up water. Let's compute the osmotic pressure that draws a small-molecule solvent B into a polymer gel. Suppose the polymer has chain length $N = N_A$ and the small molecule has $N_B = 1$. Suppose the volume fraction of polymer is $\phi = \phi_A = 1 - \phi_B$. The chemical potential difference that drives solvent from its pure liquid into the polymer will be $\Delta\mu_{\text{mix}} = \mu_B(\phi_B) - \mu_B(1) = -\pi v_B$, where π is the osmotic pressure and v_B is the partial molar volume of the solvent; see Equation (16.38). Substitute the Flory-Huggins chemical potential from Equation (32.21) to get the osmotic pressure:

$$\pi = -\frac{RT}{v_B} \left[\ln(1-\phi) + \phi \left(1 - \frac{1}{N} \right) + \chi_{AB} \phi^2 \right]. \quad (33.60)$$

EXAMPLE 33.11 Osmotic pressures of dilute polymer solutions. For polymers that are at low concentration ϕ in solution, substitute the approximation $\ln(1-\phi) \approx -\phi - \phi^2/2 - \dots$ into Equation (33.60) to get the osmotic pressure of the solvent:

$$\pi = \frac{RT}{v_B} \left[\frac{\phi}{N} + \left(\frac{1}{2} - \chi_{AB} \right) \phi^2 \right]. \quad (33.61)$$

For long chains ($N \rightarrow \infty$), the first term above (ϕ/N) is zero. Equation (33.61) makes two main predictions. First, it says that the osmotic pressure will increase with the square of the concentration ϕ of the polymer. Figure 33.19 confirms this prediction. Second, Equation (33.60) predicts that in θ solvents, for which $\chi_{AB} = 1/2$, polymer solutions exert no osmotic pressure, $\pi = 0$.

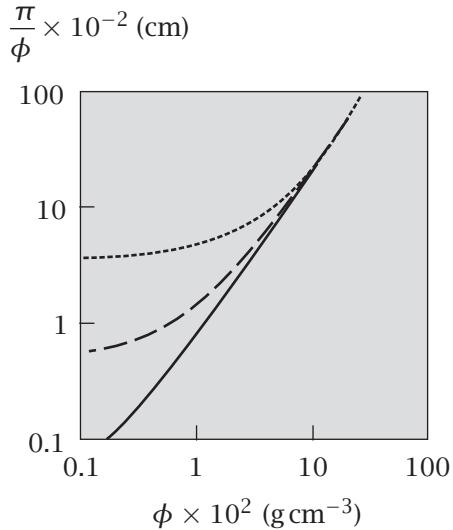


Figure 33.19 The osmotic pressure $\pi(\phi)$ as a function of polymer concentration ϕ . The experiments confirm the square dependence $\pi \propto \phi^2$ that is predicted by Equation (33.61). Source: I Noda, N Kato, T Kitano, and M Nagasawa, *Macromolecules* **14**, 668–676 (1981).

EXAMPLE 33.12 Polymer gels undergo abrupt swelling transitions. Polymer gels can often collapse or expand as abrupt functions of temperature or solvents. Such gels have applications as molecular pumps for drug delivery or for controlling insulin levels in the bodies of patients with diabetes, for

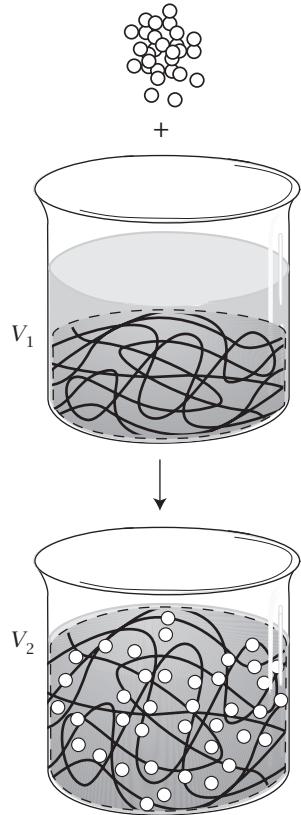


Figure 33.18 A polymer gel is a network of chains. Solvent is drawn into the polymer network region by the osmotic pressure, causing the gel to swell. This tendency toward swelling is counterbalanced by the elastic stretching of the gel, which tends to expel the solvent from the gel region. These forces balance at equilibrium, leading to a stable state of gel size. The dashed line indicates the location of the gel surface.

example. What causes a polymer gel to undergo sharp transitions in swelling or contraction? Notice that if $\chi_{AB} \approx 1/2$, Equation (33.61) says that small changes in solvent or temperature will switch the sign of the osmotic pressure. The equilibrium volume of a polymer gel is determined by a balance of two contributions to the solvent chemical potential (or correspondingly to the osmotic pressure) (see Equation (16.38)):

$$-\pi v_B = \Delta\mu_{\text{mix}} + \Delta\mu_{\text{elas}}. \quad (33.62)$$

The term $\Delta\mu_{\text{mix}}$ accounts for the tendency of solvent molecules to be drawn into the polymer to dilute out the polymer. $\Delta\mu_{\text{elas}}$ accounts for the counterbalancing effect: a gel that takes up solvent swells, stretching the chains, leading to an elastic retraction opposing further solvent uptake by the gel. Suppose the gel expands by a factor α :

$$\alpha = \frac{V}{V_0} = \frac{V_0 + n_s v_s}{V_0} = \frac{\phi_0}{\phi}, \quad (33.63)$$

where V_0 is the initial volume of the gel and V is the final volume of the gel, which includes the uptake of n_s solvent molecules having molar volume v_s . The inverse proportionality $V/V_0 = \phi_0/\phi$ is because ϕ is the polymer chain density, the number of chain segments per unit volume. The elastic free energy is given by Equation (33.43):

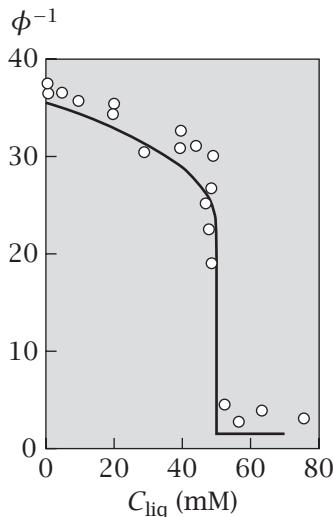


Figure 33.20 A swollen gel of poly(*N*-isopropylacrylamide) polymer undergoes a collapse transition as a sharp function of the concentration of aromatic acids in aqueous solvent. Source: K László, K Kosik, C Rochas, and E Geissler, *Macromolecules* **36**, 7771–7776 (2003).

$$\Delta F_{\text{elas}} = \frac{3mRT}{2}(\alpha^{2/3} - 1), \quad (33.64)$$

where $\lambda^3 = \alpha$ represents a uniform volume expansion of the gel in three dimensions. To get the elastic component of the chemical potential, take the derivative:

$$\Delta\mu_{\text{elas}} = \frac{\partial\Delta F_{\text{elas}}}{\partial n_s} = \frac{\partial\Delta F_{\text{elas}}}{\partial\alpha} \frac{\partial\alpha}{\partial n_s} = mRT \left(\frac{\phi}{\phi_0} \right)^{1/3} \left(\frac{v_s}{V_0} \right). \quad (33.65)$$

Sum the elastic component, Equation (33.65), with the mixing component, Equation (33.60), as indicated by Equation (33.62). At equilibrium, the net osmotic pressure is $\pi = 0$. So, to determine the dependence of gel size on different solution conditions, solve the following equation for $\phi_p(\chi_{AB})$ as a function of different values of χ_{AB} :

$$\left[\ln(1 - \phi_p) + \phi_p \left(1 - \frac{1}{N} \right) + \chi_{AB} \phi_p^2 \right] + B \phi_p^{1/3} = 0. \quad (33.66)$$

Here, $B = (mv_s/V_0)\phi_0^{-1/3}$, but experimentalists often take B as an adjustable parameter.

Figure 33.20 shows a polymer gel that undergoes a sharp 35-fold volume contraction upon increasing the concentration of weak aromatic acids, like phenol, which control the value of χ_{AB} .

The radius of a polymer chain depends not only on the temperature and the solvent character, as described earlier; the radius also depends on the polymer concentration.

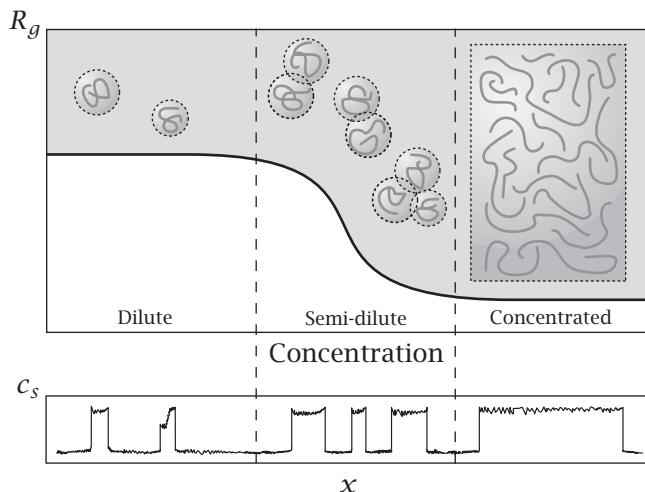


Figure 33.21 The three concentration regimes of polymers: dilute, semi-dilute, and concentrated. In a good solvent, dilute polymer chains are expanded: $R_g \sim N^{3/5}$ because chain segments contact the solvent. Concentrated polymer chains are less expanded: $R_g \sim N^{1/2}$ because chain segments contact other chain segments—both from within a chain and with other chains—so the system resembles a θ solvent. At semi-dilute concentrations, the radius R_g of a polymer chain fluctuates between large and small.

Polymer Radius Depends on Polymer Concentration: the Semi-Dilute Cross-Over Regime

Figure 33.21 shows three different polymer concentrations: dilute, concentrated, and a region between them called *semi-dilute*. In dilute solutions, the density of chain monomers in space is spotty, in the same way that people occupy countries: there are some high-density regions of chain monomers (where a polymer molecule is located—like cities, in this metaphor), and elsewhere there are zero-density regions (where you find only solvent molecules—like uninhabited countryside). In dilute polymer solutions, the spatial density fluctuations are large. In contrast, in a concentrated polymer solution (called the *melt*), chain monomers are distributed densely and uniformly throughout space. In the melt, the spatial density fluctuations are small. The intermediate density, where the chains just begin to overlap and bump into each other on average, is called the *semi-dilute* or *cross-over* regime. The semi-dilute regime occurs around the *overlap concentration* c^* , which is where (the number of polymer molecules per unit volume) \times (the volume per polymer molecule) equals one. A polymer solution has different properties above and below its overlap concentration.

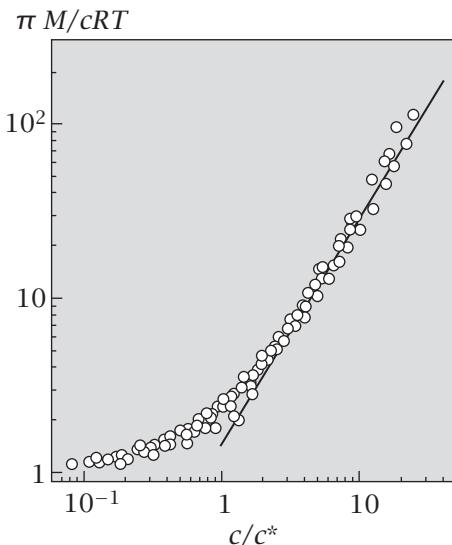
EXAMPLE 33.13 Longer polymer chains have a lower overlap concentration c^* . How does the overlap concentration c^* depend on the length of a polymer chain? The longer the polymer chain, the smaller is the concentration required for chains to overlap. To see this, suppose the chain density is c monomers per unit volume and the chain length is N monomers per chain. So, c/N is the number of chain molecules per unit volume. Compute c^* using

$$\left(\frac{c^*}{N}\right) R_g^3 = 1. \quad (33.67)$$

Now, substitute the relationship

$$R_g \sim N^\nu \quad (33.68)$$

Figure 33.22 The osmotic pressure π of a polymer solution is a steeper function of concentration c above the overlap concentration c^* than below it (M is molecular weight). Source: WW Graessley, *Polymeric Liquids and Networks: Structure and Properties*, Garland Science, New York, 2004, p. 279; I Noda, N Kato, T Kitano, and M Nagasawa, *Macromolecules* **14**, 668–676 (1981).



into Equation (33.67) to get

$$c^* = \frac{N}{R_g^3} \sim \frac{N}{N^{3\nu}} = N^{1-3\nu} = N^{-4/5}, \quad (33.69)$$

for $\nu = 3/5$ (good-solvent conditions). This shows that longer polymer chains have smaller overlap concentrations. As a point of reference, polystyrene of $N = 10^6$ has an overlap concentration of 0.5%.

A polymer solution above its overlap concentration c^* can behave differently than below it. Figure 33.22 shows that the osmotic pressure dependence on polymer concentration changes at $c = c^*$. Also, the size R_g of an individual chain molecule can be different above the overlap concentration than below it; see Figure 33.21. In a dilute solution in a good solvent, polymer chains are expanded: $R_g \sim N^{3/5}$. In the melt, a polymer chain is less expanded, $R_g \sim N^{1/2}$. Why? In 1949, PJ Flory, who won the 1974 Nobel Prize in Chemistry for this and related work, predicted that a polymer melt, despite its apparent complexity, would act like a θ solvent for any individual polymer chain inside it. The polymer melt should have this special simplicity because of *excluded-volume screening*. A given polymer molecule ‘sees’ only a sea of identical monomers, and cannot tell whether those monomers come from within its own chain or from neighboring chains. So, each chain molecule has no particular preference to expand or contract, which is the same way the chain would behave in a θ solvent. This remarkable simplicity was confirmed in the early 1970s by neutron scattering experiments [10]. In short, R_g is large in dilute polymer solutions in good solvents; R_g is small in concentrated solutions; and R_g fluctuates in the semi-dilute regime. That is, near $c = c^*$, small changes of polymer concentration can lead to large changes in R_g .

Summary

Polymers have many conformations of nearly equal energy, so they have broad distributions of conformations. Stretching or squeezing or otherwise perturbing polymers away from their equilibrium conformations leads to entropic forces that oppose the perturbations. This is the basis for rubber elasticity. One of the most important models is the random-flight theory, in which the distribution of chain end distances is Gaussian. The root-mean-square distance between the two ends of the chain increases as the square root of the chain length N : $\langle r^2 \rangle^{1/2} = (Nb^2)^{1/2}$. The random-flight theory applies when the chain is in a θ solvent. In that case, the steric tendency to expand is just balanced by the self-attraction energy causing the chain to contract. When the solvent is poor, the self-attractions between the chain monomers dominate and chains collapse to compact configurations. When the solvent is good, the self-attractions are weak and chains expand more than would be predicted by the random-flight theory.

Problems

1. Stretching a rubber sheet. What is the free energy for stretching an elastomeric material uniformly along the x and y directions, at constant volume?

2. Stretching DNA. Figure 33.6 shows that it takes about 0.1 pN of force to stretch a DNA molecule to an extension of 20 μm . Use the chain elasticity theory to estimate the undeformed size of the molecule, $\langle r^2 \rangle^{1/2}$.

3. Stretched polymers have negative thermal expansion coefficients. The thermal expansion coefficient of a material is $\alpha = (1/V)(\partial V / \partial T)_p$. Consider the corresponding one-dimensional quantity $\alpha_p = (1/x)(\partial x / \partial T)_f$ for a single polymer molecule stretched to an end-to-end length x by a stretching force f .

- (a) Compute α_p for the polymer chain.
- (b) What are the similarities and differences between a polymer and an ideal gas?

4. An ‘ideal’ solvent expands a polymer chain. If a polymer chain is composed of the same monomer units as the solvent around it, the system will be ideal in the sense that the polymer-polymer interactions will be identical to polymer-solvent interactions, so $\chi = 0$.

- (a) Write an expression for the most probable radius R for a chain in an ideal solvent.
- (b) Show that such a chain is expanded relative to a random-flight chain.
- (c) Describe the difference between an ideal solvent and a θ solvent.

5. Computing conformational averages. Using the expression for the distribution $P(r, N)$ for the end-to-end separation of a polymer chain of length N , compute $\langle r^2 \rangle$ and $\langle r^4 \rangle$.

6. Contour length of DNA. The double-stranded DNA from bacteriophage λ has a contour length $L = 17 \times 10^{-6} \text{ m}$. Each base pair has bond length $b = 3.5 \text{ \AA}$.

- (a) Compute the number of base pairs in the molecule.
- (b) Compute the molecular weight of the DNA.

7. Using elasticity to compute chain concentrations. Figure 33.14 shows the stress-strain properties of a rubber band. Using the figure and chain elasticity theory:

- (a) Estimate the number of polymer chains in a cubic volume 100 \AA on each side.
- (b) If each monomer occupies 100 \AA^3 , what is the length of each chain between junction points?

8. The polymer overlap concentration c^* . Consider a solution of polymer molecules of length N . Compute $c^*(N)$, i.e., the dependence of c^* on the chain length N , for random-flight chains (i.e., for polymers in a θ solvent).

References

- [1] PJ Flory, *Statistical Mechanics of Chain Molecules*, Wiley, New York, 1969.
- [2] R Phillips, J Kondev, and J Theriot, *Physical Biology of the Cell*, Garland Science, New York, 2009.
- [3] A Kloczkowski and A Kolinski, *Physical Properties of Polymers Handbook*, 2nd edition, JE Mark, ed., Springer, New York, 2007, Chapter 5.
- [4] WL Mattice and UW Suter, *Conformational Theory of Large Molecules: The Rotational Isomeric State Model in Macromolecular Systems*, Wiley, New York, 1994.
- [5] H Jacobson and WH Stockmayer, *J Chem Phys* **18**, 1600–1606 (1950).
- [6] B Erman and JE Mark, *Annu Rev Phys Chem* **40**, 351–374 (1989).
- [7] JE Mark and B Erman, *Rubberlike Elasticity: A Molecular Primer*, 2nd edition, Cambridge University Press, Cambridge, 2007.
- [8] PJ Flory, *Principles of Polymer Chemistry*, Cornell University Press, Ithaca, NY, 1953.
- [9] HS Chan and KA Dill, *Annu Rev Biophys Biophys Chem* **20**, 447–490 (1991).
- [10] JP Cotton, D Decker, H Benoit, et al., *Macromolecules* **7**, 863–872, (1974).

Suggested Reading

The classic texts on rotational isomeric state and random-flight models:

CR Cantor and PR Schimmel, *Biophysical Chemistry*, Part III: *The Behavior of Biological Macromolecules*, WH Freeman, San Francisco, 1980.

M Doi, *Introduction to Polymer Physics*, Oxford University Press, Oxford, 1996.

PJ Flory, *Principles of Polymer Chemistry*, Cornell University Press, Ithaca, NY, 1953.

WL Mattice and UW Suter, *Conformational Theory of Large Molecules: The Rotational Isomeric State Model in Macromolecular Systems*, Wiley, New York, 1994.

Excellent summaries of rubber elasticity:

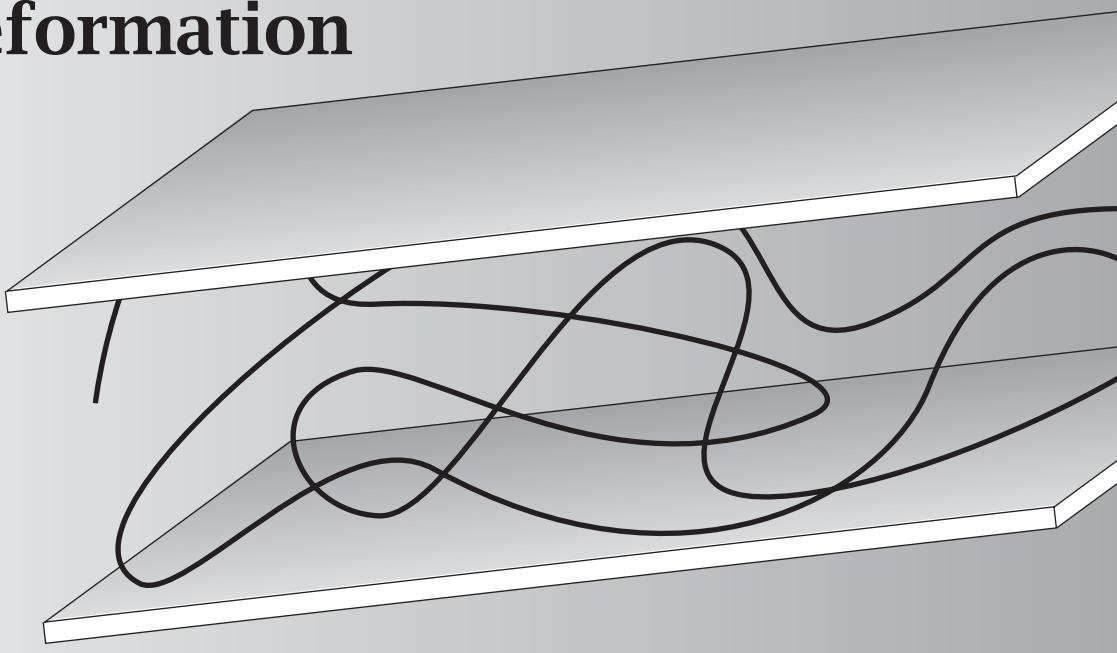
B Erman and JE Mark, *Annu Rev Phys Chem* **40**, 351–374 (1989).

JE Mark and B Erman, *Rubberlike Elasticity: A Molecular Primer*, 2nd edition, Cambridge University Press, Cambridge, 2007.

One of the first models of a polymer collapse, applied to DNA:

CB Post and BH Zimm, *Biopolymers* **18**, 1487–1501 (1979).

34 Polymers Resist Confinement & Deformation



Stretch a polymer molecule, then release it. The chain will retract. So, polymeric fluids are often *viscoelastic*, having not only fluid-like viscosity, but also rubber-like elasticity. And, if you force a polymer into a confined space, it will resist the confinement. For example, polymers will not enter into small spaces in chromatography columns and will not approach inert surfaces very closely. Polymers resist these types of deformations because chain molecules have conformational ensembles, leading to chain entropies, leading to entropic opposing forces. How do confinement and deformation affect polymers? We first consider excluded volume.

'Excluded Volume' Describes the Large Volume Inside a Polymer Conformation that Is Inaccessible to Other Chains

Two chain molecules in dilute solution rarely interpenetrate each other, even if each chain is highly expanded and solvated. They repel because of *excluded volume* (see Figure 34.1). If the volume within random-flight chain A is mostly 'empty' inside, why won't another chain B enter into that space? Here's a metaphor. A blind man walks in a forest. The trees are located randomly, but at low density. For any one random step he takes, it is rare that he will bump into a tree. But for any random walk of a hundred or a thousand steps, it becomes highly likely that in *at least one step*, he will bump into a tree. So, in a random

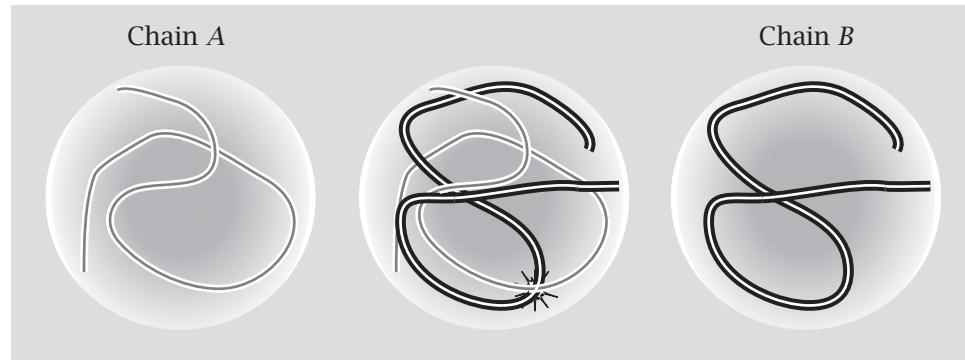


Figure 34.1 Two random-flight polymers (left and right) do not interpenetrate (center) because of *excluded volume*, even though most of the internal volume is solvent. This leads to a repulsion.

walk of many steps, it becomes very rare that he will avoid bumping into at least one tree, even if the density of obstacles (trees) is very small. For the same reason, one random-flight polymer chain is not likely to penetrate the space inside another polymer chain: one chain would have very few viable conformations inside the other chain, compared with the number it would have had free in solution. Excluded volume is also a repulsive force that pushes two noninteracting random-flight polymer molecules apart. This repulsion is the basis for the *steric stabilization* of colloids. If two compact colloidal particles are ‘fuzzy’ with surfaces covered by random-flight polymer chains, the particles will repel each other to avoid intermingling their polymer chains. The attached chains prevent particle aggregation and *stabilize* the colloidal solution. Some inks are colloidal suspensions that are stabilized in this way. Also, sugar chains that are attached to proteins and membranes may prevent their aggregation.

Random-flight polymers will also tend to repel rigid impenetrable surfaces, in the absence of attractive forces.

Chain Conformations Are Perturbed Near Surfaces

Consider a polymer chain near an impenetrable surface. As the chain’s center of mass approaches the surface, some of the chain’s extended conformations are not physically viable because those conformations would pass through the surface (see Figure 34.2). This is the basis for the entropic repulsion. Let z

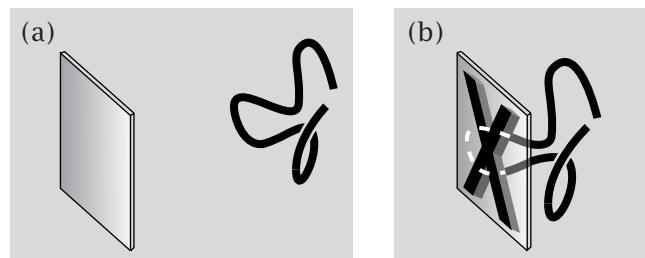


Figure 34.2 (a) Far away from a surface, chain conformations are unperturbed. (b) Close to a surface, a chain loses conformations (those that would otherwise have passed through the surface), leading to an increased free energy, and repulsion from the surface.

represent the coordinate normal to the plane of the surface. Because the conformational entropy decreases as the chain approaches the surface, the conformational free energy F increases, and this can be described as a repulsive force $f = -dF/dz$ between the surface and the chain.

There are various ways to model the distributions of chain conformations near surfaces. Here we describe the *reflectance principle*. It resembles the image charge approach in the theory of electrostatic interactions (see Chapter 21).

The DiMarzio-McCrackin Reflectance Principle Predicts Chain Distributions Near Surfaces

To describe how the distribution of chain conformations is perturbed near an impenetrable surface, we need a way to count the conformations that do not cross the plane. Consider an impenetrable plane that extends in the x and y directions, located at $z = 0$ (see Figure 34.3). The chain is on the $z > 0$ side of the plane. The beginning of the chain is at $(0, 0, z_A)$. What is the fraction $w_{\text{viable}}(z_A, z)$ of conformations that originate at z_A , end at z , and do not cross the plane? (For random flights, the x , y , and z distributions are independent of each other, so we can include multiplicative factors for the x and y dimensions later.)

If there were no surface at all, the distribution of all conformations of a random-flight chain would be given by the Gaussian function, Equation (33.22):

$$\begin{aligned} w_{\text{total}}(z_A, z) &= \left(\frac{3}{2\pi Nb^2}\right)^{1/2} \exp\left[-\frac{3(z-z_A)^2}{2Nb^2}\right] \\ &= \left(\frac{\beta}{\pi}\right)^{1/2} \exp[-\beta(z-z_A)^2], \end{aligned} \quad (34.1)$$

where $\beta = 3/(2Nb^2)$ simplifies the notation.

But the impenetrable plane eliminates some of those conformations. We want to count only the conformations that do not pass through the plane. The total number of conformations is the sum of *viable* conformations (those that don't cross the plane) and *nonviable* conformations (those that cross the plane):

$$w_{\text{total}} = w_{\text{viable}} + w_{\text{nonviable}}. \quad (34.2)$$

Equation (34.2) says that you can get the quantity you want, w_{viable} , if you know $w_{\text{nonviable}}$, because you have w_{total} from Equation (34.1).

The *reflectance principle*, developed by polymer physicists EA DiMarzio and FL McCrackin [1], gives a simple way to compute $w_{\text{nonviable}}$, the fraction of conformations that begin at A , end at B , and cross the plane at least once (see Figure 34.4). You make a mirror image of the first part of the chain conformation, from A to its first crossing point C on the surface, by reflecting the chain origin from $+z_A$ to $-z_A$ through the surface. Now count the number of conformations of these image chains ($A'CB$) that begin at $-z_A$ and end at z , and call this quantity $w_{\text{image}}(-z_A, z)$. The reflectance principle says that

$$w_{\text{nonviable}}(z_A, z) = w_{\text{image}}(-z_A, z). \quad (34.3)$$

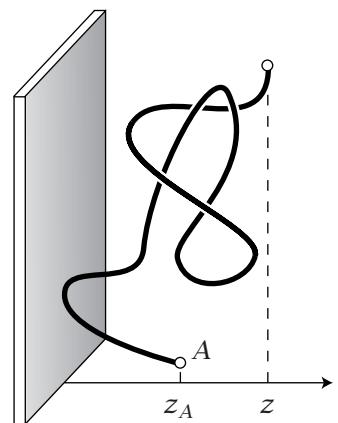
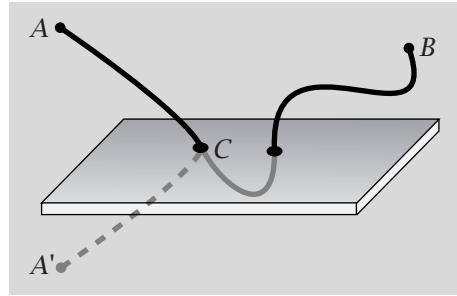


Figure 34.3 A polymer chain begins at z_A , ends at z , and does not pass through the plane.

Figure 34.4 A model for computing the loss of entropy that causes polymer repulsions from surfaces. Conformation ACB is not viable because it passes through the plane. It starts at z_A and ends at z_B . To count all the nonviable conformations, create an image of each one, $A'CB$, starting at $-z_A$ and ending at z_B .



Because $w_{\text{image}}(-z_A, z)$ is simple to compute, this expression means that $w_{\text{nonviable}}(z_A, z)$ is also simple to compute. $w_{\text{image}}(-z_A, z)$ is just the count of all the Gaussian random flights that begin at $-z_A$ and end at z . Use Equation (34.1) with z_A replaced by $-z_A$:

$$w_{\text{image}}(-z_A, z) = \left(\frac{\beta}{\pi}\right)^{1/2} \exp[-\beta(z+z_A)^2]. \quad (34.4)$$

The quantity $w_{\text{image}}(-z_A, z)$ has two remarkable properties. (1) For every chain from A to B that crosses the plane at least once, its mirror image is counted exactly once in $w_{\text{image}}(-z_A, z)$. This is because from the crossing point C to B , the chain and its image have exactly the same trajectory, and from A to C the reflected conformation is the perfect image of the crossing conformation. (2) $w_{\text{image}}(-z_A, z)$ does not count any extraneous (i.e., noncrossing) conformations, because any chain from A' to B must cross the plane.

The reflectance principle gives you a simple way to compute the fraction of all chain conformations that cross the plane. Substituting Equations (34.1), (34.3), and (34.4) into Equation (34.2) gives

$$\begin{aligned} w_{\text{viable}}(z_A, z) &= w_{\text{total}}(z_A, z) - w_{\text{image}}(-z_A, z) \\ &= \left(\frac{\beta}{\pi}\right)^{1/2} \left\{ \exp[-\beta(z-z_A)^2] - \exp[-\beta(z+z_A)^2] \right\}. \end{aligned} \quad (34.5)$$

This expression can be simplified further for chains that are tethered to the surface, where z_A is small enough that it represents only a single bond to the surface, $z_A = (b/2) \ll Nb^2$. Then the difference in exponential functions in Equation (34.5) can be approximated as a derivative:

$$\begin{aligned} &\exp\{-\beta[z-(b/2)]^2\} - \exp\{-\beta[z+(b/2)]^2\} \\ &= f(z-(b/2)) - f(z+(b/2)) = -b \frac{df}{dz} \approx 2\beta bz \exp(-\beta z^2), \end{aligned} \quad (34.6)$$

where the function $f(x) = \exp(-\beta x^2)$. Substituting Equation (34.6) into Equation (34.5) gives the distribution of the ends of chains that are tethered to an impenetrable surface:

$$w_{\text{viable, tethered}}(z) = \frac{3z}{Nb} \left(\frac{\beta}{\pi}\right)^{3/2} \exp(-\beta z^2). \quad (34.7)$$

The full three-dimensional distribution function for chains that are tethered to an impenetrable surface and that have their free ends at (x, y, z) can

be obtained by multiplying by the distributions for the independent x and y components:

$$w_{\text{viable}}(x, y, z) = \frac{3z}{Nb} \left(\frac{\beta}{\pi}\right)^{3/2} \exp[-\beta(x^2 + y^2 + z^2)]. \quad (34.8)$$

Tethered chains behave very differently than free chains: compare the front factor of z in Equation (34.8) with the front factor in the free-chain equations (33.22) and (33.25). Free chains end approximately where they begin: $\langle z \rangle = 0$. In contrast, the termini of tethered chains tend to be ‘repelled’ from the surface (see Figure 34.5): $\langle z \rangle \neq 0$, as Example 34.1 shows.

EXAMPLE 34.1 The thickness of a layer of tethered polymer chains. You can compute the average thickness $\langle z \rangle$ of a tethered polymer layer by using the distribution function, Equation (34.7):

$$\begin{aligned} \langle z \rangle &= \frac{\int_0^\infty z w_{\text{viable}}(z) dz}{\int_0^\infty w_{\text{viable}}(z) dz} = \frac{\int_0^\infty z^2 \exp(-\beta z^2) dz}{\int_0^\infty z \exp(-\beta z^2) dz} \\ &= \left[\frac{1}{4\beta} \left(\frac{\pi}{\beta} \right)^{1/2} \right] \left(\frac{1}{2\beta} \right)^{-1} = \frac{1}{2} \left(\frac{\pi}{\beta} \right)^{1/2} = b \left(\frac{\pi}{6} \right)^{1/2} N^{1/2}. \end{aligned} \quad (34.9)$$

The mean thickness of the tethered polymer layer increases with the square root of the chain length (see Figure 34.5).

Excluded volume also has a role when polymers are confined within the pores of chromatography columns, when oil is trapped underground in the microscopic cavities in rocks, or when biopolymers are confined within small spaces. The next section describes a general way to compute chain conformations that are subject to a variety of different constraints.

Polymer Conformations Can Be Described by the Diffusion Equation

A powerful insight into polymer chain conformations is an analogy with the diffusion equation. When polymers are subject to complex geometric constraints, the conformational free energy can sometimes be found as a solution of a diffusion equation.

Imagine growing a chain on a simple cubic lattice, one bond at a time. The chain has its first monomer at $(0, 0, 0)$. The distance between lattice site centers is b , the bond length. Let $p(N+1, x, y, z)$ represent the probability that the chain has its monomer $N+1$ at lattice site (x, y, z) . The probability distribution for monomer $N+1$ is related to the probability that the chain has monomer N at a neighboring site:

$$\begin{aligned} p(N+1, x, y, z) &= \frac{1}{6} [p(N, x-b, y, z) + p(N, x+b, y, z) \\ &\quad + p(N, x, y-b, z) + p(N, x, y+b, z) \\ &\quad + p(N, x, y, z-b) + p(N, x, y, z+b)], \end{aligned} \quad (34.10)$$

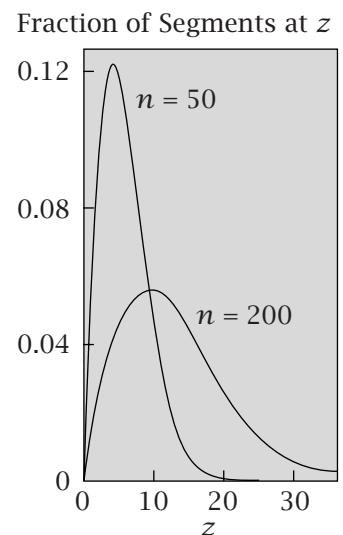


Figure 34.5 Computed distribution of the free ends of chains that are tethered by one end to a surface. The free ends tend to be distributed away from the surface. When the chain length increases fourfold, the peak (most probable location of the termini) increases twofold, supporting the predicted $N^{1/2}$ relationship. Source: EA DiMarzio, Chapter 4, in IC Sanchez, *Physics of Polymer Surfaces and Interfaces*, Butterworth-Heinemann, Stoneham, 1992.

where the factor of 1/6 represents the probability that the chain steps in the one right direction (out of six, on a cubic lattice) from the site at which monomer N is located.

Now express the difference equation (34.10) as a differential equation. You have two kinds of partial derivatives here: one is with respect to N , for the addition of each monomer, and the other is with respect to the spatial coordinates x , y , and z . For the first type of derivative, the Taylor series gives

$$\begin{aligned} p(N+1, x, y, z) &= p(N, x, y, z) + \left(\frac{\partial p}{\partial N} \right)_{x,y,z} \Delta N + \dots \\ &\approx p(N, x, y, z) + \left(\frac{\partial p}{\partial N} \right)_{x,y,z}, \end{aligned} \quad (34.11)$$

since $\Delta N = 1$. For the second type of derivative, the two x -coordinate Taylor-series terms are

$$p(N, x-b, y, z) = p(N, x, y, z) - \left(\frac{\partial p}{\partial x} \right) b + \frac{1}{2} \left(\frac{\partial^2 p}{\partial x^2} \right) b^2 + \dots \quad (34.12)$$

and

$$p(N, x+b, y, z) = p(N, x, y, z) + \left(\frac{\partial p}{\partial x} \right) b + \frac{1}{2} \left(\frac{\partial^2 p}{\partial x^2} \right) b^2 + \dots \quad (34.13)$$

Adding Equations (34.12) and (34.13) gives

$$p(N, x-b, y, z) + p(N, x+b, y, z) = 2p(N, x, y, z) + \left(\frac{\partial^2 p}{\partial x^2} \right) b^2 + \dots \quad (34.14)$$

To express Equation (34.14) in terms of a single differential equation, compute the corresponding terms in y and z , and insert all six terms into the right-hand side of Equation (34.10). Then insert Equation (34.11) into the left-hand side of Equation (34.10) to get

$$\begin{aligned} \left(\frac{\partial p}{\partial N} \right) &= \frac{b^2}{6} \left(\frac{\partial^2 p}{\partial x^2} + \frac{\partial^2 p}{\partial y^2} + \frac{\partial^2 p}{\partial z^2} \right) \\ &= \frac{b^2}{6} \nabla^2 p. \end{aligned} \quad (34.15)$$

Equation (34.15) is identical to the diffusion equation (17.13), but the polymer chain length N replaces the diffusion time t , and the site probability p for monomer N replaces the concentration c of diffusing particles at time t . Example 17.5 shows that the solution to the diffusion equation from a point source is a Gaussian distribution function. Similarly, the Gaussian function in Equation (33.26) is a solution to Equation (34.15) for unconstrained chains that begin at the origin $(0, 0, 0)$. For the diffusion equation, the width of the concentration distribution is given by $\langle x^2 \rangle = 2Dt$, while for polymers, the width in one dimension is given by $\langle x^2 \rangle = Nb^2/3$. For diffusion from a point source at $(0, 0, 0)$, the width of the spatial distribution of the particles increases with the square root of the time that the particles diffuse. For polymers that begin at $(0, 0, 0)$, the width of the spatial distribution of the chain ends increases with the square root of the chain length.

A polymer chain tends to avoid partitioning into small confined spaces, because doing so would diminish its conformational entropy and increase its free energy. For example, chain molecules tend not to enter pores in chromatography columns, even when the pores are several-fold bigger than the radius of gyration of the chain. You can compute this confinement free energy from the diffusion-equation analogy.

Polymers Tend to Avoid Confined Spaces

Consider a chain molecule that is confined within a box of dimensions $\ell_x = \ell$, ℓ_y , and ℓ_z (see Figure 34.6). To count the conformations that are fully contained within the box, you can solve Equation (34.15), subject to the boundary conditions. We won't give the details here, just the main results. At impenetrable walls, the concentration of chain ends is zero. This gives six boundary conditions for a box: $p(N, x = 0) = 0$, $p(N, x = \ell_x) = 0$, $p(N, y = 0) = 0$, $p(N, y = \ell_y) = 0$, $p(N, z = 0) = 0$, and $p(N, z = \ell_z) = 0$. These are called *absorbing conditions*. Solving Equation (34.15) subject to absorbing boundary conditions [2-4] gives $w(x_1, x_2, N)$, the x -axis distribution of chains that begin at point x_1 and end at x_2 after N steps:

$$w(x_1, x_2, N) = \frac{2}{\ell} \sum_{j=1}^{\infty} \sin\left(\frac{j\pi x_1}{\ell}\right) \sin\left(\frac{j\pi x_2}{\ell}\right) e^{-(j^2\pi^2Nb^2)/(6\ell^2)}. \quad (34.16)$$

The same distribution function applies to y and z , since all chain steps are independent in the random-flight model.

To get the full x -direction partition function q_x , you must integrate the distribution over all possible beginnings and endings of the chain:

$$q_x = \int_0^\ell dx_1 \int_0^\ell dx_2 w(x_1, x_2, N). \quad (34.17)$$

For each sine function, integration gives

$$\int_0^\ell \sin\left(\frac{j\pi x}{\ell}\right) dx = \begin{cases} \frac{2\ell}{j\pi} & \text{for } j = 1, 3, 5, \dots, \\ 0 & \text{for } j = 2, 4, 6, \dots. \end{cases} \quad (34.18)$$

Therefore

$$q_x = \frac{8\ell}{\pi^2} \sum_{j=1,3,5,\dots}^{\infty} \left(\frac{1}{j^2}\right) e^{-(\pi^2Nb^2j^2)/(6\ell^2)},$$

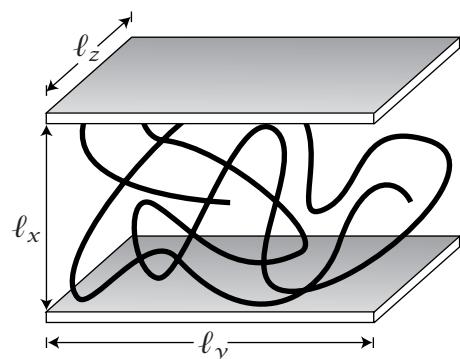


Figure 34.6 A polymer is confined in a box of dimensions ℓ_x , ℓ_y , and ℓ_z . If the box dimensions are small relative to the chain radius, there is a cost in confinement entropy owing to the loss of conformations that the chain would have had if it were in a larger space.

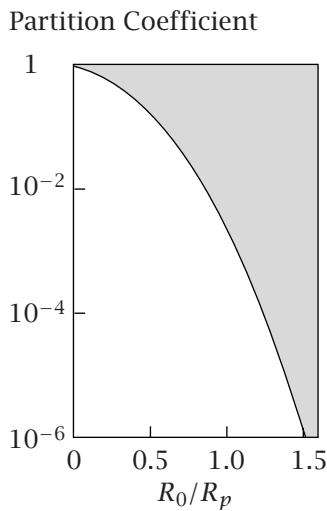


Figure 34.7 The partitioning of a chain molecule decreases as the size of the box diminishes. R_0 is the radius of gyration and R_p is the pore radius. Source: MG Davidson, UW Suter, and WM Deen, *Macromolecules* **20**, 1141–1146 (1987).

which is a sum of Gaussian functions. When the box is smaller than the polymer, $\ell^2 \ll Nb^2$, the first term ($j = 1$) dominates this sum and

$$q_x = \frac{8\ell}{\pi^2} e^{-(\pi^2 Nb^2)/(6\ell^2)}. \quad (34.19)$$

If the box is small in all directions, then q_y and q_z are identical to Equation (34.19) but with ℓ_y and ℓ_z replacing $\ell = \ell_x$.

Consider a simple case: a chain confined between infinite plates, where $\ell_x = \ell$ is small but $\ell_y, \ell_z \gg Nb^2$. Then $q_y \approx \ell_y$ and $q_z \approx \ell_z$. In this case, the full partition function is

$$\begin{aligned} q &= q_x q_y q_z \\ &= \frac{8V}{\pi^2} e^{-(\pi^2 Nb^2)/(6\ell^2)}, \end{aligned} \quad (34.20)$$

since $V = \ell_x \ell_y \ell_z$ is the volume of the box. The confinement free energy is

$$\frac{F}{kT} = -\frac{S}{k} = -\ln q = -\ln\left(\frac{8V}{\pi^2}\right) + \frac{\pi^2 Nb^2}{6\ell^2}. \quad (34.21)$$

As the confining box shrinks (ℓ becomes smaller), the conformational free energy increases. The logarithm of the partition coefficient for a polymer entering a confined space is given in terms of the dominant term, $\ln K = -\Delta F/kT \propto -Nb^2/\ell^2$ (see Figure 34.7). Equation (34.21) shows that polymers tend not to partition into small spaces.

Now let's consider a polymer molecule adsorbed on a surface. Adsorption is a form of confinement.

The Thickness of a Polymer Adsorbed on a Surface Is a Balance Between Adsorption Energy and Confinement Entropy

Let's compute the average thickness D of a polymer molecule that is adsorbed on a surface (see Figure 34.8). We'll do a simple dimensional analysis. The thickness of the polymer is determined by a balance between an adsorption energy, which tends to flatten the chain onto the surface, decreasing D , and an 'elastic' entropy that tends to expand the chain, increasing D . The confinement contribution to the free energy is derived from Equation (34.21):

$$\frac{F_{\text{elastic}}}{kT} \propto \frac{Nb^2}{D^2} \quad (34.22)$$

(leaving out constants). The adsorption free energy is the sticking energy per monomer, $\varepsilon > 0$, multiplied by the total number of monomers in the chain,

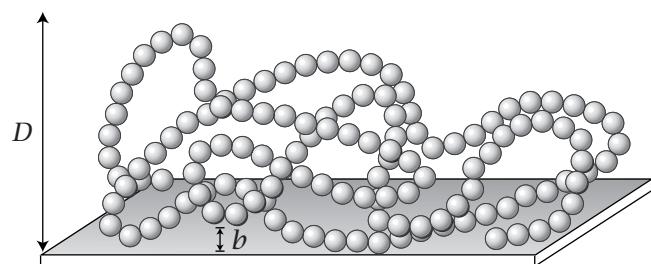


Figure 34.8 The average 'thickness' of a polymer molecule adsorbed on a surface is D . The monomer diameter is b , which defines the thickness of the first layer.

N , multiplied by the fraction b/D of the chain's monomers, on average, that are in a layer within one bond distance b from the surface, assuming that the monomers are uniformly distributed:

$$\frac{F_{\text{ads}}}{kT} \propto \left(\frac{-N\epsilon b}{D} \right). \quad (34.23)$$

The total free energy F_{total} is the sum of the elastic and adsorption free energies from Equations (34.22) and (34.23):

$$\frac{F_{\text{total}}}{kT} = \frac{F_{\text{elastic}}}{kT} + \frac{F_{\text{ads}}}{kT} = \frac{Nb^2}{D^2} - \frac{N\epsilon b}{DkT}.$$

The equilibrium thickness D_0 is the value of D that minimizes the total free energy:

$$\begin{aligned} & \left[\frac{\partial}{\partial D} \left(\frac{F_{\text{total}}}{kT} \right) \right]_{D=D_0} = 0 \\ & \Rightarrow -\frac{2Nb^2}{D_0^3} + \frac{N\epsilon b}{D_0^2 kT} = 0 \\ & \Rightarrow D_0 = \frac{2bkT}{\epsilon}. \end{aligned}$$

This dimensional analysis predicts that the thickness D_0 of the polymer layer decreases as the sticking energy increases, does not depend on the chain length, and increases with temperature.

Confinement and elastic entropies are also important for the dynamical relaxation processes of polymeric liquids and solids. For processing polymers and plastics it is often useful to know their relaxation times.

The Rouse-Zimm Model Describes the Dynamics of Viscoelastic Fluids

Some liquids, such as raw eggs, are stretchy and elastic. If you apply a shear force to a *viscoelastic fluid*, the fluid dissipates energy through viscous forces, as simpler liquids do, but also stores energy through elastic forces. Liquids are often elastic because of dissolved polymers.

To explore viscoelasticity in liquids, consider the simple shear flow experiment shown in Figure 34.9. Begin the experiment by applying a shear force to the solution until it reaches steady state. At that time, the total force f_m that is being applied to the top plate (divided by the number of molecules) equals the sum of the viscous and elastic forces exerted by the molecule:

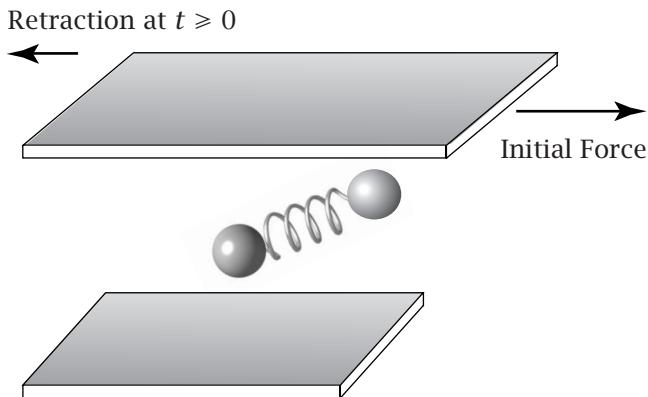
$$f_{\text{viscous}} + f_{\text{elastic}} = f_m. \quad (34.24)$$

After the applied force has been turned off, you have

$$f_{\text{viscous}} + f_{\text{elastic}} = 0. \quad (34.25)$$

After you turn off the applied force, the elastic energy that has been stored in the polymeric fluid now pulls the top plate backward (see Figure 34.9) and relaxes toward equilibrium.

Figure 34.9 The Rouse-Zimm model for viscoelasticity. Applying a shear flow to a polymer solution stretches out the chains, indicated here in terms of a beads-and-spring model, leading to an elastic entropic retractive force.



The balance of these forces for dilute solutions of polymer molecules in a small-molecule solvent was first described by PE Rouse [5] and BH Zimm [6]. Here is a simple version of their models. Assume that the polymer molecule has been deformed by the applied shear to have an end-to-end distance $x = x_0$. The elastic retractive force per molecule in a θ solvent is given by Equation (33.24), $f_{\text{elastic}} = -3kTx/R_e^2$, in terms of the mean-square end-to-end length $R_e^2 = Nb^2 = 3R_0^2/2$. The viscous force exerted per polymer chain on the surrounding medium is $f_{\text{viscous}} = \xi v$, where ξ is the friction coefficient (see Equation (17.2)) and $v = -dx/dt$ is the velocity of retraction in the direction that opposes the original applied force. You can express Equation (34.25) as a differential equation:

$$\xi \frac{dx}{dt} + \frac{3kT}{R_e^2} x = 0 \implies \frac{dx}{dt} + \frac{1}{\tau} x = 0, \quad (34.26)$$

where the relaxation time τ is

$$\tau = \frac{\xi R_e^2}{kT}. \quad (34.27)$$

We leave out small numerical factors such as 3 for this dimensional analysis. As another simplification, here we consider only the slowest relaxation time. Because the Rouse and Zimm theories treat each polymer chain as having multiple springs in series, there are many different relaxation *modes*, leading to a sum of exponentials. We treat only the slowest mode here. The solution to Equation (34.26) is

$$x(t) = x_0 e^{-t/\tau}. \quad (34.28)$$

The displacement of the top plate changes exponentially with time as $x \rightarrow 0$ after the applied force is turned off.

How does the relaxation time τ depend on the chain length? In the *Rouse model*, the friction coefficient ξ is assumed to be the sum of frictional contributions ξ_1 from each monomer unit:

$$\xi = N\xi_1, \quad (34.29)$$

where ξ_1 is a constant. Substituting $R_e^2 = Nb^2$ and Equation (34.29) into Equation (34.27) gives the Rouse-model relaxation time τ :

$$\tau \propto \frac{\xi_1 b^2}{kT} N^2. \quad (34.30)$$

The Rouse model predicts that longer polymer chains have slower relaxations: the relaxation time increases as N^2 .

The *Zimm model* treats the friction differently. In the Zimm model, the interior monomers are shielded (or *screened*) from the flow, so not all the monomers contribute to the friction. This is called *hydrodynamic screening*. In this case, the friction is treated as if the whole polymer molecule were a Stokes-Einstein sphere:

$$\xi = 6\pi\eta_0 R_e \propto \eta_0 b N^{1/2}. \quad (34.31)$$

The Zimm model, Equation (34.31), predicts that the overall friction only increases as the square root of the molecular weight. To obtain the relaxation time, substitute $R_e^2 = Nb^2$ and Equation (34.31) into Equation (34.27) to get

$$\tau \propto \frac{\eta_0 b^3}{kT} N^{3/2}. \quad (34.32)$$

The Zimm model predicts $\tau \propto N^{3/2}$, in contrast to the Rouse-model prediction $\tau \propto N^2$. Experiments have confirmed the Zimm-model prediction for dilute polymers in small-molecule solvents. These models also predict that the dynamics is slower in more viscous solvents or at lower temperatures.

In an important application, dilute-solution polymer relaxation measurements were used to prove that chromosomal DNA is a single long contiguous molecule [7]. It was not previously clear whether a chromosome was a single long DNA molecule or many small DNA molecules held together by proteins. R Kavenoff and BH Zimm digested away the proteins and measured the viscoelastic relaxation times of fruit-fly chromosomes. They found that the relaxation times were consistent with a molecule having a molecular weight of 20×10^9 daltons to 80×10^9 daltons, exactly the size expected from the number of nucleotide base pairs known to be contained in the chromosome. Earlier methods had not been able to measure the sizes of such huge single molecules.

EXAMPLE 34.2 Measuring the size of DNA by dynamical relaxation times. Let's analyze the Kavenoff-Zimm experiment. Rearranging the Zimm equation (34.32) gives the number of chain monomers N as

$$N = \left(\frac{\tau k T}{\eta_0 b^3} \right)^{2/3}.$$

Kavenoff and Zimm put a dilute solution of chromosomal DNA into water at $T = 300\text{ K}$, which has a viscosity of $\eta_0 = 1\text{ cP}$ ($1\text{ P} = 1\text{ g cm}^{-1}\text{ s}^{-1}$). They applied a shear force to stretch the DNA molecules. Then they turned off the shear force and observed an exponential relaxation time of $\tau = 1\text{ h}$. So, their experiments show that the number N of base pairs per molecule is

$$N = \left[\frac{(3600\text{ s})(1.38 \times 10^{-23}\text{ JK}^{-1})(300\text{ K})(1\text{ g cm}^2\text{ s}^{-2})(10^{-7}\text{ J})}{(10^{-2}\text{ g cm}^{-1}\text{ s}^{-1})(2 \times 10^{-7}\text{ cm})^3} \right]^{2/3}$$

$$= 7.4 \times 10^7 \text{ base pairs.}$$

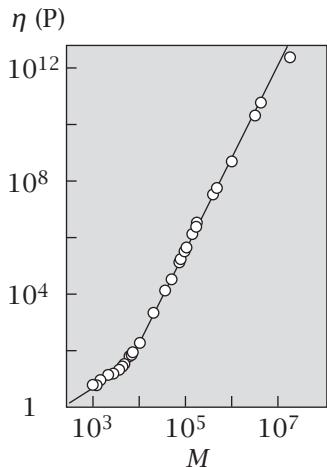


Figure 34.10 The dependence of the viscosity η on the molecular weights M of polybutadiene in the melt: $\propto N$ for small molecular weights and $\propto N^{3.4}$ for large molecular weights. The viscosity and relaxation time are linearly related. Source: RH Colby, LJ Fetter, and WW Graessley, *Macromolecules* **20**, 2226–2237 (1987).

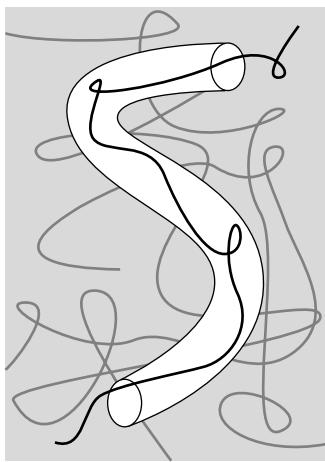


Figure 34.11 A polymer reptates by moving inside a ‘tube,’ defined by its neighboring chains.

For this calculation, we used $b = 20 \text{ \AA}$. Kavenoff and Zimm obtained b by independent measurements of the viscosity of the DNA solution. Since an average base pair has a molecular weight of 660 daltons, their experiment gives $N \approx 48 \times 10^9$ daltons per DNA molecule, consistent with the hypothesis that a *Drosophila* chromosome is composed of a single long DNA molecule.

Of great industrial importance for the processing of plastics, textile fibers, and paints are *highly concentrated* polymer solutions, where there is little or no solvent, in which the molecular organization resembles tangled spaghetti. Such concentrated solutions are called *melts*. For short-chain melts, Rouse dynamics often applies. For long-chain melts, however, the chain dynamics follows a different physical process, called *reptation*.

Reptation Describes the Snakelike Motions of Chains in Melts and Gels

In melts of long-chain polymers, the chain dynamics is much more sluggish ($\tau \propto N^{3.4}$) than would be predicted by the Rouse or Zimm theories (see Figure 34.10). This is explained by the *reptation model* of PG de Gennes, a French physicist who won the 1991 Nobel Prize in Physics for his work on polymers and other complex liquids. According to this model, the dynamics is slow because the chains must snake through a medium of surrounding chains. Think about one chain molecule A in a melt. Its surrounding chains define a sort of *tube* within which A can move (see Figure 34.11). In the reptation model, the tube remains in place during the time τ that it takes for the molecule A to slither out of the end of the tube.

The reptation model, like the Rouse model, supposes that the friction involved in dragging the chain through its tube is proportional to the chain length, $\xi = N\xi_1$, Equation (34.29). The diffusion constant D_{tube} for the chain moving through the tube is given by the Einstein-Smoluchowski relation, Equation (17.42):

$$D_{\text{tube}} = \frac{kT}{\xi} = \frac{kT}{N\xi_1} = \frac{D_1}{N}, \quad (34.33)$$

where D_1 is a constant with units of a diffusion coefficient. This predicts that the rate at which the chain snakes through a tube is inversely proportional to the tube length N . Now, to compute the *time required* for the chain to escape from the tube, note that the length of the confining tube is approximately the contour length of the chain, $L = Nb$, and combine Equation (34.33) with the diffusion expression $\langle x^2 \rangle = 2Dt$ (Equation 18.11), to get

$$\tau \propto \frac{\langle x^2 \rangle}{D_{\text{tube}}} \propto \frac{L^2}{D_{\text{tube}}} \propto \frac{NL^2}{D_1} \propto N^3. \quad (34.34)$$

Experiments give $\tau \propto N^{3.4}$ (see Figure 34.10), which is slightly steeper than this prediction $\tau \propto N^3$ of the reptation model. There is some evidence that this discrepancy is due to the finite lengths of reptation tubes. This model accounts

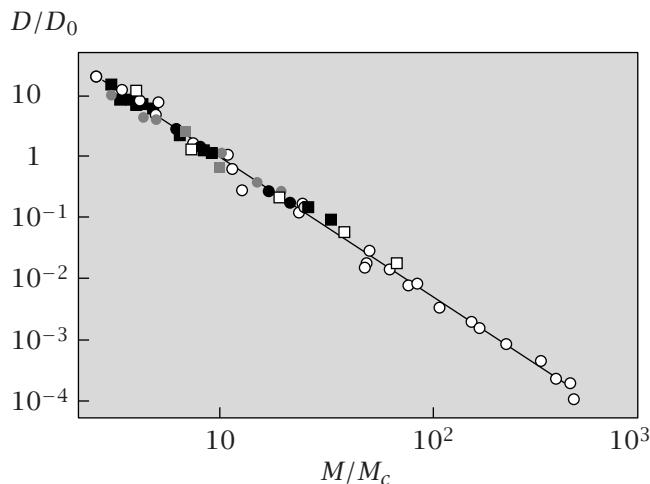


Figure 34.12 Self-diffusion coefficient D of seven different types of polymers—polystyrene, polyisoprene, polybutadiene, poly(dimethylsiloxane), poly(ethylene oxide), poly(methylstyrene), and hydrogenated polybutadiene—versus molecular weight shows that $D \propto N^{-2.28 \pm 0.05}$, which differs somewhat from the N^{-2} dependence expected from simple reptation theory. D_0 is the diffusion coefficient for $M/M_c = 10$. Source: H Tao, TP Lodge, and ED von Meerwall, *Macromolecules* **33**, 1747–1758 (2000).

in a simple way for the steep chain-length dependence of the relaxation times of polymers in bulk melts.

The reptation model also predicts a second type of diffusion constant. The diffusion constant D_{tube} cannot be readily measured, because experiments cannot track how the chain moves along its tube axis. But you can measure a diffusion constant D that describes how the center of mass of the polymer chain moves in space over time. In this case, the average distance moved is $\langle x^2 \rangle \propto R_e^2$, where $R_e^2 = Nb^2$. The reptation model, Equation (34.34), predicts that chain diffusion slows as the square of the chain length:

$$D \propto \frac{R_e^2}{\tau} \propto \frac{N}{N^3} \propto N^{-2}. \quad (34.35)$$

Experiments give $D \propto N^{-2.28}$, which differs somewhat from this reptation-model prediction of $D \propto N^{-2}$ (see Figure 34.12).

Summary

Chain molecules resist deformation or confinement, owing to elasticity and excluded volume. Chain conformations can be perturbed by the presence of a surface. For polymers that are tethered to an impenetrable surface, the free ends tend not to be at the surface but terminate at a distance proportional to $N^{1/2}$ away from the surface. Confining a chain molecule to small spaces leads to a reduction in the number of available conformations. So there is a free energy that opposes the entrance of polymer chains into confined spaces. After stretching in shear flow, polymers retract, causing elasticity in solutions, called viscoelasticity. The Rouse-Zimm theory explains the dependence of the relaxation time on the polymer molecular weight in dilute solutions. For longer chains at higher concentrations, or in gels, the relaxation and diffusional motions have been modeled in terms of snakelike reptation processes.

Problems

1. Polymer adsorption on a surface.

- (a) Write an expression to show how many monomer-sized layers M there are if a polymer of length N sticks to a surface with sticking energy ε_0 per monomer.
(b) Compute M when $\varepsilon_0 = (1/5)kT$.

2. Rouse-Zimm dynamics in a good solvent.

Derive the dependence of the relaxation time τ on the chain length N for a dilute polymer solution in a good solvent.

3. A tethered chain returns to the surface.

A random-flight chain of length N is tethered at one end to an impenetrable surface. Write an expression for the probability $P(N)$ that the free end also happens to land on the surface.

4. Relaxation times for DNA molecules.

The relaxation time is $\tau \approx 1$ s for a dilute solution of T2 DNA molecules having a molecular weight of 10^8 daltons. What is the relaxation time for *Drosophila* (fruit fly) chromosomal DNA having a molecular weight of 20×10^9 daltons?

5. The free energy of localizing DNA in a cell nucleus.

Human chromosomal DNA molecules have about 3×10^9 nucleotides, each separated from its neighbor by a bond distance of about $b = 3.4\text{ \AA}$.

- (a) Compute the contour length L if this were all a single DNA molecule.
(b) Compute its end-to-end length $R_e = (Nb^2)^{1/2}$ if this DNA were a single random-flight chain.
(c) If the cell's nucleus has radius $r_c = 4\text{ }\mu\text{m}$, compute R_e/r_c from (b).
(d) Compute the excluded-volume free energy cost of localizing this 'virtual' single molecule into the human cell nucleus at $T = 300\text{ K}$.

6. Distribution of the ends of tethered chains.

Derive $\langle z^2 \rangle$ for the termini of chains that are tethered to an impenetrable surface.

References

- [1] EA DiMarzio and FL McCrackin, *J Chem Phys* **43**, 539–547 (1965).

- [2] EF Casassa, *J Polymer Sci, Part B: Polymer Lett* **5**, 773–778 (1967).
[3] M Doi and SF Edwards, *The Theory of Polymer Dynamics*, Oxford University Press, Oxford, 1986.
[4] MG Davidson, UW Suter, and WM Deen, *Macromolecules* **20**, 1141–1146 (1987).
[5] PE Rouse, *J Chem Phys* **21**, 1272–1280 (1953).
[6] BH Zimm, *J Chem Phys* **24**, 269–278 (1956).
[7] R Kavenoff and BH Zimm, *Chromosoma* **41**, 1–27 (1973).

Suggested Reading

Excellent fundamental texts on excluded volume and conformational changes in polymers:

PG de Gennes, *Scaling Concepts in Polymer Physics*, Cornell University Press, Ithaca, NY, 1979.

M Doi, *Introduction to Polymer Physics*, Oxford University Press, Oxford, 1996.

WW Graessley, *Polymeric Liquids and Networks: Structure and Properties*, Garland Science, New York, 2004.

Good general texts on polymer theory are:

RH Boyd and PJ Phillips, *The Science of Polymer Molecules*, Cambridge University Press, Cambridge, 1993.

EF Casassa, *J Polymer Sci, Part B, Polymer Lett* **5**, 773–778 (1967).

A Yu Grosberg and AR Khokhlov, *Giant Molecules: Here, There, and Everywhere...*, Academic Press, San Diego, 1997.

EG Richards, *An Introduction to the Physical Properties of Large Molecules in Solution*, Cambridge University Press, Cambridge, 1980.

SF Sun, *Physical Chemistry of Macromolecules: Basic Principles and Issues*, 2nd edition, Wiley, New York, 2004.

General overview of polymer dynamics in dilute solutions:

J Ferry, *Viscoelastic Properties of Polymers*, 3rd edition, Wiley, New York, 1980.

Appendix A: How To Make Approximations

Physical Modeling Often Involves Making Approximations

Here are the basic mathematical tools you need for making approximations. A physical theory can involve many steps of mathematics. If one step involves a complicated function $f(x)$, having exponentials, logarithms, sines, or cosines, then further steps can become too complicated to analyze. In such cases, you can approximate the function by a simpler mathematical expression, making the analysis manageable. *Taylor series expansions* give the mathematics you need for making approximations. Typically, you will start with knowledge of the mathematical form of $f(x)$. You want to approximate that function near some particular point $x = a$.

Taylor Series

A function $f(x)$ near the point $x = a$ can be expressed in terms of its derivatives:

$$\begin{aligned} f(x) \approx f(a) + & \left(\frac{df}{dx} \right)_{x=a} (x-a) + \frac{1}{2} \left(\frac{d^2 f}{dx^2} \right)_{x=a} (x-a)^2 \\ & + \frac{1}{6} \left(\frac{d^3 f}{dx^3} \right)_{x=a} (x-a)^3 + \dots + \frac{1}{n!} \left(\frac{d^n f}{dx^n} \right)_{x=a} (x-a)^n + \dots \quad (\text{A.1}) \end{aligned}$$

Making approximations in physical theories is often a matter of truncating this series and keeping only the largest terms. If x is near enough to the point a , the quantity $(x-a)$ will be small. Then the Taylor series converges, meaning that the first-derivative term is the largest, the second-derivative term is smaller, the third-derivative term is even smaller than that, and so on. Often you need only the one or two largest terms to give useful approximations. If you keep only the constant and first-derivative terms, the approximation is called *first-order*. If you also keep the second-derivative term, it is called *second-order*.

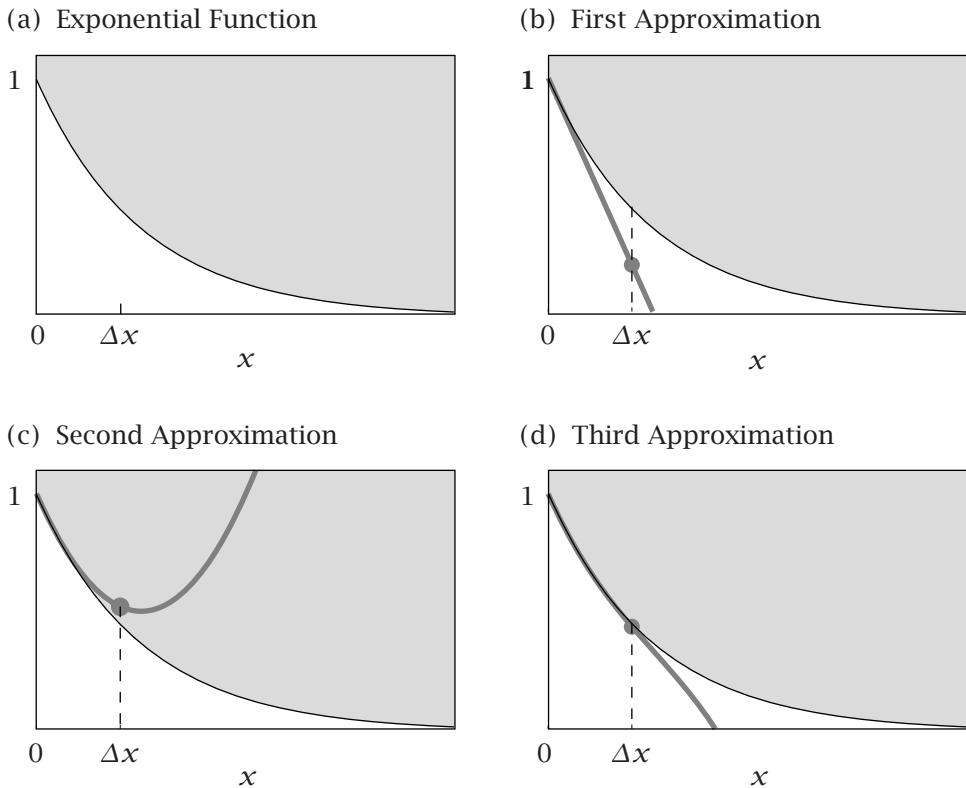


Figure A.1 Use the Taylor series to find values of $f(x) = e^{-x}$ for $x = \Delta x$ near $x = 0$. (a) The first term in the Taylor series is $f(0) = 1$. (b) The gray line is the first approximation, $f(x) \approx 1 - x$. (c) The second approximation is $f(x) \approx 1 - x + (x^2/2)$. (d) The third approximation is $f(x) \approx 1 - x + (x^2/2) - (x^3/6)$.

The Taylor series gives increasingly better approximations for $f(x)$ near $x = a$ as a polynomial expansion in $(x-a)$, $(x-a)^2$, $(x-a)^3$, etc.

Let's find the Taylor series approximation for the exponential function e^{-bx} .

EXAMPLE A.1 **The Taylor series expansion for $f(x) = e^{-bx}$ near $x = 0$.** To find the Taylor series for this function, you evaluate $f(a)$ where $a = 0$, and you evaluate the derivatives of the function at $x = a = 0$ according to Equation (A.1). For the first term, you have $f(0) = e^{-0} = 1$.

To compute the second term, you find the derivative $f^{(1)}(x)$ and evaluate it at $x = 0$:

$$\frac{d}{dx}(e^{-bx}) = -be^{-bx} \Big|_{x=0} = -b.$$

So the second term of the Taylor series is $f^{(1)}(a)(x-a) = -bx$ (see Figure A.1(b)). To compute the third term, you need the second derivative $f^{(2)}(x)$ evaluated at $x = 0$:

$$\frac{d^2}{dx^2}(e^{-bx}) = b^2 e^{-bx} \Big|_{x=0} = b^2.$$

The third term of the Taylor series is $(1/2)f''(a)(x-a)^2 = (bx)^2/2$.

To second-order approximation, the Taylor series expansion is $f(x) = e^{-bx} \approx 1 - bx + (bx)^2/2$, shown in Figure A.1(c). Taylor series expansions for various functions are given in Appendix J.

For some approximations, it is sufficient to keep first-order terms. In other cases, all the first-order terms cancel, so the most important nonzero contributions are the second-order terms. Then, you need to keep both first-order and second-order terms. Here is an example.

EXAMPLE A.2 A second-order approximation. Suppose you have a function $y(x) = bx + e^{-bx}$ and you want to know its dependence on x for small x , $bx \ll 1$. Since $e^{-bx} \approx 1 - bx + (bx)^2/2$ (see Example A.1), the first-order terms cancel, so you keep the second-order term and $y(x) \approx 1 + (bx)^2/2$.

EXAMPLE A.3 Two first-order approximations. Find a first-order approximation for $1/(1+x)$. From Appendix J, use Equation (J.6) with $p = -1$ to get

$$\frac{1}{1+x} \approx 1-x.$$

To approximate $1/(1+x)^{1/5}$, use Equation (J.6) with $p = -1/5$:

$$\frac{1}{(1+x)^{1/5}} \approx 1 - \frac{1}{5}x.$$

EXAMPLE A.4 Expanding a logarithm. Derive the first three terms of the Taylor Series expansion for $f(x) = \ln(1+2x^3)$ around $x \approx 1$. Evaluating the function and its derivatives at $a = 1$ gives:

$$f(1) = \ln(1+2) = \ln 3, \quad (\text{A.2})$$

$$f'(1) = \left[\frac{6x^2}{1+2x^3} \right]_{x=1} = \frac{6}{3} = 2, \quad (\text{A.3})$$

and

$$f''(1) = \left[\frac{12x}{1+2x^3} - \left(\frac{6x^2}{1+2x^3} \right)^2 \right]_{x=1} = \frac{12}{3} - \frac{36}{9} = 0. \quad (\text{A.4})$$

So the Taylor series approximation is

$$f(x) \approx \ln 3 + 2(x-1) + \frac{0}{2!}(x-1)^2 = \ln 3 + 2(x-1). \quad (\text{A.5})$$

This page is intentionally left blank.

Appendix B: Stirling's Approximation

Combinatoric expressions are often given in terms of factorials, which can be difficult to manipulate mathematically. Factorials are made more tractable by using **Stirling's approximation**:

$$n! \approx (2\pi n)^{1/2} \left(\frac{n}{e}\right)^n, \quad (\text{B.1})$$

which becomes quite accurate for large n . To simplify the derivation of Equation (B.1), let's work instead with the logarithm,

$$\ln n! \approx \frac{1}{2} \ln(2\pi) + \left(n + \frac{1}{2}\right) \ln n - n. \quad (\text{B.2})$$

It is often useful to simplify this even further. When n is larger than about 10, the first term $(1/2) \ln(2\pi) = 0.92$ is negligible compared with the second term. Then, $n + 1/2 \approx n$, so you have

$$\ln n! \approx n \ln n - n \implies n! \approx \left(\frac{n}{e}\right)^n. \quad (\text{B.3})$$

(Try it. For $n = 100$, the better approximation is $\ln 100! \approx 0.92 + (100.5) \ln 100 - 100 = 363.74$, and the simpler approximation is $\ln 100! \approx 100 \ln 100 - 100 = 360.5$.)

Now let's derive Equation (B.3). Since $n! = (1)(2)(3) \cdots (n-1)(n)$, the logarithm of $n!$ is the sum

$$\ln n! = \ln 1 + \ln 2 + \cdots + \ln n = \sum_{m=1}^n \ln m. \quad (\text{B.4})$$

For large values of n , approximate the sum in Equation (B.4) with an integral:

$$\ln n! = \sum_{m=1}^n \ln m \approx \int_1^n \ln m \, dm = (m \ln m - m) \Big|_1^n. \quad (\text{B.5})$$

For $n \gg 1$, this gives Equation (B.3).

This page is intentionally left blank.

Appendix C: A Math Trick for Getting Averages from Derivatives

In many problems of statistical mechanics, such as in computing polymer length distributions, or energies (Chapter 11), or fractional helicity in polymers (Chapter 26), or numbers of ligands bound to proteins or surfaces (Chapters 27 and 28), you will have a polynomial function such as

$$f(x) = 1 + x + x^2 + x^3 + \dots + x^n = \sum_{i=0}^n x^i, \quad (\text{C.1})$$

and you will want to compute the average exponent, $\langle i \rangle$:

$$\langle i \rangle = \frac{\sum_{i=0}^n ix^i}{\sum_{i=0}^n x^i} = \frac{x + 2x^2 + 3x^3 + \dots + nx^n}{1 + x + x^2 + \dots + x^n}. \quad (\text{C.2})$$

You can compute this using a simple trick. Notice that if you take the derivative of Equation (C.1), you get

$$\frac{df}{dx} = 1 + 2x + 3x^2 + 4x^3 + \dots + nx^{n-1} = \sum_{i=0}^n ix^{i-1}. \quad (\text{C.3})$$

Now multiply Equation (C.3) by x and divide by f to get $\langle i \rangle$:

$$\frac{x}{f} \frac{df}{dx} = \frac{x + 2x^2 + 3x^3 + \dots + nx^n}{1 + x + x^2 + \dots + x^n} = \langle i \rangle. \quad (\text{C.4})$$

By using the relations $d \ln x = dx/x$ and $d \ln f = df/f$, Equation (C.4) can be expressed as

$$\langle i \rangle = \frac{x}{f} \frac{df}{dx} = \frac{d \ln f}{d \ln x}. \quad (\text{C.5})$$

If you know $f(x)$, Equation (C.4) provides a much more convenient way to calculate an average than Equation (C.2) does.

Also, you can get the second moment $\langle i^2 \rangle$ by taking another derivative,

$$\begin{aligned}\langle i^2 \rangle &= \frac{\sum_{i=0}^n i^2 x^i}{\sum_{i=0}^n x^i} = \frac{x + 4x^2 + 9x^3 + \dots + n^2 x^n}{1 + x + x^2 + \dots + x^n} \\ &= \left(\frac{x}{f} \right) \frac{d}{dx} \left(x \frac{df}{dx} \right).\end{aligned}\quad (\text{C.6})$$

Equations (C.4) and (C.6) also apply to the more general polynomial, $f = a_0 + a_1 x + a_2 x^2 + \dots + a_n x^n$. Sometimes, it is useful to compute the variance

$$\sigma^2 = \langle i^2 \rangle - \langle i \rangle^2 = \frac{x}{f} (f' + x f'') - \left(\frac{x f'}{f} \right)^2, \quad (\text{C.7})$$

where ' and '' indicate the first and second derivatives respectively. Equation (C.7) comes from taking the derivatives indicated in Equations (C.4) and (C.6). Or, you can express Equation (C.7) more compactly as

$$\sigma^2 = \frac{d^2 \ln f}{d(\ln x)^2} = x \frac{d}{dx} \left(\frac{x f'}{f} \right). \quad (\text{C.8})$$

Performing the operations in Equation (C.8) will give the same result as in Equation (C.7).

Appendix D: Proof of the Euler Reciprocal Relationship

In Chapter 9, we use the Euler relationship to establish Maxwell's relations between the thermodynamic quantities. Here we derive the Euler relationship. Figure D.1 shows four points at the vertices of a rectangle in the (x, y) plane. Using a Taylor series expansion, Equation (A.1), compute the change in a function Δf through two different routes. First integrate from point A to point B to point C . Then integrate from point A to point D to point C . Compare the results to find the Euler reciprocal relationship. For $\Delta f = f(x + \Delta x, y + \Delta y) - f(x, y)$, the first terms of the Taylor series are

$$\begin{aligned} \Delta f &= \left(\frac{\partial f}{\partial x} \right) \Delta x + \left(\frac{\partial f}{\partial y} \right) \Delta y \\ &+ \frac{1}{2} \left[\left(\frac{\partial^2 f}{\partial x^2} \right) (\Delta x)^2 + \left(\frac{\partial^2 f}{\partial y^2} \right) (\Delta y)^2 + 2 \left(\frac{\partial^2 f}{\partial x \partial y} \right) \Delta x \Delta y \right] \\ &+ \dots \end{aligned} \quad (\text{D.1})$$

Along the path from point A to point B in Figure D.1, x changes, $\Delta x = x_B - x_A$, and y is constant. From point A to B , Equation (D.1) gives

$$f(B) = f(A) + \left(\frac{\partial f(A)}{\partial x} \right) \Delta x + \frac{1}{2} \left(\frac{\partial^2 f(A)}{\partial x^2} \right) (\Delta x)^2 + \dots \quad (\text{D.2})$$

From B to C , only y changes, $\Delta y = y_C - y_B$, while x remains fixed:

$$f(C) = f(B) + \left(\frac{\partial f(B)}{\partial y} \right) \Delta y + \frac{1}{2} \left(\frac{\partial^2 f(B)}{\partial y^2} \right) (\Delta y)^2 + \dots \quad (\text{D.3})$$

Substitute $f(B)$ given by Equation (D.2) into Equation (D.3), and keep terms up to second order because those are the lowest-order terms that do not cancel.

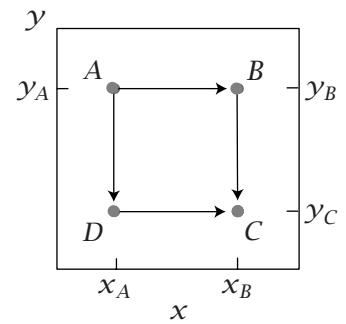


Figure D.1 For proving the Euler reciprocal relationship, two paths from A to C : through B , or through D .

Higher-order terms will be negligible. We find $f(C)$ in terms of $f(A)$:

$$\begin{aligned} f(C) &= f(A) + \left(\frac{\partial f(A)}{\partial x} \right) \Delta x + \frac{1}{2} \left(\frac{\partial^2 f(A)}{\partial x^2} \right) (\Delta x)^2 \\ &\quad + \left(\frac{\partial f(A)}{\partial y} \right) \Delta y + \frac{\partial}{\partial y} \left(\frac{\partial f(A)}{\partial x} \right) \Delta y \Delta x \\ &\quad + \frac{1}{2} \left(\frac{\partial^2 f(A)}{\partial y^2} \right) (\Delta y)^2 + \dots \end{aligned} \quad (\text{D.4})$$

Now evaluate the other path, from point A through point D to point C .

Although the path differs, the values $\Delta x = x_C - x_D = x_B - x_A$ and $\Delta y = y_D - y_A = y_C - y_B$ are the same as for path ABC . Now you have

$$\begin{aligned} f(C) &= f(A) + \left(\frac{\partial f(A)}{\partial y} \right) \Delta y + \frac{1}{2} \left(\frac{\partial^2 f(A)}{\partial y^2} \right) (\Delta y)^2 \\ &\quad + \left(\frac{\partial f(A)}{\partial x} \right) \Delta x + \frac{\partial}{\partial x} \left(\frac{\partial f(A)}{\partial y} \right) \Delta y \Delta x \\ &\quad + \frac{1}{2} \left(\frac{\partial^2 f(A)}{\partial x^2} \right) (\Delta x)^2 + \dots \end{aligned} \quad (\text{D.5})$$

Equations (D.4) and (D.5) both give $f(C)$, so you can set the right-hand sides equal and cancel identical terms. The remaining terms give the Euler reciprocal relationship we want (Equation (4.39)):

$$\left(\frac{\partial^2 f(A)}{\partial y \partial x} \right) = \left(\frac{\partial^2 f(A)}{\partial x \partial y} \right).$$

Appendix E: A Principle of Fair Apportionment Leads to the Function $S = k \ln W$

Here we explore the foundations of entropy. Why should the entropy function have the form $S/k = \ln W = -\sum_i p_i \ln p_i$? Why is the entropy $S(U, V, N)$ only useful as an extremum principle when its arguments, such as U , V , or N , are *extensive* quantities? What is the reason that S itself is also extensive? And, why is entropy useful beyond thermodynamics, for example in information theory and signal processing?

The Entropy Principle Is General

The entropy-maximization principle applies to many different situations. The situations all involve collections of equivalent ‘particles’ or ‘elemental units’ that can each take on some ‘value’ or ‘score.’ Think of a series of dice rolls with different numbers on the faces. Or, think of a collection of identical atoms or molecules, each of which has an energy. Or, think of a string of digits or letters in information transmission. Or, think of an image as a string of pixels, each of which has a particular brightness. The elementary unit of interest, which we call an *event*, is the individual coin flip, dice roll, atom or molecule, or pixel. Each event can have a ‘score,’ taken from a menu of values. Before we show why maximizing entropy predicts probability distributions, we first give some examples.

Dice problem, Chapter 5; the average score is known. In Chapter 5, you roll a die. The faces are indexed $j = 1, 2, 3, \dots, t$. On any one roll, you observe face j , which has a score ε_j . Roll the die N times. You know one quantity, the average score per roll, U/N . The entropy-maximization principle and the given value of U/N are sufficient to predict the probabilities p_j^* of observing each particular outcome. The predicted probability distribution is exponential in general, or flat if the die is unbiased (i.e., when $U/N = 3.5$ for a normal 6-sided die).

Thermal system, canonical ensemble; the temperature is known. In Chapter 10, you are given N atoms or molecules in a thermodynamic system. The

'score' is the energy state ε_j of each molecule. In the canonical ensemble, you are given the temperature T , which is identical to knowing the average energy $\langle \varepsilon \rangle$, per particle of the system. Entropy maximization subject to the constraint of fixed T predicts the Boltzmann distribution, $p_j^* \propto \exp[-\varepsilon_j/(kT)]$, which depends exponentially on the energy level.

Thermal system, isobaric ensemble; the density is known. Suppose you are given N atoms or molecules in a thermodynamic system. Now, the 'score' is the volume V_j occupied by that molecule. You are given the density of the system, which is the inverse of the average volume per molecule, V/N . This is equivalent to knowing p/kT , where p is the pressure and T is the temperature. Maximizing the multiplicity in this case predicts that it is exponentially unlikely, $\exp[-pV_j/(kT)]$, that a given particle will occupy a large volume, for example.

Thermal system, grand canonical ensemble; the average number of bound particles is known. You are given a macromolecule in solution. Now the 'score' is the number M_j of ligand molecules that are bound to each macromolecule. You are given the average number, $\langle M \rangle$ of ligands bound per macromolecule, which is equivalent to knowing $\mu/(kT)$, the chemical potential of the ligand in solution divided by kT . Maximizing the multiplicity predicts populations proportional to $\exp[-\mu M_j/(kT)]$, i.e., it is exponentially unlikely that a macromolecule will have a large number of ligands bound to it. This is the basis for the grand canonical ensemble in Chapter 28.

Time series; the average waiting time is known. Consider a *waiting time process*; see Chapter 18. For example, you sit beside a road, you start a stopwatch when one car passes by, and then you measure the time required until the next car passes by. Now each event is one waiting time. You watch N waiting events. The 'score' is the waiting time τ_j for each waiting event. You are given the average waiting time $T/N = \langle \tau \rangle$, where T is the total of all the waiting times. Maximizing the multiplicity predicts that it is exponentially unlikely, $\exp(-\lambda \tau_j)$, that you will wait a long time until the next car passes.

Single-molecule dynamical trajectories; average transition rates are known. Consider a trajectory in which a single particle jumps back and forth between two states $A \leftrightarrow B$; see Chapter 18. Divide the time course into N discrete time events, during which a transition may or may not occur. The score is one of the four statistical weights a, b, d , or u for staying in A , staying in B , or jumping from B to A or from A to B , respectively, depending on the state of the box before it in the time series; see Chapter 18. You are given, say, $\langle N_{AA} \rangle = \sum_j p_j N_{AA,j}$, the number of time steps that started in A and ended in A , over the whole trajectory. Maximizing the trajectory entropy and satisfying this constraint predicts an exponentially decreasing probability, $\exp(-\lambda N_{AA,j})$, that a trajectory will have a large number of such transitions.

In all these cases, maximizing entropy subject to a constraint predicts probability distributions. In all these cases, you have N elemental units, each of which randomly adopts a score ε_j among a set of possible options. In all these

cases, an extensive quantity U of the whole system is known or fixed. U is an observable of the system. There is a multiplicity W of ways that this value of U can be realized by microscopic states of the system. The key point is that the distribution p_j^* of probabilities that you will observe is that which maximizes W for the given value of U . Here is the basis for this approach.

First, let's define the *probabilities* of the states of the system. There are different ways to define probabilities: the *frequency interpretation* and the *subjective interpretation*. In the frequency interpretation, a probability is defined as the fraction of times you see the particular outcome of interest upon performing some event multiple times. The frequency interpretation readily applies to dice rolls and coin flips. A different and broader way to think of probabilities is the subjective interpretation, which applies to events that you can't replicate. What is the probability that your birthday is March 4? This event is not repeatable. Subjective probabilities describe either a state of knowledge or some underlying symmetry. Because you know your birthday, to you this probability will be either 0 or 1. However, to people who don't know your birthday, the probability that you were born on March 4 is 1/365. Subjective probabilities are statements either of someone's state of knowledge or of an underlying symmetry in the problem. Below, for concreteness, we use the frequency interpretation, but we note that for the mathematical machinery of entropy, it doesn't matter whether your probabilities arose from frequencies, information, or symmetries. This area is called the theory of *random variables*.

Roll a die N times. Call this sequence of outcomes an *N -string*. A single N -string may not be sufficient for you to collect accurate statistics. So, roll another N times, then another, until you have M independent strings of N rolls each. You will now have performed a total of MN individual dice rolls. Suppose that the number of times that face j appears is Mn_j . Provided M is sufficiently large, the probability p_j of the appearance of face j is well defined. It is

$$p_j = \frac{Mn_j}{MN} = \frac{n_j}{N}, \quad (\text{E.1})$$

which sums to one, $\sum_j p_j = 1$.

Now, because each of the N dice rolls could have had any of t different outcomes, the number of different outcomes you could have observed in one N -string is t^N , where t is the total number of different faces on the die. We define the *multiplicity* as the number of different sequences that have a given composition:

$$\begin{aligned} W(\{n_j\}) &= \frac{(MN)!}{(Mn_1)!(Mn_2)!(Mn_3)!\cdots(Mn_t)!} \\ &\approx (p_1^{-p_1} p_2^{-p_2} p_3^{-p_3} \cdots p_t^{-p_t})^N = W(\{p_j\}), \end{aligned} \quad (\text{E.2})$$

where we use the bracket notation $\{\dots\}$ to indicate the full set of probabilities $\{p_j\}$ or numbers $\{n_j\}$ that you observed. We used Stirling's approximation (page 703), which is valid here because we are assuming that M is large.

We are not interested in any arbitrary composition $\{n_j\}$ of outcomes. We are interested in only those compositions of outcomes that lead to a particular total score U for an N -string, or, equivalently, a particular average score per

outcome, $\langle \varepsilon \rangle_0$, because

$$U = \sum_{j=1}^t n_j \varepsilon_j = N \sum_{j=1}^t p_j \varepsilon_j = N \langle \varepsilon \rangle_0. \quad (\text{E.3})$$

The multiplicity $W(U)$ of states is different for different values of U .

We want to formulate a function $S[W(p_j^*, U)]$, which we will call the *entropy*, the extremum value of which picks out the one particular set of probabilities, $\{p_j^*\}$, that maximizes the multiplicity $W(U)$ for the given value, U . We first show that since the quantity U is *extensive*, i.e., proportional to the system size N (see Equation (E.3)), the function S must also be extensive, i.e., a sum over the entropy of each event:

$$S(p_1, p_2, p_3, \dots, p_t) = s(p_1) + s(p_2) + s(p_3) + \dots + s(p_t). \quad (\text{E.4})$$

Then, we show that the only function that satisfies these requirements is $S(W) = k \ln W$, where k is a constant.

The Entropy Is Extensive

We seek a function $S(p_1, p_2, \dots, p_t)$ that is maximal where W is maximal and satisfies two constraints: the probabilities must sum to one and the average score per event must satisfy the particular given value $\langle \varepsilon \rangle_0 = U/N$:

$$\begin{aligned} g(p_1, p_2, \dots, p_t) &= \sum_{j=1}^t p_j = 1, \\ h(p_1, p_2, \dots, p_t) &= \sum_{j=1}^t \varepsilon_j p_j = \langle \varepsilon \rangle_0, \end{aligned} \quad (\text{E.5})$$

or, in differential form, $\sum_{j=0}^t dp_j = 0$ and $\sum_{j=0}^t \varepsilon_j dp_j = 0$. To find the maximum of S subject to these constraints, use the method of Lagrange multipliers. Multiply $(\partial g / \partial p_j) = 1$ and $(\partial h / \partial p_j) = \varepsilon_j$ by the corresponding Lagrange multipliers α and λ to get the extremum:

$$\frac{\partial S(p_1, p_2, \dots, p_t)}{\partial p_j} = \alpha + \lambda \varepsilon_j. \quad (\text{E.6})$$

When this equation is satisfied, S is at an extremum subject to the two constraints.

Now, divide a system into two subsystems a and b ; see Figure E.1. Subsystem a has total score U_a , event numbers N_a , the set of score values $\{\varepsilon_{ja}\}$ for $j = 1, 2, 3, \dots, t_a$, multiplicity $W_a(U_a)$, and entropy S_a . The probabilities we seek are $\{p_{ja}^*\}$. Express the entropy for subsystem a as $S_a = S_a(p_{1a}^*, p_{2a}^*, \dots, p_{ta}^*)$. To find the total differential for the variation of S_a resulting from any change in the p 's, sum Equation (E.6) over all dp_{ja} 's:

$$dS_a = \sum_{j=1}^{t_a} \left(\frac{\partial S_a}{\partial p_{ja}} \right) dp_{ja} = \sum_{j=1}^{t_a} (\alpha_a + \lambda_a \varepsilon_{ja}) dp_{ja}. \quad (\text{E.7})$$

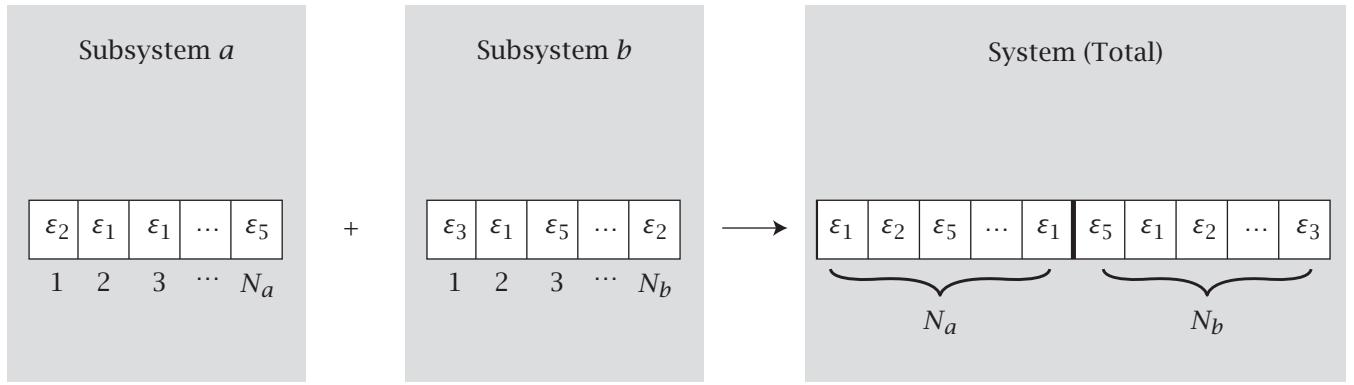


Figure E.1 Combining subsystems into a total system. Subsystem *a* has N_a particles and has a given total score U_a . That score can occur in $W_a(U_a)$ different ways. Subsystem *b* has N_b total particles and has a given total score U_b . That score can occur in $W_b(U_b)$ different ways. For the given values U_a and U_b , the total multiplicity of the combined subsystems is $W_t = W_a W_b$. We seek a mathematical function, $S(W(U))$, that, for a given value $U = U_a$, will, at its maximum, $dS_a = 0$, predict the multiplicity $W = W_a$. The text shows that such a function S : (1) is additive, $S_t = S_a + S_b$, and (2) has the form $S = k \ln W$.

Equation (E.7) does not make any assumption about the mathematical form of $S_a = S_a(p_{1a}^*, p_{2a}^*, \dots, p_{ta}^*)$. It is just the definition of the total differential. Similarly, for subsystem *b*, you have U_b, N_b , score levels ε_{jb} for $j = 1, 2, \dots, t_b$, multiplicity $W_b(U_b)$, and entropy S_b . The score levels ε_{jb} and the numbers of them t_b do not need to be the same as for subsystem *a*. To get the probabilities p_{jb}^* , use the constraints

$$\alpha_b \sum_{j=1}^{t_b} dp_{jb} = 0 \quad \text{and} \quad \lambda_b \sum_{j=1}^{t_b} \varepsilon_{jb} dp_{jb} = 0,$$

with Lagrange multipliers α_b and λ_b . The total differential for S_b is

$$dS_b = \sum_{j=1}^{t_b} \left(\frac{\partial S_b}{\partial p_{jb}} \right) dp_{jb} = \sum_{j=1}^{t_b} (\alpha_b + \lambda_b \varepsilon_{jb}) dp_{jb}. \quad (\text{E.8})$$

Now, define the total system as the combination of subsystems *a* and *b* with exactly the same given constraints on each subsystem as above. The entropy of the total system will be $S_t(p_{1a}^*, p_{2a}^*, p_{3a}^*, \dots, p_{ta}^*, p_{1b}^*, p_{2b}^*, p_{3b}^*, \dots, p_{tb}^*)$. According to the Lagrange method, applying the four subsystem constraints to the full system will give

$$dS_t = \sum_{j=1}^{t_a} (\alpha_a + \lambda_a \varepsilon_{ja}) dp_{ja} + \sum_{j=1}^{t_b} (\alpha_b + \lambda_b \varepsilon_{jb}) dp_{jb} = dS_a + dS_b. \quad (\text{E.9})$$

The last equality follows from using Equations (E.7) and (E.8). Equation (E.9) shows that the entropy of the combined system equals the sum of the entropies of the subsystems. So, the entropy is extensive because the constraint quantities n_j and U themselves are extensive.

The way we divided the system into subsystems a and b was completely arbitrary (provided only that the values of U_a and U_b are given). We could have parsed the system in many different ways, down to an individual event at a time. Following this reasoning to its logical conclusion, and integrating the differentials, leads to Equation (E.4), which shows that the entropy of the whole grid is the sum of entropies of the individual events.

The Entropy Function Must Have the Form $k \ln W$

The values U_a and U_b of two subsystems a and b are fixed independently of each other. The multiplicities are $W_a(U_a)$ and $W_b(U_b)$. The multiplicity of the full system will be the product $W_{\text{tot}}(U_{\text{tot}}) = W_a \times W_b$, where $U_{\text{tot}} = U_a + U_b$. To obtain the mathematical form of the entropy function, let's first simplify the notation. Let $u = W_a$, let $v = W_b$, and let $r = uv = W_a W_b$. In this notation, integrating Equation (E.9) gives

$$S_t(r) = S_a(u) + S_b(v) + \text{constant}. \quad (\text{E.10})$$

We want to show that $S = k \ln W$, where k is a constant, is the only function that can satisfy Equation (E.10). First, taking the derivative of the left side of Equation (E.10) with respect to v gives

$$\frac{dS_t}{dv} = \frac{dS_t}{dr} \frac{\partial r}{\partial v} = \frac{dS_t}{dr} u, \quad (\text{E.11})$$

since $r = uv$ and $(\partial r / \partial v)_u = u$. Now, taking the derivative of the right side of Equation (E.10) with respect to v gives dS_b/dv , since S_a does not depend on v . Setting these two expressions equal gives

$$\frac{dS_t}{dr} u = \frac{dS_b}{dv}. \quad (\text{E.12})$$

Now, taking the corresponding derivatives of Equation (E.10) with respect to u instead gives

$$\frac{dS_t}{dr} v = \frac{dS_a}{du}. \quad (\text{E.13})$$

Setting equal (dS_t/dr) in Equations (E.12) and (E.13) gives

$$u \frac{dS_a(u)}{du} = v \frac{dS_b(v)}{dv} = k, \quad (\text{E.14})$$

where k is a constant. The reason that both terms in Equation (E.14) must equal an identical constant is because the only way that two functions of different variables can equal each other, for arbitrary values of the variables u and v , is if they equal a constant. Rearrange Equation (E.14), to give $dS_a(u) = k du/u$ and $dS_b(v) = k dv/v$, or

$$S_a = k \ln W_a \quad \text{and} \quad S_b = k \ln W_b,$$

so

$$S_t = k \ln(W_a W_b).$$

We are free to choose the constant of integration to be $S(1) = 0$. We conclude that $S/k = \ln W = -\sum p_i \ln p_i$ is the only function that is extremal where W is

maximal and that otherwise ensures that $\sum p_i = 1$ and that $N \sum p_i \varepsilon_i$ equals the given value of U .

Two additional conclusions follow from this logic. First, this reasoning defines another extremum function $Y(p_j)$,

$$dY = dS - \lambda \langle \varepsilon \rangle, \quad (\text{E.16})$$

that does both mathematical jobs at the same time: maximizing W (or S) and satisfying the constraint $U = N\langle \varepsilon \rangle$. In the canonical ensemble in thermal physics, Y is the free energy function. Y has other meanings. For example, in dynamical problems, Y is called the *caliber* [1, 2].

Second, as shown in Chapter 5, this reasoning predicts that probability distributions are exponential in the quantities ε_j :

$$p_j^* = q^{-1} e^{-\lambda \varepsilon_j}. \quad (\text{E.17})$$

Our derivation here uses the frequency interpretation of probabilities. We have imagined replicating the N -string experiment M different times, where M is large. But our derivation does not depend on whether the p_j quantities you use come from some replicable experiment, as in dice rolls, or whether they come from some symmetry or information that underlies some model of interest to you.

References

- [1] K Ghosh, KA Dill, MM Inamdar, E Seitaridou, and R Phillips, *Am J Phys* **74**, 123–133 (2006).
- [2] G Stock, K Ghosh, and KA Dill, *J Chem Phys* **128**, 194102 (2008).

Suggested Reading

- TL Hill, *An Introduction to Statistical Thermodynamics*, Addison-Wesley, Reading, MA, 1960 (reprinted by Dover Publications, New York, 1986), Chapter 1.
- ET Jaynes, Where do we stand on maximum entropy? *The Maximum Entropy Formalism*, RD Levine and M Tribus, eds, MIT Press, Cambridge, MA, 1979, pp. 15–118.

This page is intentionally left blank.

Appendix F: Legendre Transforms

Here, we show the basis for Legendre transforms. A function $y = f(x)$ can be described as a list of pairs $(x_1, y_1), (x_2, y_2), \dots$. Our aim here is to show that you can express the same function instead as a list of different pairs: the slopes $c(x)$ and the intercepts $b(x)$: $(c_1, b_1), (c_2, b_2), \dots$

For a small change dx , the change dy in the function can be described by the slope $c(x)$ at that point:

$$dy = \left(\frac{dy}{dx} \right) dx = c(x) dx. \quad (\text{F.1})$$

Figure F.1 shows how the full function $y(x)$ (not just small changes) can be regarded as a set of slopes and intercepts—one slope $c(x)$ and one intercept $b(x)$ for each point x :

$$y(x) = c(x)x + b(x) \implies b(x) = y(x) - c(x)x. \quad (\text{F.2})$$

We are interested in the function that expresses the series of intercepts versus slopes, $b(c)$. To see how small changes in the slope c lead to small changes in the intercept b , take the differential of Equation (F.2) and substitute Equation (F.1), $dy = c dx$, to get

$$db = dy - c dx - x dc = -x dc. \quad (\text{F.3})$$

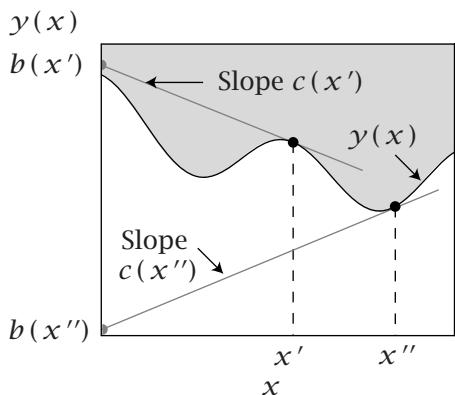


Figure F.1 To create the Legendre transform, a function $y(x)$ is expressed as a tangent slope function $c(x)$, and a tangent intercept function $b(x)$. The tangent slopes and intercepts of points x' and x'' are shown here.

EXAMPLE F.1 Express $y(x)$ in terms of $b(c)$. Suppose $y = 3x^2 + 5$. The slope is $c(x) = dy/dx = 6x$. From Equation (F.2), the intercept is $b(x) = 3x^2 + 5 - (6x)x = -3x^2 + 5$. Substitute $x = c/6$ into $b = -3x^2 + 5$ to get $b(c) = -c^2/12 + 5$. This is an alternative way to express $y = f(x)$. You can check Equation (F.3) for this example by taking the derivative: $db = -(c/6)dc = -xdc$.

Now generalize to a multivariate function $y = y(x_1, x_2, x_3)$. The differential element is

$$dy = c_1 dx_1 + c_2 dx_2 + c_3 dx_3, \quad (\text{F.4})$$

where

$$c_1 = \left(\frac{\partial y}{\partial x_1} \right)_{x_2, x_3}, \quad c_2 = \left(\frac{\partial y}{\partial x_2} \right)_{x_1, x_3}, \quad \text{and} \quad c_3 = \left(\frac{\partial y}{\partial x_3} \right)_{x_1, x_2}.$$

We want the intercept function b_1 along the x_1 axis:

$$b_1(c_1, x_2, x_3) = y - c_1 x_1. \quad (\text{F.5})$$

Take the differential of Equation (F.5) and substitute Equation (F.4) to get

$$db_1 = dy - c_1 dx_1 - x_1 dc_1 = -x_1 dc_1 + c_2 dx_2 + c_3 dx_3. \quad (\text{F.6})$$

From Equation (F.6), you can see that

$$x_1 = -\left(\frac{\partial b_1}{\partial c_1} \right)_{x_2, x_3}, \quad c_2 = \left(\frac{\partial b_1}{\partial x_2} \right)_{c_1, x_3}, \quad \text{and} \quad c_3 = \left(\frac{\partial b_1}{\partial x_3} \right)_{c_1, x_2}. \quad (\text{F.7})$$

Any of the original independent variables x_i can be exchanged with their conjugate variables c_i in this way. The transformation can be performed on any combination of conjugate pairs, so there are a total of $2^n - 1$ possible transformations. Table 8.1 shows useful relations that can be derived from transformed fundamental equations.

Suggested Reading

MM Abbott and HC Van Ness, *Schaum's Outline of Thermodynamics, with Chemical Applications*, 2nd edition,

McGraw-Hill, New York, 1989. Good discussion of Legendre transforms and many problems with solutions.

HB Callen, *Thermodynamics and an Introduction to Thermostatistics*, 2nd edition, Wiley, New York, 1985.

Appendix G: Vectors Describe Forces & Flows

Vector algebra and calculus are useful for describing the forces and flows of fluids, particles, and electrical currents as functions of spatial position and time. The following section gives some of the math.

Spatial Distributions Can Be Described by Scalar and Vector Fields

A field is a physical quantity that can take on different values at different positions in space. A weather map of the temperatures (a scalar quantity) is a *scalar field* (see Figure G.1); a *vector field* may be the set of velocities of a flowing fluid at all points throughout the fluid. Another vector field is the set of wind velocities in the USA, shown in Figure G.2. To find the forces and gradients of scalar or vector fields, we need a calculus for vectors.

Defining the Gradient Vector and the ∇ Operator

Consider a weather map with a distribution of temperatures $T(x, y, z)$ where x and y are the East-West and North-South coordinates on the map. To be general, let's also consider the altitude z . We want to know the difference in temperatures from one point $\mathbf{r} = (x, y, z)$ to another $\mathbf{r} + d\mathbf{r}$, where $d\mathbf{r}$ is given by the vector

$$d\mathbf{r} = \mathbf{i} dx + \mathbf{j} dy + \mathbf{k} dz. \quad (\text{G.1})$$

Recall from Chapter 4 that a change such as dT can be computed as

$$dT = \left(\frac{\partial T}{\partial x}\right) dx + \left(\frac{\partial T}{\partial y}\right) dy + \left(\frac{\partial T}{\partial z}\right) dz. \quad (\text{G.2})$$

You can write Equation (G.2) more simply using the vector dot product. Let one vector be the temperature gradient, written as ∇T , where ∇ is the gradient

Figure G.1 Temperature is a scalar field, as indicated on a weather map. Temperature is a function of coordinates (x, y) .

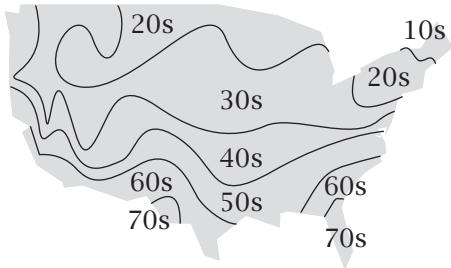
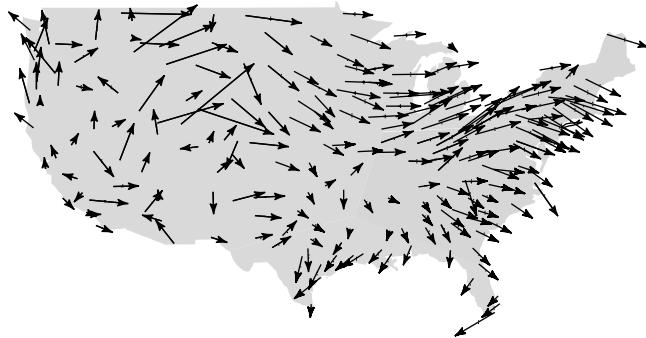


Figure G.2 A vector field: average January wind velocities in the USA. Each vector has an origin, indicating where the wind velocity was measured, and a magnitude and direction, indicating how strongly and in which direction the wind blows there. Source: *Supercomputer Institute Research Bulletin Newsletter of the University of Minnesota* **13**, 6 (1997). Data are from Professor Katherine Klunk of the Department of Geography, University of Minnesota.



operator called ‘del’ or ‘grad’:

$$\nabla T = \mathbf{i} \left(\frac{\partial T}{\partial x} \right) + \mathbf{j} \left(\frac{\partial T}{\partial y} \right) + \mathbf{k} \left(\frac{\partial T}{\partial z} \right). \quad (\text{G.3})$$

For example, suppose $T(x, y, z) = 1 - ax$: the temperature decreases linearly along the x axis, and is independent of y and z . Then $\nabla T = -\mathbf{i}a$ is the slope of the function along the x axis. Now dT in Equation (G.2) is the dot product of the temperature gradient with the change-of-position vector $d\mathbf{r}$ given by Equation (G.1):

$$dT = \nabla T \cdot d\mathbf{r}. \quad (\text{G.4})$$

The gradient of a scalar function is a vector that tells how the function varies in every direction in the neighborhood of a point. For example, a vector pointing from Dallas to Fort Worth, Texas might have a magnitude of -10 , indicating that the temperature is ten degrees lower in Fort Worth than in Dallas. The gradient is perpendicular to lines of constant temperature.

∇ is a *vector operator*, which itself can be manipulated as a vector. The **gradient vector** ∇ is defined by

$$\nabla = \mathbf{i} \left(\frac{\partial}{\partial x} \right) + \mathbf{j} \left(\frac{\partial}{\partial y} \right) + \mathbf{k} \left(\frac{\partial}{\partial z} \right). \quad (\text{G.5})$$

∇ operates on the quantity that follows it. The notation ∇T means to perform the ∇ operation on T . How the ∇ operator is applied depends on whether it is operating on a scalar or on a vector function. For a scalar function, Equation (G.5) defines ∇ as the following operation: take the partial derivative of the function with respect to each Cartesian coordinate and multiply it by the unit vector in the direction of the coordinate, then sum the products. The result is the *gradient* vector. Now let’s find how the ∇ operator acts on a vector field.

Defining the Divergence

Consider a field of vectors, rather than a field of a scalar quantity. A vector field is a set of vectors distributed throughout space, pointing in various directions. For example, Figure G.2 shows the vector field of wind velocities throughout the USA. Each vector on this diagram conveys three pieces of information: (1) the location (indicated by the spatial position of the vector), (2) the magnitude of the wind velocity at that location (indicated by the vector length), and (3) the direction in which the wind blows (indicated by the vector direction).

Now we want to know the rates of change at any point for a field of vectors. This function is called the *divergence*. The divergence, which is a scalar quantity, is the dot product of the vector operator ∇ with the vector on which it operates. In Cartesian coordinates, the divergence of the vector \mathbf{u} is

$$\begin{aligned}\nabla \cdot \mathbf{u} &= \left[\mathbf{i} \left(\frac{\partial}{\partial x} \right) + \mathbf{j} \left(\frac{\partial}{\partial y} \right) + \mathbf{k} \left(\frac{\partial}{\partial z} \right) \right] \cdot (\mathbf{i}u_x + \mathbf{j}u_y + \mathbf{k}u_z) \\ &= \left(\frac{\partial u_x}{\partial x} \right) + \left(\frac{\partial u_y}{\partial y} \right) + \left(\frac{\partial u_z}{\partial z} \right).\end{aligned}\quad (G.6)$$

Figures G.3 and G.4 illustrate the concept of divergence. In a constant vector field (Figure G.3), all the vectors point in the same direction and have the same magnitude. The divergence is zero, because there is no change in the vectors from one position in space to another. This might represent the water flow in the middle of a river. In contrast, as shown in Figure G.4, an idealized point source of fluid (e.g., a wellspring or fountainhead) or point sink (e.g., a drain-hole) causes the velocity vectors to diverge, to point in different directions from one point in space to the next. Near a point source or sink, the divergence is not zero. The divergence describes the extent to which a vector field behaves like a source or sink, i.e., the extent to which more flow exits than enters a local area.

EXAMPLE G.1 Computing the divergence. Compute the divergence $\nabla \cdot \mathbf{u}$ for the vector $\mathbf{u} = x^2\mathbf{i} + y^2\mathbf{j} + z^2\mathbf{k}$. Equation (G.6) gives

$$\nabla \cdot \mathbf{u} = 2x + 2y + 2z.$$

$$\nabla \cdot \mathbf{u} = 0$$

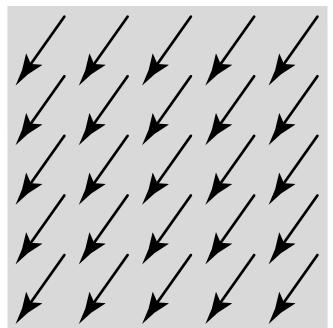
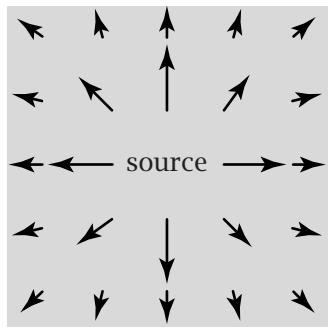


Figure G.3 The divergence is zero for a constant vector field.

$$(a) \nabla \cdot \mathbf{u} > 0$$



$$(b) \nabla \cdot \mathbf{u} < 0$$

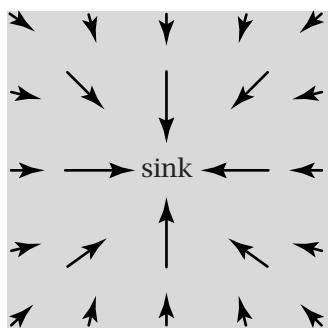


Figure G.4 The divergence of a vector field is (a) positive for a point source, and (b) negative for a point sink.

The Laplacian Operator

The Laplacian operator is the dot product of ∇ with itself. It is written ∇^2 and pronounced ‘del squared.’ For Cartesian coordinates, the Laplacian operator is

$$\nabla \cdot \nabla = \left(\mathbf{i} \frac{\partial}{\partial x} + \mathbf{j} \frac{\partial}{\partial y} + \mathbf{k} \frac{\partial}{\partial z} \right) \cdot \left(\mathbf{i} \frac{\partial}{\partial x} + \mathbf{j} \frac{\partial}{\partial y} + \mathbf{k} \frac{\partial}{\partial z} \right)$$

$$\implies \nabla^2 = \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial z^2}. \quad (\text{G.7})$$

If $T(x, y, z)$ is a scalar field, $\nabla^2 T$ is another scalar field given by

$$\begin{aligned}\nabla^2 T &= \left(\mathbf{i} \frac{\partial}{\partial x} + \mathbf{j} \frac{\partial}{\partial y} + \mathbf{k} \frac{\partial}{\partial z} \right) \left(\mathbf{i} \frac{\partial T}{\partial x} + \mathbf{j} \frac{\partial T}{\partial y} + \mathbf{k} \frac{\partial T}{\partial z} \right) \\ &= \frac{\partial^2 T}{\partial x^2} + \frac{\partial^2 T}{\partial y^2} + \frac{\partial^2 T}{\partial z^2}.\end{aligned}$$

EXAMPLE G.2 Computing the Laplacian. For the function $f(x, y, z) = x^2 + y^3 + z^4$, compute $\nabla^2 f$. Equation (G.7) gives

$$\nabla^2 f = 2 + 6y + 12z^2.$$

Operators in Different Coordinate Systems

Choosing a coordinate system appropriate to the symmetry of the problem at hand can make problems easier to solve. Cartesian coordinates are usually easiest for problems involving planes and flat surfaces. For problems involving spheres or cylinders, spherical or cylindrical coordinates are most convenient. The advantage of the operator notation, ∇ , $\nabla \cdot$, and ∇^2 , is that it is independent of the coordinate system. But when it comes to performing specific calculations, you can use Table G.1 at the end of this appendix to translate from the vector notation, such as ∇T , to the specific expressions that you need for a particular problem.

Integrating Vector Quantities: The Path Integral

When you have a function $f(x)$, you may need to integrate over x , $\int f(x) dx$. Similarly, you sometimes need to integrate vector quantities. For vectors, you integrate by components. A *path integral* (also called a *line integral*) involves a dot product of two vectors over a particular vector pathway. For example, if you have a vector $\mathbf{u} = (u_x, u_y, u_z)$ to be integrated along $d\mathbf{r} = (dx, dy, dz)$, then the *path integral* is

$$\int \mathbf{u} \cdot d\mathbf{r} = \int u_x dx + u_y dy + u_z dz. \quad (\text{G.8})$$

(Note that the term ‘path integral’ is also used in the literature with another meaning, namely as a so-called ‘functional integral.’)

EXAMPLE G.3 A simple path integral: the work performed over a path from A to B. The work w equals the path integral of the vector force \mathbf{f} acting on an object, projected onto each differential element of the path $d\ell$:

$$w = \int_A^B \mathbf{f} \cdot d\ell. \quad (\text{G.9})$$

The work of the force on the object is positive, $w > 0$, if the force acts in the direction of positive $d\ell$, and negative if the force acts in the opposite direction (see Equation (3.9)). In Figure G.5, the path is diagonally downward.

Suppose that gravity acts downward with a force $f_g = 2 \text{ N}$. The vector force is

$$\mathbf{f} = f_x \mathbf{i} + f_y \mathbf{j} = -f_g \mathbf{j}.$$

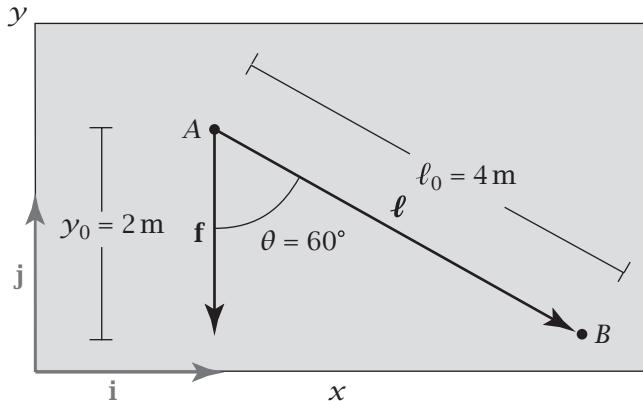


Figure G.5 An object is on a 4 m long plane inclined at 60° from vertical. The gravitational force f_g acts vertically downwards. The work exerted by the force in moving the object from A to B is calculated in Example G.3.

Suppose the downward force is pushing an object along an inclined plane at an angle $\theta = 60^\circ$ (see Figure G.5) over a distance of $\ell_0 = 4 \text{ m}$. The work is

$$\begin{aligned} w &= \int \mathbf{f} \cdot d\ell = \int (f_x \mathbf{i} + f_y \mathbf{j}) (dx \mathbf{i} + dy \mathbf{j}) \\ &= - \int_0^{-y_0} f_g dy = f_g y_0 = f_g \ell_0 \cos \theta = (2 \text{ N})(4 \text{ m})(\cos 60^\circ) \\ &= 4 \text{ N m.} \end{aligned} \quad (\text{G.10})$$

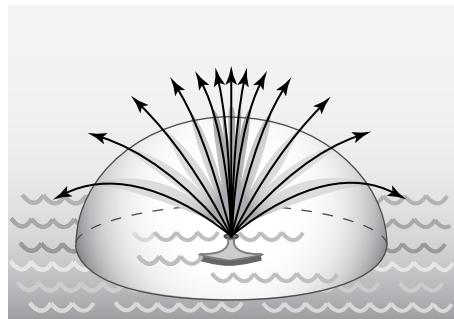
In this way, work is computed as a path integral of a force acting over a distance. The path integral also treats cases where the force varies along the trajectory.

The Flux of a Vector Field Is Its ‘Flow’ Through a Surface

Vectors are useful for describing the flows of fluids and particles. The central concept that you need is the *flux*, which is a familiar scalar quantity. The flux of water through a garden hose is the volume of water per unit time that flows through an imaginary plane of unit area that cuts perpendicularly through the hose. For fluid flow, flux has units of (volume of fluid)/[(time)(unit area)].

You can also define the flux in more complex situations, such as a water fountain in a pool (Figure G.6). You first need to define a surface through which there is flux. For the fountain, invent an imaginary hemispherical surface, or

Figure G.6 In a fountain, the water flux through an imaginary hemispherical surface is highest at the top and smallest at the sides.



'balloon,' that encloses it. The flux is the amount of water that flows through this surface per unit time, per unit area. Because water fountains are directed upwards, the flux is greatest through the 'north pole' of the hemispherical balloon, and is nearly zero through the 'equator.' To compute the total flux J , you need to integrate the different amounts of flux passing through the many different infinitesimal elements of the imaginary surface.

Let's describe the total water flux through the imaginary bounding surface mathematically. Consider the flow of a fluid through a small element of a surface. Two vectors define the liquid movement. One is the fluid velocity vector \mathbf{v} : its magnitude represents the speed of the fluid and its orientation indicates the direction of fluid flow. The second vector is $d\mathbf{s}$, which defines the size and orientation of the imaginary small plane element through which the fluid is flowing. $d\mathbf{s}$ is an infinitesimal quantity but it obeys the same mathematical rules as any other vector. The magnitude of $d\mathbf{s}$ represents the infinitesimal area of the imaginary surface through which the fluid flows. The orientation of $d\mathbf{s}$ is normal to the imaginary plane, indicating how the planar element is oriented in space. Because there are two possible normals to the surface, the convention for a closed surface is that $d\mathbf{s}$ is positive when it points *outward*. Flux is positive for flow *out* of a closed surface and negative for flow *into* a closed surface.

The dot product of these two vectors—the flow velocity vector \mathbf{v} and the area element vector $d\mathbf{s}$ —gives the infinitesimal flux of fluid $dJ = \mathbf{v} \cdot d\mathbf{s}$ through the area element. If the surface area is rotated so that the vector $d\mathbf{s}$ forms an angle α with respect to the direction of fluid flow (Figure G.7), then the flux is $dJ = \mathbf{v} \cdot d\mathbf{s} = v ds \cos \alpha$, where v is the scalar magnitude of the flow velocity and ds is the scalar magnitude of the area of the small element. If the area element is normal to the flow direction, then $dJ = v ds$. Here, the flux is a scalar quantity.

To find the total flow out through the balloon surface, integrate the flux over all the surface elements. Represent the entire surface as a large number of small flat areas each with its own normal vector given by $d\mathbf{s}_i$ as shown in Figure G.8. The total flux J is the sum over each of the individual fluxes. Because the individual fluxes are infinitesimal, this sum becomes an integral, and the total flux is

$$J = \int_{\text{surface}} \mathbf{v} \cdot d\mathbf{s}. \quad (\text{G.11})$$

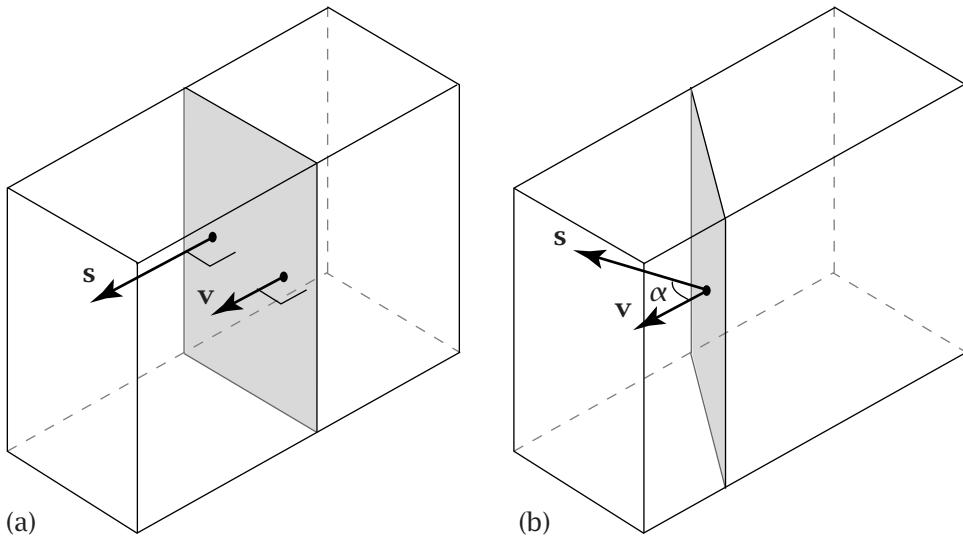


Figure G.7 The flux of a vector quantity \mathbf{v} is the dot product of \mathbf{v} with \mathbf{s} , a vector having direction normal to a planar element and magnitude equal to its area. The flux is (a) maximal if \mathbf{v} is parallel to \mathbf{s} and (b) reduced by $\cos \alpha$ if the vectors are oriented at an angle α with respect to each other.

EXAMPLE G.4 Computing the flux. Suppose you have a field of vectors that all point radially outward from an origin at $r = 0$. Each vector has magnitude $v(r) = a/r$. What is the flux through a spherical balloon with radius R centered at $r = 0$? All the volume elements have identical fluxes, so in this case Equation (G.11) becomes

$$J = v(R)4\pi R^2 = \left(\frac{a}{R}\right)(4\pi R^2) = 4\pi aR.$$

The next section describes *Gauss's theorem*, a mathematical result that sometimes simplifies the calculation of flux.

Gauss's Theorem Relates a Volume Property to a Surface Property

Flux is a *surface* property. It tells you how much of something passes through an imaginary surface. Gauss's theorem shows how to replace this surface property with a *volume* property, the divergence of the field inside the surface.

Consider a field of velocities \mathbf{v} , for example the velocities of fluid flow through an infinitesimal cube having volume $\Delta V = \Delta x \Delta y \Delta z$ (see Figure G.9). Our balloon is now a surface with a cubical shape. Suppose the cube has its left face at $x = 0$, its front face at $y = 0$, and its bottom face at $z = 0$. To compute the flux, you need the dot products of the vector field and the six unit normal surface vectors for the cube. The flux through each surface is the product of the surface area multiplied by the component of the vector field that is perpendicular to it.

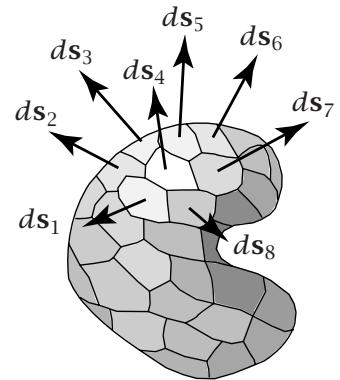


Figure G.8 The flux out of a surface is the sum of the fluxes out of infinitesimal patches of the surface.

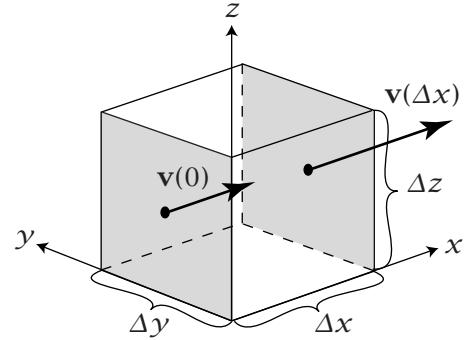


Figure G.9 To prove Gauss's theorem, construct a cube with its left face at $x = 0$, its right face at $x = \Delta x$, its front face at $y = 0$, its back face at $y = \Delta y$, its bottom face at $z = 0$, and its top face at $z = \Delta z$.

First compute the flux through the two surfaces parallel to the (y, z) plane (see Figure G.9). Because the unit normal vectors are in the directions of the coordinate axes, you need only to keep track of their direction in the plus or minus sense. The area parallel to the (y, z) plane is $\Delta y \Delta z$. The outward flux through the right-hand face at $x = \Delta x$ is $v_x(\Delta x) \Delta y \Delta z$. Through the left-hand face at $x = 0$, the outward flux is $-v_x(0) \Delta y \Delta z$. In each case, v_x is the x component of the field at the center of the face. Because Δx is taken to be very small, spatial variations of v_x can be expanded as a Taylor series. To first-order approximation, v_x varies linearly with x :

$$v_x(\Delta x) = v_x(0) + \left(\frac{\partial v_x}{\partial x} \right) \Delta x. \quad (\text{G.12})$$

Multiply both sides of Equation (G.12) by $\Delta y \Delta z$ to get the fluxes at $x = 0$ and Δx , and rearrange:

$$[v_x(\Delta x) - v_x(0)] \Delta y \Delta z = \left(\frac{\partial v_x}{\partial x} \right) \Delta x \Delta y \Delta z. \quad (\text{G.13})$$

You can compute the flux J in two different ways. First, from Equation (G.11), you have $J = \int_S \mathbf{v} \cdot d\mathbf{s}$. Second, Equation (G.13) shows that the x -direction flux is $(\partial v_x / \partial x) \Delta x \Delta y \Delta z$. Because the contributions along the y and z axes are calculated in the same way, the total flux through all six faces is

$$\begin{aligned} \int_S \mathbf{v} \cdot d\mathbf{s} &= \left[\left(\frac{\partial v_x}{\partial x} \right) + \left(\frac{\partial v_y}{\partial y} \right) + \left(\frac{\partial v_z}{\partial z} \right) \right] \Delta x \Delta y \Delta z \\ &= (\nabla \cdot \mathbf{v}) \Delta V. \end{aligned} \quad (\text{G.14})$$

Equation (G.14) gives the flux through a small volume element.

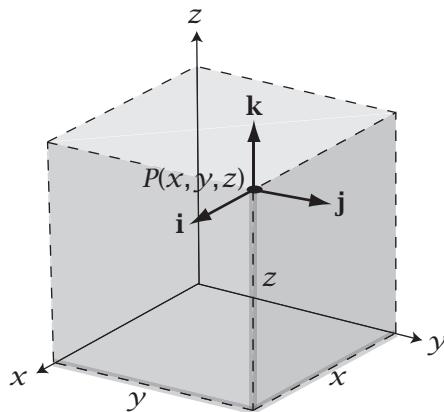
Now integrate over all the infinitesimal volume elements. The left-hand side of Equation (G.14) must be integrated over the whole surface S , and the right-hand side over the whole volume V . Because the flux into one volume element equals the flux out of an adjacent element, the only component of the surface integral that is nonzero is the one that represents the outer surface. This integration gives **Gauss's theorem**:

$$\int_S \mathbf{v} \cdot d\mathbf{s} = \int_V \nabla \cdot \mathbf{v} dV. \quad (\text{G.15})$$

Gauss's theorem equates the *flux* of a vector field *through a closed surface* with the *divergence* of that same field *throughout its volume*.

Table G.1 For a scalar f or a vector \mathbf{u} , this table gives the gradient ∇f , the divergence $\nabla \cdot \mathbf{u}$, and the Laplacian $\nabla^2 f$ at the point P in Cartesian, cylindrical, and spherical coordinates.

Cartesian Coordinates (x, y, z)



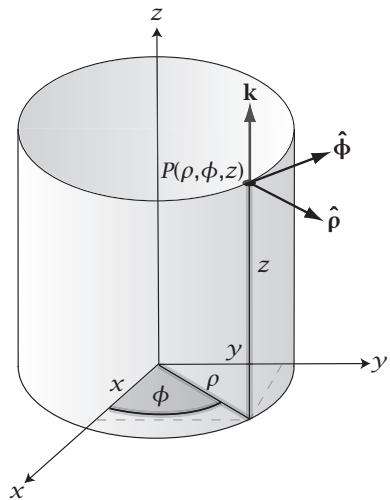
\mathbf{i} , \mathbf{j} , and \mathbf{k} are the unit vectors in the x , y , and z directions, respectively; u_x , u_y , and u_z are the components of \mathbf{u} in these directions

$$\nabla f = \frac{\partial f}{\partial x} \mathbf{i} + \frac{\partial f}{\partial y} \mathbf{j} + \frac{\partial f}{\partial z} \mathbf{k} \quad (\text{G.16})$$

$$\nabla \cdot \mathbf{u} = \frac{\partial u_x}{\partial x} + \frac{\partial u_y}{\partial y} + \frac{\partial u_z}{\partial z} \quad (\text{G.17})$$

$$\nabla^2 f = \frac{\partial^2 f}{\partial x^2} + \frac{\partial^2 f}{\partial y^2} + \frac{\partial^2 f}{\partial z^2} \quad (\text{G.18})$$

Cylindrical Coordinates (ρ, ϕ, z)



$$\begin{aligned}x &= \rho \cos \phi \\y &= \rho \sin \phi \\z &= z\end{aligned}$$

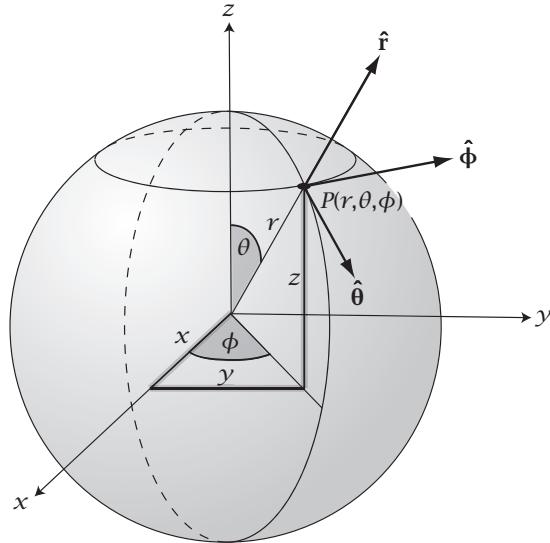
$\hat{\rho}$, $\hat{\phi}$, and \mathbf{k} are the unit vectors in the ρ , ϕ , and z directions, respectively; u_ρ , u_ϕ , and u_z are the components of \mathbf{u} in these directions

$$\nabla f = \left(\frac{\partial f}{\partial \rho} \right) \hat{\rho} + \frac{1}{\rho} \left(\frac{\partial f}{\partial \phi} \right) \hat{\phi} + \left(\frac{\partial f}{\partial z} \right) \mathbf{k} \quad (\text{G.19})$$

$$\nabla \cdot \mathbf{u} = \frac{1}{\rho} \left(\frac{\partial(\rho u_\rho)}{\partial \rho} \right) + \frac{1}{\rho} \left(\frac{\partial u_\phi}{\partial \phi} \right) + \left(\frac{\partial u_z}{\partial z} \right) \quad (\text{G.20})$$

$$\nabla^2 f = \frac{1}{\rho} \left[\frac{\partial}{\partial \rho} \left(\rho \frac{\partial f}{\partial \rho} \right) \right] + \frac{1}{\rho^2} \left(\frac{\partial^2 f}{\partial \phi^2} \right) + \left(\frac{\partial^2 f}{\partial z^2} \right) \quad (\text{G.21})$$

Spherical Coordinates (r, θ, ϕ)



$$\begin{aligned}x &= r \sin \theta \cos \phi \\y &= r \sin \theta \sin \phi \\z &= r \cos \theta \\r &= \sqrt{x^2 + y^2 + z^2}\end{aligned}$$

\hat{r} , $\hat{\theta}$, and $\hat{\phi}$ are the unit vectors in the r , θ , and ϕ directions, respectively; u_r , u_θ , and u_ϕ are the components of \mathbf{u} in these directions

$$\nabla f = \left(\frac{\partial f}{\partial r} \right) \hat{r} + \frac{1}{r} \left(\frac{\partial f}{\partial \theta} \right) \hat{\theta} + \frac{1}{r \sin \theta} \left(\frac{\partial f}{\partial \phi} \right) \hat{\phi} \quad (\text{G.22})$$

$$\nabla \cdot \mathbf{u} = \frac{1}{r^2} \left(\frac{\partial(r^2 u_r)}{\partial r} \right) + \frac{1}{r \sin \theta} \left(\frac{\partial[(\sin \theta) u_\theta]}{\partial \theta} \right) + \frac{1}{r \sin \theta} \left(\frac{\partial u_\phi}{\partial \phi} \right) \quad (\text{G.23})$$

$$\begin{aligned}\nabla^2 f = & \frac{1}{r^2} \left[\frac{\partial}{\partial r} \left(r^2 \frac{\partial f}{\partial r} \right) \right] + \frac{1}{r^2 \sin \theta} \left[\frac{\partial}{\partial \theta} \left(\sin \theta \frac{\partial f}{\partial \theta} \right) \right] \\ & + \frac{1}{r^2 \sin^2 \theta} \left(\frac{\partial^2 f}{\partial \phi^2} \right)\end{aligned}\quad (\text{G.24})$$

Note that a variety of different notations for cylindrical and spherical coordinates can be found in the literature—for example, some authors use ϕ for the angle denoted here by θ , and vice versa.

Suggested Reading

HM Schey, *Div, Grad, Curl, and All That*, 4th edition, WW Norton, New York, 2005.

MR Spiegel, S Lipschutz, and D Spellman, *Schaum's Outline of Vector Analysis*, 2nd edition, McGraw-Hill, New York, 2009.

This page is intentionally left blank.

Appendix H:

Table of Constants

Quantity	Symbol and Equivalences	Value
Gas constant	R	$8.31447 \text{ JK}^{-1} \text{ mol}^{-1}$ $8.31447 \times 10^7 \text{ erg K}^{-1} \text{ mol}^{-1}$ $1.98717 \text{ cal K}^{-1} \text{ mol}^{-1}$ $8.20575 \times 10^{-5} \text{ m}^3 \text{ atm K}^{-1} \text{ mol}^{-1}$ $8.20575 \times 10^{-2} \text{ L atm K}^{-1} \text{ mol}^{-1}$ $82.0575 \text{ cm}^3 \text{ atm K}^{-1} \text{ mol}^{-1}$
Boltzmann's constant	$k = R/\mathcal{N}$	$1.3806526 \times 10^{-23} \text{ JK}^{-1}$ $0.138 \text{ pN nm K}^{-1}$
Avogadro's constant	\mathcal{N}	$6.0221367 \times 10^{23} \text{ mol}^{-1}$
Planck's constant	h	$6.626176 \times 10^{-34} \text{ Js}$
Speed of light (vacuum)	c	$2.99792458 \times 10^8 \text{ m s}^{-1}$ (exact)
Standard acceleration of free fall	g	9.80665 m s^{-2} (exact)
Permittivity of vacuum	ϵ_0	$8.85418782 \times 10^{-12} \text{ F m}^{-1}$
Elementary unit charge	e	$1.6021892 \times 10^{-19} \text{ C}$
Atomic mass unit	u	$1.6605387 \times 10^{-27} \text{ kg}$
Mass of proton	m_p	$1.67262158 \times 10^{-27} \text{ kg}$
Rest mass of electron	m_e	$9.109534 \times 10^{-31} \text{ kg}$
Faraday's constant	$F = \mathcal{N}e$	$9.648534 \times 10^4 \text{ C mol}^{-1}$

This page is intentionally left blank.

Appendix I:

Table of Units

Physical Quantity	Unit Name	Unit Symbol	Definition
Force	newton	N	$J\text{m}^{-1} = \text{kg m s}^{-2} = 10^5 \text{ dyn}$
	dyne	dyn	g cm s^{-2}
Pressure	pascal	Pa	$\text{kg m}^{-1} \text{s}^{-2} = \text{N m}^{-2}$ $= 9.87 \times 10^{-6} \text{ atm}$
	atmosphere	atm	$1.01325 \times 10^6 \text{ g cm}^{-1} \text{s}^{-2}$ $= 1.01325 \times 10^5 \text{ Pa}$
Energy	joule	J	$\text{kg m}^2 \text{s}^{-2} = \text{N m} = 10^7 \text{ erg}$
	erg	erg	$\text{g cm}^2 \text{s}^{-2} = 10^{-7} \text{ J}$
	electron volt	eV	$1.602177 \times 10^{-19} \text{ J}$ $= 23.0605 \text{ kcal mol}^{-1}$
	calorie	cal	$4.184 \text{ J} = 4.184 \times 10^7 \text{ erg}$
Electric charge	coulomb	C	$\text{s A} = \text{J V}^{-1}$
	esu	esu	$3.00 \times 10^9 \text{ C}$
Electric potential	volt	V	$\text{kg m}^2 \text{s}^{-3} \text{A}^{-1} = \text{J A}^{-1} \text{s}^{-1}$ $= \text{J C}^{-1}$
Electric current	ampere	A	C s^{-1}
Capacitance	farad	F	$\text{m}^{-2} \text{kg}^{-1} \text{s}^4 \text{A}^2 = \text{A s V}^{-1}$ $= \text{C V}^{-1}$
Frequency	hertz	Hz	s^{-1}
Length	angstrom	Å	$10^{-10} \text{ m} = 10^{-8} \text{ cm}$

Physical Quantity	Unit Name	Unit Symbol	Definition
Volume	liter	L	1000 cm^3
Viscosity	poise	p	$\text{g cm}^{-1} \text{ s}^{-1} = \text{dyn s cm}^{-2}$
Power	watt	W	$\text{J s}^{-1} = \text{kg m}^2 \text{ s}^{-3} = 10^7 \text{ erg s}^{-1}$
	horsepower	hp	745.7 W
Useful Quantities			
Length	Bjerrum length	l_B	560 Å in vacuum 7.13 Å in water at 25°C

m = meter
 s = second
 g = gram

Appendix J: Useful Taylor Series Expansions

$$\begin{aligned} e^{ax} &= 1 + ax + \frac{(ax)^2}{2!} + \frac{(ax)^3}{3!} + \dots \\ &= \sum_{k=0}^{\infty} \frac{(ax)^k}{k!} \quad \text{for all } x \end{aligned} \tag{J.1}$$

$$\begin{aligned} \sin ax &= ax - \frac{(ax)^3}{3!} + \frac{(ax)^5}{5!} - \frac{(ax)^7}{7!} + \dots \\ &= \sum_{k=0}^{\infty} \frac{(-1)^k (ax)^{2k+1}}{(2k+1)!} \quad \text{for all } x \end{aligned} \tag{J.2}$$

$$\begin{aligned} \cos ax &= 1 - \frac{(ax)^2}{2!} + \frac{(ax)^4}{4!} - \frac{(ax)^6}{6!} + \dots \\ &= \sum_{k=0}^{\infty} \frac{(-1)^k (ax)^{2k}}{(2k)!} \quad \text{for all } x \end{aligned} \tag{J.3}$$

$$\begin{aligned} \ln(a+x) &= \ln \left[a \left(1 + \frac{x}{a} \right) \right] = \ln(a) + \frac{x}{a} - \frac{x^2}{2a^2} + \frac{x^3}{3a^3} - \dots + \dots \\ &= \ln(a) + \sum_{k=1}^{\infty} \frac{(-1)^{k-1} x^k}{ka^k} \quad \text{for all } a > 0 \\ &\quad \text{and } -a < x \leq a \end{aligned} \tag{J.4}$$

$$\begin{aligned} \tan^{-1} x &= x - \frac{x^3}{3} + \frac{x^5}{5} - \frac{x^7}{7} + \dots \\ &= \sum_{k=0}^{\infty} \frac{(-1)^k x^{2k+1}}{2k+1} \quad \text{for } |x| \leq 1 \end{aligned} \tag{J.5}$$

$$(1+x)^p = 1 + px + \frac{p(p-1)}{2!}x^2 + \frac{p(p-1)(p-2)}{3!}x^3 + \dots$$

for all p and $|x| < 1$ (J.6)

Equation (J.6) is the binomial series. When p is a positive integer, all the terms after the $(p+1)$ th equal zero, and the series becomes finite:

$$\begin{aligned}(1+x)^p &= 1 + px + \frac{p(p-1)}{2!}x^2 + \frac{p(p-1)(p-2)}{3!}x^3 + \dots + x^p \\ &= \sum_{k=0}^p \binom{p}{k} x^k \quad \text{for all } x\end{aligned}$$

Appendix K: Useful Integrals

$$\int_0^\infty e^{-ax^2} dx = \frac{1}{2} \sqrt{\frac{\pi}{a}} \quad (\text{K.1})$$

$$\int_0^\infty xe^{-ax^2} dx = \frac{1}{2a} \quad (\text{K.2})$$

$$\int_0^\infty x^2 e^{-ax^2} dx = \frac{1}{4a} \sqrt{\frac{\pi}{a}} \quad (\text{K.3})$$

$$\int \sin ax \cos ax dx = \frac{\sin^2 ax}{2a} \quad (\text{K.4})$$

$$\int \sin^n ax \cos ax dx = \frac{\sin^{n+1} ax}{(n+1)a} \quad \text{for } n \neq -1 \quad (\text{K.5})$$

$$\int \cos^n ax \sin ax dx = -\frac{\cos^{n+1} ax}{(n+1)a} \quad \text{for } n \neq -1 \quad (\text{K.6})$$

$$\int \sin^2 ax dx = \frac{x}{2} - \frac{\sin 2ax}{4a} \quad (\text{K.7})$$

$$\int \cos^2 ax dx = \frac{x}{2} + \frac{\sin 2ax}{4a} \quad (\text{K.8})$$

$$\int_0^\infty x^n e^{-ax} dx = \frac{\Gamma(n+1)}{a^{n+1}} \quad (\text{K.9})$$

$$\int_0^\infty x^m e^{-ax^2} dx = \frac{\Gamma((m+1)/2)}{2a^{(m+1)/2}} \quad (\text{K.10})$$

$$\int e^{ax} dx = \frac{1}{a} e^{ax} \quad (\text{K.11})$$

This page is intentionally left blank.

Appendix L: Multiples of Units, Their Names, & Symbols

Multiple	Prefix	Symbol	Multiple	Prefix	Symbol
10^{24}	yotta	Y	10^{-1}	deci	d
10^{21}	zetta	Z	10^{-2}	centi	c
10^{18}	exa	E	10^{-3}	milli	m
10^{15}	peta	P	10^{-6}	micro	μ
10^{12}	tera	T	10^{-9}	nano	n
10^9	giga	G	10^{-12}	pico	p
10^6	mega	M	10^{-15}	femto	f
10^3	kilo	k	10^{-18}	atto	a
10^2	hecto	h	10^{-21}	zepto	z
10	deca	da	10^{-24}	yocto	y

This page is intentionally left blank.

Index

Note: page numbers in *italic* refer to figures and tables.

- absolute entropy of argon 209–10
- absolute rate theory *see* transition-state theory
- absolute temperature scale 128
- absorbing boundary condition 315
- absorbing conditions 691
- absorption
 - of electromagnetic radiation 193
 - of heat 52
- accessible states 177
- acids
 - Bronsted 375–7
 - dissociation near oil/water interface 420, 441
- acid-base equilibria 438–9
 - association/dissociation constants 439
 - influence of electrical fields 439–41
 - apparent equilibrium constant 440
 - near oil/water interface 420, 441
- activated state
 - activation barrier 361, 379
 - activation energy 360–1
 - formation of hydrogen iodide 361
 - free energy 368
 - activation enthalpy 368–9
 - activation entropy 368–9
- activator ligand 591, 592–3
- activator protein (CAP) 596
- active transport 313
- activity coefficient 275, 287–8, 467
- Adair equation 586
- Adair, G. 586
- addition rules, probability 3–4, 8, 9
- adiabatic boundary 94
- adiabatic process 116–17
- adsorbates 541
- adsorption 541
 - BET model of multilayers of molecules 573–4
 - binding in solution 546–8
 - chromatography 548–9
 - gas molecules on a surface 541–5
 - nitrogen on silica 574
 - polymer on a surface 692–3
 - see also* binding processes
- affine network deformation model 670–3
- aggregation 456
 - as cooperative process 570
 - of oil chains 674
 - steric stabilization against 686
- alcohol-water mixtures *see* water-alcohol
- allometric scaling 44
- allosteric binding models 587–590
- allosteric regulation 586
- alloys 276, 278
 - ordering in 520–2
- alphabet
 - letters of 11–12
 - and word sequence 13–14
- ampere (A) 733
- anesthetic molecules 248
- partitioning 297
- angstrom (Å) 733
- animals, size of 43
- anode 434
- antiferromagnetism 525
- apparent equilibrium constant 440
- apparent partial molar volume 162
- apportionment, fair 90, 709–15
- approximation 699–701
 - Bragg-Williams 272–3, 526
 - first-order 699–701
 - mean-field 272–3
 - method for making 699–701
 - second-order 699–701
 - Stirling's 82, 98, 703
- argon
 - absolute entropy 209–10
 - degeneracy of states 229–30
 - energy levels for 197–8
 - packing/structure 618
 - partitioning 299
 - physical properties 621
 - translational partition function 200
- Arrhenius kinetics 360–3
 - Arrhenius plot 361
- Arrhenius, S. 360
- aspirin 439
- assembling charges, method of 444–5
- association
 - constant 439
 - of molecules 302
 - nonpolar solutes in water 635–6
- asymmetric phase diagrams 653
- atmosphere (atm) 733
- atmospheric pressure 173–4
- atomic energy states 206
- atomic mass unit 731
- atomic spectroscopy 193
- ATP hydrolysis 598
- ATPase engine/motor protein 598–600
- average exponent 705
- average particle energy 183
- average values 18–19, 20–1
- averages, from derivatives 705–6
- Avogadro's constant 731
- barometric pressure 173–4
- basal metabolic rate 43
- base, Bronsted 375–7
- bath
 - heat 132, 228
 - particle 132
- batteries 432–4
 - common types 434–6
 - made from salt solutions 436–7
- Bayes' rule 8, 9
- Beams, J.W. 317
- bending correlations, chain 662–5
- benzene
 - in perfluoroheptane 493
 - physical properties 621

benzene (*continued*)

 solubility in water 631, 632

 vapor pressure 259–60

 Bernoulli distribution 22

 Bernoulli trials 16

 BET adsorption model 573–4

 bias, evidence of 33

 coins 89

 dice rolls 84–5, 88–9

 bimolecular reaction 358

 binding polynomials 559–61, 581

 binding polynomial Q 561

 computing 567–70

 and grand canonical ensemble 578–80

 binding processes 541

 binding affinity 544, 546

 binding equilibrium 239–40, 247–8

 binding sites 543

 and Brownian ratchet 348–9

 cooperative ligand binding 559–60

 Hill plot 569–70

 and ligation states 565–6

 modeling 562–5

 Scatchard plot 568–9

 coupled binding 585

 linkage theory 593–4

 multisite ligands 575–8

 see also adsorption

 binodal curve 502

 binomial distribution 16–17, 23

 biochemical machines 585

 F1-ATPase motor protein 598–600

 hemoglobin–oxygen binding 586–9

 kinesin molecular motor 601–3

 biochemical switching 595–8

 biological membranes *see* cell membranes;

 semipermeable membranes

 biomolecular machines *see* biochemical machines

 birth processes 321

 Bjerrum length 391–2, 734

 Black, J. 48

 boiling 253–8, 505–8

 at high altitude 260

 boiling point

 computing molecular attractions from 257

 elevation of 289–90, 291

 enthalpies of 260, 261

 liquid mixtures 510–13

 order parameter 520, 521

 Boltzmann constant 81, 98, 709, 731

 Boltzmann distribution law 87, 173, 710

 Boltzmann distribution 23, 86, 173

 Boltzmann factors 176

 derivation 169–73

 see also partition function

 Boltzmann, L. 81

 bomb calorimeter 111

 bonds *see* covalent bonding; electrostatic bonding

 Born energy 447

 Bose–Einstein statistics 14

 bosons 14

 bound complex 546

 boundaries 93–4, 132

 adiabatic 94

 dielectric 420

 gas–liquid phase 257

 boundary conditions 311

 Bragg–Williams approximation 272–3, 526

 brass 520

 Brodie, B.B. 439

 Brønsted law 375–7

 Evans–Polanyi model 377–9

 bronze 276

 Brownian motion 329, 351

 Brownian ratchet model 348–50

 butane, partitioning 296

 calorie (cal) 733

 calorimetry 111–12, 137, 631

 calorique theory 48–9

 cAMP 596

 canonical ensemble 188, 709–10

 capacitance

 and cell membrane 309

 parallel-plate capacitor 408–9, 446

 unit of 386, 733

 capsaicin 604, 605, 607

 carbon dioxide, supercritical 506, 509

 carbon–hydrogen bond breaking 371

 card games 10, 14

 Carnot cycle 119–20

 hurricane example 120–2

 reversibility of 124–5

 Carnot, NLS 119

 Cartesian coordinates 720–2, 727

 Laplacian operator 721, 722

 catabolite activator protein (CAP) 596

 catalysis 372–3

 acid–base 375–7

 Brønsted law 375–9

 solvent effects 373–5

 surface catalysis 551–4

 see also enzyme

 cataracts 495

 cathode 434

 Cavendish, H. 385

 cavities

 and dimerization 301

 in solids and liquids 261–2

 water molecules around 634

 cell membrane 150

 capacitance of 309

 ion channels 437–8

 saturable kinetics 551

 cells, chemical *see* batteries

 Celsius scale 128

 centrifuge 317

 chain, polymer *see* polymers

 chain rule 75–6

 chain segments, polymer 646

 change-of-symmetry reaction 240–1

 chaotic ions 636–7

 Chapman, D.L. 457

 characteristic ratio 662

 charge distribution 444–5

 charging processes 446–7

 neutral molecules 473–4

 charge shielding 455

 charged planes 446

 charged spheres/spherical shells 415–17

 free energy of 445–6

 charges, assembling 444–5

 charging processes 446–7

 chemical equilibria 235–6

 equilibrium constant 235, 237

 change-of-symmetry reaction 240–1

 pressure-based 241–2

 Le Chatelier’s principle 243–4

 partition functions for 236–9

pressure dependence 247–8
 principle of detailed balance 358–9
 temperature dependence 244–6
 chemical potential 35, 105–6
 and chemical equilibria 236
 compared with electrochemical potential 428
 and diffusion 35–6
 from partition function 184
 for ideal gas 210–11
 lattice model 35–6, 256
 μ 99
 of neutral salts 428–9
 as partial molar free energy 163–4
 particle exchange 105–6
 solutions 274–6
 standard-state 242, 288
 surrogate for osmotic pressure 295
 chemical reactions
 equilibrium 358–9
 mass action 357–8
 temperature dependence 360–3
 see also catalysis; transition-state theory
 chromatographic separations 295, 549, 555
 chromatography 548–9
 Clapeyron, E. 98
 Clapeyron equation 259
 clathrates 632
 Clausius, R. 338
 Clausius–Clapeyron equation 259
 closed system 94
 co-ions 456
 coefficients *see individual coefficients*
 coexistence curve 494, 500–1, 502
 and critical phenomena theories 520
 coexistence region 494
 coil state 527
 coil-to-globule transitions 678
 coin flips
 bias 89
 and combinatorics 10–11, 13
 distribution 17
 heads/tails and maximum multiplicity 32–3
 mean and variance 20
 sequences 4
 collectively exhaustive events 2–3
 colligative properties 283
 polymer solutions 645
 colloids 456–7
 steric stabilization 686
 combinatorics 10–14
 common ion 431
 common tangency, points of 499, 652
 competitive binding 591
 competitive inhibition 590
 complex, bound 546
 composite events 5–7, 16–17
 composite functions 75–6
 compressibility, solid and liquid 156–7
 compression ratio 120
 concentration, units of 290–2
 concentration gradients 309–18, 325
 concerted-allostery model 589
 condensed phase 255–6
 conditional probabilities 7–10
 conditional statistical weights 532
 conductivity
 electrical 308
 thermal 309, 336
 confined spaces, polymers in 691–2
 confinement entropy and free energy
 692, 693

conformations, polymer chain
 analogy with diffusion equation 689–91
 conformational free energy 674–5, 687, 689,
 692
 conformational state and stretching 36, 37
 four-bead model 169, 170, 178–9
 and Gaussian distribution function 665–6
 perturbation near surfaces 686–9
 six-bead model 179, 180
 conjugate pairs 139
 conjugate terms 97
 conservation laws 45
 conserved properties and flow 47
 conservative forces 47
 constant(s)
 table listing 731
 see also individual constants
 constant-pressure process 115
 constant-temperature process 116
 constant-volume processes 115
 constraints 30
 and extrema of multivariate functions
 66–72
 contact energies 271–2, 385–6
 contact lens, permeability of 314
 continuity 417
 continuous distribution function 15–16, 20
 continuum model 620
 contour length 659
 conventions, solvent/solute 288–9
 cooperativity 489, 519
 binding *see* binding processes
 cooperative transitions 527
 hydrogen bonding in water 616–17
 coordinate system, operators 722, 727
 coordination number, lattice
 256
 correlation 7
 degree of 8
 coordinate systems 722
 correlation time 351–3
 corresponding states, law of 509–10
 Coulomb, C. A. 385
 coulomb (unit) 386, 733
 Coulomb's law 385–6
 counterions 456
 counting sequences 11–13
 coupled binding 559
 coupled flows 324–5
 coupling coefficients 324
 covalent bonding 144–5, 474
 bond energies 474
 bond enthalpies 144–5
 C–H bond breaking 371
 crickets, chirping 363
 critical exponent 523–4
 critical micelle concentration 571
 critical nucleus 535
 critical percolation threshold 612
 critical phenomena 519–20
 Ising vs Bragg–Williams predictions 526
 critical point 520, 522
 phase transitions 63, 504–5
 of van der Waals gas 408
 critical temperature 494, 524
 cross-derivatives 152
 cross-link density 673
 crystallin 495
 crystals, sodium chloride 387–9
 Curie's law 186–7
 current 315

curve
 binodal 502
 curve-crossings 378
 coexistence *see* coexistence curve
 solubility 494
 spinodal 502-4
 volcano 553-4
 cycles, thermodynamic 118-22
 and enthalpy calculations 143
 cyclization, polymer 669-70
 cylindrical coordinates 722, 727

 Dahl and Anderson model 620
 Davy, H. 49
 de Gennes, P. G. 696
 death processes 321
 Debye length 458-9
 Debye, P. 464
 Debye theory of solids 216
 debye (unit) 473
 Debye-Hückel equation 458
 Debye-Hückel theory of nonidealities 464-7
 decaffeination of coffee 509
 degeneracy
 electronic 206
 of states 229-30, 231
 degrees of freedom 30, 37
 atoms and molecules 200
 multiple changing 59
 particle energy 51
 density
 anomalous of water 622
 of states 177-8, 228-30
 derivatives, averages from 705-6
 desolvation 283
 and ligand binding 302
 detailed balance, principle of 358-9
 deterministic process 2
 diatomic molecules
 properties of linear molecules 205, 214
 rotations 203-4, 214
 vibrations 201-3, 214
 dice rolls 2, 4, 709-12
 bias 84-5, 88-9
 combinatorics 13-14
 composite events 5-6
 and orientation of outcomes 83
 dielectric boundary 420
 dielectric constant 389-90
 and solvation 449
 dielectric interfaces 420-1
 differential, exact/inexact 74
 diffusion
 and concentration gradient 309
 coupled flows 325
 dog-flea model 334-6
 and drug reaction rate 323
 irreversibility of 125
 lattice model 35
 particles toward a sphere 314-15
 random-walk model 329-33
 rates of
 fluctuations 336-7
 Gaussian distribution 337-8
 diffusion coefficient 320, 362
 diffusion constants 310, 319
 diffusion equation 310-11
 applied to polymer conformations 689-91
 diffusion potential 444
 diffusion-controlled reaction 315
 and energy landscapes 380-1

 DiMarzio-McCrackin reflectance endash
 principle 687-8
 dimers
 dimerization 134-5
 in solution 299-302
 migration on surface 369
 dipolar molecule 414
 dipole moment 409, 473-4
 units 473
 water 615
 dipoles
 definition of 409
 energy of 409-10
 induced and permanent 473
 orientation 83-4
 in an electric field 409-10
 orientational averaging 474-6
 potential field around 412-14
 preorganized 375
 dipole-dipole interaction 475, 476
 dislocations 278
 disorder 10
 simple models 33-4
 dispersion forces 477
 dissipation 114
 dissociated state 134-5
 dissociation
 electrolytes 462-4
 energy 205, 240
 enthalpy 245-6
 of iodine 242-3
 van't Hoff plot 255
 dissociation constant, acid 436
 distillation 510-13
 fractional 511
 distinguishable objects 12
 distinguishable particles 179-81
 distribution
 charge distribution 375, 390, 398, 416, 418,
 444-51, 471-73, 477
 exponential distribution law 88, 89
 Maxwell-Boltzmann of particle
 velocities 175
 peaked 84
 underdetermined 84-5
 distribution functions 14-15
 binomial 16-17, 23
 continuous 15-16, 20
 flat 20, 24, 84, 89
 Gaussian 23, 665-6
 graphical examples of 22-4
 multinomial 16, 18
 nth moment of 18
 probability 22-4
 and variance 19-20
 divergence 721, 721
 DLVO model 456-7
 DNA
 bending 664-5
 dynamical relaxation times 695-6
 twisting 216
 dog-flea model of diffusion 334-6
 double-reciprocal plots 550-1
 driving force, energy uptake 227-8,
 232
 drugs
 anesthetics 248, 297
 binding activities 542
 diffusion and reaction 323
 dry-cell battery 434
 Dulong and Petit law 213

dwell time *see* waiting time
 dynamical partition functions 343–4
 dynamical statistical weight 343
 dynamical trajectories, single-molecule 710–11
 dyne (dyn) 733

Earth's energy balance 42–3
 effectively accessible states 177
 efficiency 113, 126
 eigenfunctions 202
 eigenvalues 195
 Einstein model of solids 213–16
 Einstein–Smoluchowski equation 319
 elastic entropy 692
 elasticity, polymer 659
 affine network model 670–3
 bending correlations and stiffness 662–5
 lattice model 36–7
 random flight model 659–61, 665–6
 in two and three dimensions 666–9

elastomers 670
 electric charge, units of 733
 electric current, unit of 733
 electric field *see* electrostatic field
 electric potential, unit of 733
 electrical conductivity 442
 electrical double layer 456, 461–2
 electrical resistance 308
 electrochemical potential 425–6
 and chemical potential compared 428

electrodes
 half-cell reactions 430–2
 Nernst equation for 427

electrolytes
 dissociation 462–4
 strong electrolytes
 defined 462, 463
 nonideality of 463–4, 464–7
 weak electrolytes 463

electromagnetic radiation
 energy spectrum 196
 excitation of atoms and molecules 193–4
 radian heat 46, 50
 solar energy 41–2

electron, rest mass 731
 electron volt (eV) 733
 electronegativity 434
 electronic degeneracies 205, 242
 electronic excitation 194
 electronic partition function 205–6
 electroplating 432
 electrostatic bonding 386–7
 sodium chloride crystal 387–9

electrostatic field 393–4
 from line charge 398–9
 from charged planar surface 399–401
 from point charge 394–5
 Gauss's law and field flux 395–8
 as gradient of electrostatic potential 404–5
 induced 418–20

electrostatic forces
 Coulombic attractions 385–7
 unit of charge 386
 vector summing of forces 392–3
 in neutral molecules 472
 short- and long-range attractions 387, 389
 in polarizable media 389–91

electrostatic potential 403–5
 around a line charge 417–18
 around a single point charge 405–6
 around two point charges 407

around/inside spherical shell 415–17
 and ion distribution 426–8
 gradient 441–3
 in a parallel-plate capacitor 408–9
 in salt solutions
 charged plane 459–61
 spherical double layer 461–2
 surfaces and equipotential contours 406–7

electrostriction 162
 elementary events 5
 elementary reaction 358
 elementary unit charge 731
 empirical energy functions 479
 endothermic processes 136
 end-to-end length 660, 661

energy 29
 change with volume 157
 definition and units 39
 of a dipole 409–10
 exchange 53–4
 fluctuations 230–2
 fundamental thermodynamic equations
 96, 97

ground state 170
 ideal gas 207
 law of conservation of 45–6

mechanical
 kinetic energy 45
 potential energy 46
 total energy 46, 47

of mixing 271–2
 non-ideal solution 271–3
 polymer solutions 649–50

quantization of *see* energy-levels
 units of 733
 see also global energy resources; heat

energy ladders 193
 chemical equilibrium 237
 one-dimensional particle in a box 197

energy landscapes 363–4, 380
 and phase transitions 490–2, 497–9

energy levels 170, 193–4
 for argon 197–8
 diagrams 51

engine
 internal combustion 126–8
 Stirling 129
 see also heat engine; steam engine

ensembles
 canonical 188–9, 709–10
 grand canonical 188, 578–80, 710
 isobaric 710
 microcanonical 188, 189
 semi-grand 593

enthalpy 131, 137–8
 of activation 378–9
 computing for different conditions 143
 of fusion and vaporization 261
 of ion solvation 448–51

entropy 1–2, 29, 709–15
 basic principles/Boltzmann's law 81–2, 90
 justification of the law 84–5, 89
 maximum entropy 81, 85–9, 709–12
 change with pressure 155
 of combined system 133, 712–14
 subsystems and total 133, 712–714, 713

computing
 by measuring heat capacity 114
 from partition function 183, 184, 208–9

fundamental thermodynamic equations
 96, 97

entropy (*continued*)

 gas entropies

 absolute of argon 209–10

 tabulated 210

 hydrophobic 632–3

 ideal gas 208–9

 orientational/rotational 82, 83

 of solutions/mixing 269–71

 polymers 646–9

 of a system 95

 thermodynamics of activated state 368–9

 translational 82

 enzyme catalysis 342–3, 374–5

 substrate bound 549–50, 594–5

 equations

 of state 96, 114, 480

 thermodynamic *see* fundamental equations

 equilibria

 definition of 30

 mechanical 29, 104

 metastable 31–2

 multiple binding 560

 neutral 31

 stable 31

 states of 29

 unstable 32

 see also chemical equilibria; phase equilibria

 equipartition theorem 208, 212–13

 DNA twisting 216

 vibrations 213

 erg 733

 Ericsson, J. 48

 escaping tendency 106, 243

 esu 733

 ethanol 477

 see also water-alcohol mixtures

 ether, the 50

 Euler test/reciprocal relationship 74–5, 77

 proof 707–8

 Evans, M.G. 377

 Evans-Polanyi model 377–9

 events

 combinatorics/counting 10–14

 combining 5

 correlation of 7–10

 elementary and composite 5–7

 intersection of 8

 relationships among 2–3

 exact differential 74

 exchange parameter 275–6, 277, 298

 excited state 170

 electronic partition function 205–6

 excluded volume 647, 682, 685–7

 expansion

 free of gas 125

 multipole 473

 thermal 153

 exponential distribution 23, 86–7

 law 84, 88, 89

 extensive properties 94–5, 99

 extensive variables 150

 extent of reaction 238

 extremum principles 29–30

 maxima and minima of functions 63–6

 and minimum free energy 132–4

 F1-ATPase motor protein 598–600

 factorials, and Stirling's approximation 703

 Fahrenheit scale 128

 fair apportionment 90

 farad (F) 386, 733

 Faraday's constant 731

 ferromagnetism 525

 Fick's law

 basis for 333–6

 first law 309

 second law 311

 fictitious states 117

 field, electrostatic *see* electrostatic field

 final state 236

 Filmer, D. 589

 First Law of Thermodynamics 45, 49–50,

 55, 106

 interrelation of heat, work and energy

 110

 first-order approximation 699, 700

 first-order homogeneous function 99

 first-order transition 523

 Fisher, R.A. 322

 flat distribution 20, 24

 and entropy 85–8

 flocculation 457

 Flory exponent 677

 Flory, P.J. 647

 Flory theorem 655–6

 Flory-Huggins theory 645, 646–51

 flow 47, 307–8

 coupled flows 324–5

 flux of vector field as 723–5

 steady-state 312

 through a surface 723

 vectors describing 719–29

 fluctuation theorem 338–9

 fluctuations 132

 diffusion rates 336–7

 energy 230–2

 fluctuation-dissipation theorem 351–2

 fluids

 incompressible 156

 supercritical 506, 509

 viscoelastic 693–6

 flux 207–8, 723–5

 computing 725

 field 395–8

 as surface property 725

 vector field 723–5

 folding

 polymer 6, 379–80

 protein 675

 RNA 342

 food energy 40

 adult equivalence to 100 W bulb 40

 and basal metabolic rate 43

 force correlation function 353

 forces

 conservative vs nonconservative 46

 definition of 46

 units 733

 vectors describing 719–29

 Fourier's law 309

 basis for 336

 fractional distillation 511

 free energy 131

 of activation 368

 charged spheres 445–6

 charging parallel-plate capacitor 446

 conformational 674–5, 687, 689, 692

 derived from partition function 184

 and dimerization 134–5

 and extremum principle 132–4

- Gibbs free energy 138
 Helmholtz free energy 132–4, 184
 ideal gas 207
 ion solvation 639
 lattice model liquid 256
 partial molar 163–4
 and polymer collapse 135–6
 polymer confinement 692
 solutions/mixing 273–4
 - ideal solution 271
 - tending to a minimum 131
 - of transfer 138
 - of solvation 639
 free expansion 125
 freedom, degrees of *see* degrees of freedom
 freely jointed chain model 660
 freezing point depression 291–3, 463
 frequency, unit of 733
 frequency interpretation, probabilities 711, 715
 friction 125
 friction coefficient 308, 310
 - relationship to diffusion constant 319
 fuel cell 434–5
 functional integral 722
 functions 137, 145
 - composite 75–6
 - constructing 139–40, 150–1
 - empirical energy 479
 - first-order homogeneous 99
 - graphical transformations 225–7
 - multivariate 59–60
 - natural 132
 - path dependent 72, 117
 - state 72, 74–5, 77
 - time-correlated 351–3
 - velocity autocorrections 351–2, 353*see also* distribution functions
 fundamental equations 95–6
 - constructing functions 139–40, 150–1
 - for energy and entropy 96
 - differential forms 97
 funnel landscapes 379–80
 fusion, enthalpies of 661

 gambling equation 10
 gas
 - entropies tabulated 210
 - free expansion of 125
 - kinetic theory of 49–50
 - lattice model 34–5
 - see also* ideal gas
 gas constant 82, 731
 gas pressure 34–5
 - equilibration 103–4
 - ideal gas 98, 207
 - standard state for helium 211
 gauche states 644
 Gaussian distribution 23
 - and polymer chain conformation 665–6
 Gauss's law 397–8
 - differential form 415
 - and electrostatic field flux 395–8
 Gauss's theorem 414–15, 725–9
 - proof 725–7
 Gibbs free energy 138
 - and chemical equilibria 236
 - and condensed phases 164–5
 Gibbs–Helmholtz equation 246
 global energy resources
 - animal size and basal metabolic rate 43–4
 - consumption and GDP 39–40
 - solar radiation and Earth's energy
 - balance 41–3
 - world sources 40
 - global minimum 67
 - global stability 502
 - global warming 43, 287
 - Gouy, L.-G. 457
 - gradient operator 719–20
 - gradient vector 719–20
 - grain boundaries 278
 - grand canonical ensemble 188, 710
 - and binding polynomials 578–80
 - graphical transformation 225–6
 - greenhouse effect 42–3
 - Green–Kubo relationship 353
 - gross domestic product (GDP) 39–40
 - ground state 170
 - ground-state energy 240, 241
 - Guggenheim, E.A. 520
 - Guldberg, C. 358
 - gyration, radius of 661
 half-cells 432
 halothane 248
 Hamiltonian operator 195, 200–1
 harmonic oscillator model 201–3
 Haynes, F.B. 317
 heat 47–8
 - early ideas about 47–9
 - calorique model 48
 - heat flow 50–4, 100–1
 - and calorimetry 111–12
 - driving force 227–8, 232
 - irreversibility of 125
 - interconversion with work 49
 - see also* heat engine
 - and molecular motion 50
 - radiant 49, 50
 heat bath 132, 228
 heat capacity 48
 - and computing entropy 114
 - dependence on temperature 112
 - and energy fluctuations 228, 230–2
 - iceberg model and hydrophobicity 633, 634
 - for ideal gas 112–13, 226
 - proportional to energy fluctuations 228
 - and Shottky two-state model 185, 226
 - and thermal equilibrium 141–2
 - various, tabulated 142, 214
 heat engines 118–20
 - Carnot cycle 119–20
 - hurricane and 120–2
 - efficiency and waste heat 126–8
 - Otto cycle 123
 - internal combustion engine 124–8
 - see also* steam engine
 heat exchange 227–8, 232
 heat flow equation 316
 heat pumps 262–4
 helicase 348, 559, 609–10
 helium
 - coexistence curve 525
 - standard state pressure 211
 helix-coil transitions, polymer 527
 - noncooperative model 527–9
 - two-state/maximum cooperativity model 529–32
 Zimm–Bragg model 532–5

helix content 527
 Helmholtz free energy 132–4, 136–7
 and condensed phases 164–5
 fundamental equation 136–7
 ideal gas 207
 hemoglobin-oxygen binding 586
 MWC allosteric model 587–90
 Pauling model 586–7
 pH linkage 594
 PKNF model 589, 590
 Henderson-Hasselbalch equation 439
 Henn and Kauzmann model 620–1
 Henry's law
 constants 286, 299
 and dissolved oxygen 287
 Hermite polynomials 201
 hertz (Hz) 733
 Hessian test 66
 Hess's law 144
 heteronuclear diatomic molecules *see* diatomic molecules
 hexadecane 156
 higher-order transition 523
 Hildebrand's principle 276
 Hill exponent 569
 Hill plot for cooperative binding 569–70
 histidine 441
 Hofmeister series 640
 homogeneous functions 99
 homonuclear diatomic molecules 205
 Hooke's law 46
 polymer elasticity 66
 horsepower (hp) 734
 Hückel, E. 464
 hurricanes 120–2
 hydration
 enthalpy of 450
 hydrophobic *see* hydrophobic solvation
 hydrodynamic screening 695
 hydrogen bonds 478–9
 acceptor and donor atom 616
 bond strength and temperature 619–20
 cooperativity of 616–17
 near to planar nonpolar surfaces 635
 and nonpolar solutes *see* hydrophobic solvation
 tetrahedral coordination 616, 617–19
 in water 615–20
 hydrogen bromide 193
 hydrogen iodide 361
 hydrophobic entropy 632–3
 hydrophobic interaction 635
 hydrophobic solvation 629–30, 635
 definition of 635
 heat capacity change 634–5
 interaction of two nonpolar solutes 635–6
 near critical temperature of water 634–5
 in salt solutions 640
 temperature dependence 630–2
 water structuring around solutes 623–4
 ice
 density 622
 skating and pressure melting 624–5
 structure 617–19
 iceberg model of heat capacity 633, 634
 ideal gas
 adiabatic process 116–17
 computing entropy 208–9
 free energy 207
 heat capacity 112–13
 internal energy 207–8
 partition function 210–11
 pressure 207
 p, V, T relationships 114–17
 thermal properties compared 225–7
 ideal gas law 51
 derivation
 from lattice model 97–9
 from quantum mechanics 207
 and partial derivatives 61–2
 ideal solutions 271, 285
 image charges, method of 418–21
 immiscibility 507–8, 520
 of polymers 654–6
 incompressible solid/liquid 156
 independent events 3
 independent particles 179–82
 independent variables 63–4
 indifference, principle of 90
 indistinguishable objects 12
 indistinguishable particles 181–2
 induced dipole 473, 475
 induction, electrostatic 418–19
 and neutral atoms 477–8
 inertia, moment of 203
 inexact differential 74
 inferences, consistency of 3
 infinite temperature 224
 inhibition
 competitive 590–1
 noncompetitive 592
 uncompetitive 591, 592
 inhibitor ligands 590–1
 initial condition 311
 initial state 236
 insufficient reason, principle of 90
 inputs 321
 integrals, useful 737
 integration, multivariate function 72–5
 intensive properties 95, 99
 intensive variable 132
 inter-arrival time 340
 interaction energies 171, 267
 contact energies 271–2, 385–6
 lattice model liquid 256
 interactions
 dipole-dipole 475, 476
 hydrophobic 635
 local 655, 673–4
 long-ranged 255, 387, 389, 483
 nonbonded 480
 nonlocal 655
 short-range 255, 387, 389, 472
 see also intermolecular interaction
 interface, definition of 265
 dielectric 420, 441
 interfacial tension 276–8
 see also surface tension
 intermediate state 565–6
 intermolecular interactions 471–2, 486
 charge asymmetries and polarization 473–9
 computed from boiling point 257
 lattice model contact energy 385–6
 structural correlations in liquids 483–5
 van der Waals gas model 479–82
 internal energy 49–50
 computing from partition function 182–3
 ideal gas 207–8
 and intrinsic properties of particles 171

as sum of particle energies 51
 of a system 95
 intersection of events 8
 interstitial model 362
 intramolecular localization 373–4
 invariant of the motion 46
 inverted region 378
 iodine, dissociation of 242–3
 ion channels 437–8
 sensing hot and cold 604–7
 voltage-gated 411–12
 ion distribution 426–7
 ion mobilities 443, 444
 ion pairing in water 639–40
 ionic bonding *see* electrostatic bonding
 ionic salts *see* salts
 ionic strength 459
 ions
 mobility and flow
 electrostatic potential gradients 441–3
 ion distributions 426–7
 through membranes 443–4
 near interfaces
 conducting plane 418–20
 dielectric boundary 420–1
 potential determining 431
 shielding effect 455–6
 solvation 447–8, 636–40
 enthalpy of 448–51
 irreversible processes 125
 Ising, E. 525
 Ising model 520, 524
 and magnetization 525–7
 isobaric process 115
 isobaric-isothermal ensemble 188, 710
 isochoric process 115
 isodichroic point 530
 isolated system 94
 isomeric state model 665
 isosbestic point 530, 531–3, 619, 620
 isothermal compression 156
 isotherms 116, 506
 isotope effects 370–2, 621–2

 Jacobson–Stockmayer theory 669
 joint probability 8, 9
 Jones–Dole coefficients 637–8
 Joule, J.P. 49
 joules (J) 733

 Kavenoff, R. 695
 Kelvin scale 128
 Khun segments 662
 kinesin protein 348, 601–3
 kinetic energy 45–6
 kinetic mechanisms 358
 kinetic theory of gases 50–1, 175, 176
 kinetics, chemical *see* reaction rate
 Kohlrausch, F. 463
 Koschland, D. 589
 kosmotrophic ions 636–7
 Krypton, solvation of 635, 636
 Kuhn model 662–3
 Kuhn segments 662

lac operon 596–8
 Lagrange multipliers 68–72, 712
 Landau model 522–5
 Langevin equation 351
 Langmuir adsorption equation 544–5

 Langmuir, I. 542
 Langmuir model
 and binding polynomial 567
 gas adsorption 541–5
 Langmuir trough 159
 Laplacian operator 415, 721–2
 latent heat 48
 lattice model 34
 chemical potential and diffusion 35–6
 condensed phase 255–6
 contact energy 385–6
 coordination number 256
 critical point 505
 derivation of ideal gas law 98–9
 elastic properties 36–7
 gas pressure 34–5
 interfacial tension 276
 ligand binding 239–40
 mixtures/solutions 271, 278–9, 278, 501
 solvation 261, 629–30
 surface tension 265
 vapor pressure depression 284–6
 laurate ion binding 547
 law
 Coulomb's 385
 Fick's 309, 311
 of conservation of energy 46–7
 of corresponding states 509–10
 of mass action 357–8
 Ohm's 442
 Stokes's 319
 Stokes–Einstein 320
 Le Chatelier's principle 243–4
 lead-acid battery 434
 least biased 85
 Legendre polynomials 211
 Legendre transforms 137, 139, 151, 717–18, 717
 length, unit of 733, 734
 Lennard–Jones potential 478
 letters *see* alphabet
 level curve 68
 lever rule 495–7
 lifting a weight 47
 ligands 541
 activator 591, 592–3
 binding equilibrium 239–40, 247–8
 competitive inhibition 590–1
 ligation states 565, 566
 multisite and crowding 575–8
 protein binding 348–9
 shape and packing 576
 surface stacking 573–4
 see also binding processes
 light, speed of 731
 like dissolves like 285, 486
 limit of metastability 502
 line (path) integral 722–3
 linear free energy relationships 376, 377–9
 Lineweaver–Burk plots 550–1
 linkage theory 593–4
 lipid bilayer 150, 313, 512
 liquid benzene phase diagram 493
 liquid, lattice model 255
 chemical potential 256
 creating/filling cavities in 261–2
 free energy 256
 vapor pressure 256–7
 liquid–gas coexistence 507, 520
 liquid mixtures, boiling 510–13
 litre (L) 734
 lithium-ion battery 434

local instability 503
 local stability 502, 503
 localization, intermolecular 374–5
 logic, biochemical 595–7
 logic, thermodynamic 109
 London dispersion forces 477
 long-range interactions 387, 389, 472
 long-range order 255, 483
 Lorentzian distribution 24

 McCrackin, F.L. 687
 McGhee–von Hippel model 575–8
 macrostates 35, 55, 170
 and entropy 90
 Madelung constant 389
 magnesium sulfate 162
 magnetic moment 150, 186–7, 525
 magnetism 186–7, 525–7
 Malthus, T.R. 321
 Marcus, R. 379
 Markov model 343
 mass action law 357–8
 master equation 347
 mathematical operators *see* operators
 matrix multiplication 346
 matter 50
 maximum density 525, 623, 625
 maximum entropy 37, 85–90
 maximum multiplicity, principle of 32–3, 37
 and heat flow 50–4
 Maxwell–Boltzmann distribution 175
 Maxwell's demon 339
 Maxwell's relations 152–4
 Euler relationship to establish
 707–8
 mean force, potentials of 484–5
 mean-field approximation 272–3
 mean-square velocity 175
 mechanical equilibrium 29–30
 mechanisms, chemical kinetic 357–8
 melting
 enthalpies of fusion 261
 polymer melts 696
 pressure-induced 624–5
 solid mixtures 512
 membranes *see* cell membranes; semipermeable membranes
 menthol 606–7
 metabolic rate, basal 43
 metal alloys 276, 278, 520–2
 metastability 31–2
 limit of 502–4
 micelles 570–1
 micellization 570–3
 shapes 572–3
 sizes 571–2
 Michaelis constant 550
 Michaelis–Menten model 549–52
 equation 550
 microcanonical ensemble 188, 189
 microscopic reversibility 325
 microstates 35, 55, 170
 and entropy 90
 miscible system 498
 mixtures and mixing 125
 boiling a liquid mixture 510–13
 energy of mixing 271–3, 649–50
 entropy of mixing 269–71, 648–9
 lattice model approximations 278–9
 miscibility
 and critical temperature 494
 of polymers 654–5
 oil and water 274, 492–4, 501
 of polymers 655
 see also solutions
 mobile phase 548–9
 mobility 310, 442–3
 models, role and value of 33–4
 Ising 520
 micellization 537
 nearest-neighbor, 520, 524–7, 532, 536
 molality 290–1
 molar heat capacity 142
 molar volume 160
 molarity 290–1
 molecular interactions *see* intermolecular molecular logic 595–6
 transcriptional regulation and *lac* operon 596–8
 molecular machines *see* biochemical machines
 molecular motors 348
 molecular weights, of polymers 654
 molecules
 and kinetic theory 50–1
 neutral 473–4
 see also covalent bonding; intermolecular interactions
 moments
 of a distribution 18
 see also dipole moments
 of inertia 661
 momentum 45
 Monod, J. 587
 Monod–Wyman–Changeux (MWC) model 589
 monopole 473
 most probable radius, polymer 670
 multinomial distribution 16, 18
 multiple binding equilibria 560
 multiple variables 59–60
 multiplication rules 3–4
 multiplicity 3
 see also maximum multiplicity
 multipole expansion 473
 multipoles 472
 multisite binding 576
 multisite ligands 575–8
 multivariate functions 59–60, 77
 chain rule for composite functions 75–6
 extrema of functions 63–4
 subject to constraints 66–72
 integration of 72–5, 77
 partial derivatives 60–2
 small change and total differential 62–3
 muscle 438
 mutually exclusive events 2
 MWC allosteric model 587–9

 natural function 152
 natural gas clathrate 632
 natural variable 132
 nearest-neighbor model 524
 magnetic interactions 525
 polymer helix-coil transitions 527–35
 negative activation energies 362
 negative cooperativity 563
 negative temperature 221, 224–5
 Nemethy, G. 589, 619
 neon, solvation of 634, 636
 Nernst equation 427
 Nernst, W.H. 427

- Nernst–Planck equation 318
 neutral equilibrium 31
 neutral salts *see* salts
 newton (N) (unit) 733
 Newton's second law 46
 non-fundamental component functions 140
 nonpolar solutes 631, 632
 nonbonded interactions 480
 noncompetitive binding 591
 noncompetitive inhibition 592
 nonconservative forces 47
 nonequilibrium statistical mechanics 329
 nonidealities of strong electrolytes 463–4, 464–7
 nonpolar solvation *see* hydrophobic effects
 nonrenewable resources 40
 normalization 16
 N -string 711
 nucleation 535–6
 nucleation parameter 532
 nucleus, critical 535

 octupole 473
 Ohm's law 308
 oil supplies 40
 oil–water
 miscibility 297, 492–4, 629
 see also hydrophobic effect
 one-state transition 565
 Onsager, L. 520, 526
 Onsager reciprocal relations 324
 open system 94
 operators, mathematical 195
 Hamiltonian 195, 200–1
 Laplacian 415
 vector 415
 orbital steering 218
 orbitals 200
 order, concept of 10
 order parameter 520, 521, 522
 ordered sequences, permutations of 11
 ordering in metal alloys 520–2
 order–disorder transition 520, 521
 orientation of dipoles 83–4
 and intermolecular interactions 472–6
 orientational entropy 82, 83
 orientations of a vector 21–2
 osmotic pressure 293–4
 molecular masses from 294–5
 polymer solutions 678–9, 682
 surrogate for chemical potential 295
 Otto cycle 123, 124
 outcomes
 composition of 32
 orientation of 83
 and principle of maximum multiplicity 32–3
 sequence of 32
 outputs 321
 overlap concentration, polymer 681, 681–2
 oxidation 436
 see also redox
 oxygen
 binding *see* hemoglobin
 dissolved 287
 molecular 203
 sensor 436

 pair correlation function 483–4
 pair potential 471
 parabolic law 201
 paraboloid minima and maxima 65, 67, 69, 70–1

 parallel-plate capacitor 408–9
 free energy of charging 446
 paramagnetism 186–7
 partial derivatives 60–3
 determination of 139–40
 and Maxwell relations 152
 partial molar free energy 163–4
 partial molar volumes 160–2
 balls and sand analogy 162–3
 particle, average energy 183
 particle bath 132
 particle-in-a-box model 195–7
 degeneracy of states 229–30, 231
 three-dimensional box 199–200
 particles
 computing average energy 183
 distinguishable and indistinguishable 180–1
 exchange and chemical potential 105–6
 particle velocities 175–6
 partitioning of 176–7
 of a system 95
 test particle 480
 partition coefficient 295
 butane–water 296
 oil–water 297
 and solute size 299
 partition functions 87, 173, 176–7
 chemical reactions 236–7
 computing thermodynamic properties from 182–8
 quantum mechanics 207–11
 and density of states 178
 dynamical 344
 independent particles
 distinguishable 179–81
 indistinguishable 181–2
 q 87
 rotational 204–5
 total 207
 translational 198, 200
 vibrational 203
 Zimm–Bragg 532, 533
 partitioning equilibrium 439
 partitioning processes 283–4
 ion into oil 448
 Pascal (Pa) 733
 Pascal's triangle 17
 passive transport 313
 path integral 404, 722–3
 path-dependent function 72, 117
 path-dependent quantity 110
 pathways
 as combinations of processes 117–18
 of integration 73, 77, 117
 Pauling, L. 372
 Pauling model 586–7
 peaked distributions 84
 Pearson, K. 329–30
 Peltier effect 324, 325
 pencil spins 3
 percolation threshold, critical 621
 permanent dipole 473
 permanent dipole moment 390, 473, 473, 615
 permeability 313, 314
 permittivity 386, 731
 perpetual motion machine 49
 persistence length 663–4

pH
 pH-partition theory 439
 propagation through water 620-1
 titration 548
 phase 94
 prediction composition of 499
 phase boundary 257, 258, 493, 494
 phase diagrams 257-8, 592-5
 benzene 493
 boiling liquid mixtures 511
 and Clapeyron equation 257-9
 crystallin 495
 lattice model of solutions 501
 lipid bilayer 512
 polymer solutions 652-5
 asymmetry 653
 phase equilibria 253, 267
 cavities in liquids and solids 261-2
 Clapeyron equation 257-9
 energy and entropy of condensed phase 255-6
 surface tension 265-7
 vaporization 253-5
 vapor pressure 256-7
 phase ratio 549
 phase separations 651
 phase transitions 489, 512
 first- and higher-order 490-2, 523
 kinetics of 535-6
 and order parameter 520
 phenomenological flow coefficients 325
 phosphorylation 601
 physical properties of some typical liquids 561
 PKNF binding model 589, 590
 Planck's constant 193, 731
 plate capacitor, parallel, *see* capacitance
 poise (p) 320, 734
 Poisson distribution 23, 340
 Poisson's equation 415
 Poisson-Boltzmann equation 458
 Poisson-Boltzmann model 456-8
 Polanyi, M. 377
 polarizability 389-90
 of atoms 477-8
 volumes 474
 polyethylene 155, 159, 644
 Kuhn model of 663
 polymer radius 681-2
 most probable 670
 polymer solutions 644
 chain length effects 645
 solvent-sized chain segment 646
 concentration and semi-dilute/cross-over regime 681
 conformational distribution 643-5, 655-6
 energy of mixing 649-50
 entropy of mixing 646-9
 free energy and chemical potential 650-1
 nonideal colligative properties 644, 645-6,
 651
 osmotic pressure 678-9, 682
 phase behaviour 651-4
 immiscibility 654-6
 see also polymer melts
 polymerization 250, 305
 polymerization equilibrium 582
 polymers
 chains
 contour length 331, 332
 four-bead model 179
 end-to-end length 660, 661
 length distribution 643-4
 length effects in solution 645-6
 simple lattice model 37
 six-bead model 179
 collapse and expansion 673-6
 density of states 178-9
 good solvents 673, 677
 polymer gels 679-80
 poor solvents 673, 678
 θ solvents 673, 676
 toy model 135-6
 in confined spaces 691-2
 conformations *see* conformations
 cyclization 669-70
 elasticity *see* elasticity
 folding 379-80
 four-bead chain model 135-6, 169-71,
 177-9
 gels 679-80
 reptation 696-7
 linear vs branched chains 643
 melts 696
 and reptation 696-7
 stretching and elasticity 36-7
 tethered chains 688-9
 see also conformations; helix-coil transitions; polymer solutions
 polynomials
 binding 559
 Hermite 201
 Legendre 204
 population biology 321-2
 population inversion 224
 positive temperature 223-4, 224
 potassium potential 438
 potential energy 29-30, 46
 potential field around a dipole 412
 potentials of mean force 484-5
 power law 24, 523
 power, unit of 39, 734
 prebinding 591
 preorganized solvent 374-5
 pressure
 atmospheric 173-4
 effect on equilibria 247
 entropy changes with 155-6
 equilibration and volume change 103
 maximum multiplicity 104-5
 gas lattice model 34-5
 ideal gas 98
 p 99, 184, 480
 pressure-induced melting 624-5
 units of 733
 see also vapor pressure
 pressure-area isotherm 159
 Priestley, J. 385
 primary kinetic isotope effect 370
 principle
 of detailed balance 358
 of fair apportionment 709
 of indifference 90
 of insufficient reason 90
 of microscopic reversibility 325
 probability 1-2, 22, 711
 a posteriori 8, 10
 a priori 8, 10
 conditional 7, 8
 definitions 2, 711
 events 2-3
 composite 5-7

frequency interpretation 711
 joint 8
 rules
 addition 3–4, 8
 multiplication 3–4, 8
 subjective interpretation 711
 see also combinatorics; distribution functions
 probability density 16
 probability distribution functions 22–4
 processes
 ideal gas 114–17
 irreversible 125
 quasi-static 110, 118
 see also reversible processes
 progress variable 238
 propane 454
 protein properties
 histidine dissociation 441
 modelling protein loop 187–9
 predicting 9
 protein folding 675
 proton, mass 731

 quadrupoles 472, 473
 quantization of energies 193
 quantum harmonic oscillator 202
 quantum mechanics 51, 193–200
 quantum number 195, 197
 quasi-static process 110
 heat plus work 118

 radial distribution function 483
 radiant heat 46, 50
 radius
 of gyration 661
 polymer 670, 681–2
 Raman spectra 619, 620
 random network model 620
 random variables, theory 711
 random-walk/random-flight model 329–33
 Raoult's law 285
 rate coefficient 320–1
 from transition-state theory 364–8
 temperature dependence of 360–3
 rate of reaction *see* reaction rate
 Rayleigh, Lord 330
 reaction rate
 diffusion-controlled 315
 drug 323
 equilibrium and detailed balance 358–9
 mass action 357–8
 rate coefficient 320–1
 calculation from theory 364–8
 temperature dependence of 360–3
 salts affecting 466–7
 saturable kinetics 549–50
 solvent effects 373–5
 see also catalysis
 reciprocal relationship (Euler), *see* Euler test
 Redlich-Kwong equation 482
 redox potentials 431, 430
 and electronegativity 434
 redox reactions 429–30, 436
 see also batteries
 reduced mass 202
 reduction *see* redox
 reduction potential 431, 430
 reflectance principle 687–8
 refrigerators 262–4
 regular solution model 274

 relaxation time 110
 DNA 695–6
 renormalization group theory 520
 repressor protein 596
 reptation model 696–7
 resistance
 electrical 308
 of a membrane 313
 respiration 436
 rest mass of electron 731
 retraction force 158–9
 reversible processes 113–14
 combined into pathways 117–18
 describing efficient processes 122, 123–4
 for ideal gas 114–17
 reversibility, problem of 333
 rigid rotor model 203–4
 river delta 456
 RNA folding 342
 Röntgen, W. K. 619
 root mean square (RMS) 350, 661, 683
 rotation of diatomic molecule 203–4
 rotational partition function 204–5
 rotational energy 194, 196, 200
 rotational entropy 82, 83
 rotational isomeric state model 665
 Rouse model 694–5
 Rouse-Zimm model 693–6
 routes of flow 336
 rubber band
 stretching and conformational states 36–7
 thermodynamics of 158–9
 rules of probability
 addition 3–4
 multiplication 4–5
 Rumford, Count 49
 rusting 436

 $S = k \ln W$ 81, 709–15
 $S = k \log W$ 81, 709–15
 Sabatier, P. 552
 Sabatier's principle 552–4
 Sackur-Tetrode equation 209–10
 saddle point 65, 66
 and transition state 364
 salt bridge 432
 salts
 affecting reaction rates 466–7
 chemical potential of sodium chloride 428–9
 ionization of 392
 shielding effects 455–6
 solubility in water 639–40
 see also electrolytes; electrostatic bonding
 saturated solution 493
 saturation of myoglobin 609
 scalar fields 719–23, 720
 Scatchard plot 568–9
 Schellman helix-coil model 537
 Schottky model *see* two-state model
 Schrödinger equation 194–5, 200–1
 second-order approximation 699, 700
 Second Law of Thermodynamics 55, 99–100, 106
 and heat transfer 51–2, 53–4
 and perpetual motion 49
 possible violation of 339
 Seebeck effect 324
 self-diffusion coefficient of polymers 697
 self-energy 445
 semi-grand ensemble 593

semipermeable membranes 94, 293
 contact lens 314
 diffusion through 311–12
 ion flow through 443–4
see also cell membranes
 separations
 chromatographic 295
 phase 651
 sequences of events 4, 16
 coin flips 32
 permutations of 11–14
 sequential-allostery model 589
 short-range attractions 387, 389
 intermolecular forces 472
 short-range order 255
 sigma notation 15
 signal transduction 601
 silicon tungstate 625
 simple solution 631
 simple system 94
 single-particle kinetics 343–4
 single-phase solution 492
 sinks, population biology 321
 site-based binding 563
 six-state system 223, 224
 slopes 60
 Smoluchowski equation 317
 soap films 149
 sodium chloride
 bonding and crystals 386, 387–8
 diffusion potential 444
 enthalpy of solvation 450
 ionization in solution 392
 chemical potential 428–9
 solar energy 41
 and Earth's energy balance 42–3
 solder 276
 solids
 compressibility 156–7
 creating/filling cavities in 261–2
 Einstein model 213–16
 lattice model 255
 solubility curve 494
 solubility limit 493–5, 514
 solute 284
 convention 288, 289
 transfer and partition 295–9
 solutions 269, 279
 chemical potential 274–6
 energy of
 contact/interaction energies 271–2
 free energy 273–4
 mean-field approximation
 272–3
 entropy of 269–71
 ideal solution 271, 285
 metastable 503
 nonidealities 645
 supercooled 503
 supersaturated 503
 see also polymer solutions
 solvation 261, 629–30
 enthalpy of 448–51
 ions 447–8
 lattice model 284–6
 see also hydrophobic solvation
 solvent convention 288–9
 solvent preorganization 374–5
 solvent-separated minimum 484
 solvent-sized chain segments 646
 solvents
 and catalysis 373–5
 good/poor/θ 673
 polymer collapse and expansion 673–5
 sources, population biology 321
 spatial extent 95
 specific heat capacity 142, 215
 specific ion effect 637
 specific volume 155
 spectroscopy 193–4
 spheres/spherical shells, charged 415–17
 free energy 445–6
 spherical coordinates 314, 722, 728–9
 spherical double layer 461–2
 spherical harmonics 203
 spinodal curve 502–4
 spinodal decomposition 535
 springs, stretching 46–7
 square-law 201
 stability 502
 global 502
 local 502
 see metastability
 stable equilibrium 31
 standard acceleration of free fall 731
 standard deviation 20
 standard-state chemical potential 242, 288
 standard-state pressure 211
 Stanley and Teixeira model 620–1
 statcoulomb 386
 state, equations of 96, 114, 480
 state functions 72, 74–5, 77
 pathways and cycles 117, 128
 state probability 345
 computing 345–8
 state transition 565
 states 170
 degeneracy of 230–1, 231
 density of 177–8, 228–30
 effectively accessible 177
 ligation 565, 566
 statfarad 386
 stationary phase 548–9
 statistical mechanics 81, 169, 193
 statistical weights 3–4, 532–5
 conditional 532
 sequence 532
 Zimm-Bragg model 532
 statistics, Bose-Einstein 14
 steady-state flow 312
 steam engines 48, 49
 efficiency 113
 steric stabilization of colloids 686
 stiffness, chain 662–5
 Stirling's approximation 82, 98, 703
 Stirling engine 129
 stoichiometry 238
 Stokes's law 319
 Stokes-Einstein law of diffusion 320
 strains 671
 stresses 671
 stretched water 626–7
 strong electrolytes 456, 463
 nonideality of 463–4, 464–7
 subjective probabilities 711
 sublimation 265
 summations 15
 sun's energy *see* solar energy
 supercooled solution 503
 supercooled water 626–7

supercritical fluids 506, 509
 supersaturated solution 503
 surface
 definition of 265
 and interfacial tension 276–7
 surface binding *see* adsorption
 surface catalysis 551–4
 surface tension 151, 265–7
 and surfactants 159–60
 values tabulated 266
 surfactant molecules 159–60
 micellization 570–3
 phase transitions 513, 512
 surroundings 114
 susceptibilities 154
 Svedberg, T. 317
 switching, biochemical 595–7
 symmetry, change of 240–1
 symmetry-breaking 520
 systems 93
 extensive and intensive properties 94–5
 types defined 94

 tangent, common 499, 652
 Taylor series 62, 187, 699–701, 735–6
 temperature 48, 54
 absolute 128
 boiling 143, 257–8, 289–90, 634
 critical 494, 524
 defined and described 101–3
 freezing 291–2
 and kinetic theory 51
 positive 223
 and rate of reactions 360–3
 rotational 204
 scales of 128
 T and heat flow 99–101
 θ 673
 translational 198
 two-state system 221–3
 infinite temperature 224
 negative temperature 224–5
 positive temperature 223–4
 vibrational 203
 tension *see* interfacial tension;
 surface tension
 term symbols 206
 test particle 393–4, 480
 test tube system 132–4
 tethered polymer chains 688–9
 tetrafluoroethane 264
 tetrahedral coordination of water
 616, 617–19
 theory of random variables 711
 thermal conductivity 309, 336
 thermal diffusivity 316
 thermal expansion coefficient 153–4
 thermal phase equilibria 500–2
 thermal pressure coefficient 157
 thermochemistry 144–5
 thermodynamic cycles 118–22, 143
 thermodynamic equations *see* fundamental equations
 thermodynamic systems *see* systems
 thermoelectricity 324
 θ conditions 673, 676
 θ solvents 673
 θ temperatures 673
 Third Law of Thermodynamics 141–3
 tie line 494

 time-correlation function 351–3
 time constant 352
 time per event 340
 time's arrow 338
 titratable protons on glycine 564–5
 torsional freedom 644
 total differential 152, 712–13
 total energy 45, 46
 total partition function 207
 trajectories 343–4, 710–11
 computing average properties 344–5
 two-state exponential relaxation
 347–8
 trans conformations 644
 transcendental equation 501
 transcription regulation 596–8
 transducers, molecular
 energy 585, 598
 force and motion 601–3, 604
 signal 601
 transition state 361, 364
 calculation of rate coefficients 364–8
 catalysis 372–5
 isotope effects 370–1
 rate coefficients 364–8
 stabilization 372
 theory 363–8
 thermodynamic quantities 368–9
 transitions
 abrupt 519–20
 cooperative 527, 565
 first-order 523
 higher-order 523
 one- and two-state 531–2, 565
 order-disorder 520
 see also helix-coil; phase transitions
 translational energy 194, 200
 translational entropy 82
 translational freedom 82
 translational partition function 198, 200
 transmission coefficient 366
 transport 313 *see* active transport,
 passive transport
 trimer
 helicity 534–6
 partition function 534
 truncating series 699
 truth tables 595
 two binding sites model 563–6
 two-phase solution 493
 two-state exponential relaxation 347
 two-state (Schottky) model 184–6
 and heat exchange 227–8
 temperature of system 221–5
 thermal properties compared to ideal gas
 225–7
 two-state transitions 531, 565
 spectroscopic isosbestic point
 531–2

 uncompetitive inhibition 591, 592
 underdetermined distributions 84–5
 unimolecular reaction 358
 uniqueness 418
 unit vector 720, 727–28
 units
 multiples of 739
 table of 733–4
 universality 520, 524
 unstable equilibrium 32

- vacuum, permittivity of 386, 731
 van der Waals equation 482
 of state 480
 van der Waals forces 478
 van der Waals gas model 479-82
 van der Waals, J. D. 480
 van't Hoff equation 244
 van't Hoff plots 245
 vapor pressure 256-7
 depression of 284-7
 polymer solutions 645
 vaporization 253-4
 enthalpies of 260-1
 variables
 independent 63-4
 natural 132
 state 75
 variance 19-20
 vector fields 719-23, 720
 divergence 721, 721
 flux 723-5
 vector operator 720
 vectors
 averaging orientations of 21-2
 forces and flows 719-29
 integrating 722-3
 and matrix multiplication 346
 velocity 45
 mean square 175
 uncorrelated 352-3
 velocity autocorrelation function 351-2, 353
 Verhulst, P.F. 322
 Verhulst model 322
 vibrational energy 194, 196, 200
 vibrational partition function 203
 vibrational temperature 203
 vibrations
 equipartition 213
 harmonic oscillator 201-3
 virtual bond 659, 662
 viscoelastic fluids 693-6
 viscosity 319, 320, 734
 volcano curves 553-4
 volt (V) 733
 voltage-gated ion channels 411-12
 sensing of hot and cold 604-7
 volume
 excluded 647, 682, 685-7
 molar 160
 and pressure 103
 specific 155
 surface property relationship 725-9
 unit of 734
 volume fractions 646
 volumetric anomalies of water 622-4
 von Mayer, J.R. 49
- Waage, P. 358
 waiting times 341, 710
 for A-to-B trajectories 348
 enzyme catalysis 342-3
 RNA folding 342
- Wallis, J. 45
 water
 anomalous properties 155, 615, 620-2
 density and volume 622-6
 heat capacity 620, 621, 623
 isothermal compressibility 625
 negative thermal expansion coefficient 625
 pressure-dependent viscosity 625-6
 stretched water 626-7
 supercooled water 626-7
 boiling
 at high altitude 260
 boiling point elevation 291
 and molecular attractions 257
 phase transitions 505-8
 dissociation 245
 flux 723-4, 724
 freezing point depression 291-3, 463
 heat capacity 142, 143, 620, 621, 623
 partial molar volume 334, 336
 as polar solvent 375
 solvation ordering ions 636-40
 as solvent for nonpolar solutes *see hydrophobic effects*
 stretched water 626-7
 structure and hydrogen bonding 616-17
 near planar nonpolar surface 635
 temperature effects 619-20
 tetrahedral coordination 616, 617-19
 supercooled water 626-7
 thermal expansion coefficient 154-6,
 623, 625-6
 see also hydrophobic effect
 water-alcohol mixtures
 boiling and distillation 511-12
 and partial apparent molar volume 162
 volume constriction 636
- Watt, J. 48
 watt, unit of power 39, 734
 wavefunction 194
 waiting time 710
 weak electrolytes 456, 463
 weather maps 719, 720
 work
 constant pressure/temperature/volume
 115-16
 interconversion with heat 49, 113-14
 see also heat engine
 mechanical energy transfer 46, 110-11
 lifting an weight 47
 stretching a spring 47
 as path-dependent function 117-18
 as state function 118
 wormlike chain model 665
 Wyman-Gill linkage theory 593-4
- zeolite pores 454
 zero-point energy 201, 202
 Zimm model 695
 Zimm-Bragg model 532-5
 zirconium tungstate 625
 zone refining 512