THIRD EDITION

Lehninger Principles of Biochemistry

David L. Nelson
Professor of Biochemistry
University of Wisconsin–Madison

Michael M. Cox
Professor of Biochemistry
University of Wisconsin–Madison

WORTH PUBLISHERS

BPI Labs, LLC, EX1014
BPI Labs, LLC v. Eli Lilly and Company
IPR2025-01346

Lehninger Principles of Biochemistry Third Edition

David L. Nelson and Michael M. Cox

Copyright © 2000, 1993, 1982 by Worth Publishers

All rights reserved

Printed in the United States of America

Library of Congress Cataloging-in-Publication Data

Nelson, David L.
Lehninger principles of biochemistry / David L. Nelson, Michael M. Cox.— 3rd ed.
p. cm.
Includes index.
ISBN 1-57259-153-6
1. Biochemistry. I. Nelson, David L. (David Lee), 1942- II. Cox, Michael M. III. Title.

QD415 .L44 2000
572—dc21 99-049137

Printing: 5 4 3 2 1 Year: 04 03 02 01 00

Development Editor: Morgan Ryan, with Linda Strange and Valerie Neal
Project Editor: Elizabeth Geller
Art Director: Barbara Rusin
Design: Paul Lacy
Production Supervisor: Bernadine Richey
Layout: York Graphic Services and Paul Lacy
Photo Editor: Deborah Goodsite
Illustrations: Susan Tilberry (with Alan Landau and Joan Waites), J.B. Woolsey & Associates, Laura Pardi Duprey, and York Graphic Services
Molecular Graphics: Jean-Yves Sgro
Composition: York Graphic Services
Printing and Binding: R.R. Donnelley and Sons

Cover (from top to bottom): Cut-away view of GroEL, a protein complex involved in protein folding; cut-away view of tobacco mosaic virus, an RNA virus; ribbon model of a β-barrel structural domain from UDP *N*-acetylglucosamine acyltransferase; cut-away view of the F_1 subunit of ATP synthase, with bound ATP shown as a stick structure; mesh surface image of the electron-transfer protein cytochrome *c*, with its heme group shown as a stick structure.

Cover images created by Jean-Yves Sgro.

Illustration credits begin on p. IC-1 and constitute a continuation of the copyright page.

Steenbock Memorial Library
University of Wisconsin-Madison
550 Babcock Drive
Madison, WI 53706-1293

Worth Publishers
41 Madison Avenue
New York, NY 10010

chapter

4 Water



This view of Earth from space shows that most of the planet's surface is covered with water. The seas, where life probably first arose, are today the habitat of countless organisms.

Water is the most abundant substance in living systems, making up 70% or more of the weight of most organisms. The first living organisms doubtless arose in an aqueous environment, and the course of evolution has been shaped by the properties of the aqueous medium in which life began.

This chapter begins with descriptions of the physical and chemical properties of water, to which all aspects of cell structure and function are adapted. The attractive forces between water molecules and the slight tendency of water to ionize are of crucial importance to the structure and function of biomolecules. We will review the topic of ionization in terms of equilibrium constants, pH, and titration curves, and consider how aqueous solutions of weak acids or bases and their salts act as buffers against pH changes in biological systems. The water molecule and its ionization products, H^+ and OH^-, profoundly influence the structure, self-assembly, and properties of all cellular components, including proteins, nucleic acids, and lipids. The noncovalent interactions responsible for the strength and specificity of "recognition" among biomolecules are decisively influenced by the solvent properties of water.

Weak Interactions in Aqueous Systems

Hydrogen bonds between water molecules provide the cohesive forces that make water a liquid at room temperature and that favor the extreme ordering of molecules that is typical of crystalline water (ice). Polar biomolecules dissolve readily in water because they can replace water-water interactions with more energetically favorable water-solute interactions. In contrast, nonpolar biomolecules interfere with water-water interactions but are unable to form water-solute interactions—consequently, nonpolar molecules are poorly soluble in water. In aqueous solutions, nonpolar molecules tend to cluster together.

Hydrogen bonds and ionic, hydrophobic (Greek, "water-fearing"), and van der Waals interactions are individually weak, but collectively they have a very significant influence on the three-dimensional structures of proteins, nucleic acids, polysaccharides, and membrane lipids.

Hydrogen Bonding Gives Water Its Unusual Properties

Water has a higher melting point, boiling point, and heat of vaporization than most other common solvents (Table 4–1). These unusual properties

table 4–1

Melting Point, Boiling Point, and Heat of Vaporization of Some Common Solvents

	Melting point (°C)	Boiling point (°C)	Heat of vaporization (J/g)*
Water	0	100	2,260
Methanol (CH_3OH)	−98	65	1,100
Ethanol (CH_3CH_2OH)	−117	78	854
Propanol ($CH_3CH_2CH_2OH$)	−127	97	687
Butanol ($CH_3(CH_2)_2CH_2OH$)	−90	117	590
Acetone (CH_3COCH_3)	−95	56	523
Hexane ($CH_3(CH_2)_4CH_3$)	−98	69	423
Benzene (C_6H_6)	6	80	394
Butane ($CH_3(CH_2)_2CH_3$)	−135	−0.5	381
Chloroform ($CHCl_3$)	−63	61	247

**The heat energy required to convert 1.0 g of a liquid at its boiling point, at atmospheric pressure, into its gaseous state at the same temperature. It is a direct measure of the energy required to overcome attractive forces between molecules in the liquid phase.*

are a consequence of attractions between adjacent water molecules that give liquid water great internal cohesion. A look at the electron structure of the H_2O molecule reveals the cause of these intermolecular attractions.

Each hydrogen atom of a water molecule shares an electron pair with the oxygen atom. The geometry of the molecule is dictated by the shapes of the outer electron orbitals of the oxygen atom, which are similar to the bonding orbitals of carbon (see Fig. 3–4a). These orbitals describe a rough tetrahedron, with a hydrogen atom at each of two corners and unshared electron pairs at the other two corners (Fig. 4–1a). The H—O—H bond angle is 104.5°, slightly less than the 109.5° of a perfect tetrahedron because of crowding by the nonbonding orbitals of the oxygen atom.

The oxygen nucleus attracts electrons more strongly than does the hydrogen nucleus (a proton); oxygen is more electronegative (see Table 3–2). The sharing of electrons between H and O is therefore unequal; the electrons are more often in the vicinity of the oxygen atom than of the hydrogen. The result of this unequal electron sharing is two electric dipoles in the water molecule, one along each of the H—O bonds; the oxygen atom bears a partial negative charge ($2\delta^-$), and each hydrogen a partial positive charge (δ^+). As a result, there is an electrostatic attraction between the oxygen atom of one water molecule and the hydrogen of another (Fig. 4–1c), called a **hydrogen bond.** Throughout this book, we will represent hydrogen bonds with three parallel blue lines, as in Figure 4–1c.

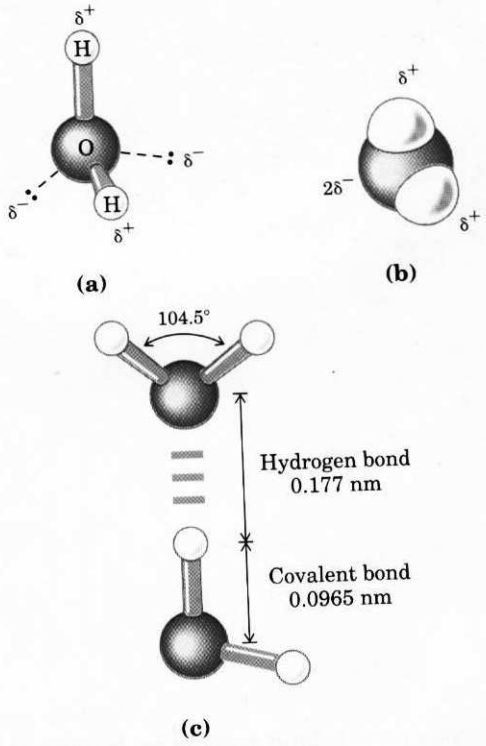


figure 4–1
Structure of the water molecule. The dipolar nature of the H_2O molecule is shown by **(a)** ball-and-stick and **(b)** space-filling models. The dashed lines in **(a)** represent the nonbonding orbitals. There is a nearly tetrahedral arrangement of the outer-shell electron pairs around the oxygen atom; the two hydrogen atoms have localized partial positive charges (δ^+) and the oxygen atom has a partial negative charge ($2\delta^-$). **(c)** Two H_2O molecules joined by a hydrogen bond (designated here, and throughout this book, by three blue lines) between the oxygen atom of the upper molecule and a hydrogen atom of the lower one. Hydrogen bonds are longer and weaker than covalent O—H bonds.

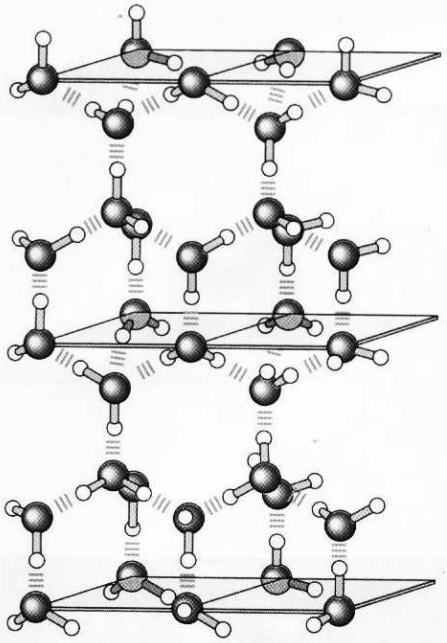


figure 4–2
Hydrogen bonding in ice. Each water molecule forms the maximum of four hydrogen bonds, creating a regular crystal lattice. In liquid water at room temperature and atmospheric pressure, by contrast, each water molecule hydrogen bonds with an average of 3.4 other water molecules. The crystal lattice of ice occupies more space than that occupied by the same number of H_2O molecules in liquid water; ice is less dense than—and thus floats on—liquid water.

Hydrogen bonds are weaker than covalent bonds. The hydrogen bonds in liquid water have a **bond dissociation energy** (the energy required to break a bond) of about 20 kJ/mol, compared with 348 kJ/mol for the covalent C—C bond. At room temperature, the thermal energy of an aqueous solution (the kinetic energy of motion of the individual atoms and molecules) is of the same order of magnitude as that required to break hydrogen bonds. When water is heated, the increase in temperature reflects the faster motion of individual water molecules. Although at any given time most of the molecules in liquid water are engaged in hydrogen bonding, the lifetime of each hydrogen bond is less than 1×10^{-9} s. The apt phrase "flickering clusters" has been applied to the short-lived groups of hydrogen-bonded molecules in liquid water. The sum of all the hydrogen bonds between molecules nevertheless confers great internal cohesion on liquid water.

The nearly tetrahedral arrangement of the orbitals about the oxygen atom (Fig. 4–1a) allows each water molecule to form hydrogen bonds with as many as four neighboring water molecules. In liquid water at room temperature and atmospheric pressure, however, water molecules are disorganized and in continuous motion, so that each molecule forms hydrogen bonds with an average of only 3.4 other molecules. In ice, on the other hand, each water molecule is fixed in space and forms hydrogen bonds with four other water molecules to yield a regular lattice structure (Fig. 4–2). Breakage of a sufficient number of hydrogen bonds to destabilize the crystal lattice of ice requires much thermal energy, which accounts for the relatively high melting point of water (Table 4–1). When ice melts or water evaporates, heat is taken up by the system:

$$H_2O(s) \longrightarrow H_2O(l) \qquad \Delta H = +5.9\ \text{kJ/mol}$$

$$H_2O(l) \longrightarrow H_2O(g) \qquad \Delta H = +44.0\ \text{kJ/mol}$$

During melting or evaporation, the entropy of the aqueous system increases as more highly ordered arrays of water molecules relax into the less orderly hydrogen-bonded arrays in liquid water or the wholly disordered gaseous state. At room temperature, both the melting of ice and the evaporation of water occur spontaneously; the tendency of the water molecules to associate through hydrogen bonds is outweighed by the energetic push toward randomness. Recall that the free-energy change (ΔG) must have a negative value for a process to occur spontaneously: $\Delta G = \Delta H - T\,\Delta S$, where ΔG represents the driving force, ΔH the enthalpy change from making and breaking bonds, and ΔS the change in randomness. Because ΔH is positive for melting and evaporation, it is clearly the increase in entropy (ΔS) that makes ΔG negative and drives these transformations.

Water Forms Hydrogen Bonds with Polar Solutes

Hydrogen bonds are not unique to water. They readily form between an electronegative atom (the hydrogen acceptor, usually oxygen or nitrogen with a lone pair of electrons) and a hydrogen atom covalently bonded to another electronegative atom (the hydrogen donor) in the same or another molecule (Fig. 4–3). Hydrogen atoms covalently bonded to carbon atoms (which are not electronegative) do not participate in hydrogen bonding. The distinction explains why butanol ($CH_3(CH_2)_2CH_2OH$) has a relatively high boiling point of 117 °C, whereas butane ($CH_3(CH_2)_2CH_3$) has a boiling point of only −0.5 °C. Butanol has a polar hydroxyl group and thus can form intermolecular hydrogen bonds.

Uncharged but polar biomolecules such as sugars dissolve readily in water because of the stabilizing effect of hydrogen bonds between the hydroxyl groups or carbonyl oxygen of the sugar and the polar water molecules. Alcohols, aldehydes, ketones, and compounds containing N—H bonds all form hydrogen bonds with water molecules (Fig. 4–4) and tend to be soluble in water.

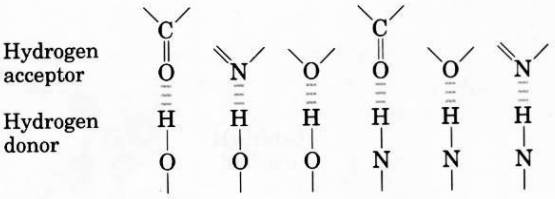


figure 4–3
Common hydrogen bonds in biological systems. The hydrogen acceptor is usually oxygen or nitrogen.

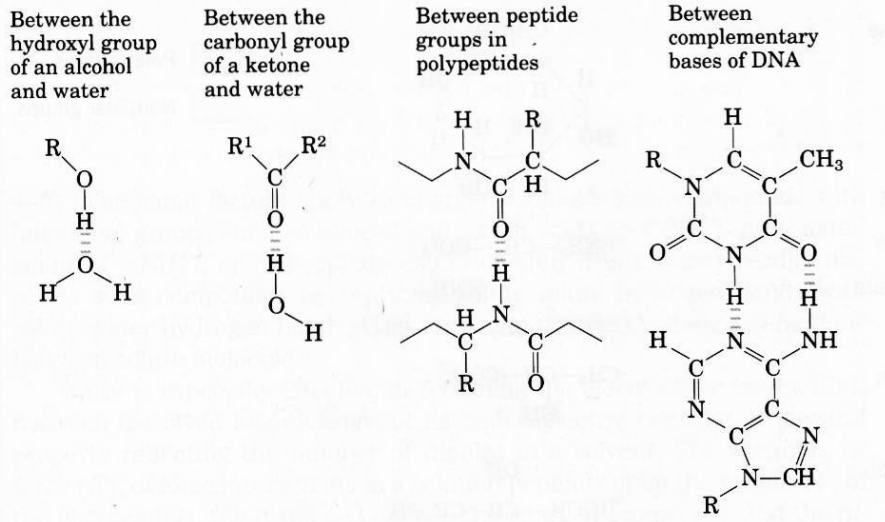


figure 4–4
Some biologically important hydrogen bonds.

Hydrogen bonds are strongest when the bonded molecules are oriented to maximize electrostatic interaction, which occurs when the hydrogen atom and the two atoms that share it are in a straight line—that is, when the acceptor atom is in line with the covalent bond between the donor atom and H (Fig. 4–5). Hydrogen bonds are thus highly directional and capable of holding two hydrogen-bonded molecules or groups in a specific geometric arrangement. As we shall see later, this property of hydrogen bonds confers very precise three-dimensional structures on protein and nucleic acid molecules, which have many intramolecular hydrogen bonds.

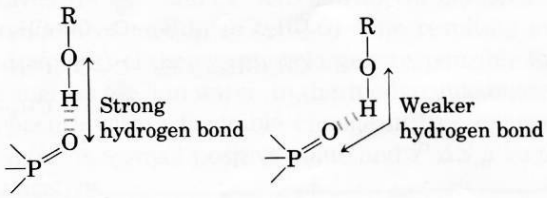


figure 4–5
Directionality of the hydrogen bond. The attraction between the partial electric charges (see Fig. 4–1) is greatest when the three atoms involved (in this case O, H, and O) lie in a straight line. When the hydrogen-bonded moieties are structurally constrained (as when they are parts of a single protein molecule, for example), this ideal geometry may not be possible and the resulting hydrogen bond is weaker.

Water Interacts Electrostatically with Charged Solutes

Water is a polar solvent. It readily dissolves most biomolecules, which are generally charged or polar compounds (Table 4–2); compounds that dissolve easily in water are **hydrophilic** (Greek, "water-loving"). In contrast, nonpolar solvents such as chloroform and benzene are poor solvents for polar biomolecules but easily dissolve those that are **hydrophobic**—nonpolar molecules such as lipids and waxes.

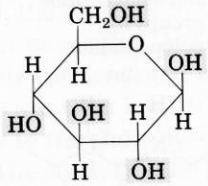
Water dissolves salts such as NaCl by hydrating and stabilizing the Na^+ and Cl^- ions, weakening the electrostatic interactions between them and thus counteracting their tendency to associate in a crystalline lattice (Fig.

table 4–2

Some Examples of Polar, Nonpolar, and Amphipathic Biomolecules (Shown as Ionic Forms at pH 7)

Polar

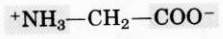
Glucose



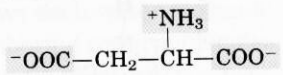
Polar groups

Nonpolar groups

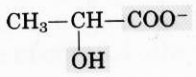
Glycine



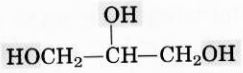
Aspartate



Lactate

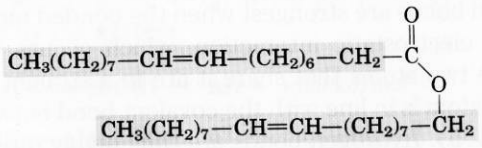


Glycerol



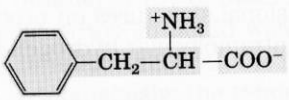
Nonpolar

Typical wax

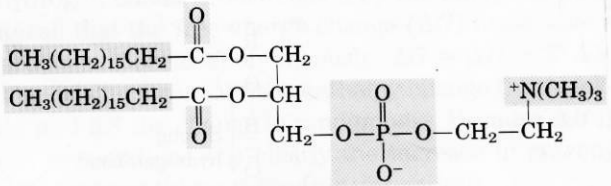


Amphipathic

Phenylalanine



Phosphatidylcholine



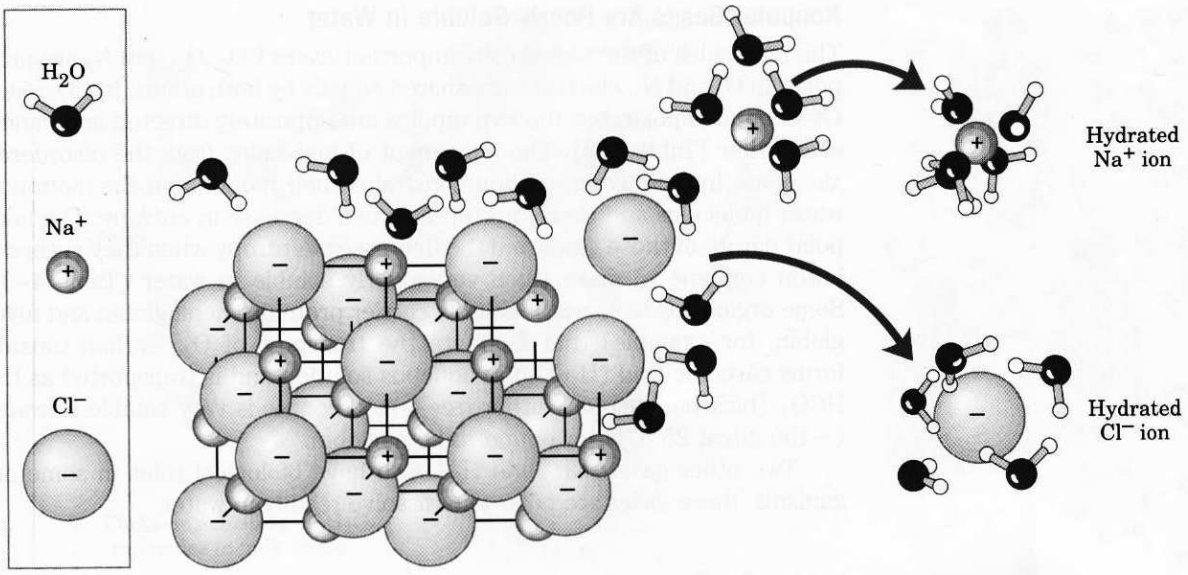


figure 4–6
Water dissolves many crystalline salts by hydrating their component ions. The NaCl crystal lattice is disrupted as water molecules cluster about the Cl^- and Na^+ ions. The ionic charges are partially neutralized, and the electrostatic attractions necessary for lattice formation are weakened.

4–6). The same factors apply to charged biomolecules, compounds with functional groups such as ionized carboxylic acids (—COO^-), protonated amines (—NH_3^+), and phosphate esters or anhydrides. Water readily dissolves such compounds by replacing solute-solute hydrogen bonds with solute-water hydrogen bonds, thus screening the electrostatic interactions between solute molecules.

Water is especially effective in screening the electrostatic interactions between dissolved ions because of its high dielectric constant, a physical property reflecting the number of dipoles in a solvent. The strength, or force (F), of ionic interactions in a solution depends upon the magnitude of the charges (Q), the distance between the charged groups (r), and the dielectric constant (ϵ) of the solvent in which the interactions occur:

$$F = \frac{Q_1 Q_2}{\epsilon r^2}$$

For water at 25 °C, ϵ (which is dimensionless) is 78.5, and for the very nonpolar solvent benzene, ϵ is 4.6. Thus, ionic interactions are much stronger in less polar environments. The dependence on r^2 is such that ionic attractions or repulsions operate only over short distances—in the range of 10 to 40 nm (depending on the electrolyte concentration) when the solvent is water.

Entropy Increases as Crystalline Substances Dissolve

As a salt such as NaCl dissolves, the Na^+ and Cl^- ions leaving the crystal lattice acquire far greater freedom of motion (Fig. 4–6). The resulting increase in the entropy (randomness) of the system is largely responsible for the ease of dissolving salts such as NaCl in water. In thermodynamic terms, formation of the solution occurs with a favorable change in free energy: $\Delta G = \Delta H - T\,\Delta S$, where ΔH has a small positive value and $T\,\Delta S$ a large positive value; thus ΔG is negative.

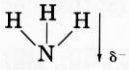
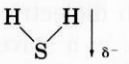
Nonpolar Gases Are Poorly Soluble in Water

The molecules of the biologically important gases CO_2, O_2, and N_2 are nonpolar. In O_2 and N_2, electrons are shared equally by both atoms. In CO_2, each C=O bond is polar, but the two dipoles are oppositely directed and cancel each other (Table 4–3). The movement of molecules from the disordered gas phase into aqueous solution constrains their motion and the motion of water molecules and therefore represents a decrease in entropy. The nonpolar nature of these gases and the decrease in entropy when they enter solution combine to make them very poorly soluble in water (Table 4–3). Some organisms have water-soluble carrier proteins (hemoglobin and myoglobin, for example) that facilitate the transport of O_2. Carbon dioxide forms carbonic acid (H_2CO_3) in aqueous solution and is transported as the HCO_3^- (bicarbonate) ion, either free—bicarbonate is very soluble in water (~100 g/L at 25 °C)—or bound to hemoglobin.

Two other gases, NH_3 and H_2S, also have biological roles in some organisms; these gases are polar and dissolve readily in water.

table 4–3

Solubilities of Some Gases in Water

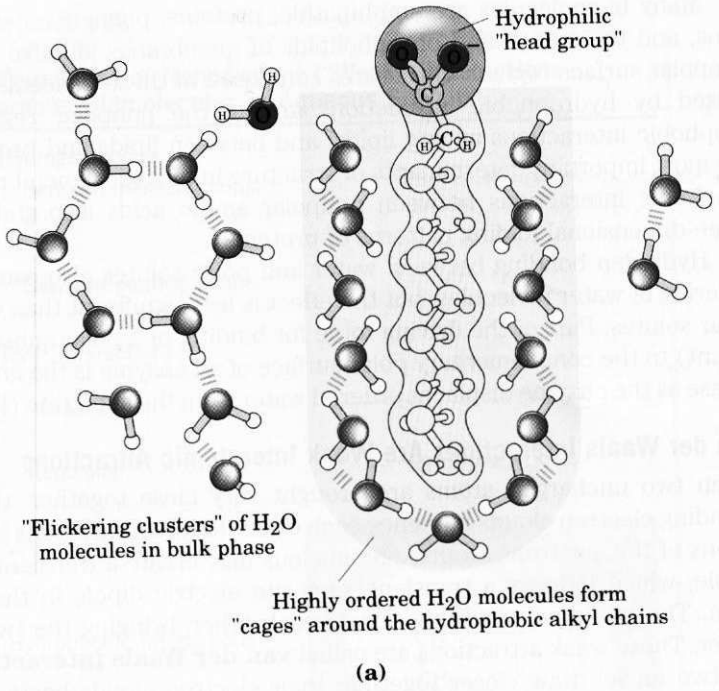
Gas	Structure*	Polarity	Solubility in water (g/L)†
Nitrogen	N≡N	Nonpolar	0.018 (40 °C)
Oxygen	O=O	Nonpolar	0.035 (50 °C)
Carbon dioxide	$\overset{\delta^-}{\leftarrow}\ \overset{\delta^-}{\rightarrow}$ O=C=O	Nonpolar	0.97 (45 °C)
Ammonia		Polar	900 (10 °C)
Hydrogen sulfide		Polar	1,860 (40 °C)

*The arrows represent electric dipoles; there is a partial negative charge (δ^-) at the head of the arrow, a partial positive charge (δ^+; not shown here) at the tail.

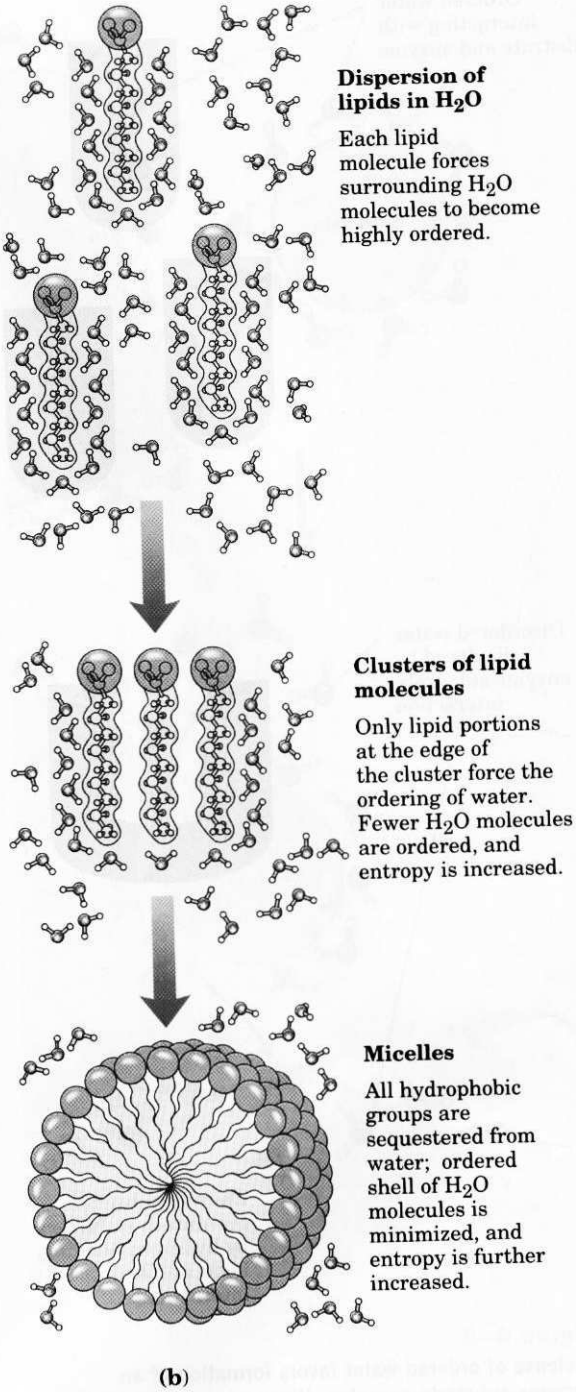
†Note that polar molecules dissolve far better even at low temperatures than do nonpolar molecules at relatively high temperatures.

Nonpolar Compounds Force Energetically Unfavorable Changes in the Structure of Water

When water is mixed with benzene or hexane, two phases form; neither liquid is soluble in the other. Nonpolar compounds such as benzene and hexane are hydrophobic—they are unable to undergo energetically favorable interactions with water molecules, and they actually interfere with the hydrogen bonding among water molecules. All molecules or ions in aqueous solution interfere with the hydrogen bonding of some water molecules in their immediate vicinity, but polar or charged solutes (such as NaCl) compensate for lost water-water hydrogen bonds by forming new solute-water interactions. The net change in enthalpy (ΔH) for dissolving these solutes is generally small. Hydrophobic solutes, however, offer no such compensation, and their addition to water may therefore result in a small gain of enthalpy; the breaking of hydrogen bonds between water molecules takes up



(a)



(b)

energy from the system. Furthermore, dissolving hydrophobic compounds in water produces a measurable decrease in entropy. Water molecules in the immediate vicinity of a nonpolar solute are constrained in their possible orientations as they form a highly ordered cagelike shell around each solute molecule. These water molecules are not as highly ordered as those in the crystalline compound of a nonpolar solute and water (a **clathrate**), but the effect is the same in both cases: the ordering of water molecules reduces entropy. The number of ordered water molecules, and therefore the magnitude of the entropy decrease, is proportional to the surface area of the hydrophobic solute enclosed within the cage of water molecules. The free-energy change for dissolving a nonpolar solute in water is thus unfavorable: $\Delta G = \Delta H - T\,\Delta S$, where ΔH has a positive value, ΔS has a negative value, and ΔG is positive.

Amphipathic compounds contain regions that are polar (or charged) and regions that are nonpolar (Table 4–2). When an amphipathic compound is mixed with water, the polar, hydrophilic region interacts favorably with the solvent and tends to dissolve, but the nonpolar, hydrophobic region tends to avoid contact with the water (Fig. 4–7a). The nonpolar regions of the molecules cluster together to present the smallest hydrophobic area to the aqueous solvent, and the polar regions are arranged to maximize their interaction with the solvent (Fig. 4–7b). These stable structures of amphipathic compounds in water, called **micelles,** may contain hundreds or thousands of molecules. The forces that hold the nonpolar regions of the molecules together are called **hydrophobic interactions.** The strength of hydrophobic interactions is not due to any intrinsic attraction between nonpolar moieties. Rather, it results from the system's achieving greatest thermodynamic stability by minimizing the number of ordered water molecules required to surround hydrophobic portions of the solute molecules.

figure 4–7
Amphipathic compounds in aqueous solution. (a) Long-chain fatty acids have very hydrophobic alkyl chains, each of which is surrounded by a layer of highly ordered water molecules. **(b)** By clustering together in micelles, the fatty acid molecules expose the smallest possible hydrophobic surface area to the water, and fewer water molecules are required in the shell of ordered water. The energy gained by freeing immobilized water molecules stabilizes the micelle.

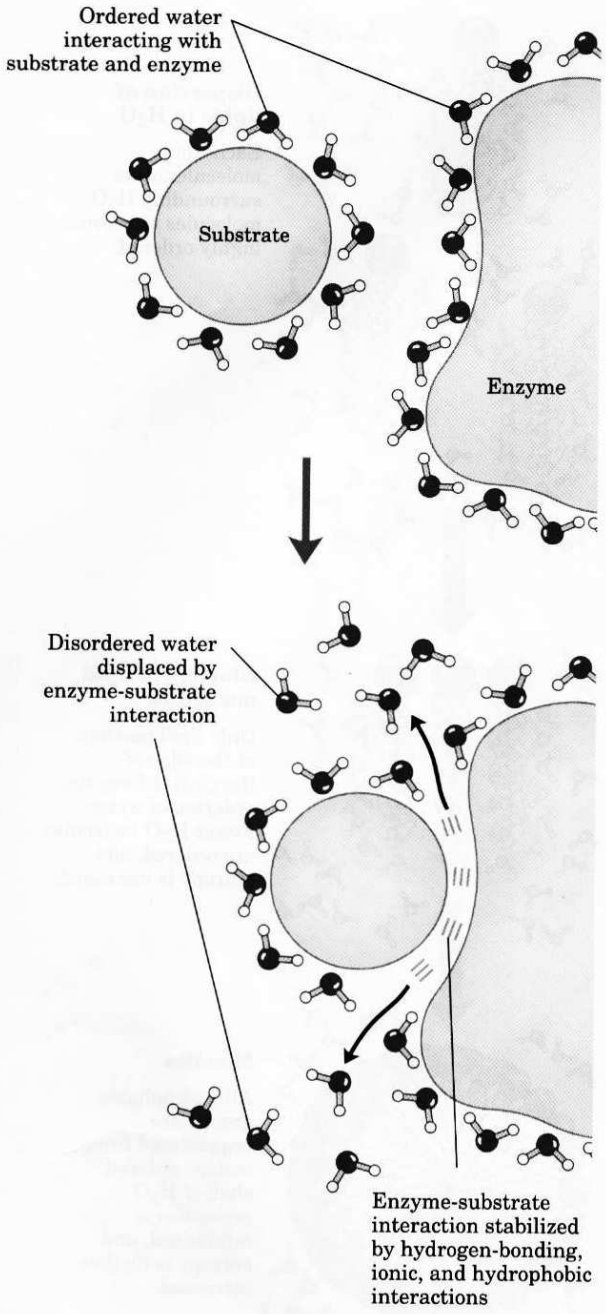


figure 4–8

Release of ordered water favors formation of an enzyme-substrate complex. While separate, both enzyme and substrate force neighboring water molecules into an ordered shell. Binding of substrate to enzyme releases some of the ordered water, and the resulting increase in entropy provides a thermodynamic push toward formation of the enzyme-substrate complex.

Many biomolecules are amphipathic; proteins, pigments, certain vitamins, and the sterols and phospholipids of membranes all have polar and nonpolar surface regions. Structures composed of these molecules are stabilized by hydrophobic interactions among the nonpolar regions. Hydrophobic interactions among lipids, and between lipids and proteins, are the most important determinants of structure in biological membranes. Hydrophobic interactions between nonpolar amino acids also stabilize the three-dimensional folding patterns of proteins.

Hydrogen bonding between water and polar solutes also causes some ordering of water molecules, but the effect is less significant than with nonpolar solutes. Part of the driving force for binding of a polar substrate (reactant) to the complementary polar surface of an enzyme is the entropy increase as the enzyme displaces ordered water from the substrate (Fig. 4–8).

Van der Waals Interactions Are Weak Interatomic Attractions

When two uncharged atoms are brought very close together, their surrounding electron clouds influence each other. Random variations in the positions of the electrons around one nucleus may create a transient electric dipole, which induces a transient, opposite electric dipole in the nearby atom. The two dipoles weakly attract each other, bringing the two nuclei closer. These weak attractions are called **van der Waals interactions.** As the two nuclei draw closer together, their electron clouds begin to repel each other. At the point when the van der Waals attraction exactly balances this repulsive force, the nuclei are said to be in van der Waals contact. Each atom has a characteristic **van der Waals radius,** a measure of how close that atom will allow another to approach (see Table 3–1). In the "space-filling" molecular models shown throughout this book (e.g., Fig. 3–7c) the atoms are depicted in sizes proportional to their van der Waals radii.

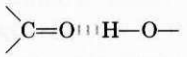
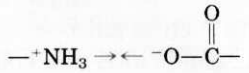
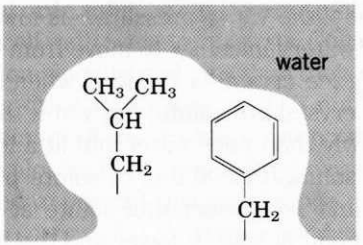
Weak Interactions Are Crucial to Macromolecular Structure and Function

The noncovalent interactions we have described (hydrogen bonds and ionic, hydrophobic, and van der Waals interactions) (Table 4–4) are much weaker than covalent bonds. An input of about 350 kJ of energy is required to break a mole of (6×10^{23}) C—C single bonds, and about 410 kJ to break a mole of C—H bonds, but as little as 4 kJ is sufficient to disrupt a mole of typical van der Waals interactions. Hydrophobic interactions are also much weaker than covalent bonds, although they are substantially strengthened by a highly polar solvent (a concentrated salt solution, for example). Ionic interactions and hydrogen bonds are variable in strength, depending on the polarity of the solvent, but they are always significantly weaker than covalent bonds. In aqueous solvent at 25 °C, the available thermal energy can be of the same order of magnitude as the strength of these weak interactions, and the interaction between solute and solvent (water) molecules is nearly as favorable as solute-solute interactions. Consequently, hydrogen bonds and ionic, hydrophobic, and van der Waals interactions are continually formed and broken.

Although these four types of interactions are individually weak relative to covalent bonds, the cumulative effect of many such interactions with a protein or nucleic acid can be very significant. For example, the noncovalent binding of an enzyme to its substrate may involve several hydrogen bonds and one or more ionic interactions, as well as hydrophobic and van der Waals interactions. The formation of each of these weak bonds contributes to a net decrease in the free energy of the system. The stability of a noncovalent interaction such as that of a small molecule hydrogen-bonded to its macromolecular partner is calculable from the binding energy. Stabil-

table 4–4

Four Types of Noncovalent ("Weak") Interactions among Biomolecules in Aqueous Solvent

Hydrogen bonds	
Between neutral groups	
Between peptide bonds	
Ionic interactions	
Attraction	
Repulsion	
Hydrophobic interactions	
Van der Waals interactions	Any two atoms in close proximity

ity, as measured by the equilibrium constant (see below) of the binding reaction, varies *exponentially* with binding energy. The dissociation of two biomolecules associated noncovalently by multiple weak interactions (such as an enzyme and its bound substrate) requires all these interactions to be disrupted at the same time. Because the interactions fluctuate randomly, such simultaneous disruptions are very unlikely. The molecular stability bestowed by two or five or 20 weak interactions is therefore much greater than would be expected intuitively from a simple summation of small binding energies.

Macromolecules such as proteins, DNA, and RNA contain so many sites of potential hydrogen bonding or ionic, van der Waals, or hydrophobic interactions that the cumulative effect of the many small binding forces is enormous. For macromolecules, the most stable (native) structure is usually that in which weak-bonding possibilities are maximized. The folding of a single polypeptide or polynucleotide chain into its three-dimensional shape is determined by this principle. The binding of an antigen to a specific antibody depends on the cumulative effects of many weak interactions. *As noted earlier, the energy released when an enzyme binds noncovalently to its substrate is the main source of the enzyme's catalytic power. The binding of a hormone or a neurotransmitter to its cellular receptor protein is the result of weak interactions. One consequence of the large size of enzymes and receptors is that their extensive surfaces provide many opportunities for weak interactions. At the molecular level, the complementarity* between interacting biomolecules reflects the complementarity and weak interactions between polar, charged, and hydrophobic groups on the surfaces of the molecules.

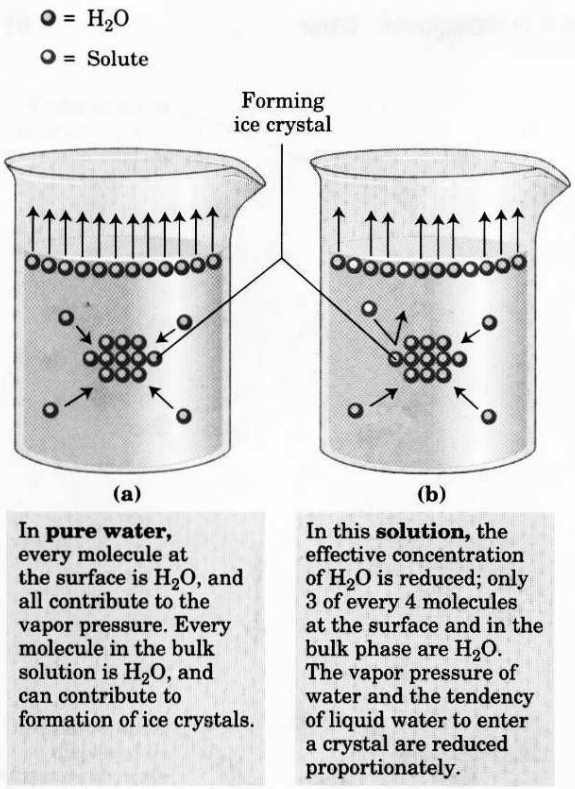


figure 4–9

Solutes alter the colligative properties of aqueous solutions. (a) At 101 kPa (1 atm) pressure, pure water boils at 100 °C and freezes at 0 °C. **(b)** The presence of solute molecules reduces the probability of a water molecule leaving the solution and entering the gas phase, thereby reducing the vapor pressure of the solution and increasing the boiling point. Similarly, the probability of a water molecule colliding with and joining a forming ice crystal is reduced when some of the molecules colliding with the crystal are solute, not water, molecules. The effect is depression of the freezing point.

Solutes Affect the Colligative Properties of Aqueous Solutions

Dissolved solutes of all kinds alter certain physical properties of the solvent, water: its vapor pressure, boiling point, melting point (freezing point), and osmotic pressure. These are called **colligative** ("tied together") **properties** because the effect of solutes on all four properties has the same basis: the concentration of water is lower in solutions than in pure water. The effect of solute concentration on the colligative properties of water is independent of the chemical properties of the solute; it depends only on the *number* of solute particles (molecules, ions) in a given amount of water. A compound such as NaCl, which dissociates in solution, has twice the effect on osmotic pressure, for example, as an equal number of moles of a nondissociating solute such as glucose.

Dissolved solutes alter the colligative properties of aqueous solutions by lowering the effective concentration of water. For example, when a significant fraction of the molecules at the surface of an aqueous solution are not water but solute, the tendency of water molecules to escape into the vapor phase—the vapor pressure—is lowered (Fig. 4–9). Similarly, the tendency of water molecules to move from the aqueous phase to the surface of a forming ice crystal is reduced when some of the molecules that collide with the crystal are solute, not water. In that case, the solution will freeze more slowly than pure water and at a lower temperature. For a 1.00 molal aqueous solution (1.00 mol of solute per 1,000 g water) of an ideal, nonvolatile, and nondissociating solute at 101 kPa (1 atm) of pressure, the freezing point is 1.86 °C lower and the boiling point is 0.543 °C higher than for pure water. For 0.100 molal solutions of the same solute, the changes are one-tenth as large.

Water molecules tend to move from a region of higher water concentration to one of lower water concentration. When two different aqueous solutions are separated by a semipermeable membrane (one that allows the passage of water but not solute molecules), water molecules diffusing from the region of higher water concentration to that of lower water concentration produce osmotic pressure (Fig. 4–10). This pressure, Π, measured as the force necessary to resist water movement (Fig. 4–10c), is approximated by the van't Hoff equation:

$$\Pi = icRT$$

in which R is the gas constant and T is the absolute temperature. The term ic is the **osmolarity** of the solution, the product of the solute's molar concentration c and the van't Hoff factor i, which is a measure of the extent to

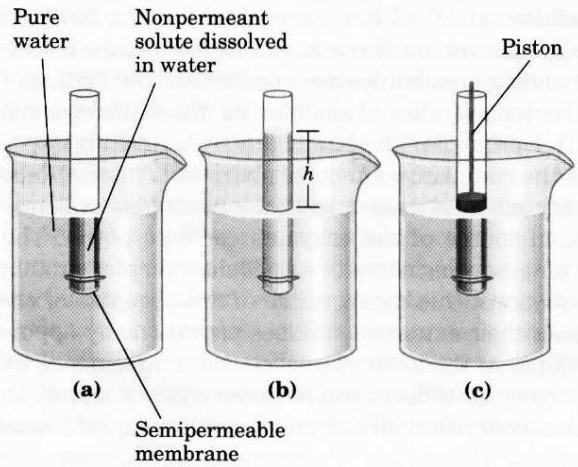


figure 4–10

Osmosis and the measurement of osmotic pressure. (a) The initial state. The tube contains an aqueous solution, the beaker contains pure water, and the semipermeable membrane allows the passage of water but not solute. Water flows from the beaker into the tube to equalize its concentration across the membrane. **(b)** The final state. Water has moved into the solution of the nonpermeant compound, diluting it and raising the column of water within the tube. At equilibrium, the force of gravity operating on the solution in the tube exactly balances the tendency of water to move into the tube, where its concentration is lower. **(c)** Osmotic pressure (Π) is measured as the force that must be applied to return the solution in the tube to the level of that in the beaker. This force is proportional to the height, h, of the column in **(b).**

which the solute dissociates into two or more ionic species. In dilute NaCl solutions, the solute completely dissociates into Na^+ and Cl^-, doubling the number of solute particles, and $i = 2$. For nonionizing solutes, i is always 1. For solutions of several (n) solutes, Π is the sum of the contributions of each species:

$$\Pi = RT(i_1c_1 + i_2c_2 + \cdots + i_nc_n)$$

Osmosis, water movement across a semipermeable membrane driven by differences in osmotic pressure, is an important factor in the life of most cells. Plasma membranes are more permeable to water than to most other small molecules, ions, and macromolecules. This permeability is due partly to simple diffusion of water through the lipid bilayer and partly to protein channels (aquaporins) in the membrane that selectively permit the passage of water. Solutions of equal osmolarity are said to be **isotonic.** Surrounded by an isotonic solution, a cell neither gains nor loses water (Fig. 4–11). In a **hypertonic** solution, one with higher osmolarity than the cytosol, the cell shrinks as water flows out. In **hypotonic** solution (of lower osmolarity), the cell swells and, if unsupported by a cell wall, eventually bursts. Cells generally contain higher concentrations of biomolecules and ions than their surroundings, so osmotic pressure tends to drive water into cells. If not somehow counterbalanced, this inward movement of water would distend the plasma membrane and eventually cause explosion of the cell (osmotic lysis).

Three mechanisms have evolved to prevent this catastrophe. In bacteria and plants, the plasma membrane is surrounded by a nonexpandable cell wall of sufficient rigidity and strength to resist osmotic pressure and prevent osmotic lysis. Certain freshwater protozoans, which live in a highly hypotonic medium, have an organelle (contractile vacuole) that pumps water out of the cell. In multicellular animals, blood plasma and interstitial fluid (the extracellular fluid of tissues) are maintained at an osmolarity close to that of the cytosol. The high concentration of albumin and other proteins in blood plasma contributes to its osmolarity. Cells also actively pump out ions such as Na^+ into the interstitial fluid to stay in osmotic balance with their surroundings.

Because the effect of solutes on osmolarity depends on the *number* of dissolved particles, not their *masses*, macromolecules (proteins, nucleic acids, polysaccharides) have far less effect on the osmolarity of a solution than would an equal mass of their monomeric components. For example, a *gram* of a polysaccharide composed of 1,000 glucose units has the same effect on osmolarity as a *milligram* of glucose. One effect of storing fuel as polysaccharides (starch or glycogen) rather than as glucose or other simple sugars is prevention of an enormous increase in osmotic pressure within the storage cell.

Plants use osmotic pressure to achieve mechanical rigidity. The very high solute concentration in the vacuole draws water into the cell (see Fig. 2–10). The resulting osmotic pressure against the cell wall (turgor pressure) stiffens the cell, the tissue, and the plant body. When the lettuce in your salad wilts, it is because loss of water has reduced turgor pressure. Dramatic alterations in turgor pressure produce the movement of plant parts seen in touch-sensitive plants such as the Venus flytrap and mimosa (Box 4–1).

Osmosis also has consequences for laboratory protocols. Mitochondria, chloroplasts, and lysosomes, for example, are bounded by semipermeable membranes. In isolating these organelles from broken cells (see Fig. 2–20), biochemists must perform the fractionations in isotonic solutions. Buffers used in cellular fractionations commonly contain sufficient concentrations (about 0.2 M) of sucrose or some other inert solute to protect the organelles from osmotic lysis.

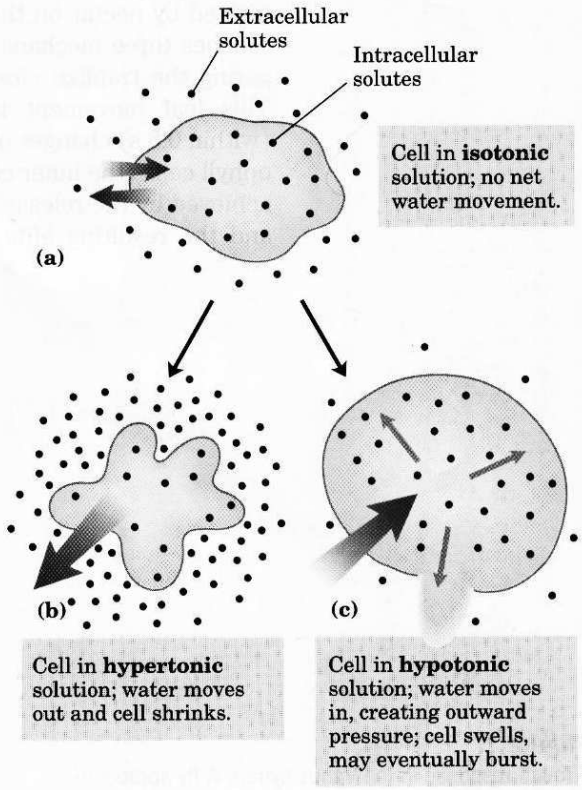


figure 4–11
The effect of extracellular osmolarity on water movement across a plasma membrane. When a cell in osmotic balance with its surrounding medium (that is, in an isotonic medium) **(a)** is transferred into a hypertonic solution **(b)** or hypotonic solution **(c)**, water moves across the plasma membrane in the direction that tends to equalize osmolarity outside and inside the cell.

box 4–1

Touch Response in Plants: An Osmotic Event

The highly specialized leaves of the Venus flytrap (*Dionaea muscipula*) rapidly fold together in response to a light touch by an unsuspecting insect, entrapping the insect for later digestion. Attracted by nectar on the leaf surface, the insect touches three mechanically sensitive hairs, triggering the traplike closing of the leaf (Fig. 1). This leaf movement is produced by sudden (within 0.5 s) changes of turgor pressure in mesophyll cells (the inner cells of the leaf), probably achieved by the release of K^+ ions from the cells and the resulting efflux, by osmosis, of water. Digestive glands in the leaf's surface release enzymes that extract nutrients from the insect.

The sensitive plant (*Mimosa pudica*) also undergoes a remarkable change in leaf shape triggered by mechanical touch (Fig. 2). A light touch or vibration produces a sudden drooping of the leaves, a result of a dramatic reduction in turgor pressure in cells at the base of each leaflet and leaf. As in the Venus flytrap, the drop in turgor pressure results from K^+ release followed by the efflux of water.



(a) (b)

figure 1
Touch response in the Venus flytrap. A fly approaching an open leaf **(a)** is trapped for digestion by the plant **(b).**



(a) (b)

figure 2
The feathery leaflets of the sensitive plant **(a)** close and drop **(b)** to protect the plant from structural damage by wind.

Ionization of Water, Weak Acids, and Weak Bases

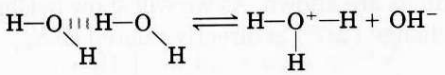
Although many of the solvent properties of water can be explained in terms of the uncharged H_2O molecule, the small degree of ionization of water to hydrogen ions (H^+) and hydroxide ions (OH^-) must also be taken into account. Like all reversible reactions, the ionization of water can be described by an equilibrium constant. When weak acids are dissolved in water, they contribute H^+ by ionizing; bases consume H^+ by being protonated. These processes are also governed by equilibrium constants. The total hydrogen ion concentration from all sources is experimentally measurable, and is expressed as the pH of the solution. To predict the state of ionization of solutes in water, we must take into account the relevant equilibrium constants for each ionization reaction. We therefore turn now to a brief discussion of the ionization of water, and of weak acids and bases dissolved in water.

Pure Water Is Slightly Ionized

Water molecules have a slight tendency to undergo reversible ionization to yield a hydrogen ion (proton) and a hydroxide ion, giving the equilibrium

$$H_2O \rightleftharpoons H^+ + OH^- \qquad (4\text{–}1)$$

Although we commonly show the dissociation product of water as H^+, free protons do not exist in solution; hydrogen ions formed in water are immediately hydrated to **hydronium ions** (H_3O^+). Hydrogen bonding between water molecules makes the hydration of dissociating protons virtually instantaneous:



The ionization of water can be measured by its electrical conductivity; pure water carries electrical current as H^+ migrates toward the cathode and OH^- toward the anode. The movement of hydronium and hydroxide ions in the electric field is anomalously fast compared with that of other ions such as Na^+, K^+, and Cl^-. This high ionic mobility results from the kind of "proton hopping" shown in Figure 4–12. No individual proton moves very far through the bulk solution, but a series of proton hops between hydrogen-bonded water molecules causes the net movement of a proton over a long distance in a remarkably short time. As a result of the high ionic mobility of H^+ (and of OH^-, which also moves rapidly by proton hopping, but in the opposite direction), acid-base reactions in aqueous solutions are generally exceptionally fast. Proton hopping very likely also plays a role in biological proton transfer reactions.

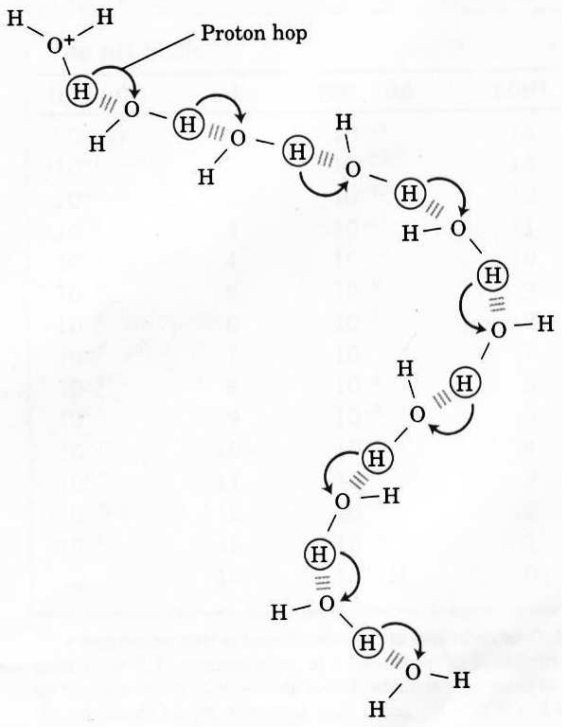


figure 4–12
Proton hopping. Short "hops" of protons between a series of hydrogen-bonded water molecules effects an extremely rapid net movement of a proton over a long distance. As a hydronium ion (upper left) gives up a proton, a water molecule some distance away (lower right) acquires one, becoming a hydronium ion. Proton hopping is much faster than true diffusion and explains the remarkably high ionic mobility of hydrogen ions compared with other monovalent cations such as Na^+ or K^+.

Because reversible ionization is crucial to the role of water in cellular function, we must have a means of expressing the extent of ionization of water in quantitative terms. A brief review of some properties of reversible chemical reactions will show how this can be done.

The position of equilibrium of any chemical reaction is given by its **equilibrium constant, K_{eq}** (sometimes expressed simply as K). For the generalized reaction

$$A + B \rightleftharpoons C + D \tag{4–2}$$

an equilibrium constant can be defined in terms of the concentrations of reactants (A and B) and products (C and D) at equilibrium:

$$K_{eq} = \frac{[C][D]}{[A][B]}$$

Strictly speaking, the concentration terms should be the *activities,* or effective concentrations in nonideal solutions, of each species. Except in very accurate work, the equilibrium constant may be approximated by measuring the *concentrations* at equilibrium. For reasons beyond the scope of this discussion, equilibrium constants are dimensionless. However, we have generally retained the concentration units (M) in the equilibrium expressions used in this book to remind you that molarity is the unit of concentration used in calculating K_{eq}.

The equilibrium constant is fixed and characteristic for any given chemical reaction at a specified temperature. It defines the composition of the final equilibrium mixture, regardless of the starting amounts of reactants and products. Conversely, one can calculate the equilibrium constant for a given reaction at a given temperature if the equilibrium concentrations of all its reactants and products are known. As we will show in Chapter 14, the standard free-energy change ($\Delta G°$) is directly related to K_{eq}.

The Ionization of Water Is Expressed by an Equilibrium Constant

The degree of ionization of water at equilibrium (Eqn 4–1) is small; at 25 °C only about one of every 10^7 molecules in pure water is ionized at any instant. The equilibrium constant for the reversible ionization of water (Eqn 4–1) is

$$K_{eq} = \frac{[H^+][OH^-]}{[H_2O]} \tag{4–3}$$

In pure water at 25 °C, the concentration of water is 55.5 M (grams of H_2O in 1 L divided by its gram molecular weight: (1,000 g/L)/(18.015 g/mol)) and is essentially constant in relation to the very low concentrations of H^+ and OH^-, namely, 1×10^{-7} M. Accordingly, we can substitute 55.5 M in the equilibrium constant expression (Eqn 4–3) to yield

$$K_{eq} = \frac{[H^+][OH^-]}{55.5\ \text{M}},$$

which, on rearranging, becomes

$$(55.5\ \text{M})(K_{eq}) = [H^+][OH^-] = K_w \tag{4–4}$$

where K_w designates the product $(55.5\ \text{M})(K_{eq})$, the **ion product of water** at 25 °C.

The value for K_{eq}, determined by electrical-conductivity measurements of pure water, is 1.8×10^{-16} M at 25 °C. Substituting this value for K_{eq} in Equation 4–4 gives the ion product of water:

$$K_w = [H^+][OH^-] = (55.5\ \text{M})(1.8 \times 10^{-16}\ \text{M}) = 1.0 \times 10^{-14}\ \text{M}^2$$

box 4–2 The Ion Product of Water: Two Illustrative Problems

The ion product of water makes it possible to calculate the concentration of H^+, given the concentration of OH^-, and vice versa; the following problems demonstrate this.

1. What is the concentration of H^+ in a solution of 0.1 M NaOH?

$$K_w = [H^+][OH^-]$$

Solving for $[H^+]$ gives

$$[H^+] = \frac{K_w}{[OH^-]} = \frac{1 \times 10^{-14}\ M^2}{0.1\ M} = \frac{10^{-14}\ M^2}{10^{-1}\ M}$$
$$= 10^{-13}\ M \quad \textit{(answer)}$$

2. What is the concentration of OH^- in a solution in which the H^+ concentration is 0.00013 M?

$$K_w = [H^+][OH^-]$$

Solving for $[OH^-]$ gives

$$[OH^-] = \frac{K_w}{[H^+]} = \frac{1 \times 10^{-14}\ M^2}{0.00013\ M} = \frac{10^{-14}\ M^2}{1.3 \times 10^{-4}\ M}$$
$$= 7.7 \times 10^{-11}\ M \quad \textit{(answer)}$$

Thus the product $[H^+][OH^-]$ in aqueous solutions at 25 °C always equals $1 \times 10^{-14}\ M^2$. When there are exactly equal concentrations of both H^+ and OH^-, as in pure water, the solution is said to be at **neutral pH.** At this pH, the concentration of H^+ and OH^- can be calculated from the ion product of water as follows:

$$K_w = [H^+][OH^-] = [H^+]^2$$

Solving for $[H^+]$ gives

$$[H^+] = \sqrt{K_w} = \sqrt{1 \times 10^{-14}\ M^2}$$

$$[H^+] = [OH^-] = 10^{-7}\ M$$

As the ion product of water is constant, whenever $[H^+]$ is greater than 1×10^{-7} M, $[OH^-]$ must become less than 1×10^{-7} M, and vice versa. When $[H^+]$ is very high, as in a solution of hydrochloric acid, $[OH^-]$ must be very low. From the ion product of water we can calculate $[H^+]$ if we know $[OH^-]$, and vice versa (Box 4–2).

The pH Scale Designates the H^+ and OH^- Concentrations

The ion product of water, K_w, is the basis for the **pH scale** (Table 4–5). It is a convenient means of designating the concentration of H^+ (and thus of OH^-) in any aqueous solution in the range between 1.0 M H^+ and 1.0 M OH^-. The term **pH** is defined by the expression

$$pH = \log \frac{1}{[H^+]} = -\log [H^+]$$

The symbol p denotes "negative logarithm of." For a precisely neutral solution at 25 °C, in which the concentration of hydrogen ions is 1.0×10^{-7} M, the pH can be calculated as follows:

$$pH = \log \frac{1}{1.0 \times 10^{-7}} = \log (1.0 \times 10^7) = \log 1.0 + \log 10^7 = 0 + 7 = 7$$

The value of 7 for the pH of a precisely neutral solution is not an arbitrarily chosen figure; it is derived from the absolute value of the ion product of water at 25 °C, which by convenient coincidence is a round number. Solutions

table 4–5

The pH Scale

$[H^+]$ (M)	pH	$[OH^-]$ (M)	pOH*
10^0 (1)	0	10^{-14}	14
10^{-1}	1	10^{-13}	13
10^{-2}	2	10^{-12}	12
10^{-3}	3	10^{-11}	11
10^{-4}	4	10^{-10}	10
10^{-5}	5	10^{-9}	9
10^{-6}	6	10^{-8}	8
10^{-7}	7	10^{-7}	7
10^{-8}	8	10^{-6}	6
10^{-9}	9	10^{-5}	5
10^{-10}	10	10^{-4}	4
10^{-11}	11	10^{-3}	3
10^{-12}	12	10^{-2}	2
10^{-13}	13	10^{-1}	1
10^{-14}	14	10^0 (1)	0

*The expression pOH is sometimes used to describe the basicity, or OH^- concentration, of a solution; pOH is defined by the expression $pOH = -\log [OH^-]$, which is analogous to the expression for pH. Note that in all cases, pH + pOH = 14.

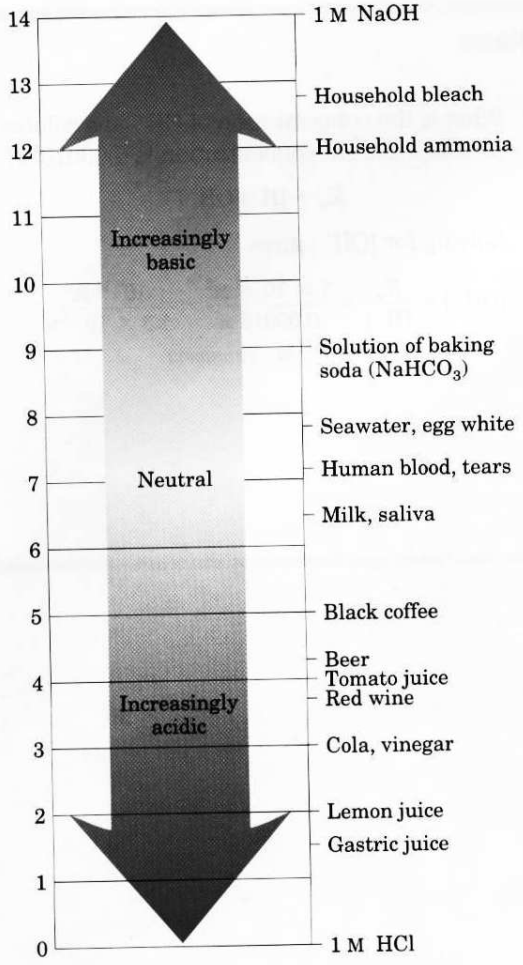


figure 4–13
The pH of some aqueous fluids.

having a pH greater than 7 are alkaline or basic; the concentration of OH^- is greater than that of H^+. Conversely, solutions having a pH less than 7 are acidic.

Note that the pH scale is logarithmic, not arithmetic. To say that two solutions differ in pH by 1 pH unit means that one solution has ten times the H^+ concentration of the other, but it does not tell us the absolute magnitude of the difference. Figure 4–13 gives the pH of some common aqueous fluids. A cola drink (pH 3.0) or red wine (pH 3.7) has an H^+ concentration approximately 10,000 times that of blood (pH 7.4).

The pH of an aqueous solution can be approximately measured using various indicator dyes, including litmus, phenolphthalein, and phenol red, which undergo color changes as a proton dissociates from the dye molecule. Accurate determinations of pH in the chemical or clinical laboratory are made with a glass electrode that is selectively sensitive to H^+ concentration but insensitive to Na^+, K^+, and other cations. In a pH meter the signal from such an electrode is amplified and compared with the signal generated by a solution of accurately known pH.

Measurement of pH is one of the most important and frequently used procedures in biochemistry. The pH affects the structure and activity of biological macromolecules; for example, the catalytic activity of enzymes is strongly dependent on pH (see Fig. 4–19). Measurements of the pH of blood and urine are commonly used in medical diagnoses. The pH of the blood plasma of severely diabetic people, for example, is often below the normal value of 7.4; this condition is called acidosis. In certain other disease states the pH of the blood is higher than normal, the condition of alkalosis.

Weak Acids and Bases Have Characteristic Dissociation Constants

Hydrochloric, sulfuric, and nitric acids, commonly called strong acids, are completely ionized in dilute aqueous solutions; the strong bases NaOH and KOH are also completely ionized. Of more interest to biochemists is the behavior of weak acids and bases—those not completely ionized when dissolved in water. These are common in biological systems and play important roles in metabolism and its regulation. The behavior of aqueous solutions of weak acids and bases is best understood if we first define some terms.

Acids may be defined as proton donors and bases as proton acceptors. A proton donor and its corresponding proton acceptor make up a **conjugate acid-base pair** (Fig. 4–14). Acetic acid (CH_3COOH), a proton donor, and the acetate anion (CH_3COO^-), the corresponding proton acceptor, constitute a conjugate acid-base pair, related by the reversible reaction

$$CH_3COOH \rightleftharpoons H^+ + CH_3COO^-$$

Each acid has a characteristic tendency to lose its proton in an aqueous solution. The stronger the acid, the greater its tendency to lose its proton. The tendency of any acid (HA) to lose a proton and form its conjugate base (A^-) is defined by the equilibrium constant (K_{eq}) for the reversible reaction

$$HA \rightleftharpoons H^+ + A^-$$

which is

$$K_{eq} = \frac{[H^+][A^-]}{[HA]} = K_a$$

Equilibrium constants for ionization reactions are usually called ionization or **dissociation constants,** often designated K_a. The dissociation constants of some acids are given in Figure 4–14. Stronger acids, such as phosphoric and carbonic acids, have larger dissociation constants; weaker acids, such as

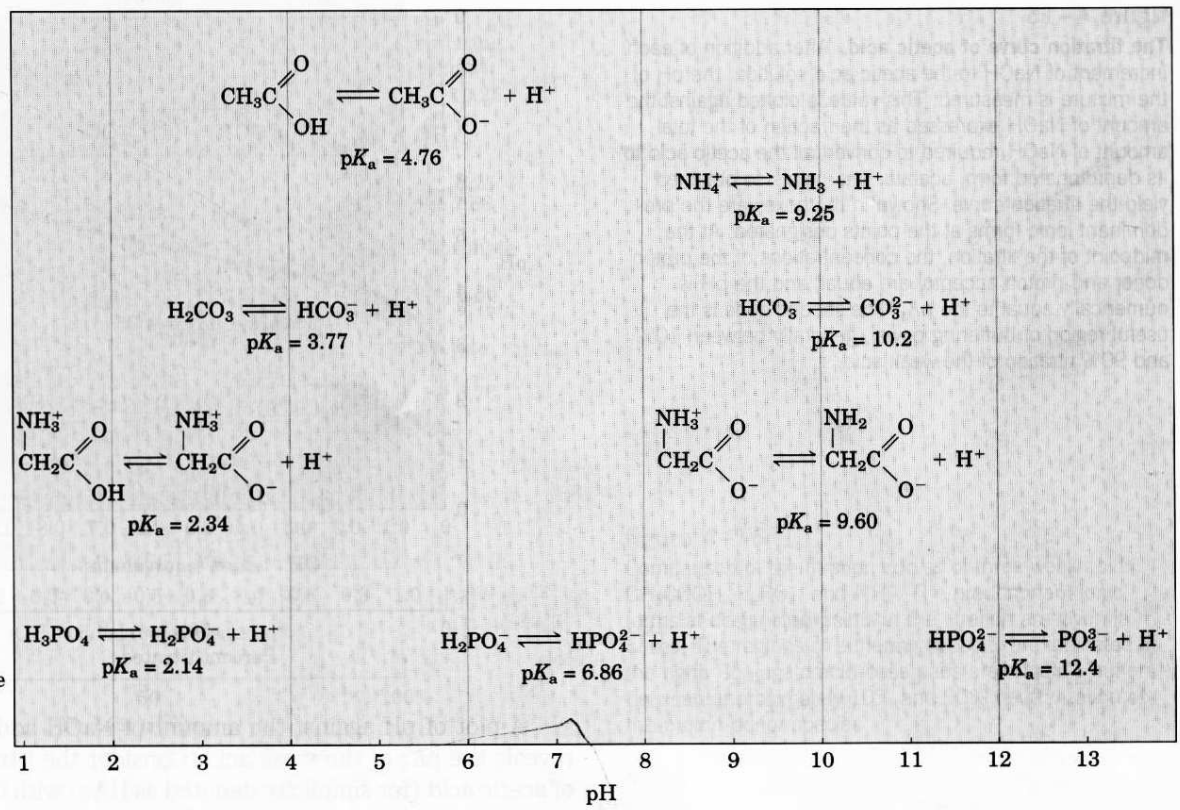


figure 4–14

Conjugate acid-base pairs consist of a proton donor and a proton acceptor. Some compounds, such as acetic acid and ammonia, are monoprotic; they can give up only one proton. Others are diprotic (H_2CO_3 and glycine) or triprotic (H_3PO_4). The dissociation reactions for each pair are shown where they occur along a pH gradient. The equilibrium or dissociation constant (K_a) and its negative logarithm, the pK_a, are shown for each reaction.

monohydrogen phosphate (HPO_4^{2-}), have smaller dissociation constants.

Also included in Figure 4–14 are values of **pK_a**, which is analogous to pH and is defined by the equation

$$pK_a = \log \frac{1}{K_a} = -\log K_a$$

The stronger the tendency to dissociate a proton, the stronger is the acid and the lower its pK_a. As we shall now see, the pK_a of any weak acid can be determined quite easily.

Titration Curves Reveal the pK_a of Weak Acids

Titration is used to determine the amount of an acid in a given solution. A measured volume of the acid is titrated with a solution of a strong base, usually sodium hydroxide (NaOH), of known concentration. The NaOH is added in small increments until the acid is consumed (neutralized), as determined with an indicator dye or a pH meter. The concentration of the acid in the original solution can be calculated from the volume and concentration of NaOH added.

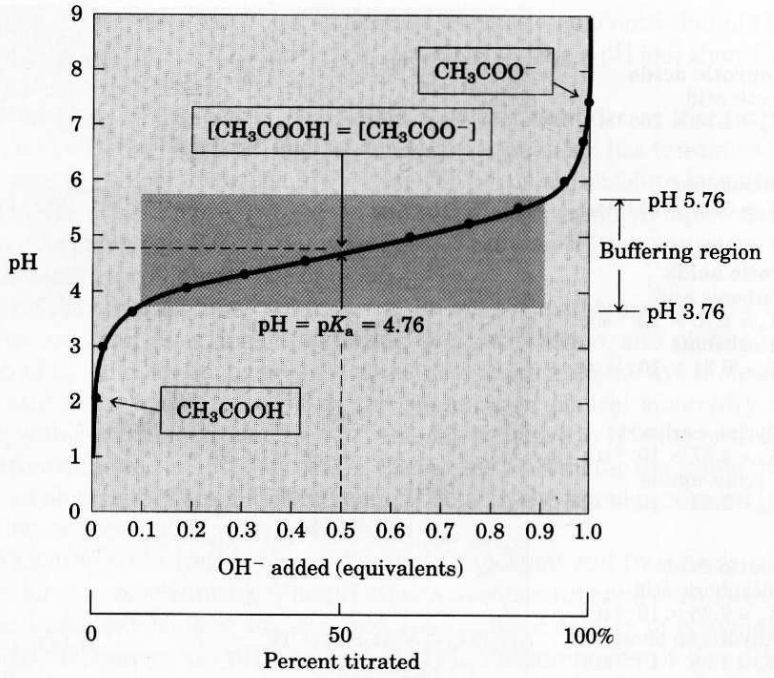


figure 4–15

The titration curve of acetic acid. After addition of each increment of NaOH to the acetic acid solution, the pH of the mixture is measured. This value is plotted against the amount of NaOH expressed as the fraction of the total amount of NaOH required to convert all the acetic acid to its deprotonated form, acetate. The points so obtained yield the titration curve. Shown in the boxes are the predominant ionic forms at the points designated. At the midpoint of the titration, the concentrations of the proton donor and proton acceptor are equal, and the pH is numerically equal to the pK_a. The shaded zone is the useful region of buffering power, generally between 10% and 90% titration of the weak acid.

A plot of pH against the amount of NaOH added (a **titration curve**) reveals the pK_a of the weak acid. Consider the titration of a 0.1 M solution of acetic acid (for simplicity denoted as HAc) with 0.1 M NaOH at 25 °C (Fig. 4–15). Two reversible equilibria are involved in the process:

$$H_2O \rightleftharpoons H^+ + OH^- \qquad (4\text{–}5)$$

$$HAc \rightleftharpoons H^+ + Ac^- \qquad (4\text{–}6)$$

The equilibria must simultaneously conform to their characteristic equilibrium constants, which are, respectively,

$$K_w = [H^+][OH^-] = 1 \times 10^{-14}\ M^2 \qquad (4\text{–}7)$$

$$K_a = \frac{[H^+][Ac^-]}{[HAc]} = 1.74 \times 10^{-5}\ M \qquad (4\text{–}8)$$

At the beginning of the titration, before any NaOH is added, the acetic acid is already slightly ionized, to an extent that can be calculated from its dissociation constant (Eqn 4–8).

As NaOH is gradually introduced, the added OH^- combines with the free H^+ in the solution to form H_2O, to an extent that satisfies the equilibrium relationship in Equation 4–7. As free H^+ is removed, HAc dissociates further to satisfy its own equilibrium constant (Eqn 4–8). The net result as the titration proceeds is that more and more HAc ionizes, forming Ac^-, as the NaOH is added. At the midpoint of the titration, at which exactly 0.5 equivalent of NaOH has been added, one-half of the original acetic acid has undergone dissociation, so that the concentration of the proton donor, [HAc], now equals that of the proton acceptor, $[Ac^-]$. At this midpoint a very important relationship holds: the pH of the equimolar solution of acetic acid and acetate is exactly equal to the pK_a of acetic acid ($pK_a = 4.76$; see Figs 4–14, 4–15). The basis for this relationship, which holds for all weak acids, will soon become clear.

As the titration is continued by adding further increments of NaOH, the remaining undissociated acetic acid is gradually converted into acetate. The end point of the titration occurs at about pH 7.0: all the acetic acid has lost its protons to OH^-, to form H_2O and acetate. Throughout the titration the

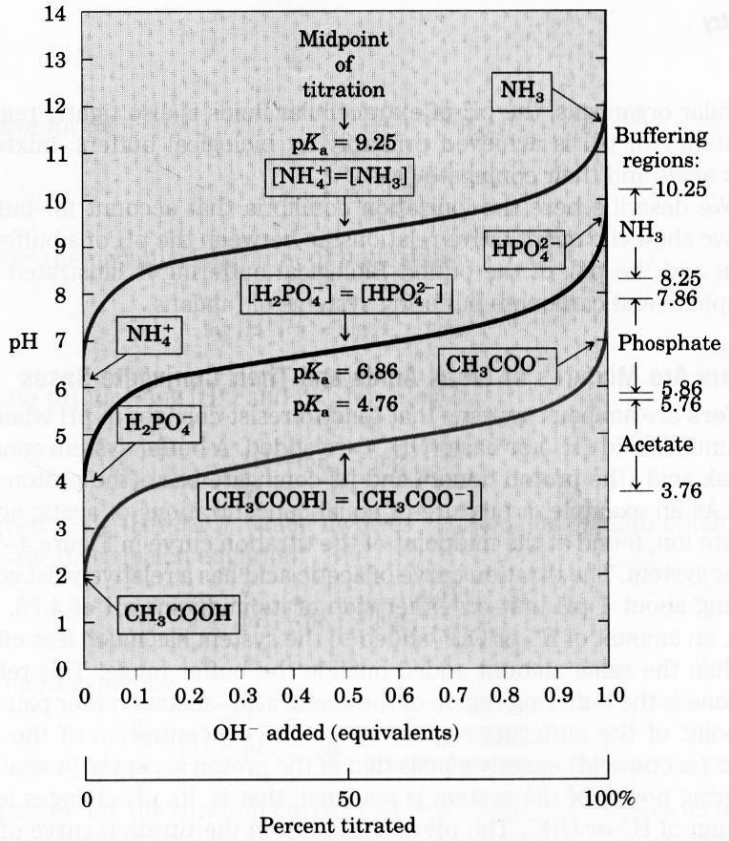


figure 4–16
Comparison of the titration curves of three weak acids, CH_3COOH, $H_2PO_4^-$, and NH_4^+. The predominant ionic forms at designated points in the titration are given in boxes. The regions of buffering capacity are indicated at the right. Conjugate acid-base pairs are effective buffers between approximately 10% and 90% neutralization of the proton-donor species.

two equilibria (Eqns 4–5, 4–6) coexist, each always conforming to its equilibrium constant.

Figure 4–16 compares the titration curves of three weak acids with very different dissociation constants: acetic acid ($pK_a = 4.76$); dihydrogen phosphate, $H_2PO_4^-$ ($pK_a = 6.86$); and ammonium ion, NH_4^+ ($pK_a = 9.25$). Although the titration curves of these acids have the same shape, they are displaced along the pH axis because the three acids have different strengths. Acetic acid is the strongest (loses its proton most readily) because its K_a is highest (pK_a lowest) of the three. Acetic acid is already half dissociated at pH 4.76. Dihydrogen phosphate loses a proton less readily, being half dissociated at pH 6.86. Ammonium ion is the weakest acid of the three and does not become half dissociated until pH 9.25.

The most important point about the titration curve of a weak acid is that it shows graphically that a weak acid and its anion—a conjugate acid-base pair—can act as a buffer.

Buffering against pH Changes in Biological Systems

Almost every biological process is pH dependent; a small change in pH produces a large change in the rate of the process. This is true not only for the many reactions in which the H^+ ion is a direct participant, but also for those in which there is no apparent role for H^+ ions. The enzymes that catalyze cellular reactions, and many of the molecules on which they act, contain ionizable groups with characteristic pK_a values. The protonated amino and carboxyl groups of amino acids and the phosphate groups of nucleotides, for example, function as weak acids; their ionic state depends on the pH of the surrounding medium. As we noted above, ionic interactions are among the forces that stabilize a protein molecule and allow an enzyme to recognize and bind its substrate.

Cells and organisms maintain a specific and constant cytosolic pH, keeping biomolecules in their optimal ionic state, usually near pH 7. In mul-

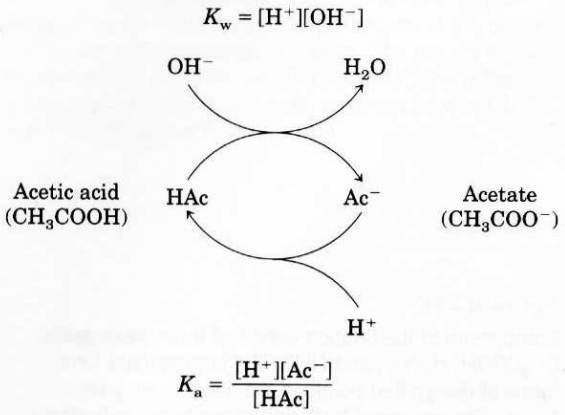


figure 4–17

The acetic acid–acetate pair as a buffer system. The system is capable of absorbing either H^+ or OH^- through the reversibility of the dissociation of acetic acid. The proton donor, acetic acid (HAc), contains a reserve of bound H^+, which can be released to neutralize an addition of OH^- to the system, forming H_2O. This happens because the product $[H^+][OH^-]$ transiently exceeds K_w (1×10^{-14} M^2). The equilibrium quickly adjusts so that this product equals 1×10^{-14} M^2 (at 25 °C), thus transiently reducing the concentration of H^+. But now the quotient $[H^+][Ac^-]/[HAc]$ is less then K_a, so HAc dissociates further to restore equilibrium. Similarly, the conjugate base, Ac^-, can react with H^+ ions added to the system; again, the two ionization reactions simultaneously come to equilibrium. Thus a conjugate acid-base pair, such as acetic acid and acetate ion, tends to resist a change in pH when small amounts of acid or base are added. Buffering action is simply the consequence of two reversible reactions taking place simultaneously and reaching their points of equilibrium as governed by their equilibrium constants, K_w and K_a.

ticellular organisms, the pH of extracellular fluids is also tightly regulated. Constancy of pH is achieved primarily by biological buffers: mixtures of weak acids and their conjugate bases.

We describe here the ionization equilibria that account for buffering, and we show the quantitative relationship between the pH of a buffered solution and the pK_a of the buffer. Biological buffering is illustrated by the phosphate and carbonate buffering systems of humans.

Buffers Are Mixtures of Weak Acids and Their Conjugate Bases

Buffers are aqueous systems that tend to resist changes in pH when small amounts of acid (H^+) or base (OH^-) are added. A buffer system consists of a weak acid (the proton donor) and its conjugate base (the proton acceptor). As an example, a mixture of equal concentrations of acetic acid and acetate ion, found at the midpoint of the titration curve in Figure 4–15, is a buffer system. The titration curve of acetic acid has a relatively flat zone extending about 1 pH unit on either side of its midpoint pH of 4.76. In this zone, an amount of H^+ or OH^- added to the system has much less effect on pH than the same amount added outside the buffer range. This relatively flat zone is the buffering region of the acetic acid–acetate buffer pair. At the midpoint of the buffering region, where the concentration of the proton donor (acetic acid) exactly equals that of the proton acceptor (acetate), the buffering power of the system is maximal; that is, its pH changes least on addition of H^+ or OH^-. The pH at this point in the titration curve of acetic acid is equal to its pK_a. The pH of the acetate buffer system does change slightly when a small amount of H^+ or OH^- is added, but this change is very small compared with the pH change that would result if the same amount of H^+ (or OH^-) were added to pure water or to a solution of the salt of a strong acid and strong base, such as NaCl, which has no buffering power.

Buffering results from two reversible reaction equilibria occurring in a solution of nearly equal concentrations of a proton donor and its conjugate proton acceptor. Figure 4–17 explains how a buffer system works. Whenever H^+ or OH^- is added to a buffer, the result is a small change in the ratio of the relative concentrations of the weak acid and its anion and thus a small change in pH. The decrease in concentration of one component of the system is balanced exactly by an increase in the other. The sum of the buffer components does not change, only their ratio.

Each conjugate acid-base pair has a characteristic pH zone in which it is an effective buffer (Fig. 4–16). The $H_2PO_4^-/HPO_4^{2-}$ pair has a pK_a of 6.86 and thus can serve as an effective buffer system between approximately pH 5.9 and pH 7.9; the NH_4^+/NH_3 pair, with a pK_a of 9.25, can act as a buffer between approximately pH 8.3 and pH 10.3.

A Simple Expression Relates pH, p*K*, and Buffer Concentration

The titration curves of acetic acid, $H_2PO_4^-$, and NH_4^+ (Fig. 4–16) have nearly identical shapes, suggesting that these curves reflect a fundamental law or relationship. This is indeed the case. The shape of the titration curve of any weak acid is described by the **Henderson-Hasselbalch equation,** which is important for understanding buffer action and acid-base balance in the blood and tissues of vertebrates. This equation is simply a useful way of restating the expression for the dissociation constant of an acid. For the dissociation of a weak acid HA into H^+ and A^-, the Henderson-Hasselbalch equation can be derived as follows:

$$K_a = \frac{[H^+][A^-]}{[HA]}$$

First solve for $[H^+]$:

$$[H^+] = K_a \frac{[HA]}{[A^-]}$$

Then take the negative logarithm of both sides:

$$-\log [H^+] = -\log K_a - \log \frac{[HA]}{[A^-]}$$

Substitute pH for $-\log [H^+]$ and pK_a for $-\log K_a$:

$$pH = pK_a - \log \frac{[HA]}{[A^-]}$$

Now invert $-\log [HA]/[A^-]$, which involves changing its sign, to obtain the Henderson-Hasselbalch equation:

$$pH = pK_a + \log \frac{[A^-]}{[HA]}$$

Stated more generally,

$$pH = pK_a + \log \frac{[\text{proton acceptor}]}{[\text{proton donor}]}$$

This equation fits the titration curve of all weak acids and enables us to deduce a number of important quantitative relationships. For example, it shows why the pK_a of a weak acid is equal to the pH of the solution at the midpoint of its titration. At that point, [HA] equals $[A^-]$, and

$$pH = pK_a + \log 1.0 = pK_a + 0 = pK_a$$

As shown in Box 4–3, the Henderson-Hasselbalch equation also makes it possible to (1) calculate pK_a, given pH and the molar ratio of proton donor and acceptor; (2) calculate pH, given pK_a and the molar ratio of proton donor and acceptor; and (3) calculate the molar ratio of proton donor and acceptor, given pH and pK_a.

box 4–3 Solving Problems Using the Henderson-Hasselbalch Equation

1. Calculate the pK_a of lactic acid, given that when the concentration of lactic acid is 0.010 M and the concentration of lactate is 0.087 M, the pH is 4.80.

$$pH = pK_a + \log \frac{[\text{lactate}]}{[\text{lactic acid}]}$$

$$pK_a = pH - \log \frac{[\text{lactate}]}{[\text{lactic acid}]}$$

$$= 4.80 - \log \frac{0.087}{0.010} = 4.80 - \log 8.7$$

$$= 4.80 - 0.94 = 3.86 \quad \textit{(answer)}$$

2. Calculate the pH of a mixture of 0.1 M acetic acid and 0.2 M sodium acetate. The pK_a of acetic acid is 4.76.

$$pH = pK_a + \log \frac{[\text{acetate}]}{[\text{acetic acid}]}$$

$$= 4.76 + \log \frac{0.2}{0.1} = 4.76 + 0.301$$

$$= 5.06 \quad \textit{(answer)}$$

3. Calculate the ratio of the concentrations of acetate and acetic acid required in a buffer system of pH 5.30.

$$pH = pK_a + \log \frac{[\text{acetate}]}{[\text{acetic acid}]}$$

$$\log \frac{[\text{acetate}]}{[\text{acetic acid}]} = pH - pK_a$$

$$= 5.30 - 4.76 = 0.54$$

$$\frac{[\text{acetate}]}{[\text{acetic acid}]} = \text{antilog } 0.54 = 3.47 \quad \textit{(answer)}$$

Weak Acids or Bases Buffer Cells and Tissues against pH Changes

The intracellular and extracellular fluids of multicellular organisms have a characteristic and nearly constant pH. The organism's first line of defense against changes in internal pH is provided by buffer systems. The cytoplasm of most cells contains high concentrations of proteins, which contain many amino acids with functional groups that are weak acids or weak bases. For example, the side chain of histidine (Fig. 4–18) has a pK_a of 6.0; proteins containing histidine residues therefore buffer effectively near neutral pH. Nucleotides such as ATP, as well as many low molecular weight metabolites, contain ionizable groups that can contribute buffering power to the cytoplasm. Some highly specialized organelles and extracellular compartments have high concentrations of compounds that contribute buffering capacity: organic acids buffer the vacuoles of plant cells; ammonia buffers urine.

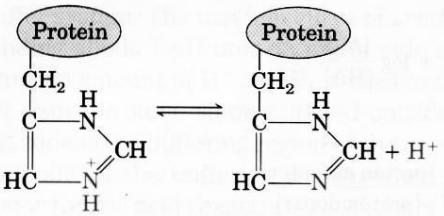


figure 4–18
The amino acid histidine, a component of proteins, is a weak acid. The pK_a of the protonated nitrogen of the side chain is 6.0.

Two especially important biological buffers are the phosphate and bicarbonate systems. The phosphate buffer system, which acts in the cytoplasm of all cells, consists of $H_2PO_4^-$ as proton donor and HPO_4^{2-} as proton acceptor:

$$H_2PO_4^- \rightleftharpoons H^+ + HPO_4^{2-}$$

The phosphate buffer system is maximally effective at a pH close to its pK_a of 6.86 (Figs 4–14, 4–16) and thus tends to resist pH changes in the range between about 5.9 and 7.9. It is therefore an effective buffer in biological fluids; in mammals, for example, extracellular fluids and most cytoplasmic compartments have a pH in the range of 6.9 to 7.4.

Blood plasma is buffered in part by the bicarbonate system, consisting of carbonic acid (H_2CO_3) as proton donor and bicarbonate (HCO_3^-) as proton acceptor:

$$H_2CO_3 \rightleftharpoons H^+ + HCO_3^-$$

$$K_1 = \frac{[H^+][HCO_3^-]}{[H_2CO_3]}$$

This buffer system is more complex than other conjugate acid-base pairs because one of its components, carbonic acid (H_2CO_3), is formed from dissolved (d) carbon dioxide and water, in a reversible reaction:

$$CO_2(d) + H_2O \rightleftharpoons H_2CO_3$$

$$K_2 = \frac{[H_2CO_3]}{[CO_2(d)][H_2O]}$$

Carbon dioxide is a gas under normal conditions, and the concentration of dissolved CO_2 is the result of equilibration with CO_2 of the gas phase:

$$CO_2(g) \rightleftharpoons CO_2(d)$$

$$K_3 = \frac{[CO_2(d)]}{[CO_2(g)]}$$

box 4–4 Blood, Lungs, and Buffer: The Bicarbonate Buffer System

In animals with lungs, the bicarbonate buffer system is an effective physiological buffer near pH 7.4 because the H_2CO_3 of blood plasma is in equilibrium with a large reserve capacity of CO_2(g) in the air space of the lungs. This buffer system involves three reversible equilibria between gaseous CO_2 in the lungs and bicarbonate (HCO_3^-) in the blood plasma (Fig. 1). When H^+ (from lactic acid produced in muscle tissue during vigorous exercise, for example) is added to blood as it passes through the tissues, reaction 1 proceeds toward a new equilibrium, in which the concentration of H_2CO_3 is increased. This increases the concentration of CO_2(d) in the blood plasma (reaction 2) and thus increases the pressure of CO_2(g) in the air space of the lungs (reaction 3); the extra CO_2 is exhaled.

Conversely, when the pH of blood plasma is raised (by NH_3 production during protein catabolism, for example), the opposite events occur: the H^+ concentration of blood plasma is lowered, causing more H_2CO_3 to dissociate into H^+ and HCO_3^-. This in turn causes more CO_2(g) from the lungs to dissolve in the blood plasma. The rate of breathing—that is, the rate of inhaling and exhaling CO_2—can quickly adjust these equilibria to keep the blood pH nearly constant.

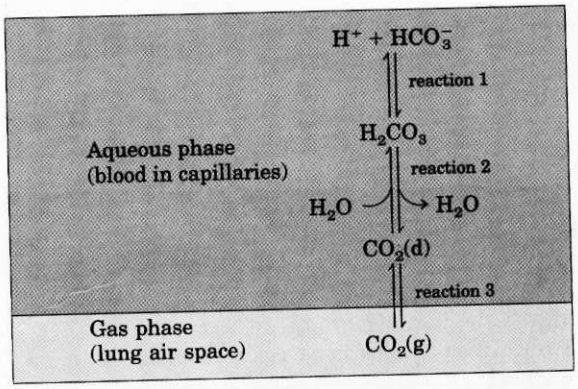


figure 1
The CO_2 in the air space of the lungs is in equilibrium with the bicarbonate buffer in the blood plasma passing through the lung capillaries. Because the concentration of dissolved CO_2 can be adjusted rapidly through changes in the rate of breathing, the bicarbonate buffer system of the blood is in near-equilibrium with a large potential reservoir of CO_2.

The pH of a bicarbonate buffer system depends on the concentration of H_2CO_3 and HCO_3^-, the proton donor and acceptor components. The concentration of H_2CO_3 in turn depends on the concentration of dissolved CO_2, which in turn depends on the concentration of CO_2 in the gas phase, called the **partial pressure** of CO_2. Thus the pH of a bicarbonate buffer exposed to a gas phase is ultimately determined by the concentration of HCO_3^- in the aqueous phase and the partial pressure of CO_2 in the gas phase (Box 4–4).

Human blood plasma normally has a pH close to 7.4. Should the pH-regulating mechanisms fail or be overwhelmed, as may happen in severe uncontrolled diabetes when an overproduction of metabolic acids causes acidosis, the pH of the blood can fall to 6.8 or below, leading to irreparable cell damage and death. In other diseases the pH may rise to lethal levels. Although many aspects of cell structure and function are influenced by pH, it is the catalytic activity of enzymes that is especially sensitive. Enzymes typically show maximal catalytic activity at a characteristic pH, called the **pH optimum** (Fig. 4–19). On either side of the optimum pH their catalytic activity often declines sharply. Thus, a small change in pH can make a large difference in the rate of some crucial enzyme-catalyzed reactions. Biological control of the pH of cells and body fluids is therefore of central importance in all aspects of metabolism and cellular activities.

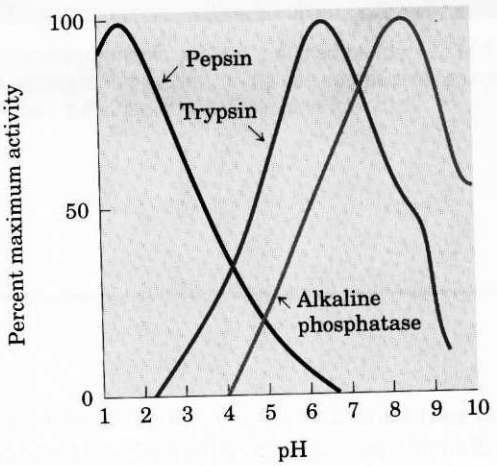


figure 4–19
The pH optima of some enzymes. Pepsin is a digestive enzyme secreted into gastric juice; trypsin, a digestive enzyme that acts in the small intestine; alkaline phosphatase of bone tissue, a hydrolytic enzyme thought to aid in bone mineralization.

Water as a Reactant

Water is not only the solvent in which the chemical reactions of living cells occur; it is very often a direct participant in those reactions. The formation of ATP from ADP and inorganic phosphate is an example of a **condensation reaction** (p. 69) in which the elements of water are eliminated (Fig. 4–20a). The reverse of this reaction—cleavage accompanied by the addition of the elements of water—is a **hydrolysis reaction.** Hydrolysis reactions are also responsible for the enzymatic depolymerization of proteins, carbohydrates, and nucleic acids. Hydrolysis reactions, catalyzed by enzymes called **hydrolases,** are almost invariably exergonic. The formation of cellular polymers from their subunits by simple reversal of hydrolysis would be endergonic and therefore does not occur. As we shall see, cells circumvent this thermodynamic obstacle by coupling endergonic condensation reactions to exergonic processes, such as breakage of the anhydride bond in ATP.

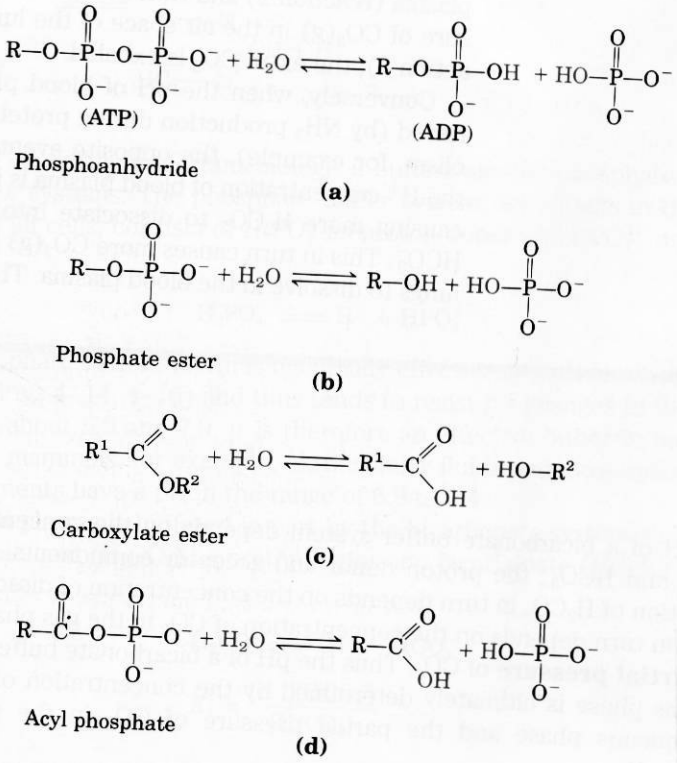


figure 4–20
Participation of water in biological reactions. (a) ATP is a phosphoanhydride formed by a condensation reaction (loss of the elements of water) between ADP and phosphate. R represents adenosine monophosphate (AMP). This condensation reaction requires energy. The hydrolysis of (addition of the elements of water to) ATP releases an equivalent amount of energy. Also shown are some other condensation and hydrolysis reactions common in biological systems **(b), (c), (d).**

You are (we hope!) consuming oxygen as you read. Water and carbon dioxide are the end products of the oxidation of fuels such as glucose. The overall reaction can be summarized as

$$\underset{\text{Glucose}}{C_6H_{12}O_6} + 6O_2 \rightleftharpoons 6CO_2 + 6H_2O$$

The "metabolic water" thus formed from solid food and stored fuels is actually enough to allow some animals in very dry habitats (gerbils, kangaroo rats, camels) to survive without drinking water for extended periods.

Green plants and algae use the energy of sunlight to split water in the process of photosynthesis:

$$2H_2O + 2A \xrightarrow{\text{light}} O_2 + 2AH_2$$

In this reaction, A is an electron-accepting species, which varies with the type of photosynthetic organism.

The Fitness of the Aqueous Environment for Living Organisms

Organisms have effectively adapted to their aqueous environment and have even evolved means of exploiting the unusual properties of water. The high specific heat of water (the heat energy required to raise the temperature of 1 g of water by 1 °C) is useful to cells and organisms because it allows water to act as a "heat buffer," permitting the temperature of an organism to remain relatively constant as the temperature of the air fluctuates and as heat is generated as a byproduct of metabolism. Furthermore, some vertebrates exploit the high heat of vaporization of water (Table 4–1) by using (thus losing) excess body heat to evaporate sweat. The high degree of internal cohesion of liquid water, due to hydrogen bonding, is exploited by plants as a means of transporting dissolved nutrients from the roots to the leaves during the process of transpiration. Even the density of ice, lower than that of liquid water, has important biological consequences in the life cycles of aquatic organisms. Ponds freeze from the top down, and the layer of ice at the top insulates the water below from frigid air, preventing the pond (and the organisms in it) from freezing solid. Most fundamental to all living organisms is the fact that many physical and biological properties of cell macromolecules, particularly the proteins and nucleic acids, derive from their interactions with water molecules of the surrounding medium. The influence of water on the course of biological evolution has been profound and determinative. If life forms have evolved elsewhere in the universe, it is unlikely that they resemble those of Earth unless their extraterrestrial origin is also a place in which plentiful liquid water is available.



Aqueous environments support a myriad of species. Soft corals, sponges, bryozoans, and algae compete for space on this reef substrate off the Philippine Islands.

summary

Water is the most abundant compound in living organisms. Its relatively high freezing point, boiling point, and heat of vaporization are the result of strong intermolecular attractions in the form of hydrogen bonding between adjacent water molecules. Liquid water has considerable short-range order and consists of short-lived hydrogen-bonded clusters. The polarity and hydrogen-bonding properties of water make it a potent solvent for many ionic compounds and other polar molecules. Nonpolar compounds, including the gases CO_2, O_2, and N_2, are poorly soluble in water.

Four types of weak interactions occur within and between biomolecules in an aqueous solvent: ionic, hydrophobic, and van der Waals interactions, and hydrogen bonds. Although weak individually, these interactions collectively cre-

ate a very strong stabilizing force for proteins, nucleic acids, and membranes. Weak (noncovalent) interactions are also at the heart of enzyme catalysis, antibody function, and receptor-ligand interactions.

When aqueous solutions of different concentration are separated by a semipermeable membrane, water crosses the membrane in the direction of lower water concentration. This tendency toward movement of water across a semipermeable membrane (osmosis) creates osmotic pressure. For cells in hypotonic solutions, inward osmotic movement of water across the plasma membrane causes swelling, producing turgor pressure or, if the cell is not protected by a rigid wall, osmotic lysis. The colligative properties of aqueous solutions (melting and boiling points, vapor pressure, and osmotic pressure) depend on the number of dissolved particles (ions, molecules), not on their molecular mass or chemical properties.

Water ionizes very slightly to form H^+ and OH^- ions. The rapid hopping of protons along strings of hydrogen-bonded water molecules gives the appearance of exceptionally fast diffusion of protons in water. In dilute aqueous solutions, the concentrations of H^+ and OH^- ions are inversely related by the expression $K_w = [H^+][OH^-] = 1 \times 10^{-14}\ M^2$ (at 25 °C). The hydrogen-ion concentration of biological systems is usually expressed in terms of pH, defined as $pH = -\log [H^+]$.

Acids are defined as proton donors and bases as proton acceptors. A conjugate acid-base pair consists of a proton donor (HA) and its corresponding proton acceptor (A^-). The tendency of an acid HA to donate protons is expressed by its dissociation constant ($K_a = [H^+][A^-]/[HA]$) or by the function pK_a, defined as $-\log K_a$, which can be determined from an experimental titration curve. The pH of a solution of a weak acid is quantitatively related to its pK_a and to the ratio of the concentrations of its proton-donor and proton-acceptor species by the Henderson-Hasselbalch equation.

A conjugate acid-base pair can act as a buffer and resist changes in pH; its capacity to do so is greatest at a pH equal to its pK_a. Many types of biomolecules have functional groups that contribute buffering capacity. Proteins, H_2CO_3/HCO_3^-, and $H_2PO_4^-/HPO_4^{2-}$ are important biological buffers. The catalytic activity of enzymes is strongly influenced by pH, and the environments in which enzymes function must be buffered against large pH changes.

Water is not only the solvent in which metabolic reactions occur; it participates directly in many biochemical processes, including hydrolysis and condensation reactions. The physical and chemical properties of water are central to biological structure and function. The evolution of life on Earth has been influenced greatly by both the solvent and reactant properties of water.

further reading

General

Denny, M.W. (1993) *Air and Water: The Biology and Physics of Life's Media,* Princeton University Press, Princeton, NJ.

A wonderful investigation of the biological relevance of the properties of water.

Eisenberg, D. & Kauzmann, W. (1969) *The Structure and Properties of Water,* Oxford University Press, New York.

An advanced treatment of the physical chemistry of water.

Franks, F. & Mathias, S.F. (eds) (1982) *Biophysics of Water,* John Wiley & Sons, Inc., New York.

A large collection of papers on the structure of pure water and of the cytoplasm.

Gerstein, M. & Levitt, M. (1998) Simulating water and the molecules of life. *Sci. Am.* **279** (November), 100–105.

A well-illustrated description of the use of computer simulation to study the biologically important association of water with proteins and nucleic acids.

Gronenborn, A. & Clore, M. (1997) Water in and around proteins. *The Biochemist* **19** (3), 18–21.

A brief discussion of protein-bound water as detected by crystallography and NMR.

Kornblatt, J. & Kornblatt, J. (1997) The role of water in recognition and catalysis by enzymes. *The Biochemist* **19** (3), 14–17.

A short, useful summary of the ways in which bound water influences the structure and activity of proteins.

Kuntz, I.D. & Zipp, A. (1977) Water in biological systems. *N. Engl. J. Med.* **297,** 262–266.

A brief review of the physical state of cytosolic water and its interactions with dissolved biomolecules.

Ladbury, J. (1996) Just add water! The effect of water on the specificity of protein-ligand binding sites and its potential application to drug design. *Chem. Biol.* **3,** 973–980.

Stillinger, F.H. (1980) Water revisited. *Science* **209,** 451–457.

A short review of the physical structure of water, including the importance of hydrogen bonding and the nature of hydrophobic interactions.

Westhof, E. (ed.) (1993) *Water and Biological Macromolecules,* CRC Press, Inc., Boca Raton, FL.

Fourteen chapters, each by a different author, cover (at an advanced level) the structure of water and its interactions with proteins, nucleic acids, polysaccharides, and lipids.

Wiggins, P.M. (1990) Role of water in some biological processes. *Microbiol. Rev.* **54,** 432–449.

A review of water in biology, including discussion of the physical structure of liquid water, its interaction with biomolecules, and the state of water in living cells.

Weak Interactions in Aqueous Systems

Fersht, A.R. (1987) The hydrogen bond in molecular recognition. *Trends Biochem. Sci.* **12,** 301–304.

A clear, brief, quantitative discussion of the contribution of hydrogen bonding to molecular recognition and enzyme catalysis.

Frieden, E. (1975) Non-covalent interactions: key to biological flexibility and specificity. *J. Chem. Educ.* **52,** 754–761.

Review of the four kinds of weak interactions that stabilize macromolecules and confer biological specificity, with clear examples.

Jeffrey, G.A. (1997) *An Introduction to Hydrogen Bonding,* Oxford University Press, New York.

A detailed and advanced discussion of the structure and properties of hydrogen bonds, including those in water and biomolecules.

Schwabe, J.W.R. (1997) The role of water in protein-DNA interactions. *Curr. Opin. Struct. Biol.* **7,** 126–134.

Examines the important role of water in both the specificity and affinity of protein-DNA interactions.

Tanford, C. (1978) The hydrophobic effect and the organization of living matter. *Science* **200,** 1012–1018.

A review of the chemical and energetic basis for hydrophobic interactions between biomolecules in aqueous solutions.

Weak Acids, Weak Bases, and Buffers: Problems for Practice

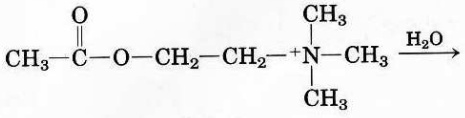
Segel, I.H. (1976) *Biochemical Calculations,* 2nd edn, John Wiley & Sons, Inc., New York.

problems

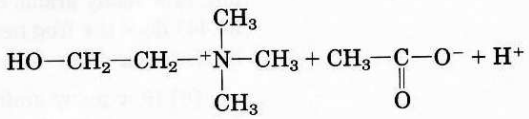
1. Simulated Vinegar One way to make vinegar (*not* the preferred way) is to prepare a solution of acetic acid, the sole acid component of vinegar, at the proper pH (see Fig. 4–13) and add appropriate flavoring agents. Acetic acid (M_r 60) is a liquid at 25 °C with a density of 1.049 g/mL. Calculate the volume that must be added to distilled water to make 1 L of simulated vinegar (see Fig. 4–14).

2. Acidity of Gastric HCl In a hospital laboratory, a 10.0 mL sample of gastric juice, obtained several hours after a meal, was titrated with 0.1 M NaOH to neutrality; 7.2 mL of NaOH was required. The patient's stomach contained no ingested food or drink, thus assume that no buffers were present. What was the pH of the gastric juice?

3. Measurement of Acetylcholine Levels by pH Changes The concentration of acetylcholine (a neurotransmitter) in a sample can be determined from the pH changes that accompany its hydrolysis. When the sample is incubated with the enzyme acetylcholinesterase, acetylcholine is quantitatively converted into choline and acetic acid, which dissociates to yield acetate and a hydrogen ion:



Acetylcholine



Choline Acetate

In a typical analysis, 15 mL of an aqueous solution containing an unknown amount of acetylcholine had a pH of 7.65. When incubated with acetylcholinesterase, the pH of the solution decreased to 6.87. Assuming that there was no buffer in the assay mixture, determine the number of moles of acetylcholine in the 15 mL sample.

4. Osmotic Balance in a Marine Frog The crab-eating frog of Southeast Asia, *Rana cancrivora*, is born and matures in fresh water but searches for its food in coastal mangrove swamps (80% to full-strength seawater). Consequently, when the frog moves from its freshwater home to seawater it experiences a large change in the osmolarity of its environment (from hypotonic to hypertonic).

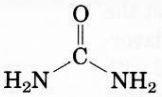
(a) Eighty percent seawater contains 460 mM NaCl, 10 mM KCl, 10 mM $CaCl_2$ and 50 mM $MgCl_2$. What are the concentrations of the various ionic species in this seawater? Assuming that these salts account for nearly all the solutes in seawater, what is the osmolarity of the seawater?

(b) The chart below lists the cytoplasmic concentrations of ions in *Rana cancrivora*. Ignoring dissolved proteins, amino acids, nucleic acids, and other small metabolites, what is the osmolarity of the frog's cells based solely on the ionic concentrations given below?

	Na^+ (mM)	K^+ (mM)	Cl^- (mM)	Ca^{2+} (mM)	Mg^{2+} (mM)
Rana cancrivora	122	10	100	2	1

(c) Like all frogs, the crab-eating frog can exchange gases through its permeable skin, allowing it to stay underwater for long periods of time without breathing. How does the high permeability of frog skin affect the frog's cells when it moves from fresh water to seawater?

(d) The crab-eating frog uses two mechanisms to maintain its cells in osmotic balance with its environment. First, it allows the Na^+ and Cl^- concentrations in its cells to slowly increase as the ions diffuse down their concentration gradients. Second, like many elasmobranchs (sharks), it retains the waste product urea in its cells. The addition of both NaCl and urea increases the osmolarity of the cytosol to a value that is nearly equal to that of the surrounding environment.



Urea (CH_4N_2O)

Assuming the volume of water in a typical frog is 100 mL, how many grams of NaCl (formula weight (FW) 58.44) does the frog need to take up in order to make its tissues isotonic with seawater?

(e) How many grams of urea (FW 60) must it retain to accomplish the same thing?

5. Properties of a Buffer The amino acid glycine is often used as the main ingredient of a buffer in biochemical experiments. The amino group of glycine, which has a pK_a of 9.6, can exist either in the protonated form ($-NH_3^+$) or as the free base ($-NH_2$) because of the reversible equilibrium

$$R-NH_3^+ \rightleftharpoons R-NH_2 + H^+$$

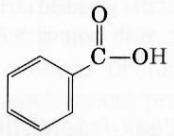
(a) In what pH range can glycine be used as an effective buffer due to its amino group?

(b) In a 0.1 M solution of glycine at pH 9.0, what fraction of glycine has its amino group in the $-NH_3^+$ form?

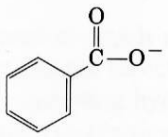
(c) How much 5 M KOH must be added to 1.0 L of 0.1 M glycine at pH 9.0 to bring its pH to exactly 10.0?

(d) When 99% of the glycine is in its $-NH_3^+$ form, what is the numerical relation between the pH of the solution and the pK_a of the amino group?

6. The Effect of pH on Solubility The strongly polar, hydrogen-bonding properties of water make it an excellent solvent for ionic (charged) species. By contrast, nonionized, nonpolar organic molecules, such as benzene, are relatively insoluble in water. In principle, the aqueous solubility of any organic acid or base can be increased by conversion of the molecules to charged species. For example, the solubility of benzoic acid in water is low. The addition of sodium bicarbonate to a mixture of water and benzoic acid raises the pH and deprotonates the benzoic acid to form benzoate ion, which is quite soluble in water.

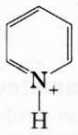


Benzoic acid
$pK_a \approx 5$



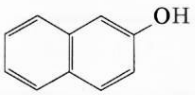
Benzoate ion

Are the following compounds more soluble in an aqueous solution of 0.1 M NaOH or 0.1 M HCl? (The dissociable protons are shown in red.)



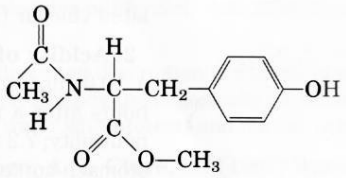
Pyridine ion
$pK_a \approx 5$

(a)



β-Naphthol
$pK_a \approx 10$

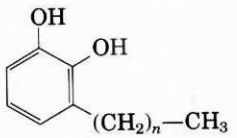
(b)



N-Acetyltyrosine methyl ester
$pK_a \approx 10$

(c)

7. Treatment of Poison Ivy Rash The components of poison ivy and poison oak that produce the characteristic itchy rash are catechols substituted with long-chain alkyl groups.



$pK_a \approx 8$

If you were exposed to poison ivy, which of the treatments below would you apply to the affected area? Justify your choice.

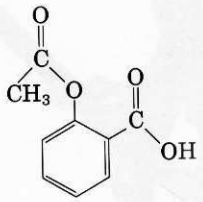
(a) Wash the area with cold water.

(b) Wash the area with dilute vinegar or lemon juice.

(c) Wash the area with soap and water.

(d) Wash the area with soap, water, and baking soda (sodium bicarbonate).

8. pH and Drug Absorption Aspirin is a weak acid with a pK_a of 3.5.



It is absorbed into the blood through the cells lining the stomach and the small intestine. Absorption requires passage through the plasma membrane, the rate of which is determined by the polarity of the molecule: charged and highly polar molecules pass slowly, whereas neutral hydrophobic ones pass rapidly. The pH of the stomach contents is about 1.5, and the pH of the contents of the small intestine is about 6. Is more aspirin absorbed into the bloodstream from the stomach or from the small intestine? Clearly justify your choice.

9. Preparation of Standard Buffer for Calibration of a pH Meter The glass electrode used in commercial pH meters gives an electrical response proportional to the concentration of hydrogen ion. To convert these responses into pH, glass electrodes must be calibrated against standard solutions of known H^+ concentration. Determine the weight in grams of sodium dihydrogen phosphate ($NaH_2PO_4 \cdot H_2O$; formula weight (FW) 138.01) and disodium hydrogen phosphate (Na_2HPO_4; FW 141.98) needed to prepare 1 L of a standard buffer at pH 7.00 with a total phosphate concentration of 0.100 M (see Fig. 4–14).

10. Control of Blood pH by Respiration Rate

(a) The partial pressure of CO_2 in the lungs can be varied rapidly by the rate and depth of breathing. For example, a common remedy to alleviate hiccups is to increase the concentration of CO_2 in the lungs. This can be achieved by holding one's breath, by very slow and shallow breathing (hypoventilation), or by breathing in and out of a paper bag. Under such conditions, the partial pressure of CO_2 in the air space of the lungs rises above normal. Qualitatively explain the effect of these procedures on the blood pH.

(b) A common practice of competitive short-distance runners is to breathe rapidly and deeply (hyperventilation) for about half a minute to remove CO_2 from their lungs just before running in, say, a 100 m dash. Their blood pH may rise to 7.60. Explain why the blood pH increases.

(c) During a short-distance run the muscles produce a large amount of lactic acid ($CH_3CH(OH)COOH$, $K_a = 1.38 \times 10^{-4}$) from their glucose stores. In view of this fact, why might hyperventilation before a dash be useful?

chapter

5

Amino Acids, Peptides, and Proteins

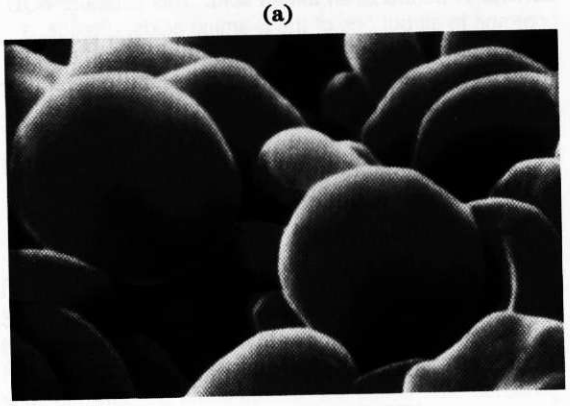
Proteins are the most abundant biological macromolecules, occurring in all cells and all parts of cells. Proteins also occur in great variety; thousands of different kinds, ranging in size from relatively small peptides to huge polymers with molecular weights in the millions, may be found in a single cell. Moreover, proteins exhibit enormous diversity of biological function and are the most important final products of the information pathways discussed in Part IV of this book. Proteins are the molecular instruments through which genetic information is expressed. It is appropriate to begin our study of biological macromolecules with the proteins, whose name derives from the Greek *prōtos*, meaning "first" or "foremost."

Relatively simple monomeric subunits provide the key to the structure of the thousands of different proteins. All proteins, whether from the most ancient lines of bacteria or from the most complex forms of life, are constructed from the same ubiquitous set of 20 amino acids, covalently linked in characteristic linear sequences. Because each of these amino acids has a side chain with distinctive chemical properties, this group of 20 precursor molecules may be regarded as the alphabet in which the language of protein structure is written.

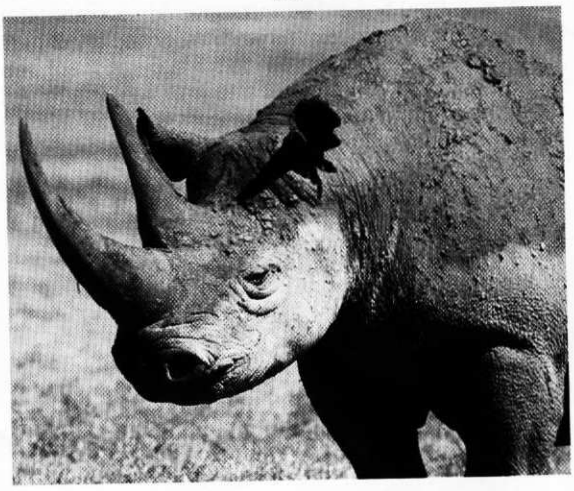
What is most remarkable is that cells can produce proteins with strikingly different properties and activities by joining the same 20 amino acids in many different combinations and sequences. From these building blocks different organisms can make such widely diverse products as enzymes, hormones, antibodies, transporters, muscle, the lens protein of the eye, feathers, spider webs, rhinoceros horn, milk proteins, antibiotics, mushroom poisons, and a myriad of other substances having distinct biological activities (Fig. 5–1). Among these protein products, the enzymes are the most varied and specialized. Virtually all cellular reactions are catalyzed by enzymes.



(a)



(b)



(c)

figure 5–1

Some functions of proteins. (a) The light produced by fireflies is the result of a reaction involving the protein luciferin and ATP, catalyzed by the enzyme luciferase (see Box 14–3). **(b)** Erythrocytes contain large amounts of the oxygen-transporting protein hemoglobin. **(c)** The protein keratin, formed by all vertebrates, is the chief structural component of hair, scales, horn, wool, nails, and feathers. The black rhinoceros is nearing extinction in the wild because of the myth prevalent in some parts of the world that a powder derived from its horn has aphrodisiac properties. In reality, the chemical properties of powdered rhinoceros horn are no different from those of powdered bovine hooves or human fingernails.

Protein structure and function are the topics of this and the next three chapters. We begin with a description of the fundamental chemical properties of amino acids, peptides, and proteins.

Amino Acids

Proteins are dehydration polymers of amino acids, with each **amino acid residue** joined to its neighbor by a specific type of covalent bond. (The term "residue" reflects the loss of the elements of water when one amino acid is joined to another.) Proteins can be broken down (hydrolyzed) to their constituent amino acids by a variety of methods, and the earliest studies of proteins naturally focused on the free amino acids derived from them. The first to be discovered was asparagine, in 1806. The last of the 20 to be found, threonine, was not identified until 1938. All the amino acids have trivial or common names, in some cases derived from the source from which they were first isolated. Asparagine was first found in asparagus, and glutamate in wheat gluten; tyrosine was first isolated from cheese (its name is derived from the Greek *tyros,* "cheese"); and glycine (Greek *glykos,* "sweet") was so named because of its sweet taste.

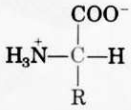
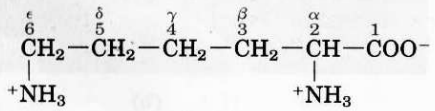


figure 5–2
General structure of an amino acid. This structure is common to all but one of the α-amino acids. (Proline, a cyclic amino acid, is the exception.) The R group or side chain (red) attached to the α carbon (blue) is different in each amino acid.



Lysine

Amino Acids Share Common Structural Features

All 20 standard amino acids found in proteins are α-amino acids. They have a carboxyl group and an amino group bonded to the same carbon atom (the α carbon) (Fig. 5–2). They differ from each other in their side chains, or **R groups,** which vary in structure, size, and electric charge, and which influence the solubility of the amino acids in water. The 20 amino acids of proteins are often referred to as the standard amino acids, to distinguish them from less common amino acids that are residues modified after a protein has been synthesized, and from the many other kinds of amino acids present in living organisms but not in proteins. The standard amino acids have been assigned three-letter abbreviations and one-letter symbols (Table 5–1, p. 118), which are used as shorthand to indicate the composition and sequence of amino acids polymerized in proteins.

In a practice that can be confusing, two conventions are used to identify the carbons within an amino acid. The additional carbons in an R group are commonly designated β, γ, δ, ϵ, and so forth, proceeding out from the α carbon. For most other organic molecules, carbon atoms are simply numbered from one end, giving highest priority to carbons with substitutions containing atoms with the highest atomic numbers. Within this latter convention, the carboxyl group of an amino acid would be C-1 and the α carbon would be C-2. In some cases, such as amino acids with heterocyclic R groups, the Greek lettering system is ambiguous and the numbering convention is therefore used.

For all the standard amino acids except glycine, the α carbon is bonded to four different groups: a carboxyl group, an amino group, an R group, and a hydrogen atom (Fig. 5–2; in glycine, the R group is another hydrogen atom). The α-carbon atom is thus a **chiral center** (see Fig. 3–9). Because of the tetrahedral arrangement of the bonding orbitals around the α-carbon atom, the four different groups can occupy two different spatial arrangements that are nonsuperimposable mirror images of each other (Fig. 5–3). These two forms represent a class of stereoisomers called **enantiomers** (see Fig. 3–10). All molecules with a chiral center are also **optically active**—that is, they rotate plane-polarized light (see Box 3–1).

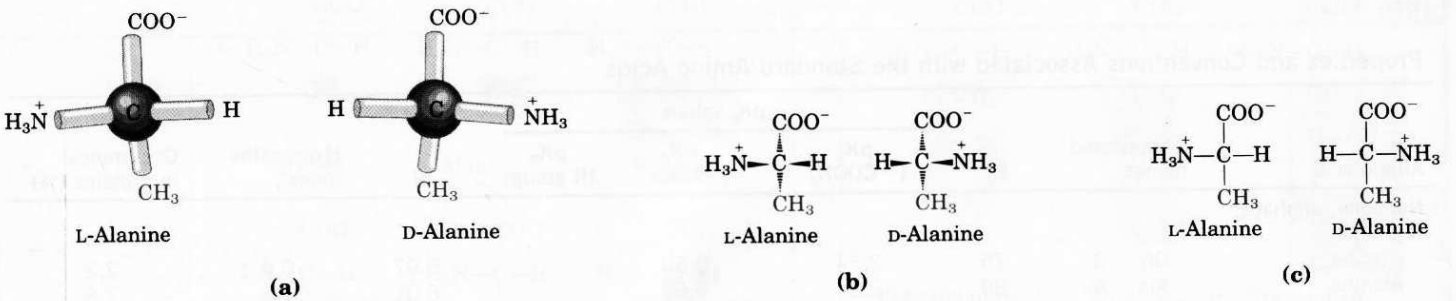


figure 5–3
Stereoisomerism in α-amino acids. (a) The two stereoisomers of alanine, L- and D-alanine, are nonsuperimposable mirror images of each other (enantiomers). **(b, c)** Two different conventions for showing the configurations in space of stereoisomers. In perspective formulas **(b)** the solid wedge-shaped bonds project out of the plane of the paper, the dashed bonds behind it. In projection formulas **(c)** the horizontal bonds are assumed to project out of the plane of the paper, the vertical bonds behind. However, projection formulas are often used casually and are not always intended to portray a specific stereochemical configuration.

Special nomenclature has been developed to specify the **absolute configuration** of the four substituents of asymmetric carbon atoms. The absolute configurations of simple sugars and amino acids are specified by the **D, L system** (Fig. 5–4), based on the absolute configuration of the three-carbon sugar glyceraldehyde, a convention proposed by Emil Fischer in 1891. (Fischer knew what groups surrounded the asymmetric carbon of glyceraldehyde but had to guess at their absolute configuration; his guess was later confirmed by x-ray diffraction analysis.) For all chiral compounds, stereoisomers having a configuration related to that of L-glyceraldehyde are designated L, and stereoisomers related to D-glyceraldehyde are designated D. The functional groups of L-alanine are related to those of L-glyceraldehyde by simple chemical conversions. Thus the carboxyl group of L-alanine occupies the same position about the chiral carbon as does the aldehyde group of L-glyceraldehyde, because an aldehyde is readily converted (oxidized) to a carboxyl group. Historically, the similar *l* and *d* designations were used for levorotatory (rotating light to the left) and dextrorotatory (rotating light to the right) (see Box 3–1). However, not all L-amino acids are levorotatory, and the convention shown in Figure 5–4 was needed to avoid potential ambiguities about absolute configuration. By Fischer's convention, L and D refer *only* to the absolute configuration of the four substituents around the chiral carbon.

Another system of specifying configuration around a chiral center is the **RS system** (explained in Chapter 3), which is used in the systematic nomenclature of organic chemistry and describes more precisely the configuration of molecules with more than one chiral center.

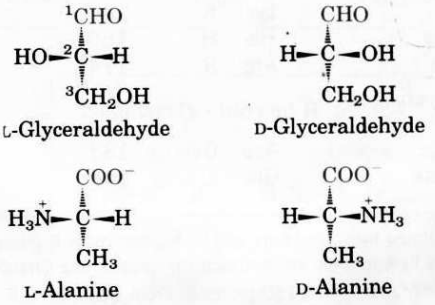


figure 5–4
Steric relationship of the stereoisomers of alanine to the absolute configuration of L- and D-glyceraldehyde. In these perspective formulas, the carbons are lined up vertically, with the chiral atom in the center. The carbons in these molecules are numbered beginning with the aldehyde or carboxyl carbons on the end (red), 1 to 3 from top to bottom as shown. When presented in this way, the R group of the amino acid (in this case the methyl group of alanine) is always below the α carbon. L-Amino acids are those with the α-amino group on the left, and D-amino acids have the α-amino group on the right.

The Amino Acid Residues in Proteins Are L Stereoisomers

Nearly all biological compounds with a chiral center occur naturally in only one stereoisomeric form, either D or L. The amino acid residues in protein molecules are exclusively L stereoisomers. D-Amino acid residues have been found only in a few, generally small peptides, including some peptides of bacterial cell walls and certain peptide antibiotics.

It is remarkable that all amino acid residues in proteins are L stereoisomers. As we noted in Chapter 3, when chiral compounds are formed by ordinary chemical reactions, the result is a racemic mixture of D and L isomers, which are difficult for a chemist to distinguish and separately isolate. But to a living system, D and L isomers are as different as the right hand and the left. The formation of stable, repeating substructures in proteins (Chapter 6) generally requires that their constituent amino acids be of one stereochemical series. Cells are able to specifically synthesize the L isomers of amino acids because the active sites of enzymes are asymmetric, causing the reactions they catalyze to be stereospecific.

table 5–1

Properties and Conventions Associated with the Standard Amino Acids

Amino acid	Abbreviated names		M_r	pK_a values: pK_1 (—COOH)	pK_2 ($—NH_3^+$)	pK_R (R group)	pI	Hydropathy index*	Occurrence in proteins (%)†
Nonpolar, aliphatic R groups									
Glycine	Gly	G	75	2.34	9.60		5.97	−0.4	7.2
Alanine	Ala	A	89	2.34	9.69		6.01	1.8	7.8
Valine	Val	V	117	2.32	9.62		5.97	4.2	6.6
Leucine	Leu	L	131	2.36	9.60		5.98	3.8	9.1
Isoleucine	Ile	I	131	2.36	9.68		6.02	4.5	5.3
Methionine	Met	M	149	2.28	9.21		5.74	1.9	2.3
Aromatic R groups									
Phenylalanine	Phe	F	165	1.83	9.13		5.48	2.8	3.9
Tyrosine	Tyr	Y	181	2.20	9.11	10.07	5.66	−1.3	3.2
Tryptophan	Trp	W	204	2.38	9.39		5.89	−0.9	1.4
Polar, uncharged R groups									
Serine	Ser	S	105	2.21	9.15		5.68	−0.8	6.8
Proline	Pro	P	115	1.99	10.96		6.48	1.6	5.2
Threonine	Thr	T	119	2.11	9.62		5.87	−0.7	5.9
Cysteine	Cys	C	121	1.96	10.28	8.18	5.07	2.5	1.9
Asparagine	Asn	N	132	2.02	8.80		5.41	−3.5	4.3
Glutamine	Gln	Q	146	2.17	9.13		5.65	−3.5	4.2
Positively charged R groups									
Lysine	Lys	K	146	2.18	8.95	10.53	9.74	−3.9	5.9
Histidine	His	H	155	1.82	9.17	6.00	7.59	−3.2	2.3
Arginine	Arg	R	174	2.17	9.04	12.48	10.76	−4.5	5.1
Negatively charged R groups									
Aspartate	Asp	D	133	1.88	9.60	3.65	2.77	−3.5	5.3
Glutamate	Glu	E	147	2.19	9.67	4.25	3.22	−3.5	6.3

*A scale combining hydrophobicity and hydrophilicity of R groups; it can be used to measure the tendency of an amino acid to seek an aqueous environment (− values) or a hydrophobic environment (+ values). See Chapter 12. From Kyte, J. & Doolittle, R.F. (1982) *J. Mol. Biol.* **157**, 105–132.

†Average occurrence in over 1150 proteins. From Doolittle, R.F. (1989) Redundancies in protein sequences. In *Prediction of Protein Structure and the Principles of Protein Conformation* (Fasman, G.D., ed) Plenum Press, NY, pp. 599–623.

Amino Acids Can Be Classified by R Group

Knowledge of the chemical properties of the standard amino acids is central to an understanding of biochemistry. The topic can be simplified by grouping the amino acids into five main classes based on the properties of their R groups (Table 5–1), in particular, their **polarity** or tendency to interact with water at biological pH (near pH 7.0). The polarity of the R groups varies widely, from totally nonpolar or hydrophobic (water-insoluble) to highly polar or hydrophilic (water-soluble).

The structures of the 20 standard amino acids are shown in Figure 5–5, and some of their properties are listed in Table 5–1. Within each class there are gradations of polarity, size, and shape of the R groups.

Nonpolar, Aliphatic R Groups The R groups in this class of amino acids are nonpolar and hydrophobic. The side chains of **alanine, valine, leucine,** and **isoleucine** tend to cluster together within proteins, stabilizing protein structure by means of hydrophobic interactions. **Glycine** has the simplest

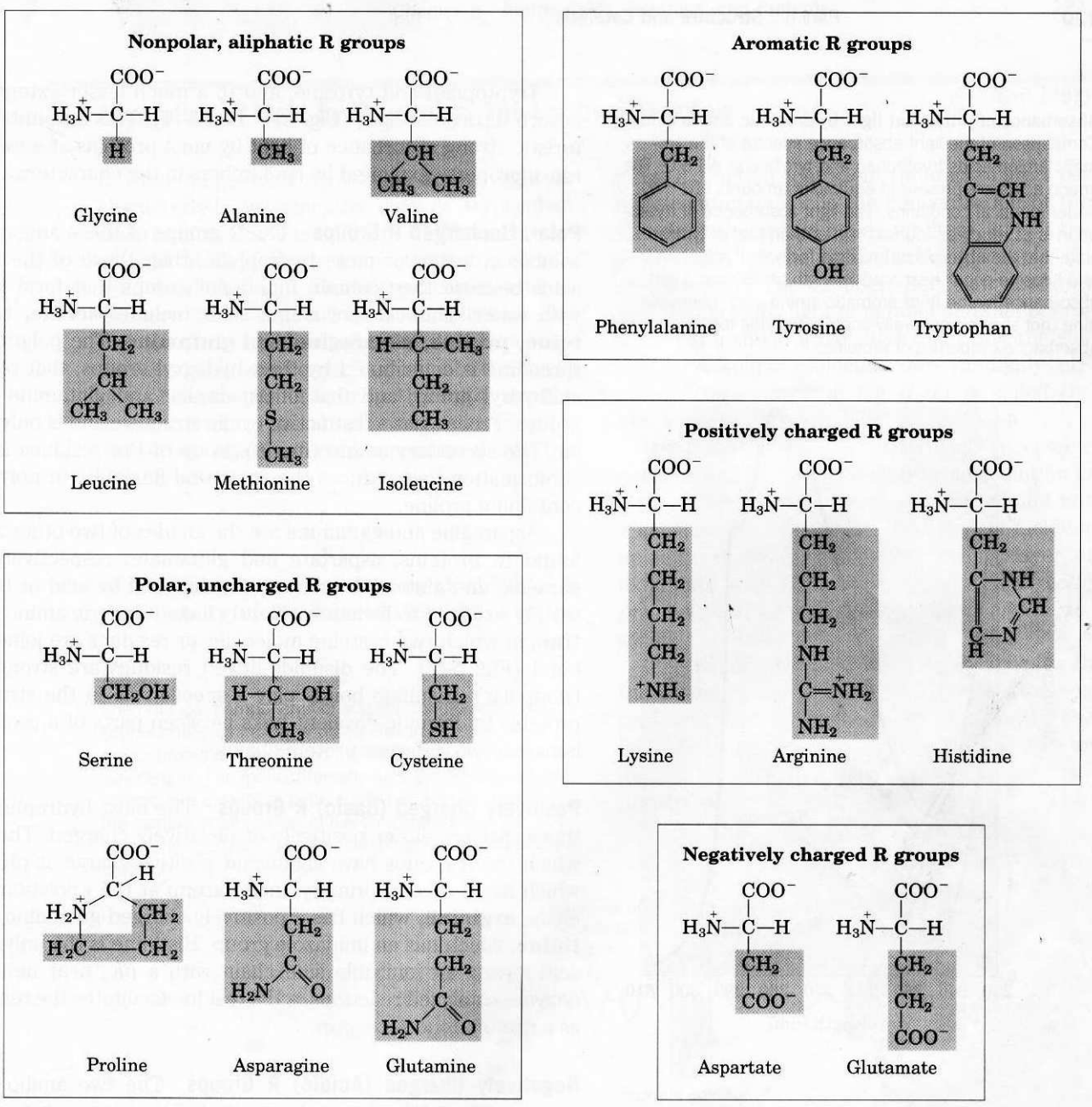
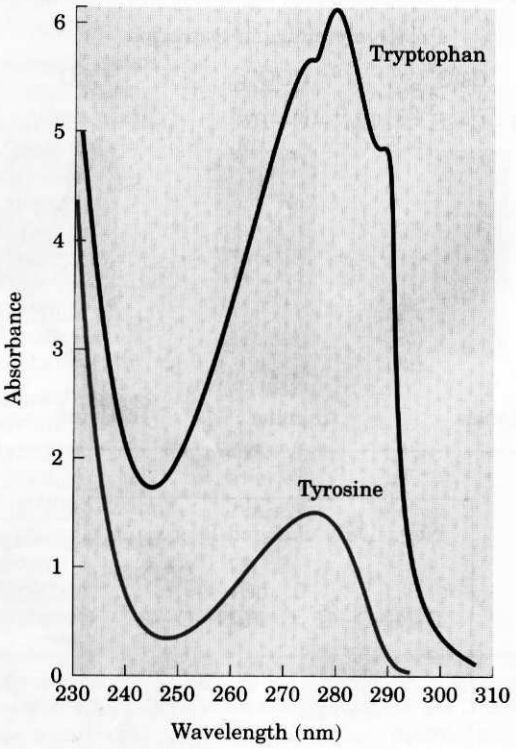


figure 5–5
The 20 standard amino acids of proteins. The structural formulas show the state of ionization that would predominate at pH 7.0. The unshaded portions are those common to all the amino acids; the portions shaded in red are the R groups. Although the R group of histidine is shown uncharged, its pK_a (see Table 5–1) is such that a small but significant fraction of these groups are positively charged at pH 7.0.

structure. Although it is formally nonpolar, its very small side chain makes no real contribution to hydrophobic interactions. **Methionine,** one of the two sulfur-containing amino acids, has a nonpolar thioether group in its side chain.

Aromatic R Groups **Phenylalanine, tyrosine,** and **tryptophan,** with their aromatic side chains, are relatively nonpolar (hydrophobic). All can participate in hydrophobic interactions. The hydroxyl group of tyrosine can form hydrogen bonds, and it is an important functional group in some enzymes. Tyrosine and tryptophan are significantly more polar than phenylalanine because of the tyrosine hydroxyl group and the nitrogen of the tryptophan indole ring.

figure 5–6
Absorbance of ultraviolet light by aromatic amino acids. Comparison of the light absorbance spectra of the aromatic amino acids tryptophan and tyrosine at pH 6.0. The amino acids are present in equimolar amounts (10^{-3} M) under identical conditions. The light absorbance of tryptophan is as much as fourfold higher than that of tyrosine. Note that the absorbance maxima for both tryptophan and tyrosine occur near a wavelength of 280 nm. Light absorbance by the third aromatic amino acid, phenylalanine (not shown), generally contributes little to the absorbance properties of proteins.



Tryptophan and tyrosine, and to a much lesser extent phenylalanine, absorb ultraviolet light (Fig. 5–6; Box 5–1). This accounts for the characteristic strong absorbance of light by most proteins at a wavelength of 280 nm, a property exploited by researchers in the characterization of proteins.

Polar, Uncharged R Groups The R groups of these amino acids are more soluble in water, or more hydrophilic, than those of the nonpolar amino acids because they contain functional groups that form hydrogen bonds with water. This class of amino acids includes **serine, threonine, cysteine, proline, asparagine,** and **glutamine.** The polarity of serine and threonine is contributed by their hydroxyl groups; that of cysteine by its sulfhydryl group; and that of asparagine and glutamine by their amide groups. Proline has a distinctive cyclic structure and is only moderately polar. The secondary amino (imino) group of Pro residues is held in a rigid conformation that reduces the structural flexibility of polypeptide regions containing proline.

Asparagine and glutamine are the amides of two other amino acids also found in proteins, aspartate and glutamate, respectively, to which asparagine and glutamine are easily hydrolyzed by acid or base. Cysteine is readily oxidized to form a covalently linked dimeric amino acid called **cystine,** in which two cysteine molecules or residues are joined by a disulfide bond (Fig. 5–7). The disulfide-linked residues are strongly hydrophobic (nonpolar). Disulfide bonds play a special role in the structures of many proteins by forming covalent links between parts of a protein molecule or between two different protein chains.

Positively Charged (Basic) R Groups The most hydrophilic R groups are those that are either positively or negatively charged. The amino acids in which the R groups have significant positive charge at pH 7.0 are **lysine,** which has a second primary amino group at the ϵ position on its aliphatic chain; **arginine,** which has a positively charged guanidino group; and **histidine,** which has an imidazole group. Histidine is the only standard amino acid having an ionizable side chain with a pK_a near neutrality. In many enzyme-catalyzed reactions, a His residue facilitates the reaction by serving as a proton donor/acceptor.

Negatively Charged (Acidic) R Groups The two amino acids having R groups with a net negative charge at pH 7.0 are **aspartate** and **glutamate,** each of which has a second carboxyl group.

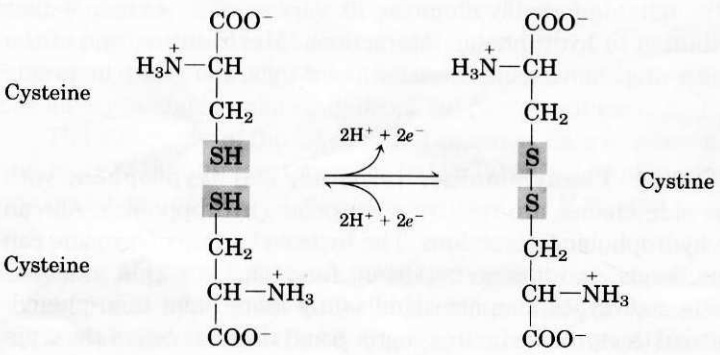


figure 5–7
Reversible formation of a disulfide bond by the oxidation of two molecules of cysteine. Disulfide bonds between Cys residues stabilize the structures of many proteins.

box 5–1 Absorption of Light by Molecules: The Lambert-Beer Law

A wide range of biomolecules absorb light at characteristic wavelengths, just as tryptophan absorbs light at 280 nm (Fig. 5–6). Measurement of light absorption by a spectrophotometer is used to detect and identify molecules and to measure their concentration in solution. The fraction of the incident light absorbed by a solution at a given wavelength is related to the thickness of the absorbing layer (path length) and the concentration of the absorbing species (Fig. 1). These two relationships are combined into the Lambert-Beer law,

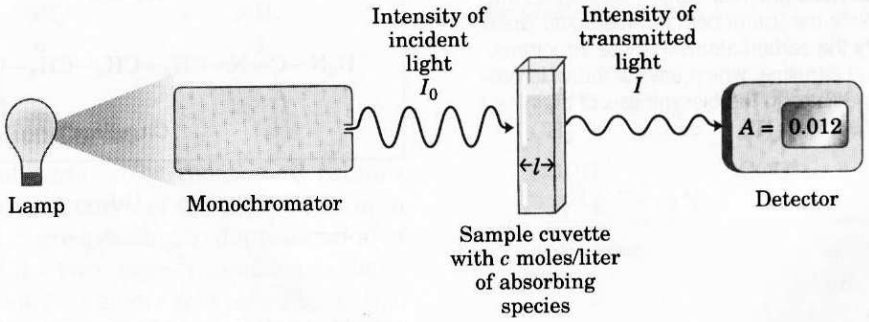
$$\log \frac{I_0}{I} = \epsilon cl$$

where I_0 is the intensity of the incident light, I is the intensity of the transmitted light, ϵ is the molar extinction coefficient (in units of liters per mole-centimeter), c is the concentration of the absorbing species (in moles per liter), and l is the path length of the light-absorbing sample (in centimeters). The Lambert-Beer law assumes that the incident light is parallel and monochromatic (of a single wavelength) and that the solvent and solute molecules are randomly oriented. The expression $\log (I_0/I)$ is called the **absorbance,** designated A.

It is important to note that each successive millimeter of path length of absorbing solution in a 1.0 cm cell absorbs not a constant amount but a constant fraction of the light that is incident upon it. However, with an absorbing layer of fixed path length, *the absorbance* A *is directly proportional to the concentration of the absorbing solute.*

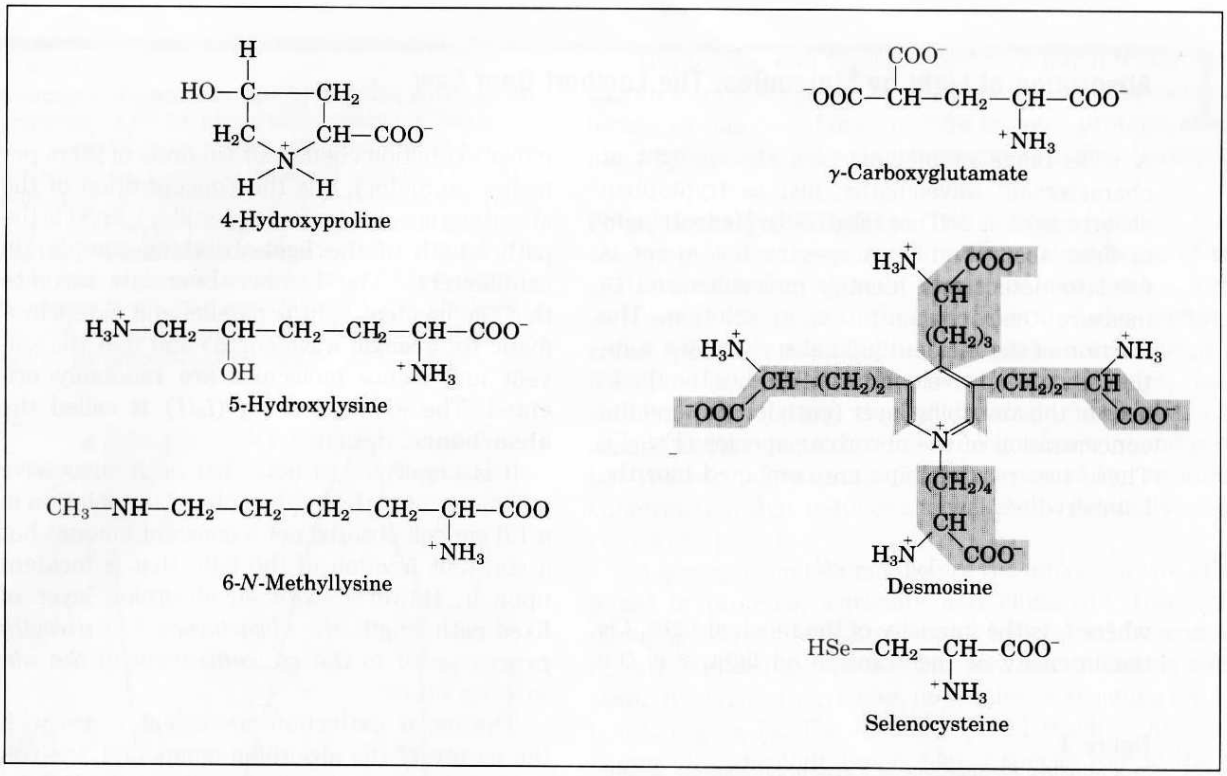
The molar extinction coefficient varies with the nature of the absorbing compound, the solvent, and the wavelength, and also with pH if the light-absorbing species is in equilibrium with an ionization state that has different absorbance properties.

figure 1
The principal components of a spectrophotometer. A light source emits light along a broad spectrum, then the monochromator selects and transmits light of a particular wavelength. The monochromatic light passes through the sample in a cuvette of path length l and is absorbed by the sample in proportion to the concentration of the absorbing species. The transmitted light is measured by a detector.

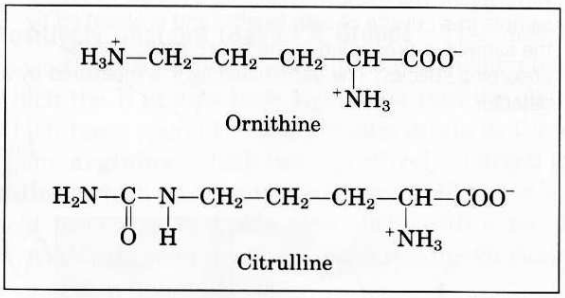


Nonstandard Amino Acids Also Have Important Functions

In addition to the 20 standard amino acids, proteins may contain nonstandard residues created by modification of standard residues already incorporated into a polypeptide (Fig. 5–8a, p. 122). Among the nonstandard amino acids are **4-hydroxyproline,** a derivative of proline, and **5-hydroxylysine,** derived from lysine. The former is found in plant cell wall proteins, and both are found in collagen, a fibrous protein of connective tissues. **6-*N*-Methyllysine** is a constituent of myosin, a contractile protein of



(a)



(b)

figure 5–8

Nonstandard amino acids. (a) Some nonstandard amino acids found in proteins. All are derived from standard amino acids. Extra functional groups added by modification reactions are shown in red. Desmosine is formed from four Lys residues (the four carbon backbones are shaded gray). Note the use of both numbers and Greek letters to identify the carbon atoms in these structures. **(b)** Ornithine and citrulline, which are not found in proteins, are intermediates in the biosynthesis of arginine and in the urea cycle.

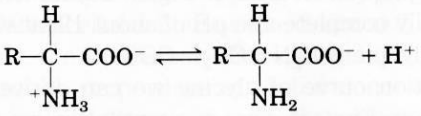
muscle. Another important nonstandard amino acid is **γ-carboxyglutamate,** found in the blood-clotting protein prothrombin and in certain other proteins that bind Ca^{2+} as part of their biological function. More complicated is **desmosine,** a derivative of four Lys residues, which is found in the fibrous protein elastin.

Selenocysteine is a special case. This rare amino acid residue is introduced during protein synthesis rather than created through a postsynthetic modification. It contains selenium rather than the sulfur of cysteine. Actually derived from serine, selenocysteine is a constituent of just a few known proteins.

Some 300 additional amino acids have been found in cells. They have a variety of functions but are not constituents of proteins. **Ornithine** and **citrulline** (Fig. 5–8b) deserve special note because they are key intermediates in the biosynthesis of arginine (Chapter 22) and in the urea cycle (Chapter 18).

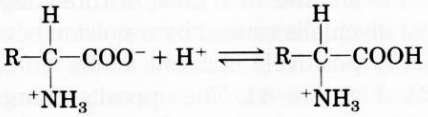
Amino Acids Can Act as Acids and Bases

When an amino acid is dissolved in water, it exists in solution as the dipolar ion, or **zwitterion** (German for "hybrid ion"), shown in Figure 5–9. A zwitterion can act as either an acid (proton donor):



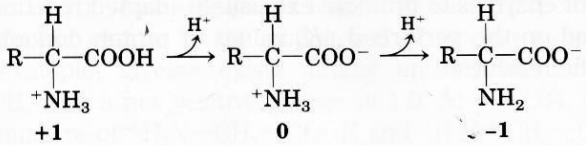
Zwitterion

or a base (proton acceptor):



Zwitterion

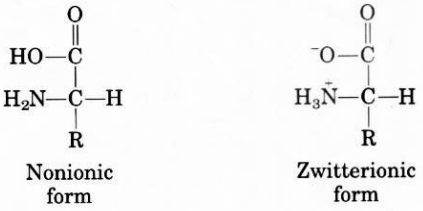
Substances having this dual nature are **amphoteric** and are often called **ampholytes** (from "amphoteric electrolytes"). A simple monoamino monocarboxylic α-amino acid, such as alanine, is a diprotic acid when fully protonated—it has two groups, the —COOH group and the $-NH_3^+$ group, that can yield protons:



Net charge: +1 0 −1

Amino Acids Have Characteristic Titration Curves

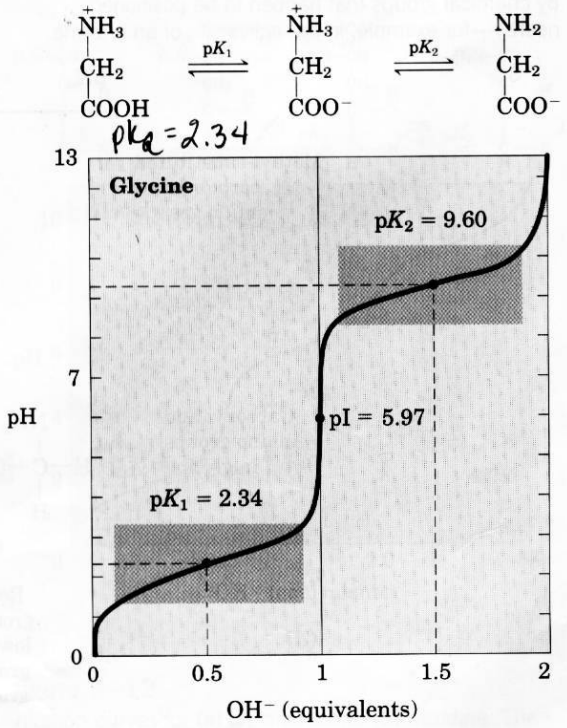
Acid-base titration involves the gradual addition or removal of protons (Chapter 4). Figure 5–10 shows the titration curve of the diprotic form of glycine. The plot has two distinct stages, corresponding to deprotonation of two different groups on glycine. Each of the two stages resembles in shape the titration curve of a monoprotic acid, such as acetic acid (see Fig. 4–15), and can be analyzed in the same way. At very low pH, the predominant ionic species of glycine is $^+H_3N—CH_2—COOH$, the fully protonated form. At the midpoint in the first stage of the titration, in which the —COOH group of glycine loses its proton, equimolar concentrations of the proton-donor ($^+H_3N—CH_2—COOH$) and proton-acceptor ($^+H_3N—CH_2—COO^-$) species are present. At the midpoint of any titration, a point of inflection is reached where the pH is equal to the pK_a of the protonated group being titrated (see Fig. 4–16). For glycine, the pH at the midpoint is 2.34, thus its —COOH group has a pK_a (labeled pK_1 in Fig. 5–10) of 2.34. (Recall from Chapter 4 that pH and pK_a are simply convenient notations for proton concentration and the equilibrium constant for ionization, respectively. The pK_a is a measure of the tendency of a group to give up a proton, with that tendency decreasing tenfold as the pK_a increases by one unit.) As the titration proceeds, another important point is reached at pH 5.97. Here there is another point of inflection, at which removal of the first proton is essentially complete and



Nonionic form Zwitterionic form

figure 5–9
Nonionic and zwitterionic forms of amino acids. The nonionic form does not occur in significant amounts in aqueous solutions. The zwitterion predominates at neutral pH.

figure 5–10
Titration of an amino acid. Shown here is the titration curve of 0.1 M glycine at 25 °C. The ionic species predominating at key points in the titration are shown above the graph. The shaded boxes, centered at about $pK_1 = 2.34$ and $pK_2 = 9.60$, indicate the regions of greatest buffering power.



removal of the second has just begun. At this pH glycine is present largely as the dipolar ion $^{+}H_3N—CH_2—COO^{-}$. We shall return to the significance of this inflection point in the titration curve (pI in Fig. 5–10) shortly.

The second stage of the titration corresponds to the removal of a proton from the $—NH_3^+$ group of glycine. The pH at the midpoint of this stage is 9.60, equal to the pK_a (labeled pK_2 in Fig. 5–10) for the $—NH_3^+$ group. The titration is essentially complete at a pH of about 12, at which point the predominant form of glycine is $H_2N—CH_2—COO^{-}$.

From the titration curve of glycine we can derive several important pieces of information. First, it gives a quantitative measure of the pK_a of each of the two ionizing groups: 2.34 for the —COOH group and 9.60 for the $—NH_3^+$ group. Note that the carboxyl group of glycine is over 100 times more acidic (more easily ionized) than the carboxyl group of acetic acid, which, as we saw in Chapter 4, has a pK_a of 4.76, about average for a carboxyl group attached to an otherwise unsubstituted aliphatic hydrocarbon. The perturbed pK_a of glycine is caused by repulsion between the departing proton and the nearby positively charged amino group on the α-carbon atom, as described in Figure 5–11. The opposite charges on the resulting zwitterion are stabilizing, nudging the equilibrium farther to the right. Similarly, the pK_a of the amino group in glycine is perturbed downward relative to the average pK_a of an amino group. This effect is due partly to the electronegative oxygen atoms in the carboxyl groups, which tend to pull electrons toward them regardless of the carboxyl group charge, increasing the tendency of the amino group to give up a proton. Hence, the α-amino group has a pK_a that is lower than that of an aliphatic amine such as methylamine (Fig. 5–11). In short, the pK_a of any functional group is greatly affected by its chemical environment, a phenomenon sometimes exploited in the active sites of enzymes to promote exquisitely adapted reaction mechanisms that depend on the perturbed pK_a values of proton donor/acceptor groups of specific residues.

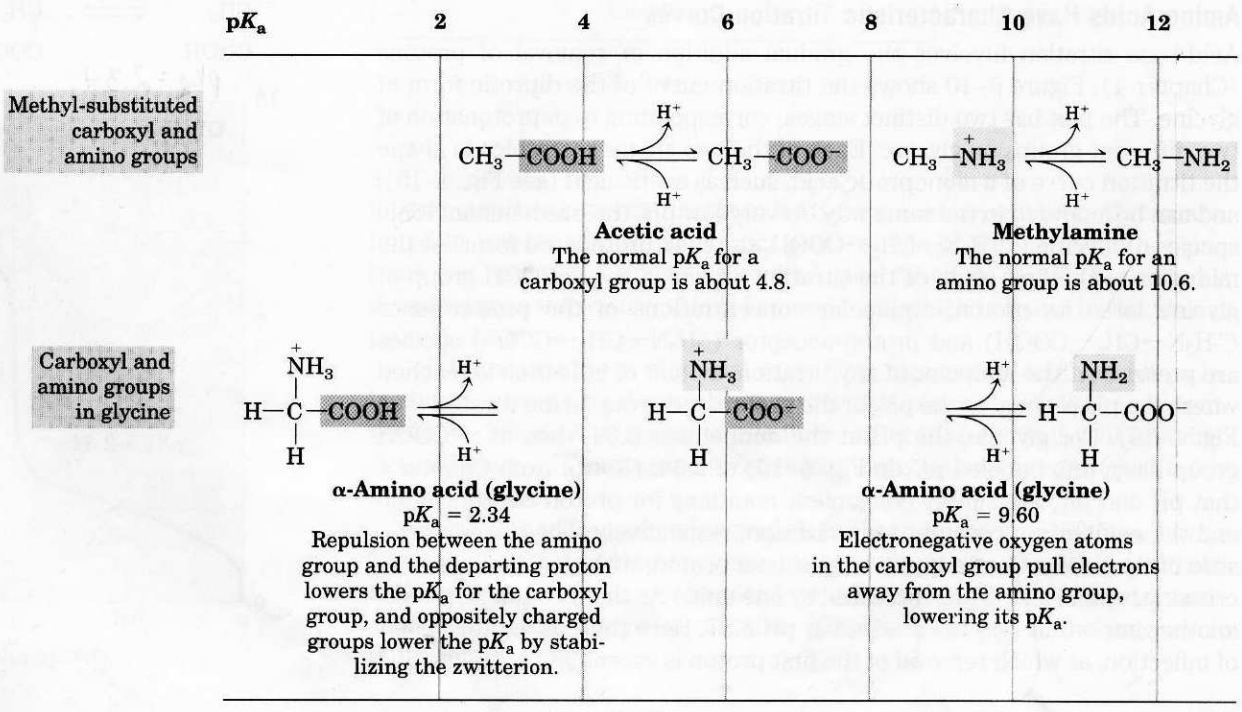


figure 5–11
Effect of the chemical environment on pK_a. The pK_a values for the ionizable groups in glycine are lower than those for simple, methyl-substituted amino and carboxyl groups. These downward perturbations of pK_a are due to intramolecular interactions. Similar effects can be caused by chemical groups that happen to be positioned nearby—for example, in the active site of an enzyme.

The second piece of information provided by the titration curve of glycine (Fig. 5–10) is that this amino acid has *two* regions of buffering power (see Fig. 4–17). One of these is the relatively flat portion of the curve, extending for approximately one pH unit on either side of the first pK_a of 2.34, indicating that glycine is a good buffer near this pH. The other buffering zone is centered around pH 9.60. Note that glycine is not a good buffer at the pH of intracellular fluid or blood, about 7.4. Within the buffering ranges of glycine, the Henderson-Hasselbalch equation (Chapter 4) can be used to calculate the proportions of proton-donor and proton-acceptor species of glycine required to make a buffer at a given pH.

Titration Curves Predict the Electric Charge of Amino Acids

Another important piece of information derived from the titration curve of an amino acid is the relationship between its net electric charge and the pH of the solution. At pH 5.97, the point of inflection between the two stages in its titration curve, glycine is present predominantly as its dipolar form, fully ionized but with no *net* electric charge (Fig. 5–10). The characteristic pH at which the net electric charge is zero is called the **isoelectric point** or **isoelectric pH,** designated **pI.** For glycine, which has no ionizable group in its side chain, the isoelectric point is simply the arithmetic mean of the two pK_a values:

$$pI = \frac{1}{2}(pK_1 + pK_2) = \frac{1}{2}(2.34 + 9.60) = 5.97$$

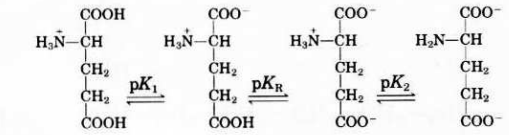
As is evident in Figure 5–10, glycine has a net negative charge at any pH above its pI and will thus move toward the positive electrode (the anode) when placed in an electric field. At any pH below its pI, glycine has a net positive charge and will move toward the negative electrode (the cathode). The farther the pH of a glycine solution is from its isoelectric point, the greater the net electric charge of the population of glycine molecules. At pH 1.0, for example, glycine exists almost entirely as the form $^+H_3N—CH_2—COOH$, with a net positive charge of 1.0. At pH 2.34, where there is an equal mixture of $^+H_3N—CH_2—COOH$ and $^+H_3N—CH_2—COO^-$, the average or net positive charge is 0.5. The sign and the magnitude of the net charge of any amino acid at any pH can be predicted in the same way.

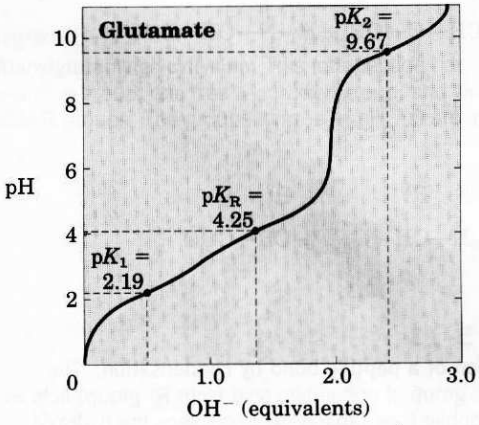
Amino Acids Differ in Their Acid-Base Properties

The shared properties of many amino acids permit some simplifying generalizations about their acid-base behaviors.

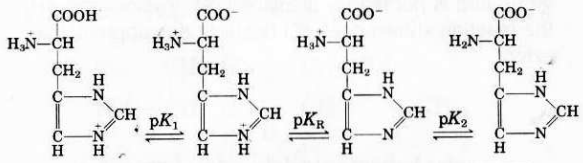
All amino acids with a single α-amino group, a single α-carboxyl group, and an R group that does not ionize have titration curves resembling that of glycine (Fig. 5–10). These amino acids have very similar, although not identical, pK_a values: pK_a of the —COOH group in the range of 1.8 to 2.4, and pK_a of the $—NH_3^+$ group in the range of 8.8 to 11.0 (Table 5–1).

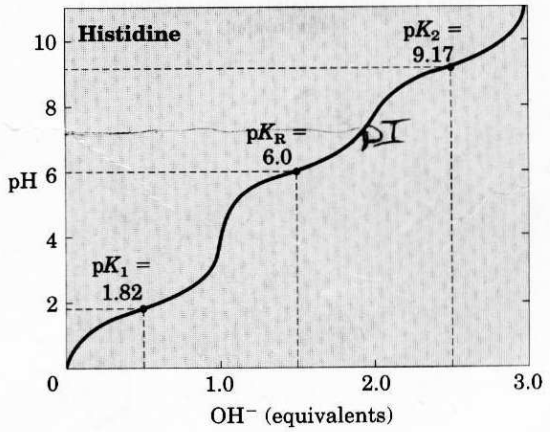
Amino acids with an ionizable R group have more complex titration curves, with *three* stages corresponding to the three possible ionization steps; thus they have three pK_a values. The additional stage for the titration of the ionizable R group merges to some extent with the other two. The titration curves for two amino acids of this type, glutamate and histidine, are shown in Figure 5–12. The isoelectric points reflect the nature of the ionizing R groups present. For example, glutamate has a pI of 3.22, considerably lower than that of glycine. This is due to the presence of two carboxyl groups which, at the average of their pK_a values (3.22), contribute a net negative charge of −1 that balances the +1 contributed by the amino group. Similarly, the pI of histidine, with two groups that are positively charged when protonated, is 7.59 (the average of the pK_a values of the amino and imidazole groups), much higher than that of glycine.





(a)





(b)

figure 5–12
Titration curves for **(a)** glutamate and **(b)** histidine. The pK_a of the R group is designated here as pK_R.

Another important generalization can be made about the acid-base behavior of the 20 standard amino acids. As pointed out earlier, under the general condition of free and open exposure to the aqueous environment, only histidine has an R group (pK_a = 6.0) providing significant buffering power near the neutral pH usually found in the intracellular and intercellular fluids of most animals and bacteria. No other amino acid has an ionizable side chain with a pK_a value near enough to pH 7.0 to be an effective physiological buffer (Table 5–1).

Peptides and Proteins

We now turn to polymers of amino acids, the **peptides** and **proteins.** Biologically occurring peptides range in size from small to very large, consisting of two or three to thousands of linked amino acid residues. The focus here is on the fundamental chemical properties of these polymers.

Peptides Are Chains of Amino Acids

Two amino acid molecules can be covalently joined through a substituted amide linkage, termed a **peptide bond,** to yield a dipeptide. Such a linkage is formed by removal of the elements of water (dehydration) from the α-carboxyl group of one amino acid and the α-amino group of another (Fig. 5–13). Peptide bond formation is an example of a condensation reaction, a common class of reaction in living cells. Under standard biochemical conditions the reaction shown in Figure 5–13 has an equilibrium that favors reactants rather than products. To make the reaction thermodynamically more favorable, the carboxyl group must be chemically modified or activated so that the hydroxyl group can be more readily eliminated. A chemical approach to this problem is outlined later in this chapter. The biological approach to peptide bond formation is a major topic of Chapter 27.

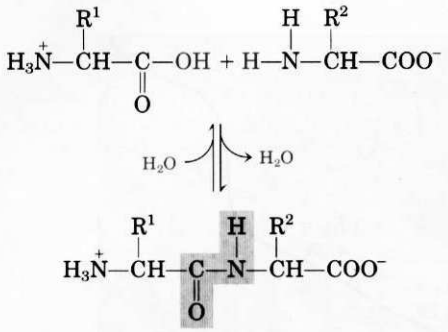


figure 5–13
Formation of a peptide bond by condensation. The α-amino group of one amino acid (with R^2 group) acts as a nucleophile (see Table 3–4) to displace the hydroxyl group of another amino acid (with R^1 group), forming a peptide bond (shaded in gray). Amino groups are good nucleophiles, but the hydroxyl group is a poor leaving group and is not readily displaced. At physiological pH, the reaction shown does not occur to any appreciable extent.

Three amino acids can be joined by two peptide bonds to form a tripeptide; similarly, amino acids can be linked to form tetrapeptides and pentapeptides. When a few amino acids are joined in this fashion, the structure is called an **oligopeptide.** When many amino acids are joined, the product is called a **polypeptide.** Proteins may have thousands of amino acid residues. Although the terms "protein" and "polypeptide" are sometimes used interchangeably, molecules referred to as polypeptides generally have molecular weights below 10,000.

Figure 5–14 shows the structure of a pentapeptide. As already noted, an amino acid unit in a peptide is often called a residue (the part left over after losing a hydrogen atom from its amino group and a hydroxyl moiety from its carboxyl group). In a peptide, the amino acid residue at the end with a free α-amino group is the **amino-terminal** (or *N*-terminal) residue; the residue at the other end, which has a free carboxyl group, is the **carboxyl-terminal** (*C*-terminal) residue.

Although hydrolysis of a peptide bond is an exergonic reaction, it occurs slowly because of its high activation energy. As a result, the peptide bonds in proteins are quite stable, with a half-life ($t_{1/2}$) of about 7 years under most intracellular conditions.

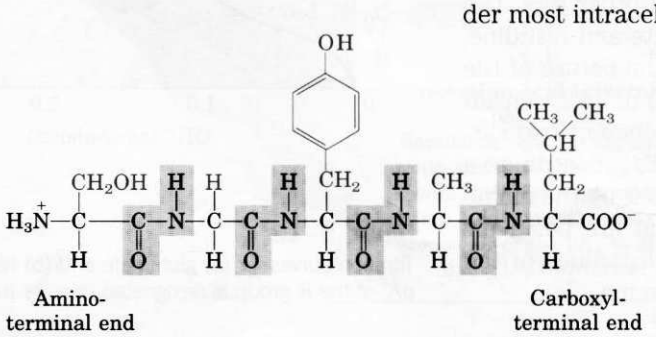


figure 5–14
The pentapeptide serylglycyltyrosylalanylleucine, or Ser–Gly–Tyr–Ala–Leu. Peptides are named beginning with the amino-terminal residue, which by convention is placed at the left. The peptide bonds are shaded in gray, the R groups are in red.

Peptides Can Be Distinguished by Their Ionization Behavior

Peptides contain only one free α-amino group and one free α-carboxyl group, one at each end of the chain (Fig. 5–15). These groups ionize as they do in free amino acids, although the ionization constants are different because the oppositely charged group is absent from the α carbon. The α-amino and α-carboxyl groups of all nonterminal amino acids are covalently joined in the form of peptide bonds, which do not ionize and thus do not contribute to the total acid-base behavior of peptides. However, the R groups of some amino acids can ionize (Table 5–1), and in a peptide these contribute to the overall acid-base properties of the molecule (Fig. 5–15). Thus the acid-base behavior of a peptide can be predicted from its free α-amino and α-carboxyl groups as well as the nature and number of its ionizable R groups. Like free amino acids, peptides have characteristic titration curves and a characteristic isoelectric pH (pI) at which they do not move in an electric field. These properties are exploited in some of the techniques used to separate peptides and proteins, as we shall see later in the chapter. It should be emphasized that the pK_a value for an ionizable R group can change somewhat when an amino acid becomes a residue in a peptide. The loss of charge in the α-carboxyl and α-amino groups, interactions with other peptide R groups, and other environmental factors can affect the pK_a. The pK_a values for R groups listed in Table 5–1 can be a useful guide to the pH range in which a given group will ionize, but they cannot be strictly applied to peptides.

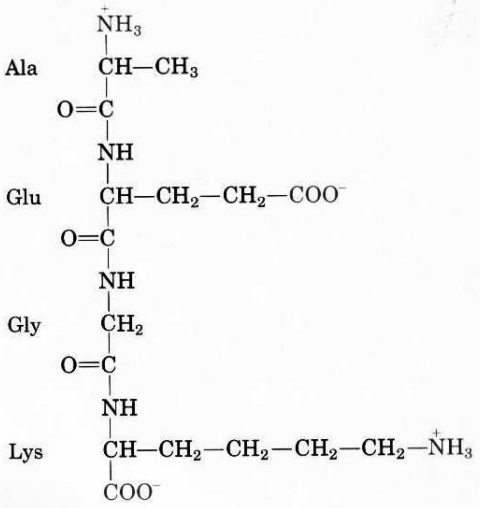
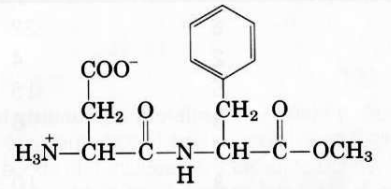


figure 5–15
Alanylglutamylglycyllysine. This tetrapeptide has one free α-amino group, one free α-carboxyl group, and two ionizable R groups. The groups ionized at pH 7.0 are in red.

Biologically Active Peptides and Polypeptides Occur in a Vast Range of Sizes

No generalizations can be made about the molecular weights of biologically active peptides and proteins in relation to their function. Naturally occurring peptides range in length from two amino acids to many thousands of residues. Even the smallest peptides can have biologically important effects. Consider the commercially synthesized dipeptide L-aspartyl-L-phenylalanine methyl ester, the artificial sweetener better known as aspartame or NutraSweet.



L-Aspartyl-L-phenylalanine methyl ester (aspartame)

Many small peptides exert their effects at very low concentrations. For example, a number of vertebrate hormones (Chapter 23) are small peptides. These include oxytocin (nine amino acid residues), which is secreted by the posterior pituitary and stimulates uterine contractions; bradykinin (nine residues), which inhibits inflammation of tissues; and thyrotropin-releasing factor (three residues), which is formed in the hypothalamus and stimulates the release of another hormone, thyrotropin, from the anterior pituitary gland. Some extremely toxic mushroom poisons, such as amanitin, are also small peptides, as are many antibiotics.

Slightly larger are small polypeptides and oligopeptides such as the pancreatic hormone insulin, which contains two polypeptide chains, one having 30 amino acid residues and the other 21. Glucagon, another pancreatic hormone, has 29 residues; it opposes the action of insulin. Corticotropin is a 39-residue hormone of the anterior pituitary gland that stimulates the adrenal cortex.

How long are the polypeptide chains in proteins? As Table 5–2 shows, lengths vary considerably. Human cytochrome *c* has 104 amino acid residues linked in a single chain; bovine chymotrypsinogen has 245 residues. At the extreme is titin, a constituent of vertebrate muscle, which has nearly 27,000 amino acid residues and a molecular weight of about 3,000,000. The vast majority of naturally occurring polypeptides are much smaller than this, containing less than 2,000 amino acid residues.

Some proteins consist of a single polypeptide chain, but others, called **multisubunit** proteins, have two or more polypeptides associated

table 5–2

Molecular Data on Some Proteins

	Molecular weight	Number of residues	Number of polypeptide chains
Cytochrome *c* (human)	13,000	104	1
Ribonuclease A (bovine pancreas)	13,700	124	1
Lysozyme (egg white)	13,930	129	1
Myoglobin (equine heart)	16,890	153	1
Chymotrypsin (bovine pancreas)	21,600	241	3
Chymotrypsinogen (bovine)	22,000	245	1
Hemoglobin (human)	64,500	574	4
Serum albumin (human)	68,500	609	1
Hexokinase (yeast)	102,000	972	2
RNA polymerase (*E. coli*)	450,000	4,158	5
Apolipoprotein B (human)	513,000	4,536	1
Glutamine synthetase (*E. coli*)	619,000	5,628	12
Titin (human)	2,993,000	26,926	1

noncovalently (Table 5–2). The individual polypeptide chains in a multisubunit protein may be identical or different. If at least two are identical the protein is said to be **oligomeric,** and identical units (consisting of one or more polypeptide chains) are referred to as **protomers.** Hemoglobin, for example, has four polypeptide subunits: two identical α chains and two identical β chains, all four held together by noncovalent interactions. Each α subunit is paired in an identical way with a β subunit within the structure of this multisubunit protein, so that hemoglobin can be considered either a tetramer of four polypeptide subunits or a dimer of $\alpha\beta$ protomers.

A few proteins contain two or more polypeptide chains linked covalently. For example, the two polypeptide chains of insulin are linked by disulfide bonds. In such cases, the individual polypeptides are not considered subunits, but are commonly referred to simply as chains.

We can calculate the approximate number of amino acid residues in a simple protein containing no other chemical group by dividing its molecular weight by 110. Although the average molecular weight of the 20 standard amino acids is about 138, the smaller amino acids predominate in most proteins; if we take into account the proportions in which the various amino acids occur in proteins (Table 5–1), the average molecular weight is nearer to 128. Because a molecule of water (M_r 18) is removed to create each peptide bond, the average molecular weight of an amino acid residue in a protein is about 128 − 18 = 110.

Polypeptides Have Characteristic Amino Acid Compositions

Hydrolysis of peptides or proteins with acid yields a mixture of free α-amino acids. When completely hydrolyzed, each type of protein yields a characteristic proportion or mixture of the different amino acids. The 20 standard amino acids almost never occur in equal amounts in a protein. Some amino acids may occur only once per molecule or not at all in a given type of protein; others may occur in large numbers. Table 5–3 shows the composition of the amino acid mixtures obtained on complete hydrolysis of bovine cytochrome *c* and chymotrypsinogen, the inactive precursor of the digestive enzyme chymotrypsin. These two proteins, with very different functions, also differ significantly in the relative numbers of each kind of amino acid they contain.

table 5–3

Amino Acid Composition of Two Proteins*

Amino acid	Number of residues per molecule of protein: Bovine cytochrome *c*	Bovine chymotrypsinogen
Ala	6	22
Arg	2	4
Asn	5	15
Asp	3	8
Cys	2	10
Gln	3	10
Glu	9	5
Gly	14	23
His	3	2
Ile	6	10
Leu	6	19
Lys	18	14
Met	2	2
Phe	4	6
Pro	4	9
Ser	1	28
Thr	8	23
Trp	1	8
Tyr	4	4
Val	3	23
Total	104	245

*Note that standard procedures for the acid hydrolysis of proteins convert Asn and Gln to Asp and Glu, respectively. In addition, Trp is destroyed. Special procedures must be employed to determine the amounts of these amino acids.

Some Proteins Contain Chemical Groups Other Than Amino Acids

Many proteins, for example the enzymes ribonuclease and chymotrypsinogen, contain only amino acid residues and no other chemical groups; these are considered simple proteins. However, some proteins contain permanently associated chemical components in addition to amino acids; these are called **conjugated proteins.** The non–amino acid part of a conjugated protein is usually called its **prosthetic group.** Conjugated proteins are classified on the basis of the chemical nature of their prosthetic groups (Table 5–4); for example, **lipoproteins** contain lipids, **glycoproteins** contain sugar groups, and **metalloproteins** contain a specific metal. A number of proteins contain more than one prosthetic group. Usually the prosthetic group plays an important role in the protein's biological function.

table 5–4

Conjugated Proteins

Class	Prosthetic group(s)	Example
Lipoproteins	Lipids	β_1-Lipoprotein of blood
Glycoproteins	Carbohydrates	Immunoglobulin G
Phosphoproteins	Phosphate groups	Casein of milk
Hemoproteins	Heme (iron porphyrin)	Hemoglobin
Flavoproteins	Flavin nucleotides	Succinate dehydrogenase
Metalloproteins	Iron	Ferritin
	Zinc	Alcohol dehydrogenase
	Calcium	Calmodulin
	Molybdenum	Dinitrogenase
	Copper	Plastocyanin

There Are Several Levels of Protein Structure

For large macromolecules such as proteins, the tasks of describing and understanding structure are approached at several levels of complexity, arranged in a kind of conceptual hierarchy. Four levels of protein structure are commonly defined (Fig. 5–16). A description of all covalent bonds (mainly peptide bonds and disulfide bonds) linking amino acid residues in a polypeptide chain is its **primary structure.** The most important element of primary structure is the *sequence* of amino acid residues. **Secondary structure** refers to particularly stable arrangements of amino acid residues giving rise to recurring structural patterns. **Tertiary structure** describes all aspects of the three-dimensional folding of a polypeptide. When a protein has two or more polypeptide subunits, their arrangement in space is referred to as **quaternary structure.**

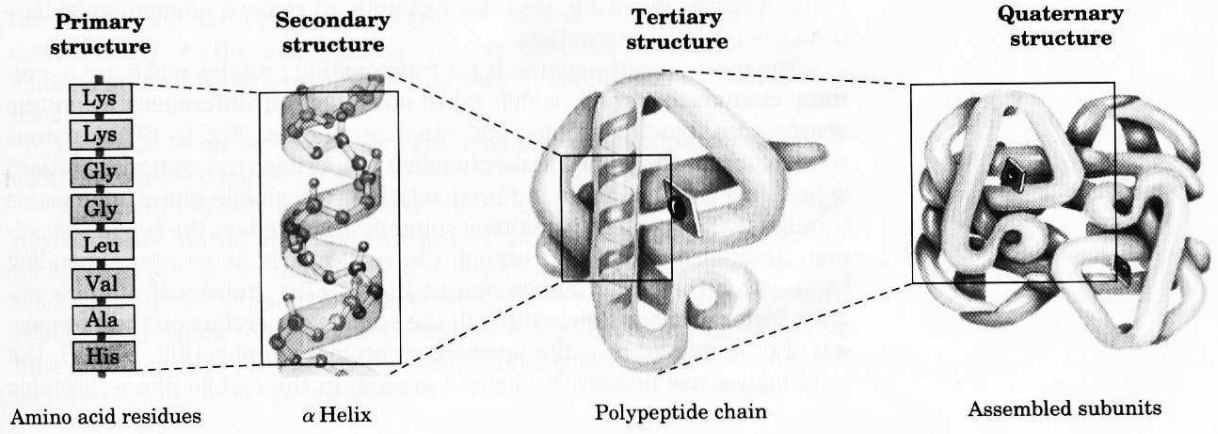


figure 5–16

Levels of structure in proteins. The *primary structure* consists of a sequence of amino acids linked together by peptide bonds and includes any disulfide bonds. The resulting polypeptide can be coiled into units of *secondary structure,* such as an α helix. The helix is a part of the *tertiary structure* of the folded polypeptide, which is itself one of the subunits that make up the *quaternary structure* of the multisubunit protein, in this case hemoglobin.

Working with Proteins

Our understanding of protein structure and function has been derived from the study of many individual proteins. To study a protein in any detail it must be separated from all other proteins, and techniques must be available to determine its properties. The necessary methods come from protein chemistry, a discipline as old as biochemistry itself and one that retains a central position in biochemical research.

Proteins Can Be Separated and Purified

A pure preparation of a protein is essential before its properties, amino acid composition, and sequence can be determined. Given that cells contain thousands of different kinds of proteins, how can one protein be purified? Methods for separating proteins take advantage of properties that vary from one protein to the next. For example, many proteins bind to other biomolecules with great specificity, and such proteins can be separated on the basis of their binding properties.

The source of a protein is generally tissue or microbial cells. The first step in any protein purification procedure is to break open these cells, releasing their proteins into a solution called a **crude extract.** If necessary, differential centrifugation can be used to prepare subcellular fractions or to isolate specific organelles (see Fig. 2–20).

Once the extract or organelle preparation is ready, various methods are available for purifying one or more of the proteins it contains. Commonly, the extract is subjected to treatments that separate the proteins into different fractions based on some property such as size or charge, a process referred to as **fractionation.** Early fractionation steps in a purification utilize differences in protein solubility, which is a complex function of pH, temperature, salt concentration, and other factors. The solubility of proteins is generally lowered at high salt concentrations, an effect called "salting out." The addition of a salt in the right amounts can selectively precipitate some proteins, while others remain in solution. Ammonium sulfate ($(NH_4)_2SO_4$) is often used for this purpose because of its high solubility in water.

A solution containing the protein of interest often must be further altered before subsequent purification steps are possible. For example, **dialysis** is a procedure that separates proteins from solvents by taking advantage of the proteins' larger size. The partially purified extract is placed in a bag or tube made of a semipermeable membrane. When this is suspended in a larger volume of buffered solution of appropriate ionic strength, the membrane allows the exchange of salt and buffer but not proteins. Thus dialysis retains large proteins within the membranous bag or tube while allowing the concentration of other solutes in the protein preparation to change until they come into equilibrium with the solution outside the membrane. Dialysis might be used, for example, to remove ammonium sulfate from the protein preparation.

The most powerful methods for fractionating proteins make use of **column chromatography,** which takes advantage of differences in protein charge, size, binding affinity, and other properties (Fig. 5–17). A porous solid material with appropriate chemical properties (the stationary phase) is held in a column, and a buffered solution (the mobile phase) percolates through it. The protein-containing solution is layered on the top of the column, then also percolates through the solid matrix as an ever-expanding band within the larger mobile phase (Fig. 5–17b). Individual proteins migrate faster or more slowly through the column depending on their properties. For example, in cation-exchange chromatography (Fig. 5–18a), the solid matrix has negatively charged groups. In the mobile phase, proteins

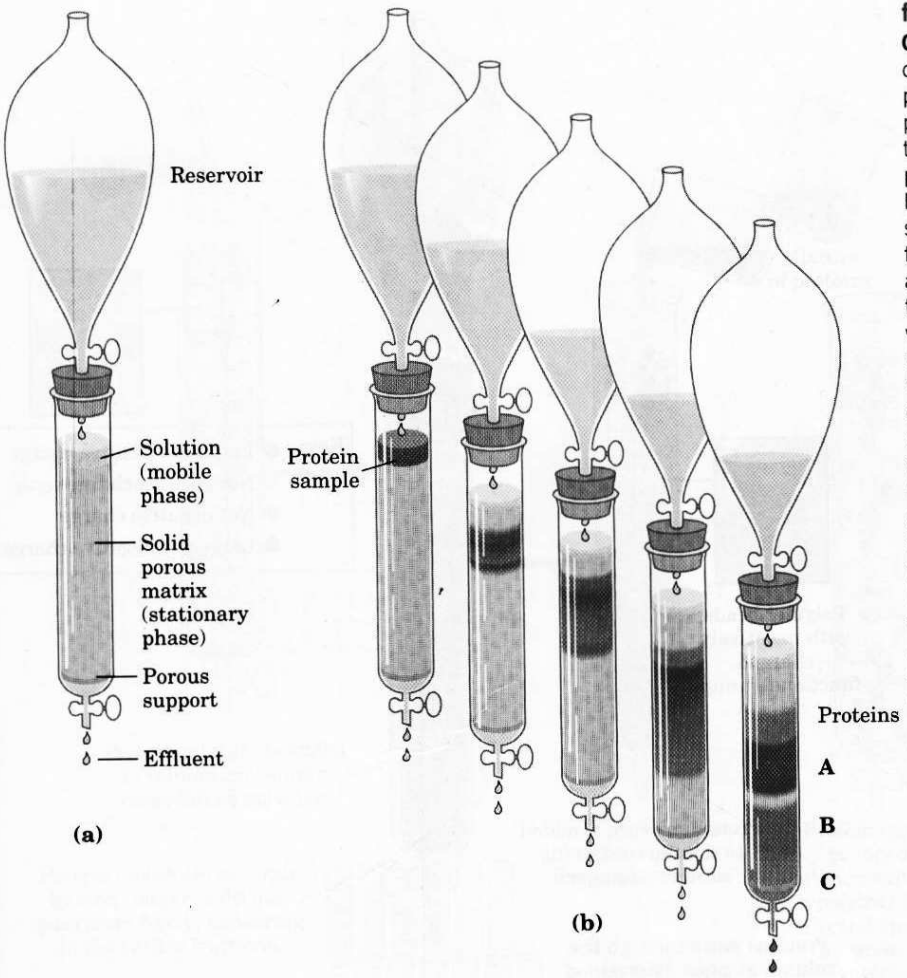


figure 5–17
Column chromatography. (a) The standard elements of a chromatographic column. A solid, porous material is supported inside a column generally made of some form of plastic. The solid material (matrix) makes up the stationary phase through which flows a solution, the mobile phase. The solution that passes out of the column at the bottom (the effluent) is constantly replaced by solution supplied from a reservoir at the top. **(b)** The protein solution to be separated is layered on top of the column and allowed to percolate into the solid matrix. Additional solution is added on top. The protein solution forms a band within the mobile phase that is initially the depth of the protein solution applied to the column. As proteins migrate through the column, they are retarded to different degrees by their different interactions with the matrix material. The overall protein band thus widens as it moves through the column. Individual types of proteins (such as A, B, and C, shown in blue, red, and green) gradually separate from each other, forming bands within the broader protein band. Separation improves (resolution increases) as the length of the column increases. However, each individual protein band also broadens with time due to diffusional spreading, a process that decreases resolution. In this example, protein A is well separated from B and C, but diffusional spreading prevents complete separation of B and C under these conditions.

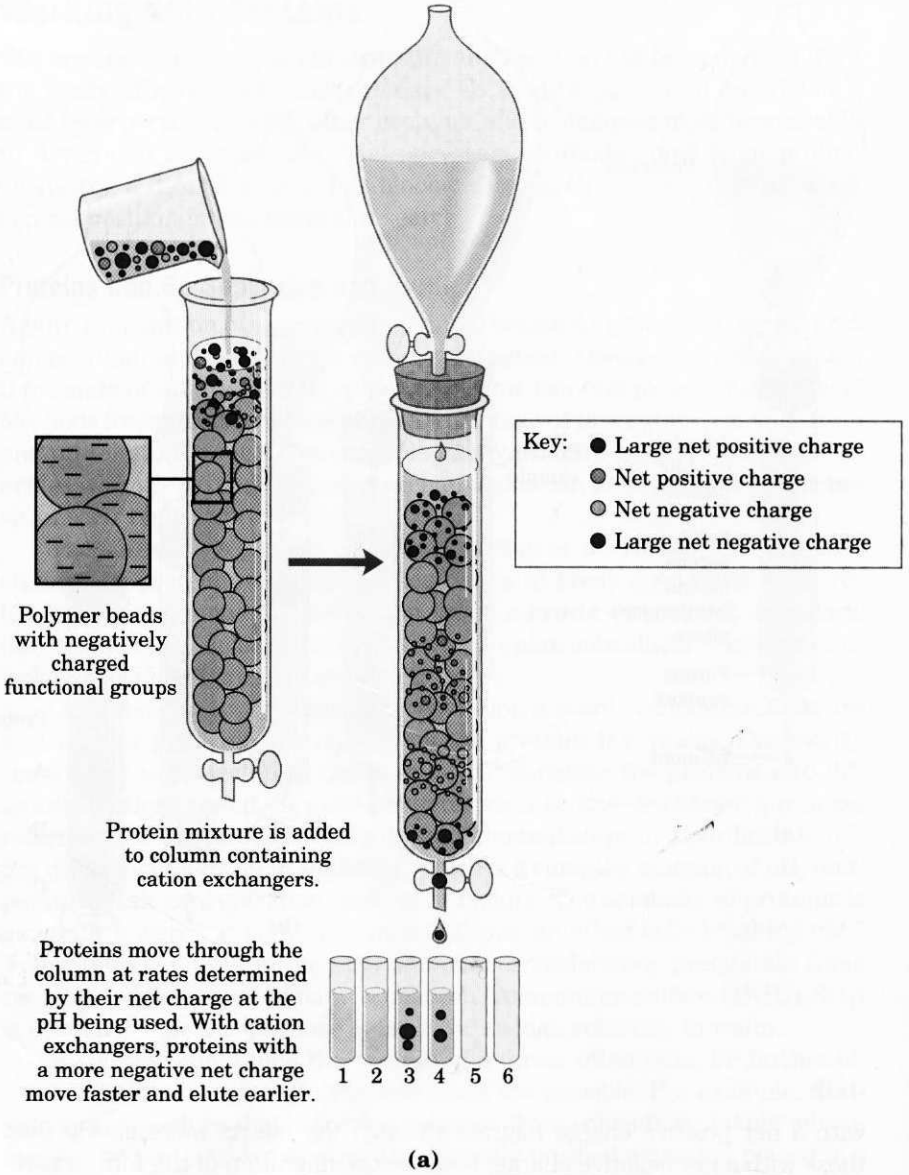
with a net positive charge migrate through the matrix more slowly than those with a net negative charge, because the migration of the former is retarded more by interaction with the stationary phase. The two types of protein can separate into two distinct bands. The expansion of the protein band in the mobile phase (the protein solution) is caused both by separation of proteins with different properties and by diffusional spreading. As the length of the column increases, the resolution of two types of protein with different net charges generally improves. However, the rate at which the protein solution can flow through the column usually decreases with column length. As the length of time spent on the column increases, the resolution can decline as a result of diffusional spreading within each protein band.

Figure 5–18 shows two other variations of column chromatography in addition to ion exchange.

A modern refinement in chromatographic methods is **HPLC,** or **high-performance liquid chromatography.** HPLC makes use of high-pressure pumps that speed the movement of the protein molecules down the column, as well as higher-quality chromatographic materials that can withstand the crushing force of the pressurized flow. By reducing the transit time on the column, HPLC can limit diffusional spreading of protein bands and thus greatly improve resolution.

figure 5–18

Three chromatographic methods used in protein purification. (a) Ion-exchange chromatography exploits differences in the sign and magnitude of the net electric charges of proteins at a given pH. The column matrix is a synthetic polymer containing bound charged groups; those with bound anionic groups are called **cation exchangers,** and those with bound cationic groups are called **anion exchangers.** Ion-exchange chromatography on a cation exchanger is shown here. The affinity of each protein for the charged groups on the column is affected by the pH (which determines the ionization state of the molecule) and the concentration of competing free salt ions in the surrounding solution. Separation can be optimized by gradually changing the pH and/or salt concentration of the mobile phase so as to create a pH or salt gradient. **(b) Size-exclusion chromatography,** also called gel filtration, separates proteins according to size. The column matrix is a cross-linked polymer with pores of selected size. Larger proteins migrate faster than smaller ones because they are too large to enter the pores in the beads and hence take a more direct route through the column. The smaller proteins enter the pores and are slowed by the more labyrinthine path they take through the column. **(c) Affinity chromatography** separates proteins by their binding specificities. The proteins retained on the column are those that bind specifically to a ligand cross-linked to the beads. (In biochemistry, the term "ligand" is used to refer to a group or molecule that binds to a macromolecule such as a protein.) After proteins that do not bind to the ligand are washed through the column, the bound protein of particular interest is eluted (washed out of the column) by a solution containing free ligand.



The approach to the purification of a protein that has not been previously isolated is guided both by established precedents and by common sense. In most cases, several different methods must be used sequentially to purify a protein completely. The choice of method is somewhat empirical, and many protocols may be tried before the most effective one is found. Trial and error can often be minimized by basing the procedure on purification techniques developed for similar proteins. Published purification protocols are available for many thousands of proteins. Common sense dictates that inexpensive procedures such as "salting out" be used first, when the total volume and number of contaminants is greatest. Chromatographic methods are often impractical at early stages because the amount of chromatographic medium needed increases with sample size. As each purification step is completed, the sample size generally becomes smaller (Table 5–5), making it feasible to use more sophisticated (and expensive) chromatographic procedures at later stages.

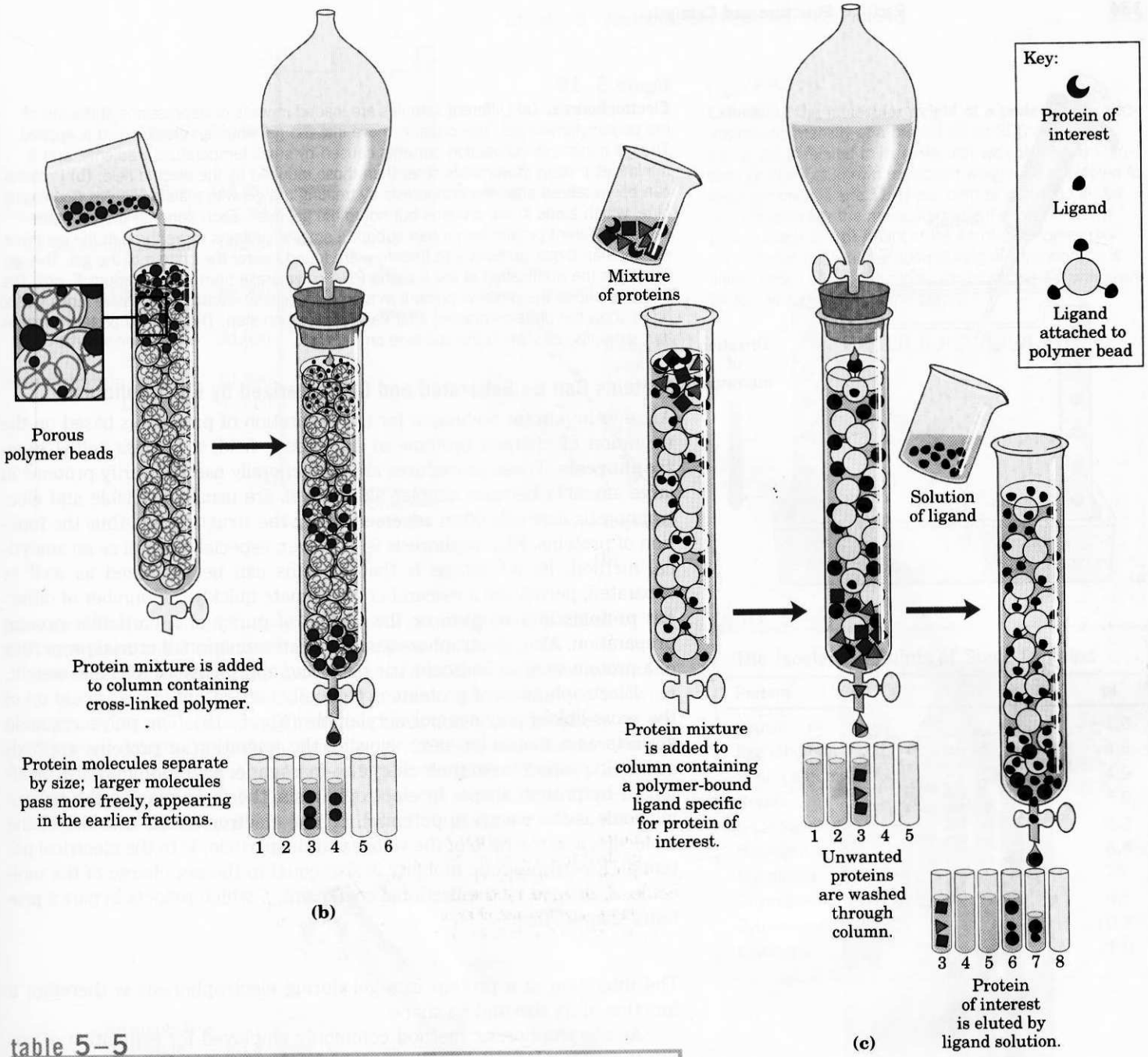
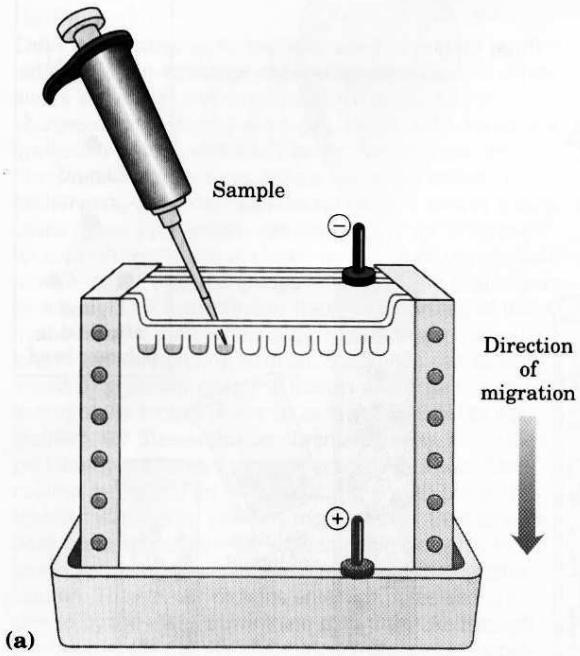


table 5–5

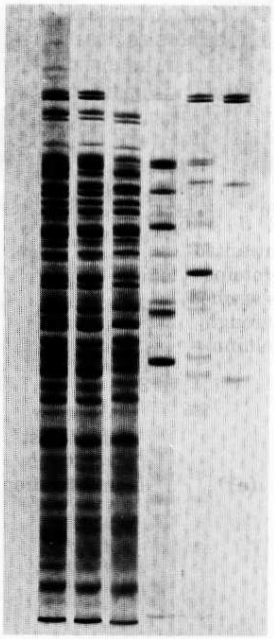
A Purification Table for a Hypothetical Enzyme*

Procedure or step	Fraction volume (ml)	Total protein (mg)	Activity (units)	Specific activity (units/mg)
1. Crude cellular extract	1,400	10,000	100,000	10
2. Precipitation with ammonium sulfate	280	3,000	96,000	32
3. Ion-exchange chromatography	90	400	80,000	200
4. Size-exclusion chromatography	80	100	60,000	600
5. Affinity chromatography	6	3	45,000	15,000

*All data represent the status of the sample *after* the designated procedure has been carried out. Activity and specific activity are defined on page 137.



(a)



(b)

figure 5–19

Electrophoresis. (a) Different samples are loaded in wells or depressions at the top of the polyacrylamide gel. The proteins move into the gel when an electric field is applied. The gel minimizes convection currents caused by small temperature gradients, and it minimizes protein movements other than those induced by the electric field. **(b)** Proteins can be visualized after electrophoresis by treating the gel with a stain such as Coomassie blue, which binds to the proteins but not to the gel itself. Each band on the gel represents a different protein (or protein subunit); smaller proteins move through the gel more rapidly than larger proteins and therefore are found nearer the bottom of the gel. This gel illustrates the purification of the enzyme RNA polymerase from the bacterium *E. coli*. The first lane shows the proteins present in the crude cellular extract. Successive lanes (left to right) show the proteins present after each purification step. The purified protein contains four subunits, as seen in the last lane on the right.

Proteins Can Be Separated and Characterized by Electrophoresis

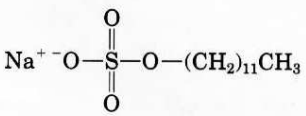
Another important technique for the separation of proteins is based on the migration of charged proteins in an electric field, a process called **electrophoresis.** These procedures are not generally used to purify proteins in large amounts because simpler alternatives are usually available and electrophoretic methods often adversely affect the structure and thus the function of proteins. Electrophoresis is, however, especially useful as an analytical method. Its advantage is that proteins can be visualized as well as separated, permitting a researcher to estimate quickly the number of different proteins in a mixture or the degree of purity of a particular protein preparation. Also, electrophoresis allows determination of crucial properties of a protein such as its isoelectric point and approximate molecular weight.

Electrophoresis of proteins is generally carried out in gels made up of the cross-linked polymer polyacrylamide (Fig. 5–19). The polyacrylamide gel acts as a molecular sieve, slowing the migration of proteins approximately in proportion to their charge-to-mass ratio. Migration may also be affected by protein shape. In electrophoresis, the force moving the macromolecule is the electrical potential, E. The electrophoretic mobility of the molecule, μ, is the ratio of the velocity of the particle, V, to the electrical potential. Electrophoretic mobility is also equal to the net charge of the molecule, Z, divided by the frictional coefficient, f, which reflects in part a protein's shape. Thus:

$$\mu = \frac{V}{E} = \frac{Z}{f}$$

The migration of a protein in a gel during electrophoresis is therefore a function of its size and its shape.

An electrophoretic method commonly employed for estimation of purity and molecular weight makes use of the detergent **sodium dodecyl sulfate (SDS).** SDS binds to most proteins (probably by hydrophobic interactions; see Chapter 4) in amounts roughly proportional to the molecular weight of the protein, about one molecule of SDS for every two amino acid residues. The bound SDS contributes a large net negative charge, rendering the intrinsic charge of the protein insignificant and conferring on each protein a similar charge-to-mass ratio. In addition, the native conformation of a protein is altered when SDS is bound, and most proteins assume a similar shape. Electrophoresis in the presence of SDS therefore separates proteins almost exclusively on the basis of mass (molecular weight), with smaller polypeptides migrating more rapidly. After electrophoresis, the proteins are visualized by adding a dye such as Coomassie blue, which binds to proteins but not to the gel itself (Fig. 5–19b). Thus one can monitor the progress of a protein purification procedure, because the number of protein bands visible on the gel should decrease after each new fractionation step. When compared with the positions to which proteins of known molecular weight migrate in the gel, the position of an unidentified protein can provide an excellent measure of its molecular weight (Fig. 5–20). If the protein has two



Sodium dodecyl sulfate (SDS)

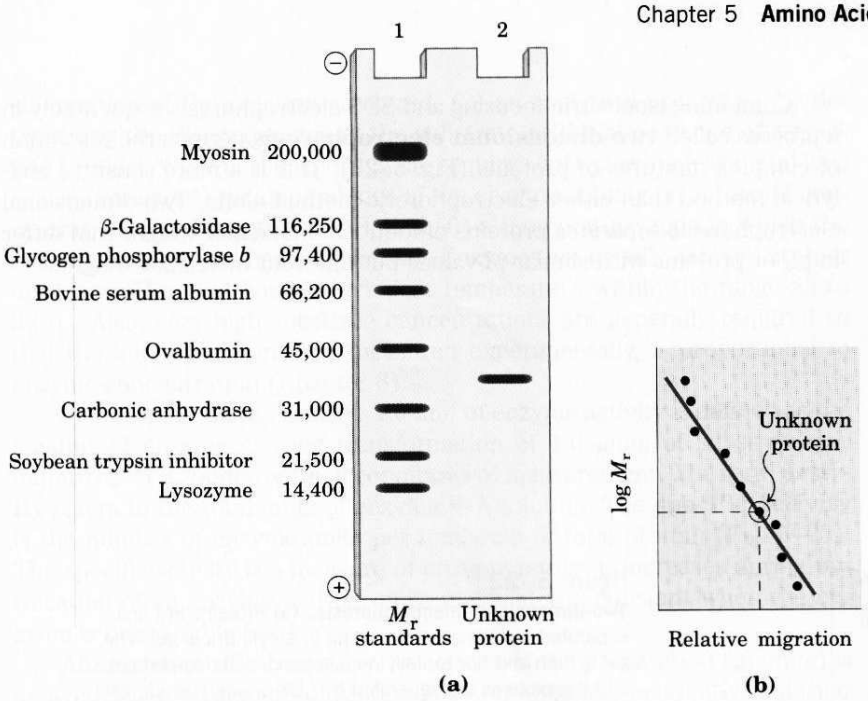


figure 5–20
Estimating the molecular weight of a protein. The electrophoretic mobility of a protein on an SDS polyacrylamide gel is related to its molecular weight, M_r. **(a)** Standard proteins of known molecular weight are subjected to electrophoresis (lane 1). These marker proteins can be used to estimate the molecular weight of an unknown protein (lane 2). **(b)** A plot of log M_r of the marker proteins versus relative migration during electrophoresis is linear, which allows the molecular weight of the unknown protein to be read from the graph.

or more different subunits, the subunits will generally be separated by the SDS treatment and a separate band will appear for each.

Isoelectric focusing is a procedure used to determine the isoelectric point (pI) of a protein (Fig. 5–21). A pH gradient is established by allowing a mixture of low molecular weight organic acids and bases (ampholytes; see p. 123) to distribute themselves in an electric field generated across the gel. When a protein mixture is applied, each protein migrates until it reaches the pH that matches its pI (Table 5–6). Proteins with different isoelectric points are thus distributed differently throughout the gel.

table 5–6

The Isoelectric Points of Some Proteins

Protein	pI
Pepsin	~1.0
Egg albumin	4.6
Serum albumin	4.9
Urease	5.0
β-Lactoglobulin	5.2
Hemoglobin	6.8
Myoglobin	7.0
Chymotrypsinogen	9.5
Cytochrome *c*	10.7
Lysozyme	11.0

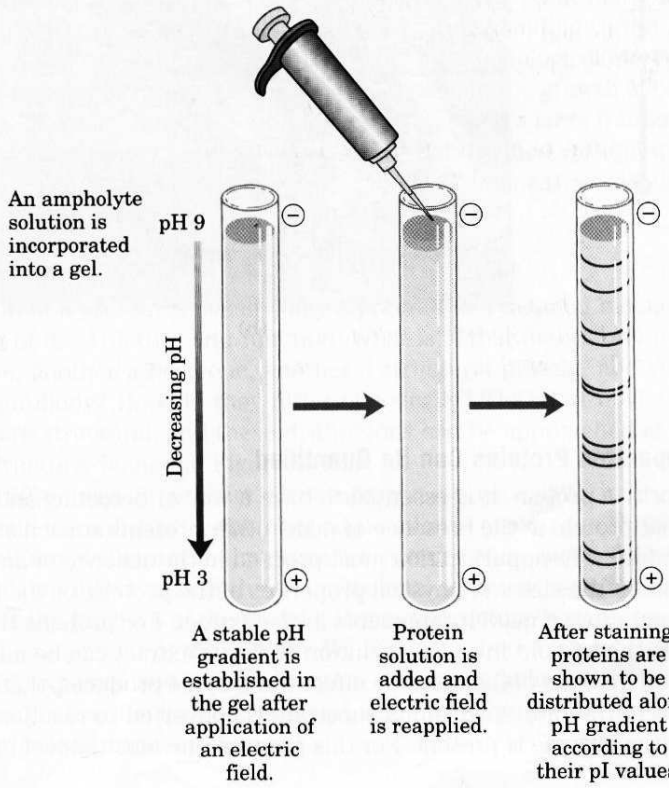
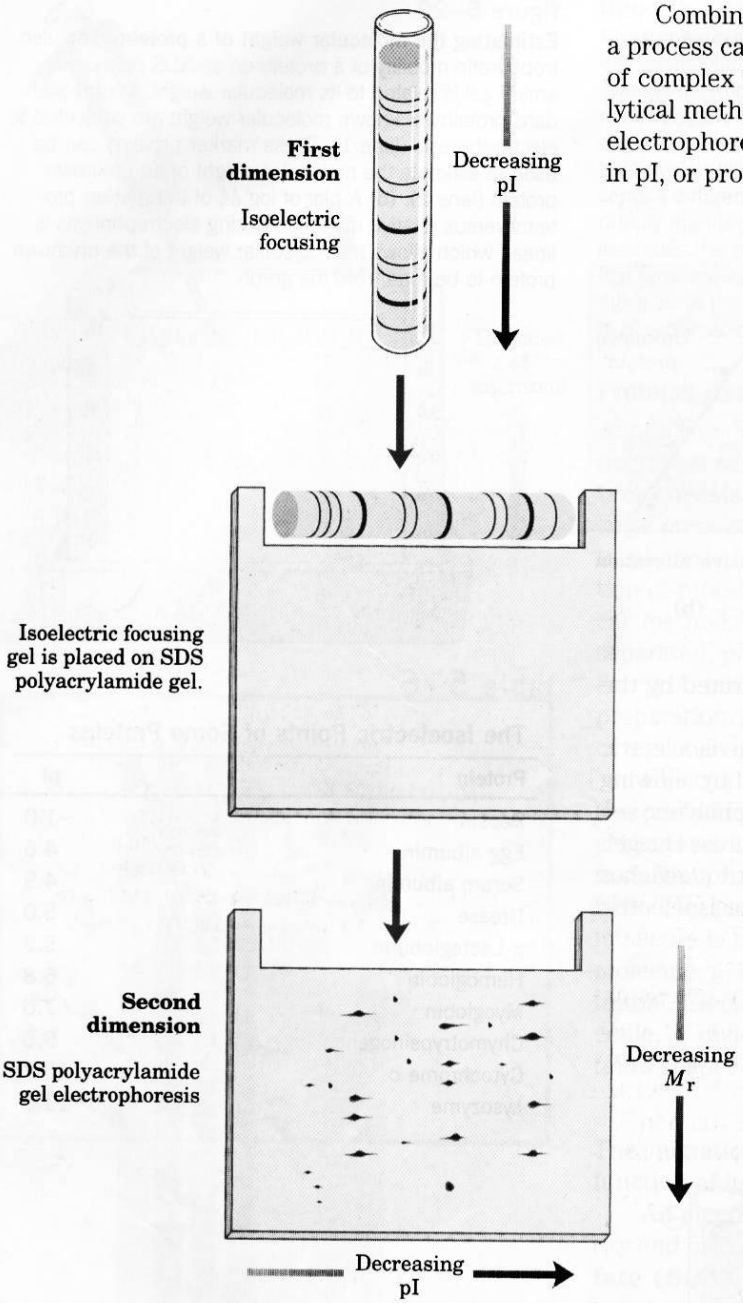


figure 5–21
Isoelectric focusing. This technique separates proteins according to their isoelectric points. A stable pH gradient is established in the gel by the addition of appropriate ampholytes. A protein mixture is placed in a well on the gel. With an applied electric field, proteins enter the gel and migrate until each reaches a pH equivalent to its pI. Remember that when pH = pI, the net charge of a protein is zero.

Combining isoelectric focusing and SDS electrophoresis sequentially in a process called **two-dimensional electrophoresis** permits the resolution of complex mixtures of proteins (Fig. 5–22). This is a more sensitive analytical method than either electrophoretic method alone. Two-dimensional electrophoresis separates proteins of identical molecular weight that differ in pI, or proteins with similar pI values but different molecular weights.



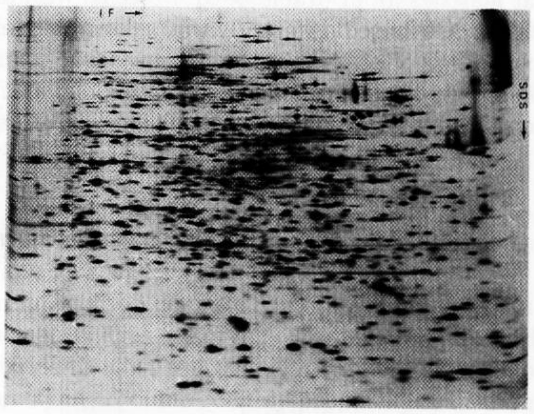


figure 5–22
Two-dimensional electrophoresis. (a) Proteins are first separated by isoelectric focusing in a cylindrical gel. The gel is then laid horizontally on a second, slab-shaped gel, and the proteins are separated by SDS polyacrylamide gel electrophoresis. Horizontal separation reflects differences in pI; vertical separation reflects differences in molecular weight. **(b)** More than 1,000 different proteins from *E. coli* can be resolved using this technique.

Unseparated Proteins Can Be Quantified

To purify a protein, it is essential to have a way of detecting and quantifying that protein in the presence of many other proteins at each stage of the procedure. Often, purification must proceed in the absence of any information about the size and physical properties of the protein, or the fraction of the total protein mass it represents in the extract. For proteins that are enzymes, the amount in a given solution or tissue extract can be measured or assayed in terms of the catalytic effect the enzyme produces, that is, the *increase* in the rate at which its substrate is converted to reaction products when the enzyme is present. For this purpose one must know (1) the over-

all equation of the reaction catalyzed, (2) an analytical procedure for determining the disappearance of the substrate or the appearance of a reaction product, (3) whether the enzyme requires cofactors such as metal ions or coenzymes, (4) the dependence of the enzyme activity on substrate concentration, (5) the optimum pH, and (6) a temperature zone in which the enzyme is stable and has high activity. Enzymes are usually assayed at their optimum pH and at some convenient temperature within the range 25 to 38 °C. Also, very high substrate concentrations are generally required so that the initial reaction rate, measured experimentally, is proportional to enzyme concentration (Chapter 8).

By international agreement, 1.0 unit of enzyme activity is defined as the amount of enzyme causing transformation of 1.0 μmol of substrate per minute at 25 °C under optimal conditions of measurement. The term **activity** refers to the total units of enzyme in a solution. The **specific activity** is the number of enzyme units per milligram of total protein (Fig. 5–23). The specific activity is a measure of enzyme purity: it increases during purification of an enzyme and becomes maximal and constant when the enzyme is pure (Table 5–5).

After each purification step, the activity of the preparation (in units) is assayed, the total amount of protein is determined independently, and their ratio gives the specific activity. Activity and total protein generally decrease with each step. Activity decreases because some loss always occurs due to inactivation or nonideal interactions with chromatographic materials or other molecules in the solution. Total protein decreases because the objective is to remove as much unwanted or nonspecific protein as possible. In a successful step, the loss of nonspecific protein is much greater than the loss of activity; therefore, specific activity increases even as total activity falls. The data are then assembled in a purification table similar to Table 5–5. A protein is generally considered pure when further purification steps fail to increase specific activity and when only a single protein species can be detected (for example, by electrophoresis).

For proteins that are not enzymes, other quantification methods are required. Transport proteins can be assayed by their binding to the molecule they transport, and hormones and toxins by the biological effect they produce; for example, growth hormones will stimulate the growth of certain cultured cells. Some structural proteins represent such a large fraction of a tissue mass that they can be readily extracted and purified without a functional assay. The approaches are as varied as the proteins themselves.

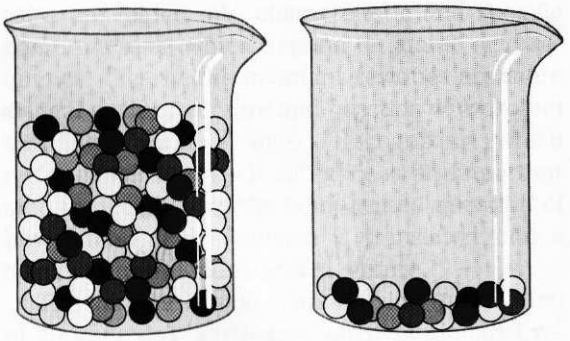
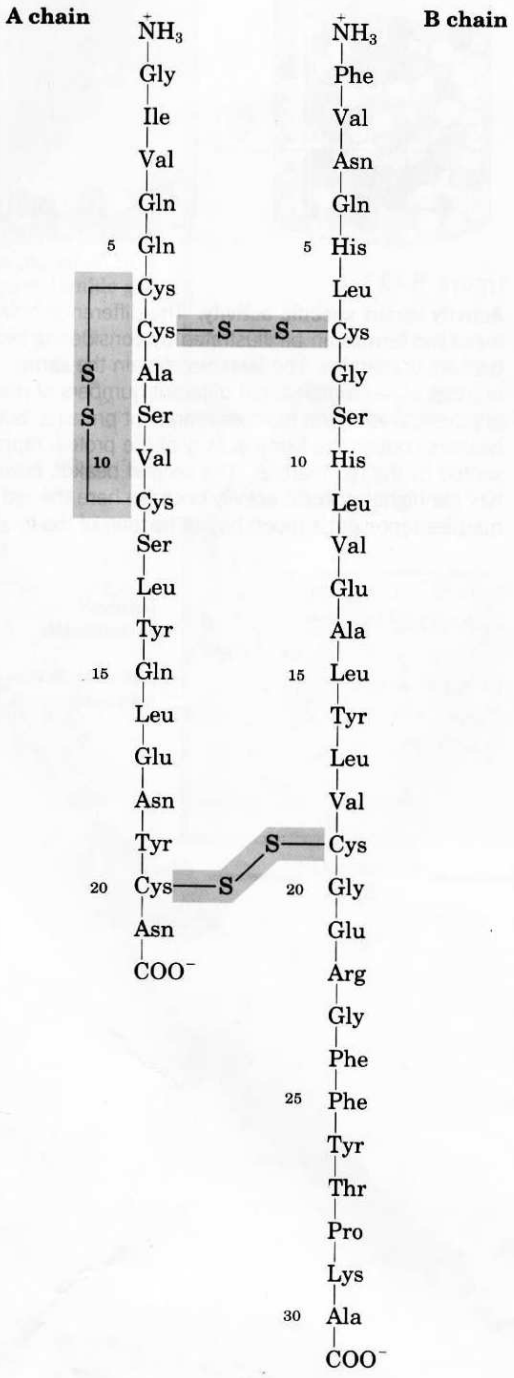


figure 5–23
Activity versus specific activity. The difference between these two terms can be illustrated by considering two beakers of marbles. The beakers contain the same number of red marbles, but different numbers of marbles of other colors. If the marbles represent proteins, both beakers contain the same *activity* of the protein represented by the red marbles. The second beaker, however, has the higher *specific activity* because here the red marbles represent a much higher fraction of the total.

The Covalent Structure of Proteins

Purification of a protein is usually only a prelude to a detailed biochemical dissection of its structure and function. What is it that makes one protein an enzyme, another a hormone, another a structural protein, and still another an antibody? How do they differ chemically? The most obvious distinctions are structural, and these distinctions can be approached at every level of structure defined in Figure 5–16.

The differences in primary structure can be especially informative. Each protein has a distinctive number and sequence of amino acid residues. As we shall see in Chapter 6, the primary structure of a protein determines how it folds up into a unique three-dimensional structure, and this in turn determines the function of the protein. Primary structure now becomes the focus of the remainder of the chapter. We first consider empirical clues that amino acid sequence and protein function are closely linked, then describe how amino acid sequence is determined, and finally outline the many uses to which this information can be put.



The Function of a Protein Depends on Its Amino Acid Sequence

The bacterium *E. coli* produces more than 3,000 different proteins; a human being produces 50,000 to 100,000. In both cases, each type of protein has a unique three-dimensional structure and this structure confers a unique function. Each type of protein also has a unique amino acid sequence. Intuition suggests that the amino acid sequence must play a fundamental role in determining the three-dimensional structure of the protein, and ultimately its function, but is this expectation correct? A quick survey of proteins and how they vary in amino acid sequence provides a number of empirical clues that help substantiate the important relationship between amino acid sequence and biological function. First, as we have already noted, proteins with different functions always have different amino acid sequences. Second, thousands of human genetic diseases have been traced to the production of defective proteins. Perhaps one-third of these proteins are defective because of a single change in their amino acid sequence; hence, if the primary structure is altered, the function of the protein may also be changed. Finally, on comparing functionally similar proteins from different species, we find that these proteins often have similar amino acid sequences (Box 5–2). An extreme case is ubiquitin, a 76-residue protein involved in regulating the degradation of other proteins. The amino acid sequence of ubiquitin is identical in species as disparate as fruit flies and humans.

Is the amino acid sequence absolutely fixed, or invariant, for a particular protein? No; some flexibility is possible. An estimated 20% to 30% of the proteins in humans are **polymorphic,** having amino acid sequence variants in the human population. Many of these variations in sequence have little or no effect on the function of the protein. Furthermore, proteins that carry out a broadly similar function in distantly related species can differ greatly in overall size and amino acid sequence.

Proteins often contain crucial regions within their amino acid sequence that are essential to their biological functions. The amino acid sequence in other regions might vary considerably without affecting these functions. The fraction of the sequence that is critical varies from protein to protein, complicating the task of relating sequence to three-dimensional structure, and structure to function. Before we can consider this problem further, however, we must examine how sequence information is obtained.

The Amino Acid Sequences of Numerous Proteins Have Been Determined

Two major discoveries in 1953 were of crucial importance in the history of biochemistry. In that year James D. Watson and Francis Crick deduced the double-helical structure of DNA and proposed a structural basis for its precise replication (Chapter 10). Their proposal illuminated the molecular reality behind the idea of a gene. In that same year, Frederick Sanger worked out the sequence of amino acid residues in the polypeptide chains of the hormone insulin (Fig. 5–24), surprising many researchers who had long thought that elucidation of the amino acid sequence of a polypeptide would be a hopelessly difficult task. It quickly became evident that the nucleotide

figure 5–24

Amino acid sequence of bovine insulin. The two polypeptide chains are joined by disulfide cross-linkages. The A chain is identical in human, pig, dog, rabbit, and sperm whale insulins. The B chains of the cow, pig, dog, goat, and horse are identical. Such identities between similar proteins of different species are discussed in Box 5–2.

box 5–2 Protein Homology among Species

Homologous proteins are proteins that are evolutionarily related. They usually perform the same function in different species; an example is **cytochrome *c*,** an iron-containing mitochondrial protein that transfers electrons during biological oxidations in eukaryotic cells. Homologous proteins from different species may have polypeptide chains that are identical or nearly identical in length. Many positions in the amino acid sequence are occupied by the same residue in all species and are thus called **invariant residues.** Other positions show considerable variation in the amino acid residue from one species to another; these are called **variable residues.**

The functional significance of sequence homology is well illustrated by cytochrome *c* (M_r ~13,000), which has about 100 amino acid residues in most species. The amino acid sequences of cytochrome *c* molecules from many different species have been determined, and 27 positions in the chain are invariant in all species tested (Fig. 1), suggesting that they are the most important residues specifying the biological activity of this protein. The residues in other positions exhibit some interspecies variation. There are clear gradations in the number of differences observed in the variable residues. In some positions, most substitutions involve similar amino acid residues (for example, positively charged Arg might replace positively charged Lys); these are called **conservative substitutions.** At other positions the substitutions are less restricted (nonconservative). As we will show in the next chapter, the polypeptide chains of proteins are folded into characteristic and specific conformations, which depend on amino acid sequence. Clearly, the invariant residues are more critical to the structure and function of a protein than the variable ones. Recognizing which residues fall into each category is an important step in deciphering the complicated question of how amino acid sequence is translated into a specific three-dimensional structure.

The variable amino acids provide information of another sort. Phylogenetic (evolutionary) relationships based on taxonomic methods have been tested and experimentally confirmed through biochemistry. The examination of sequences of cytochrome *c* and other homologous proteins has led to an important conclusion: the number of residues that differ in homologous proteins from any two species is in proportion to the phylogenetic difference between those species. For example, 48 amino acid residues differ in the cytochrome *c* molecules of the horse and of yeast, which are very widely separated species, whereas only two residues differ in the cytochrome *c* molecules of the much more closely related duck and chicken. In fact, cytochrome *c* has identical amino acid sequences in the chicken and the turkey, and in the pig, cow,

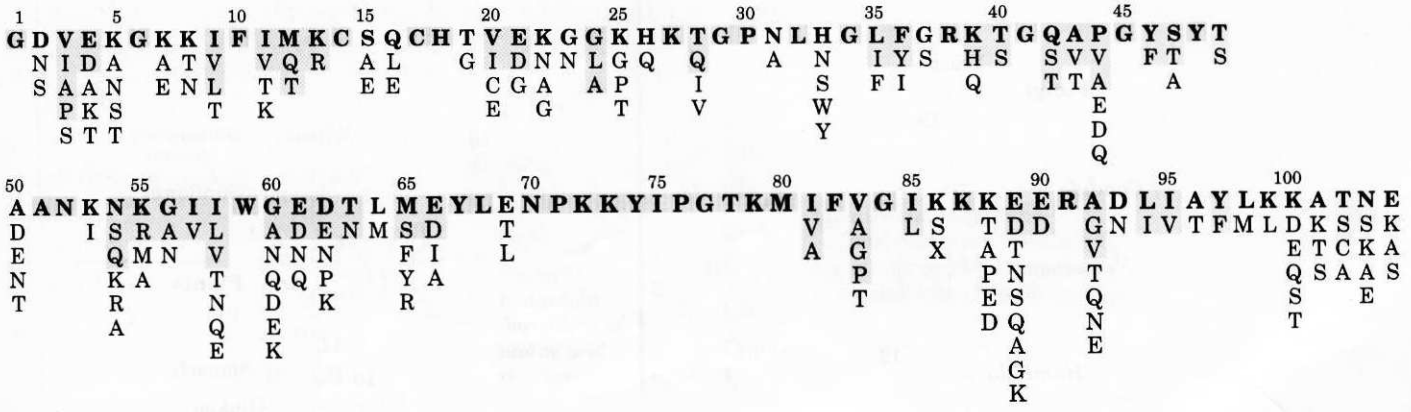


figure 1
Amino acid sequence of human cytochrome *c*. Amino acid substitutions in other species are listed below the individual residues. Invariant amino acids are shaded in yellow, conservative substitutions are shaded in blue, and nonconservative substitutions are unshaded. X is an unusual amino acid, trimethyllysine. The one-letter abbreviations for amino acids are given in Table 5–1.

Continued on next page