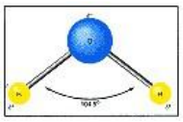
Chapter 1—
Eukaryotic Cell Structure

Thomas M. Devlin



1.1— Overview: Cells and Cellular Compartments

Over three billion years ago, under conditions not entirely clear and in a time span difficult to comprehend, elements such as carbon, hydrogen, oxygen, nitrogen, sulfur, and phosphorus formed simple chemical compounds. They combined, dispersed, and recombined to form a variety of larger molecules until a combination was achieved that was capable of replicating itself. These macromolecules consisted of simpler molecules linked together by chemical bonds. With continued evolution and formation of ever more complex molecules, the water environment around some of these self-replicating molecules became enclosed by a membrane. This development gave these primordial structures the ability to control their own environment to some extent. A form of life had evolved and a unit of three-dimensional space—a cell—had been established. With the passing of time a diversity of cells evolved, and their chemistry and structure became more complex. They could extract nutrients from the environment, chemically converting these nutrients to sources of energy or to complex molecules, control chemical processes that they catalyzed, and carry out cellular replication. Thus the vast diversity of life observed today began. The cell is the basic unit of life in all forms of living organisms, from the smallest bacterium to the most complex animal.

The limiting outer membrane of cells, the **plasma membrane,** delineates the space occupied by a cell and separates the variable and potentially hostile environment outside from the relatively constant milieu within. It is the communication link between the cell and its surroundings.

On the basis of microscopic and biochemical differences, living cells are divided into two major classes: **prokaryotes,** which include bacteria, blue-green algae, and rickettsiae, and **eukaryotes,** which include yeasts, fungi, and plant and animal cells. Prokaryotes have a variety of shapes and sizes, in most cases being 1/1000 to 1/10,000 the size of eukaryotic cells. They lack intracellular membrane-bound structures that can be visualized by a microscope (Figure 1.1). The deoxyribonucleic acid (DNA) of prokaryotes is often segregated into a discrete mass, the nucleoid region, that is not surrounded by a membrane or envelope. The plasma membrane is often invaginated. In contrast, eukaryotic cells have a well-defined membrane surrounding a central nucleus and a variety of intracellular structures and organelles (Figure 1.1*b*). Intracellular membrane systems establish distinct subcellular compartments, as described in Section 1.4, that permit a unique degree of subcellular specialization. By compartmentalization different chemical reactions that require different environments can occur simultaneously. Many reactions occur in or on specific membranes, thus creating an additional environment for the diverse functions of cells.

Besides these structural variations between prokaryotic and eukaryotic cells (Figures 1*a* and 1*b*), there are differences in chemical composition and biochemical activities. Prokaryotes lack histones, a class of proteins that complex with DNA in eukaryotes. There are major structural differences in the ribonucleic acid–protein complexes involved in biosynthesis of proteins between the cell types, as well as differences in transport mechanisms across the plasma membrane and in enzyme content. The many similarities, however, are equally striking. The emphasis throughout this book is on the chemistry of eukaryotes, particularly mammalian cells, but much of our knowledge of the biochemistry of living cells has come from studies of prokaryotic and nonmammalian eukaryotic cells. The basic chemical components and fundamental chemical reactions of all living cells are very similar. Availability of certain cell populations, for example, bacteria in contrast to human liver, has led to much of our knowledge about some cells; in some areas our knowledge is derived nearly exclusively from studies of prokaryotes. The universality of many biochemical phenomena, however, permits many extrapolations from bacteria to humans.

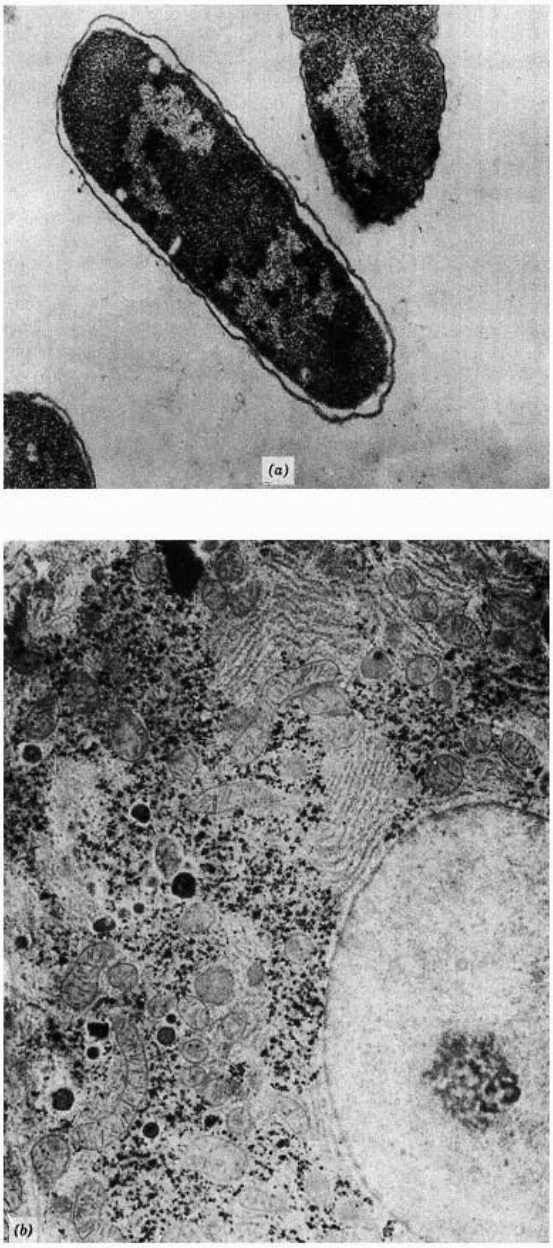


Figure 1.1
Cellular organization of prokaryotic and eukaryotic cells.
(a) Electron micrograph of *Escherichia coli*, a representative prokaryote; approximate magnification ×30,000. There is little apparent intracellular organization and no cytoplasmic organelles. Chromatin is condensed in a nuclear zone but not surrounded by a membrane. Prokaryotic cells are much smaller than eukaryotic cells. (b) Electron micrograph of a thin section of a liver cell (rat hepatocyte), a representative eukaryotic cell; approximate magnification ×7500. Note the distinct nuclear membrane, different membrane-bound organelles or vesicles, and extensive membrane systems. Various membranes create a variety of intracellular compartments. Photograph (a) generously supplied by Dr. M. E. Bayer, Fox Chase Cancer Institute, Philadelphia, PA; photograph (b) reprinted with permission of Dr. K. R. Porter, from Porter, K. R., and Bonneville, M. A. In: *Fine Structure of Cells and Tissues*. Philadelphia: Lea & Febiger, 1972.

Before we dissect the complexities of mammalian cells and tissues in the following chapters, it is appropriate to review some of the chemical and physical characteristics of the environment in which the various biochemical phenomena occur. This environment places many constraints on the cell's activities. The concluding section outlines the activities and roles of subcellular compartments.

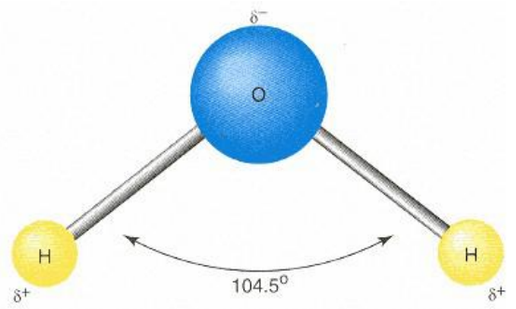


Figure 1.2
Structure of a water molecule.
The H–O–H bond angle is 104.5°. Both hydrogen atoms carry a partial positive charge and the oxygen a partial negative charge, creating a dipole.

1.2— Cellular Environment: Water and Solutes

All biological cells contain essentially the same building blocks and types of macromolecules. The general classes of substances in cells are presented in Table 1.1. There are significant variations in concentration of specific components in different cell types and in organelles of eukaryotic cells. **Microenvironments** are also created by macromolecules and membranes in which the composition differs from that of the surrounding milieu. Cells depend on the external environment for nutrients required for replacement of components, growth, and energy needs. They have a variety of mechanisms to cope with variations in composition of the external environment. **Water** is the one common component of all environments. It is the solvent in which the substances required for the cell's existence are dissolved or suspended. The unique physicochemical properties of water make life on earth possible.

Hydrogen Bonds Form between Water Molecules

Two hydrogen atoms share their electrons with an unshared pair of electrons of an oxygen atom to form a water molecule. The oxygen nucleus has a stronger attraction for shared electrons than hydrogen, and positively charged hydrogen nuclei are left with an unequal share of electrons, creating a partial positive charge on each hydrogen and a partial negative charge on oxygen. The bond angle between hydrogens and oxygen is 104.5°, making the molecule electrically asymmetric and producing an electric dipole (Figure 1.2). Water molecules interact because positively charged hydrogen atoms on one molecule are attracted to the negatively charged oxygen atom on another, with formation of a weak bond between two water molecules (Figure 1.3*a*). This bond, indicated by a dashed line, is a **hydrogen bond.** A detailed discussion of noncovalent interactions between molecules, including electrostatic, van der Waals, and hydrophobic, is presented on page 64. Five molecules of water form a tetrahedral structure (Figure 1.3*b*), because each oxygen shares its electrons with four hydrogen atoms and each hydrogen with another oxygen. A tetrahedral lattice structure is formed in ice and gives ice its crystalline structure. Some hydrogen bonds are broken as ice is transformed to liquid water. Each bond is relatively

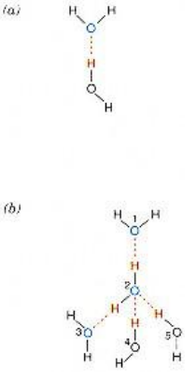


Figure 1.3
Hydrogen bonding.
(a) Hydrogen bonding, indicated by dashed lines, between two water molecules.
(b) Tetrahedral hydrogen bonding of five water molecules. Water molecules 1, 2, and 3 are in the plane of the page, 4 is below, and 5 is above.

TABLE 1.1 Chemical Components of Biological Cells

Component	*Range of Molecular Weights*
H_2O	18
Inorganic ions Na^+, K^+, Cl^-, SO_4^{2-}, HCO_3^- Ca^{2+}, Mg^{2+}, etc.	23–100
Small organic molecules	100–1200
Carbohydrates, amino acids, lipids, nucleotides, peptides	
Macromolecules	50,000–1,000,000,000
Proteins, polysaccharides, nucleic acids	

weak compared to a covalent bond but the large number of hydrogen bonds between molecules in liquid water is the reason for the stability of water. Liquid water actually has a definite structure due to hydrogen bonding that is in a dynamic state as these bonds break and reform. Hydrogen bonds in water have a half-life of less than 1×10^{-10} s. Liquid water contains a significant number of hydrogen bonds even at 100°C, which accounts for its high heat of vaporization; in the transformation from liquid to vapor state, hydrogen bonds are disrupted.

Water molecules hydrogen bond to different chemical structures. Hydrogen bonding also occurs between other molecules and within a molecule wherever electronegative oxygen or nitrogen comes in close proximity to hydrogen covalently bonded to another electronegative atom. Representative hydrogen bonds are presented in Figure 1.4. Intramolecular hydrogen bonding occurs extensively in large macromolecules such as proteins and nucleic acids and is partially responsible for their structural stability.

Many models for the structure of liquid water have been proposed, but none adequately explains all its properties.

Water Has Unique Solvent Properties

The polar nature and ability to form hydrogen bonds are the basis for the unique **solvent properties** of water. Polar molecules are readily dispersed in water. **Salts** in which a crystal lattice is held together by attraction of positive and negative groups dissolve in water because electrostatic forces in the crystal can be overcome by attraction of charges to the dipole of water. NaCl is an example where electrostatic attraction of individual Na^+ and Cl^- atoms is overcome by interaction of Na^+ with the negative charge on oxygen atoms, and Cl^- with positive charges on the hydrogen atoms. Thus a shell of water surrounds the individual ions. The number of weak charge–charge interactions between water and Na^+ and Cl^- ions is sufficient to separate the two charged ions.

Many organic molecules that contain nonionic but weakly polar groups are soluble in water because of attraction of these groups to molecules of water. Sugars and alcohols are readily soluble in water for this reason. **Amphipathic** molecules, compounds that contain both polar and nonpolar groups, disperse in water if attraction of the polar group for water can overcome hydrophobic interactions of nonpolar portions of the molecules. Very hydrophobic molecules, such as compounds that contain long hydrocarbon chains, however, do not readily disperse as single molecules in water but interact with one another to exclude the polar water molecules.

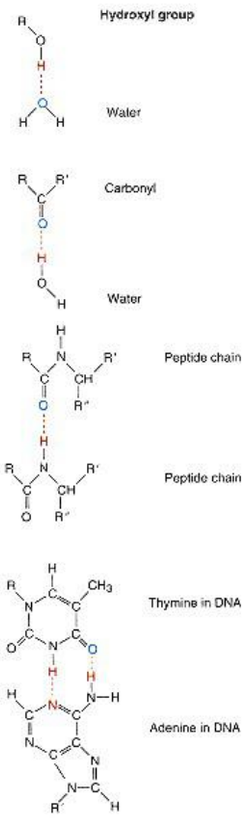


Figure 1.4
Representative hydrogen bonds of importance in biological systems.

Some Molecules Dissociate with Formation of Cations and Anions

Substances that dissociate in water into a **cation** (positively charged ion) and an **anion** (negatively charged ion) are classified as **electrolytes**. The presence of charged ions facilitates conductance of an electrical current through an aqueous solution. Sugars or alcohols, which readily dissolve in water but do not carry a charge or dissociate into charged species, are classified as **nonelectrolytes.**

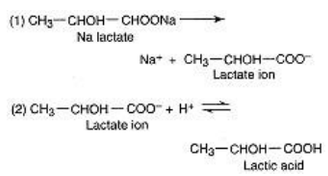


Figure 1.5
Reactions that occur when sodium lactate is dissolved in water.

Salts of alkali metals (e.g., Li, Na, and K), dissolved in water at low concentrations, dissociate completely; at high concentrations, however, there is increased potential for interaction of anions and cations. With biological systems it is customary to consider such compounds as totally dissociated because their concentrations are low. Salts of organic acids, for example, sodium lactate, also dissociate totally and are classified as electrolytes; the dissociated anion, lactate ion, reacts to a limited extent with a proton to form undissociated acid (Figure 1.5). When such salts are dissolved in water, individual ions are present in solution rather than the undissociated salt. If a solution has been prepared with

several different salts (e.g., NaCl, K_2SO_4, and Na lactate), the original molecules do not exist as such in solution, only the ions (e.g., Na^+, K^+, $SO_{4^{2-}}$ and lactate$^-$).

Many acids, however, when dissolved in water do not totally dissociate but rather establish an equilibrium between undissociated and dissociated components. Thus lactic acid, an important metabolic intermediate, partially dissociates into lactate anions and H^+ as follows:

Because of their partial dissociation, however, such compounds have a lower capacity to carry an electrical charge on a molar basis when compared to those that dissociate totally; they are termed **weak electrolytes.**

Weak Electrolytes Dissociate Partially

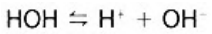
In partial dissociation of a weak electrolyte, represented by HA, the concentration of the various species can be determined from the equilibrium equation:

$$K'_{eq} = \frac{[H^+][A^-]}{[HA]}$$

A^- represents the dissociated anion and square brackets indicate concentration of each component in concentration units such as moles per liter (mol L^{-1}) or millimol L^{-1}. The **activity** of each species rather than concentration should be employed in the equilibrium equation but since most compounds of interest in biological systems are present in low concentration, the value for activity approaches that of concentration. Thus the equilibrium constant is indicated as K'_{eq} cannot be determined because at equilibrium there is no remaining undissociated solute.

Water Is a Weak Electrolyte

Water dissociates as follows:



A proton that dissociates interacts with oxygen of another water molecule to form the hydronium ion, H_3O^+. For convenience, in this book the proton will be presented as H^+ rather than H_3O^+, even though the latter is the actual chemical species. At 25°C the value of K'_{eq} for dissociation of water is very small and is about 1.8×10^{-16}:

$$K'_{eq} = 1.8 \times 10^{-16} = \frac{[H^+][OH^-]}{[H_2O]} \qquad \textbf{(1.1)}$$

With such a small K'_{eq} an insignificant number of water molecules actually dissociate relative to the number of undissociated molecules. Thus the concentration of water, which is 55.5 M, is essentially unchanged. Equation 1.1 can be rewritten as follows:

$$K'_{eq} \times [H_2O] = [H^+][OH^-] \qquad \textbf{(1.2)}$$

$K'_{eq} \times [55.5]$ is a constant and is termed the **ion product of water.** Its value at 25°C is 1×10^{-14}. In pure water the concentration of H^+ equals OH^-, and by substituting $[H^+]$ for $[OH^-]$ in the equation above, $[H^+]$ is 1×10^{-7} M. Similarly,

[OH^-] is also 1×10^{-7} M. The equilibrium of H_2O, H^+, and OH^- always exists in dilute solutions regardless of the presence of dissolved substances. If dissolved material alters either the H^+ or OH^- concentration, as occurs on addition of an acid or base, a concomitant change in the other ion must occur in order to satisfy the equilibrium relationship. By using the equation for the ion product, [H^+] or [OH^-] can be calculated if concentration of one of the ions is known.

TABLE 1.2 Relationships Between [H^+] and pH and [OH^-] and pOH

[H^+] (*M*)	*pH*	[OH^-] (*M*)	*pOH*
1.0	0	1×10^{-14}	14
0.1 (1×10^{-1})	1	1×10^{-13}	13
1×10^{-2}	2	1×10^{-12}	12
1×10^{-3}	3	1×10^{-11}	11
1×10^{-4}	4	1×10^{-10}	10
1×10^{-5}	5	1×10^{-9}	9
1×10^{-6}	6	1×10^{-8}	8
1×10^{-7}	7	1×10^{-7}	7
1×10^{-8}	8	1×10^{-6}	6
1×10^{-9}	9	1×10^{-5}	5
1×10^{-10}	10	1×10^{-4}	4
1×10^{-11}	11	1×10^{-3}	3
1×10^{-12}	12	1×10^{-2}	2
1×10^{-13}	13	0.1 (1×10^{-1})	1
1×10^{-14}	14	1.0	0

The importance of hydrogen ions in biological systems will become apparent in subsequent chapters. For convenience [H^+] is usually expressed in terms of **pH,** calculated as follows:

$$\text{pH} = \log \frac{1}{[\text{H}^+]} \qquad (1.3)$$

In pure water [H^+] and [OH^-] are both 1×10^{-7} M, and pH = 7.0. [OH^-] is expressed as the pOH. For the equation describing dissociation of water, 1×10^{-14} = [H^+][OH^-]; taking negative logarithms of both sides, the equation becomes 14 = pH + pOH. Table 1.2 presents the relationship between pH and [H^+].

The pH values of different **biological fluids** are presented in Table 1.3. In blood plasma, [H^+] is 0.00000004 M or a pH of 7.4. Other cations are between 0.001 and 0.10 M, well over 10,000 times higher than [H^+]. An increase in hydrogen ion to 0.0000001 M (pH 7.0) leads to serious medical consequences and is life threatening; a detailed discussion of mechanisms by which the body maintains intra- and extracellular pH is presented in Chapter 25.

Many Biologically Important Molecules Are Acids or Bases

The definitions of an acid and a base proposed by Lowry and Brønsted are most convenient in considering biological systems. An **acid** is a **proton donor** and a **base** is a **proton acceptor.** Hydrochloric acid (HCl) and sulfuric acid (H_2SO_4) are strong acids because they dissociate totally, and OH^- ion is a base because it accepts a proton, shifting the equilibrium

$$OH^- + H^+ \leftrightharpoons H_2O$$

When a strong acid and OH^- are combined, H^+ from the acid and OH^- interact and are in equilibrium with H_2O. Neutralization of H^+ and OH^- occurs because the ion product for water is so small.

Anions produced when strong acids dissociate totally, such as Cl^- from HCl, are not bases because they do not associate with protons in solution. When an organic acid, such as lactic acid, is dissolved in water it dissociates only partially, establishing an equilibrium between an acid (proton donor), an anion of the acid, and a proton as follows:

$$\text{Lactic acid} \leftrightharpoons \text{lactate}^- + H^+$$

Lactic acid is a **weak acid.** The anion is a base because it accepts a proton and reforms the acid. The weak acid and the base formed on dissociation are referred to as a **conjugate pair;** other examples are presented in Table 1.4. Ammonium ion (NH_4^+) is an acid because it dissociates to yield H^+ and ammonia (NH_3), an uncharged species, which is a **conjugate base.** Phosphoric acid (H_3PO_4) is an acid and PO_4^{3-} is a base, but $H_2PO_4^-$ and HPO_4^{2-} are either a base or acid depending on whether the phosphate group is accepting or donating a proton.

TABLE 1.3 pH of Some Biological Fluids

Fluid	*pH*
Blood plasma	7.4
Interstitial fluid	7.4
Intracellular fluid	
Cytosol (liver)	6.9
Lysosomal matrix	Below 5.0
Gastric juice	1.5–3.0
Pancreatic juice	7.8–8.0
Human milk	7.4
Saliva	6.4–7.0
Urine	5.0–8.0

The tendency of a conjugate acid to dissociate H^+ can be evaluated from the K'_{eq} of 1×10^{-14} at 25°C.

A convenient method of stating the K'_{eq} is in the form of **p*K′*,** as

$$pK' = \log \frac{1}{K'_{eq}} \qquad (1.4)$$

TABLE 1.4 Some Conjugate Acid–Base Pairs of Importance in Biological Systems

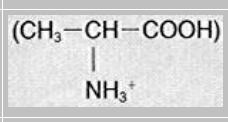
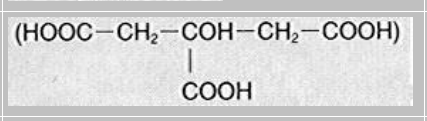
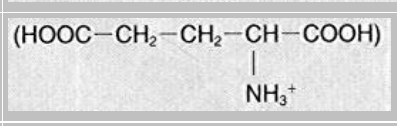
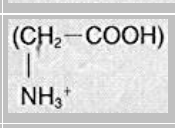
Proton Donor (*Acid*)		*Proton Acceptor* (*Base*)
$CH_3–CHOH–COOH$	⇌	$H^+ + CH_3–CHOH–COO^-$
(lactic acid)		(lactate)
$CH_3–CO–COOH$	⇌	$H^+ + CH_3–CO–COO^-$
(pyruvic acid)		(pyruvate)
$HOOC–CH_2–CH_2–COOH$	⇌	$2H^+ + {}^-OOC–CH_2–CH_2–COO^-$
(succinic acid)		(succinate)
$^+H_3NCH_2–COOH$	⇌	$H^+ + {}^+H_3N–CH_2–COO^-$
(glycine)		(glycinate)
H_3PO_4	⇌	$H^+ + H_2PO_{4-}$
H_2PO_{4-}	⇌	$H^+ + HPO_{4\,2-}$
HPO_4^{2-}	⇌	$H^+ + PO_4^{3-}$
Glucose 6-PO_3H^-	⇌	H^+ + glucose 6-$PO_{3\,2-}$
H_2CO_3	⇌	$H^+ + HCO_{3-}$
NH_4^+	⇌	$H^+ + NH_3$
H_2O	⇌	$H^+ + OH^-$

Note the similarity of this definition with that of pH; as with pH and [H^+], the relationship between pK and K'_{eq} and pK for conjugate acids of importance in biological systems are presented in Table 1.5.

A special case of a weak acid important in medicine is **carbonic acid** (H_2CO_3). Carbon dioxide when dissolved in water is involved in the following equilibrium reactions:

$$CO_2 + H_2O \underset{}{\overset{K'_2}{\rightleftharpoons}} H_2CO_3 \overset{K'_1}{\rightleftharpoons} H^+ + HCO_3^-$$

TABLE 1.5 Apparent Dissociation Constant and pK' of Some Compounds of Importance in Biochemistry

Compound		K'_{eq} (*M*)	*p* K′
Acetic acid	(CH_3—COOH)	1.74×10^{-5}	4.76
Alanine		4.57×10^{-3} 2.04×10^{-10}	2.34 (COOH) 9.69 (NH_{3+})
Citric acid		8.12×10^{-4} 1.77×10^{-5} 3.89×10^{-6}	3.09 3.74 5.41
Glutamic acid		6.45×10^{-3} 5.62×10^{-5} 2.14×10^{-10}	2.19 (COOH) 4.25 (COOH) 9.67 (NH_{3+})
Glycine		4.57×10^{-3} 2.51×10^{-10}	2.34 (COOH) 9.60 (NH_{3+})
Lactic acid	(CH_3—CHOH—COOH)	1.38×10^{-4}	3.86
Pyruvic acid	(CH_3—CO—COOH)	3.16×10^{-3}	2.50
Succinic acid	(HOOC—CH_2—CH_2—COOH)	6.46×10^{-5}	4.19
		3.31×10^{-6}	5.48
Glucose 6-PO_3H^-		7.76×10^{-7}	6.11
H_3PO_4		1×10^{-2}	2.0
H_2PO_{4-}		2.0×10^{-7}	6.7
HPO_4^{2-}		3.4×10^{-13}	12.5
H_2CO_3		1.70×10^{-4}	3.77
NH_4^+		5.62×10^{-10}	9.25
H_2O		1×10^{-14}	14.0

Carbonic acid is a relatively strong acid with a pK_1' of 3.77. The equilibrium equation for this reaction is

$$K_1' = \frac{[H^+][HCO_3^-]}{[H_2CO_3]} \qquad (1.5)$$

Carbonic acid is, however, in equilibrium with dissolved CO_2 and the equilibrium equation for this reaction is

$$K_2' = \frac{[H_2CO_3]}{[CO_2][H_2O]} \qquad (1.6)$$

Solving Eq. 1.6 for H_2CO_3 and substituting for the H_2CO_3 in Eq. 1.5, the two equilibrium reactions are combined into one equation:

$$K_1' = \frac{[H^+][HCO_3^-]}{K_2'[CO_2][H_2O]} \qquad (1.7)$$

Rearranging to combine constants, including the concentration of H_2O, simplifies the equation and yields a new combined constant, K_3', as follows:

$$K_1'K_2'[H_2O] = K_3' = \frac{[H^+][HCO_3^-]}{[CO_2]} \qquad (1.8)$$

It is common practice to refer to dissolved CO_2 as a conjugate acid; it is the acid anhydride of H_2CO_3. The term K_3' has a value of 7.95×10^{-7} and [illegible]. If the aqueous system is in contact with an air phase, dissolved CO_2 will also be in equilibrium with CO_2 in the air phase. A decrease or increase of one component—that is, CO_2 (air), CO_2 (dissolved), H_2CO_3, H^+ or [illegible] —will cause a change in all the other components.

The Henderson–Hasselbalch Equation Defines the Relationship between pH and Concentrations of Conjugate Acid and Base

A change in concentration of any component in an equilibrium reaction necessitates a concomitant change in every component. For example, an increase in $[H^+]$ will decrease the concentration of conjugate base (e.g., lactate ion) with an equivalent increase in the conjugate acid (e.g., lactic acid). This relationship is conveniently expressed by rearranging the equilibrium equation and solving for H^+, as shown for the following dissociation:

$$\text{Conjugate acid} \leftrightharpoons \text{conjugate base} + H^+$$

$$K_{eq}' = \frac{[H^+][\text{conjugate base}]}{[\text{conjugate acid}]} \qquad (1.9)$$

Rearranging Eq. 1.9 by dividing through by both $[H^+]$ and K_{eq}' leads to

$$\frac{1}{[H^+]} = \frac{1}{[K_{eq}']} \cdot \frac{[\text{conjugate base}]}{[\text{conjugate acid}]} \qquad (1.10)$$

Taking the logarithm of both sides gives

$$\log \frac{1}{[H^+]} = \log \frac{1}{K_{eq}'} + \log \frac{[\text{conjugate base}]}{[\text{conjugate acid}]} \qquad (1.11)$$

Since pH = log $1/[H^+]$ and $pK' = \log 1/K_{eq}'$, Eq. 1.11 becomes

$$\mathbf{pH = pK' + \log \frac{[conjugate\ base]}{[conjugate\ acid]}} \qquad (1.12)$$

Equation 1.12, developed by Henderson and Hasselbalch, is a convenient way of viewing the relationship between pH of a solution and relative amounts of conjugate base and acid present. Analysis of Eq. 1.12 demonstrates that when the ratio of [base]/[acid] is 1 : 1, pH equals the pK of the acid because log 1 = 0,

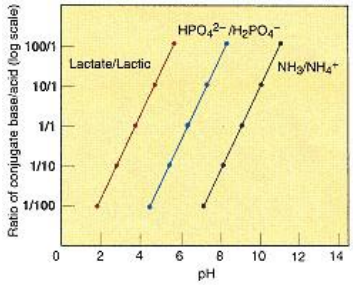


Figure 1.6
Ratio of conjugate [base]/[acid] as a function of the pH.
When the ratio of [base]/[acid] is 1, pH equals p*K* of weak acid.

and thus pH = pK . If pH is one unit less than p*K* , the [base]/[acid] ratio is 1 : 10, and if pH is one unit above p*K* , the [base]/[acid] ratio is 10 : 1. Figure 1.6 is a plot of ratios of conjugate base to conjugate acid versus pH of several weak acids; note that ratios are presented on a logarithmic scale.

Buffering Is Important to Control pH

When NaOH is added to a solution of a weak acid such as lactic acid, the ratio of [conjugate base]/[conjugate acid] changes. NaOH dissociates totally and the OH^- formed is neutralized by existing H^+ to form H_2O. The decrease in $[H^+]$ will cause further dissociation of weak acid to comply with requirements of its equilibrium reaction. The amount of weak acid dissociated will be so nearly equal to the amount of OH^- added that it is considered to be equal. Thus the decrease in amount of conjugate acid is equal to the amount of conjugate base that is formed. These series of events are represented in titration curves of two weak acids presented in Figure 1.7. When 0.5 equiv of OH^- is added, 50% of the weak acid is dissociated and the [acid]/[base] ratio is 1.0; pH at this point is equal to p*K* of the acid. Shapes of individual titration curves are similar but displaced due to differences in p*K* values. There is a rather steep rise in pH when only 0.1 equiv of OH^- are added, but between 0.1 and 0.9 equiv of added OH^-, the pH change is only ~2. Thus a large amount of OH^- is added with a relatively small change in pH. This is called **buffering** and is defined as the ability of a solution to resist a change in pH when an acid or base is added. If weak acid were not present, the pH would be very high with only a small amount of OH^- because there would be no source of H^+ to neutralize the OH^-.

The best buffering range for a conjugate pair is in the pH range near the p*K* of the weak acid. Starting from a pH one unit below to a pH one unit above p*K* , ~82% of a weak acid in solution will dissociate, and therefore an amount of base equivalent to about 82% of original acid can be neutralized with a change in pH of 2. The maximum buffering range for a conjugate pair is considered to be between 1 pH unit above and below the p*K* . Lactic acid with p*K* = 3.86 is an effective buffer in the range of pH 3 to 5 but has no buffering capacity at pH = 7.0. The $HPO_4^{2-}/H_2PO_4^-$ pair with p*K* = 6.7, however, is an effective buffer at pH = 7.0. Thus at the pH of the cell's cytosol (~7.0), the lactate–lactic acid pair is not an effective buffer but the phosphate system is.

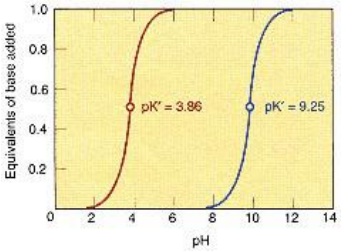


Figure 1.7
Acid–base titration curves for lactic acid (p*K′* 3.86) and NH_4^+ (p*K′* 9.25).
At pH equal to respective p*K* values, there will be an equal amount of acid and base for each conjugate pair.

Buffering capacity also depends on the concentrations of conjugate acid and base. The higher the concentration of conjugate base, the more added H^+ with which it can react. The more conjugate acid the more added OH^- can be

neutralized by the dissociation of the acid. A case in point is blood plasma at pH 7.4. For $HPO_4^{2-}/H_2PO_4^-$ the pK of 6.7 would suggest that this conjugate pair would be an effective buffer; the concentration of the phosphate pair, however, is low compared to that of the HCO_3^-/CO_2 system with a pK of 6.1, which is present at a 20-fold higher concentration and accounts for most of the buffering capacity. In considering the buffering capacity both the pK and the concentration of the conjugate pair must be taken into account. Most organic acids are relatively unimportant as buffers in cellular fluids because their pK values are more than several pH units lower than the pH of the cell, and their concentrations are too low in comparison to such buffers as $HPO_4^{2-}/H_2PO_4^-$ and the HCO_3^-/CO_2 system.

The importance of pH and buffers in biochemistry and clinical medicine will become apparent, particularly in Chapters 2, 4, and 25. Figure 1.8 presents

1. Calculate the ratio of HPO_4^{2-} / $H_2PO_4^-$ (pK=6.7) at pH 5.7, 6.7, and 8.7.

Solution:

$$pH = pK + \log \frac{[HPO_4^{2-}]}{[H_2PO_4^-]}$$

5.7 = 6.7 + log of ratio; rearranging

5.7– 6.7 = –1 = log of ratio

The antilog of –1 = 0.1 or 1/10. Thus, $HPO_4^{2-}/H_2PO_4^-$ = 1/10 at pH 5.7. Using the same procedure, the ratio at pH 6.7 = 1/1 and at pH 8.7 = 100/1.

2. If the pH of blood is 7.1 and the HCO_3^- concentration is 8 mM, what is the concentration of CO_2 in blood (pK' for HCO_3^-/CO_2 = 6.1)?

Solution:

$$pH = pK + \log \frac{[HCO_3^-]}{[CO_2]}$$

7.1 = 6.1 + log 8 mM / $[CO_2]$; rearranging

7.1 – 6.1 = 1 = log 8 mM / $[CO_2]$.

The antilog of 1 = 10. Thus, 10 = 8 mM / $[CO_2]$, or $[CO_2]$ = 8 mM/10 = 0.8 mM.

3. At a normal blood pH of 7.4, the sum of $[HCO_3^-] + [CO_2]$ = 25.2 mM. What is the concentration of HCO_3^- and CO_2 (pK' for HCO_3^-/CO_2 = 6.1)?

Solution:

$$pH = pK + \log \frac{[HCO_3^-]}{[CO_2]}$$

7.4 = 6.1 + log $[HCO_3^-]$ / $[CO_2]$; rearranging

7.4 – 6.1 = 1.3 = log $[HCO_3^-]$ / $[CO_2]$.

The antilog of 1.3 is 20. Thus $[HCO_3^-]$ / $[CO_2]$ = 20. Given $[HCO_3^-] + [CO_2]$ = 25.2, solve these two equations for $[CO_2]$ by rearranging the first equation:

$[HCO_3^-]$ = 20 $[CO_2]$.

Substituting in the second equation,

20 $[CO_2]$ + $[CO_2]$ = 25.2

or

CO_2 = 1.2 mM

Then substituting for CO_2, 1.2 + $[HCO_3^-]$ = 25.2, and solving, $[HCO_3^-]$ = 24 mM.

Figure 1.8
Typical problems of pH and buffering.

some typical problems using the Henderson–Hasselbalch equation and Clin. Corr. 1.1 is a representative problem encountered in clinical practice.

CLINICAL CORRELATION 1.1

Blood Bicarbonate Concentration in Metabolic Acidosis

Blood buffers in a normal adult control blood pH at about 7.40; if the pH should drop below 7.35, the condition is referred to as an acidosis. A blood pH of near 7.0 could lead to serious consequences and possibly death. Thus in acidosis, particularly that caused by a metabolic change, it is important to monitor the acid–base parameters of a patient's blood. Values of interest to a clinician include the pH and HCO_3^- and CO_2 concentrations. Normal values for these are pH = 7.40, $[HCO_3^-]$ = 24.0 mM, and $[CO_2]$ = 1.20 mM.

Blood values of a patient with a metabolic acidosis were pH = 7.03 and $[CO_2]$ = 1.10 mM. What is the patient's blood $[HCO_3^-]$ and how much of the normal $[HCO_3^-]$ has been used in buffering the acid causing the condition?

1. The Henderson–Hasselbalch equation is

$$pH = pK' + \log([HCO_3^-]/[CO_2])$$

The pK value for $[HCO_3^-]/[CO_2]$ is 6.10.

2. Substitute the given values in the equation.

$$7.03 = 6.10 + \log([HCO_3^-]/1.10\ mM)$$

or

$$7.03 - 6.10 = 0.93$$
$$= \log([HCO_3^-]/1.10\ mM)$$

The antilog of 0.93 is 8.5; thus

$$8.5 = [HCO_3^-]/1.10\ mM$$

or

$$[HCO_3^-] = 9.4\ mM$$

3. Since the normal value of $[HCO_3^-]$ is 24 mM, there has been a decrease of 14.6 mmol of HCO_3^- per liter of blood in this patient. If much more HCO_3^- is lost, a point would be reached when this important buffer would be unavailable to buffer any more acid in the blood and the pH would drop rapidly. In Chapter 25 there is a detailed discussion of the causes and compensations that occur in such conditions.

1.3— Organization and Composition of Eukaryotic Cells

As described above, eukaryotic cells are organized into compartments, each delineated by a membrane (Figure 1.9). These are well-defined cellular organelles such as nucleus, mitochondria, lysosomes, and peroxisomes. Membranes also form a tubule-like network throughout the cell enclosing an interconnecting space or cisternae, as is the case of the endoplasmic reticulum or Golgi complex. As described in Section 1.4, these compartments have specific functions and activities.

The semipermeable nature of **cellular membranes** prevents the ready diffusion of many molecules from one side to the other. Specific mechanisms in membranes for translocation of large and small, charged and uncharged molecules allow membranes to modulate concentrations of substances in various compartments. Macromolecules, such as proteins and nucleic acids, do not cross biological membranes unless there is a specific mechanism for their translocation or the membrane is damaged. Thus the fluid matrix of various cellular compartments has a distinctive composition of inorganic ions, organic molecules, and macromolecules. Partitioning of activities and components in membrane-enclosed compartments and organelles has a number of advantages for the economy of the cell. These include the sequestering of substrates and cofactors where they are required, and adjustments of pH and ionic composition for maximum activity of biological processes.

The activities and composition of cellular structures and organelles have been determined with intact cells by a variety of histochemical, immunological, and fluorescent staining methods. Continuous observation in real time of cellular events in intact viable cells is possible. Examples are studies that involve changes of ionic calcium concentration in the cytosol by the use of fluorescent calcium indicators. Individual organelles, membranes, and components of the cytosol can be isolated and analyzed following disruption of the plasma membrane. Permeability of the plasma membrane can be altered to permit the release of subcellular components. Techniques for disrupting membranes include use of detergents, osmotic shock, and homogenization of tissues, where shearing forces break down the plasma membrane. In an appropriate isolation medium, cell organelles and membrane systems can be separated by centrifugation because of differences in size and density. Chromatographic procedures have been employed for isolation of individual cellular fractions and components. These techniques have permitted isolation of cellular fractions from most mammalian tissues. In addition, components of organelles such as nuclei and mitochondria can be isolated following disruption of the organelle membrane.

In many instances the isolated structures and cellular fractions appear to retain the chemical and biochemical characteristics of the structure *in situ*. But biological membrane systems are very sensitive structures, subject to damage even under very mild conditions, and alterations can occur during isolation, which can lead to change in composition of the structure. The slightest damage to a membrane alters its permeability properties, allowing substances that would normally be excluded to traverse the membrane barrier. In addition, many proteins are only loosely associated with membranes and easily dissociate when damage occurs (see p. 186).

Not unexpectedly, there are differences in structure, composition, and activities of cells from different tissues due to the diverse functions of tissues. Major biochemical activities of the cellular organelles and membrane systems, however, are fairly constant from tissue to tissue. Thus biochemical pathways in liver are often present in other tissues. The differences between cell types are

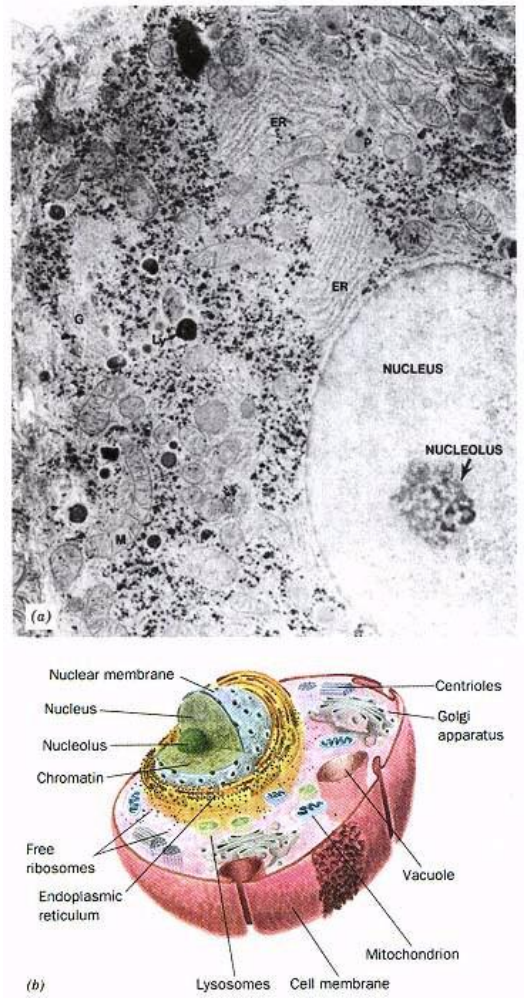


Figure 1.9
(a) Electron micrograph of a rat liver cell labeled to indicate the major structural components of eukaryotic cells and
(b) a schematic drawing of an animal cell.
Note the number and variety of subcellular organelles and the network of interconnecting membranes enclosing channels, that is, cisternae. All eukaryotic cells are not as complex in their appearance, but most contain the major structures shown in the figure. ER, endoplasmic reticulum; G, Golgi zone, Ly, lysosomes, P, peroxisomes; M, mitochondria. Photograph (a) reprinted with permission of Dr. K. R. Porter from Porter, K. R., and Bonneville, M. A. In: *Fine Structure of Cells and Tissues*. Philadelphia: Lea & Febiger, 1972; schematic (b) reprinted with permission from Voet, D., and Voet, J. G. *Biochemistry*, 2nd ed. New York: Wiley, 1995.

usually in distinctive specialized activities. Even within one tissue, cells of different origin have qualitative and quantitative differences in cell organelle composition.

Chemical Composition of Cells

Each cellular compartment has an aqueous fluid or **matrix** that contains various ions, small molecular weight organic molecules, different proteins, and nucleic acids. Localization of specific macromolecules, such as enzymes, has been

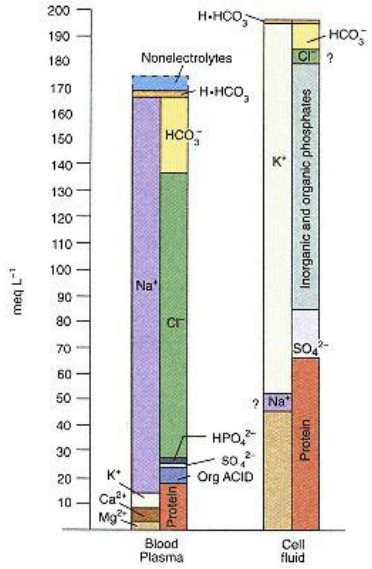


Figure 1.10
Major chemical constituents of blood plasma and cell fluid.
Height of left half of each column indicates total concentration of cations; that of right half, concentrations of anions. Both are expressed in milliequivalents per liter (meq L^{-1}) of fluid. Note that chloride and sodium values in cell fluid are questioned. It is probable that, at least in muscle, the cytosol contains some sodium but no chloride.
Adapted from Gregersen, M. I. In: P. Bard (Ed.), *Medical Physiology*, 11th ed. St Louis, MO: Mosby, 1961, p. 307.

determined but the exact ionic composition of the matrix of organelles is still uncertain. Each has a distinctly different ionic composition and pH. The overall ionic composition of intracellular fluid, considered to represent the cytosol primarily, compared to blood plasma is presented in Figure 1.10. **Na^+** is the major extracellular cation, with a concentration of ~140 meq L^{-1} (mM); very little Na^+ is present in intracellular fluid. **K^+** is the major intracellular cation. **Mg^{2+}** is present in both extra- and intracellular compartments at concentrations much lower than Na^+ and K^+. The major extracellular anions are **Cl^-** and **HCO_3^-**, with lower amounts of phosphate and sulfate. Most proteins have a negative charge at pH 7.4 (Chapter 2), being anions at the pH of tissue fluids. Major intracellular anions are inorganic phosphate, organic phosphates, and proteins. Other inorganic and organic anions and cations are present in concentrations well below the milliequivalent per liter (millimolar) level. Except for very small differences created by membranes and leading to development of membrane potentials, the *total anion concentration equals the total cation concentration in the different fluids*.

Intracellular concentrations of most small molecular weight organic molecules, such as sugars, organic acids, amino acids, and phosphorylated intermediates, are in the range of 0.01–1.0 mM but can have significantly lower concentrations. Coenzymes, organic molecules required for activity of some enzymes, are in the same range of concentration. Substrates for enzymes are present in relatively low concentration in contrast to inorganic ions, but localization in a

specific organelle or cellular microenvironment can increase their concentrations significantly.

It is not very meaningful to determine the molar concentration of individual proteins in cells. In many cases they are localized with specific structures or in combination with other proteins to create a functional unit. It is in a restricted compartment that individual proteins carry out their role, whether structural, catalytic, or regulatory.

1.4— Functional Role of Subcellular Organelles and Membrane Systems

The subcellular localization of various metabolic pathways will be described throughout this book. In some cases an entire pathway is located in a single compartment but many are divided between two locations, with the intermediates in the pathway moving or being translocated from one compartment to another. In general, organelles have very specific functions and the enzymatic activities involved are used to identify them during isolation.

The following describes briefly some major roles of eukaryotic cell structures to indicate the complexity and organization of cells. A summary of functions and division of labor within eukaryotic cells is presented in Table 1.6 and the structures are presented in Figure 1.9.

TABLE 1.6 Summary of Eukaryotic Cell Compartments and Their Major Functions

Compartment	*Major Functions*
Plasma membrane	Transport of ions and molecules
	Recognition
	Receptors for small and large molecules
	Cell morphology and movement
Nucleus	DNA synthesis and repair
	RNA synthesis
Nucleolus	RNA processing and ribosome synthesis
Endoplasmic reticulum	Membrane synthesis
	Synthesis of proteins and lipids for some organelles and for export
	Lipid synthesis
	Detoxication reactions
Golgi apparatus	Modification and sorting of proteins for incorporation into organelles and for export
	Export of proteins
Mitochondria	Energy conservation
	Cellular respiration
	Oxidation of carbohydrates and lipids
	Urea and heme synthesis
Lysosomes	Cellular digestion: hydrolysis of proteins, carbohydrates, lipids, and nucleic acids
Peroxisomes	Oxidative reactions involving O_2 Utilization of H_2O_2
Microtubules and microfilaments	Cell cytoskeleton
	Cell morphology
	Cell motility
	Intracellular movements
Cytosol	Metabolism of carbohydrates, lipids, amino acids, and nucleotides
	Protein synthesis

CLINICAL CORRELATION 1.2

Mitochondrial Diseases: Luft's Disease

A disease specifically involving mitochondrial energy transduction was first reported in 1962. A 30-year-old patient was described with general weakness, excessive perspiration, a high caloric intake without increase in body weight, and an excessively elevated basal metabolic rate (a measure of oxygen utilization). It was demonstrated that the patient had a defect in the mechanism that controls mitochondrial oxygen utilization (see Chapter 6). The condition is referred to as Luft's disease. Since that time, over 100 mitochondrial-based diseases have been identified, including those involving a variety of enzymes and transport systems required for the proper maintenance and control of energy conservation. Many involve skeletal muscle and the central nervous system. Replication of mitochondria depends on the mitochondrial DNA (mtDNA) and inheritance of mitochondria is by maternal transmission. Mutations of mtDNA as well as nuclear DNA lead to genetic diseases. Mitochondrial damage may also occur due to free-radical (superoxides) formation which can damage mtDNA. Thus age-related degenerative diseases, such as Parkinson's and Alzheimer's, and cardiomyopathies may have a component of mitochondrial damage. For details of specific diseases see Clin. Corr. 13.4 and 14.6.

Luft, R. The development of mitochondrial medicine. *Proc. Natl. Acad. Sci. USA* 91:8731,1994.

Plasma Membrane Is the Limiting Boundary of a Cell

The **plasma membrane** of every cell has a unique role in maintenance of that cell's integrity. One surface is in contact with a variable external environment and the other with a relatively constant environment provided by the cell's cytoplasm. As will be discussed in Chapter 5, the two sides of the plasma membrane, and all intracellular membranes, have different chemical compositions and functions. A major role of the plasma membrane is to permit entrance of some substances but exclude many others. With cytoskeletal elements, the plasma membrane is involved in cell shape and movements. Through this membrane cells communicate; the membrane contains many specific receptor sites for chemical signals, such as hormones (Chapter 20), released by other cells. The inner surface of plasma membranes is the site for attachment of some enzymes involved in various metabolic pathways. Plasma membranes from a variety of cells have been isolated and studied extensively; details of their structure and biochemistry and those of other membranes are presented in Chapter 5.

Nucleus Is Site of DNA and RNA Synthesis

Early microscopists divided the interior of cells into a **nucleus,** the largest membrane-bound compartment, and the **cytoplasm.** The nucleus is surrounded by two membranes, termed the **nuclear envelope,** with the outer membrane being continuous with membranes of the endoplasmic reticulum. The nuclear envelope has numerous pores about 90 Å in diameter that permit flow of all but the largest molecules between nuclear matrix and cytoplasm. The nucleus contains a subcompartment, seen clearly in electron micrographs, the **nucleolus.** The vast amount of cellular **deoxyribonucleic acid (DNA)** is located in the nucleus as a DNA–protein complex, **chromatin,** that is organized into chromosomes. DNA is the repository of genetic information and the importance of the nucleus in cell division and for controlling phenotypic expression of genetic information is well established. Biochemical reactions in the nucleus are replication of DNA during mitosis, repair of DNA following damage (Chapter 15), and transcription of the information stored in DNA into a form that can be translated into cell proteins (Chapter 16). Transcription of DNA involves synthesis of ribonucleic acid (RNA) that is processed into a variety of forms following synthesis. Part of this processing occurs in the nucleolus, which is very rich in RNA.

Endoplasmic Reticulum Has a Role in Many Synthetic Pathways

The cytoplasm of most eukaryotic cells contains a network of interconnecting membranes that enclose channels, **cisternae,** that thread from the perinuclear envelope to the plasma membrane. This extensive subcellular structure, termed **endoplasmic reticulum,** consists of membranes with a rough appearance in some areas and smooth in other places. The rough appearance is due to the presence of **ribonucleoprotein particles,** that is, **ribosomes,** attached on the cytosolic side of the membrane. Smooth endoplasmic reticulum does not contain bound ribosomes. During cell fractionation the endoplasmic reticulum network is disrupted, with the membrane resealing into small vesicles called **microsomes** that can be isolated by differential centrifugation. Microsomes per se do not occur in cells.

A major function of ribosomes on **rough endoplasmic reticulum** is biosynthesis of proteins for export to the outside of the cell and proteins for incorporation into cellular organelles such as the endoplasmic reticulum, Golgi apparatus, plasma membrane, and lysosomes. **Smooth endoplasmic reticulum** is involved in membrane lipid synthesis and contains an important class

of enzymes termed **cytochromes P450** that catalyze hydroxylation of a variety of endogenous and exogenous compounds. These enzymes are important in biosynthesis of steroid hormones and removal of toxic substances (see Chapter 23). Endoplasmic reticulum with the Golgi apparatus has a role in formation of other cellular organelles such as lysosomes and peroxisomes.

The Golgi Apparatus Is Involved in Sequestering of Proteins

The **Golgi apparatus** is a network of flattened smooth membranes and vesicles responsible for the secretion to the external environment of a variety of proteins synthesized on the endoplasmic reticulum. Golgi membranes catalyze the transfer of carbohydrate and lipid precursors to proteins to form glycoproteins and lipoproteins and is a major site of new membrane formation. Membrane vesicles are formed in the Golgi apparatus in which various proteins and enzymes are encapsulated to be secreted from the cell after an appropriate signal. Digestive enzymes synthesized by the pancreas are stored in intracellular vesicles formed by the Golgi apparatus and released when needed in the digestive process (see p. 1059). The role in membrane synthesis also includes the formation of intracellular organelles such as lysosomes and peroxisomes.

TABLE 1.7 Representative Lysosomal Enzymes and Their Substrates

Type of Substrate and Enzyme	*Specific Substrate*
POLYSACCHARIDE-	
HYDROLYZING	
ENZYMES	
α-Glucosidase	Glycogen
α-Flucosidase	Membrane fucose
	Galactosides
β-Galactosidase	Mannosides
α-Mannosidase	Glucuronides
β-Glucuronidase Hyaluronidase	Hyaluronic acid and chondroitin sulfates
Arylsulfatase	Organic sulfates
Lysozyme	Bacterial cell walls
PROTEIN-HYDROLYZING	
ENZYMES	
Cathepsins	Proteins
Collagenase	Collagen
Elastase	Elastin
Peptidases	Peptides
NUCLEIC ACID-	
HYDROLYZING	
ENZYMES	
Ribonuclease	RNA
Deoxyribonuclease	DNA
LIPID-HYDROLYZING	
ENZYMES	
Lipases	Triglyceride and cholesterol esters
Esterase	Fatty acid esters
Phospholipase	Phospholipids
PHOSPHATASES	
Phosphatase	Phospho-monoesters
Phosphodiesterase	Phosphodiesters
SULFATASES	
Heparan sulfate	
Dermatan sulfate	

Mitochondria Supply Most Cell Needs for ATP

Mitochondria appear as spheres, rods, or filamentous bodies that are usually about 0.5–1 μm in diameter and up to 7 μm in length. The internal matrix, the **mitosol**, is surrounded by two membranes, distinctively different in appearance and biochemical function. The inner membrane convolutes into the matrix to form **cristae** and contains numerous small spheres attached by stalks on the inner surface. Outer and inner membranes contain different enzymes. The components of the respiratory chain and the mechanism for ATP synthesis are part of the inner membrane and are described in detail in Chapter 6. Major metabolic pathways involved in oxidation of carbohydrates, lipids, and amino acids, and parts of special biosynthetic pathways involving urea and heme synthesis are located in the mitosol. The outer membrane is relatively permeable but the inner membrane is highly selective and contains a variety of transmembrane transport systems.

Mitochondria contain a specific DNA, with genetic information for some of the mitochondrial proteins, and the biochemical equipment for limited protein synthesis. The presence of this biosynthetic capacity indicates the unique role that mitochondria have in their own destiny. See Clin. Corr. 1.2 for descriptions of diseases attributed to deficits in mitochondrial function.

Lysosomes Are Required for Intracellular Digestion

Intracellular digestion of a variety of substances occurs inside structures designated as **lysosomes.** They have a single limiting membrane and maintain a pH lower in the lysosomal matrix than that of the cytosol. Encapsulated in lysosomes is a group of glycoprotein enzymes—hydrolases—that catalyze hydrolytic cleavage of carbon-oxygen, carbon–nitrogen, carbon–sulfur, and oxygen–phosphorus bonds in proteins, lipids, carbohydrates, and nucleic acids. A partial list of lysosomal enzymes is presented in Table 1.7. As in gastrointestinal digestion, lysosomal enzymes split complex molecules into simple low molecular weight compounds that can be utilized by metabolic pathways of the cell. Enzymes of the lysosome are characterized by being most active when the pH of the medium is acidic, that is, pH 5 and below. The relationship between pH and enzyme activity is discussed in Chapter 4. The pH of the cytosol is close to pH 7.0 and lysosomal enzymes have little activity at this pH.

CLINICAL CORRELATION 1.3

Lysosomal Enzymes and Gout

Catabolism of purines, nitrogen-containing heterocyclic compounds found in nucleic acids, leads to formation of uric acid, which is excreted in the urine (see Chapter 12 for details). Gout is an abnormality in which there is excessive uric acid production with an increase in uric acid in blood and deposition of urate crystals in joints. The consequences are clinical manifestations in the joint, particularly the big toe, including inflammation, pain, swelling, and increased warmth. Uric acid is not very soluble and some of the clinical symptoms of gout can be attributed to damage done by urate crystals. Crystals are phagocytosed by cells in the joint and accumulate in digestive vacuoles that contain lysosomal enzymes. Crystals cause physical damage to the vacuoles, releasing lysosomal enzymes into the cytosol. Even though the pH optima of lysosomal enzymes are lower than the pH of the cytosol, they have some hydrolytic activity at the higher pH. This activity causes digestion of cellular components, release of substances from the cell and autolysis.

Weissmann, G. Crystals, lysosomes and gout. *Adv. Intern. Med.* 19:239, 1974; and Burt, H. M., Kalkman, P. H., and Mauldin, D. Membranolytic effects of crystalline monosodium urate monohydrate. *J. Rheumatol.* 10:440, 1983.

The enzyme content of lysosomes in different tissues varies and depends on specific needs of individual tissues. The lysosomal membrane is impermeable to both small and large molecules and specific protein mediators in the membrane are necessary for translocation of substances. Carefully isolated lysosomes do not catalyze hydrolysis of substrates until this membrane is disrupted. The activities of lysosomal enzymes are termed "**latent**." Membrane disruption *in situ* can lead to cellular digestion, and various pathological conditions have been attributed to release of lysosomal enzymes, including arthritis, allergic responses, several muscle diseases, and drug-induced tissue destruction (see Clin. Corr. 1.3).

Lysosomes are involved in normal digestion of intra- and extracellular substances that must be removed by a cell. Through **endocytosis**, external material is taken into cells and encapsulated in membrane-bound vesicles (Figure 1.11). The plasma membrane invaginates around formed foreign substances, such as microorganisms, by **phagocytosis** and takes up extracellular fluid containing suspended material by **pinocytosis**. Vesicles containing external material fuse with lysosomes to form organelles that contain the materials to be digested and enzymes capable of carrying out the digestion. These vacuoles are identified microscopically by their size and often by the presence of partially formed structures in the process of being digested. Lysosomes in which the

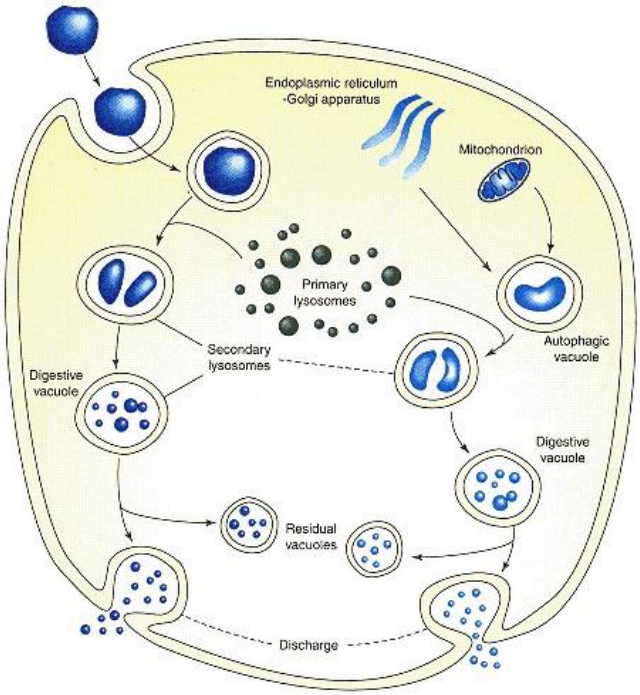


Figure 1.11
Diagrammatic representation of the role of lysosomes in intracellular digestion of substances internalized by phagocytosis (heterophagy) and of cellular components (autophagy).
In both processes substances to be digested are enclosed in a membrane vesicle, followed by fusing with a primary lysosome to form a secondary lysosome.

enzymes are not as yet involved in the digestive process are termed **primary lysosomes**, whereas those in which digestion of material is under way are **secondary lysosomes** or **digestive vacuoles** that will vary in size and appearance.

CLINICAL CORRELATION 1.4

Lysosomal Acid Lipase Deficiency

Two phenotypic forms of a genetic deficiency of lysosomal acid lipase are known. Wolman's disease occurs in infants and is usually fatal by age 1, while cholesterol ester-storage disease usually is diagnosed in adulthood. Both are autosomal recessive disorders. There is deposition of triacylglycerols and cholesterol esters in tissues, particularly the liver. In the latter disease there is early onset of severe atherosclerosis. Acid lipase catalyzes hydrolysis of mono-, di-, and triglycerols as well as cholesterol esters. It is a critical enzyme in cholesterol metabolism, serving to make available free cholesterol for cell needs.

Hegele, R. A., Little, J. A., Vezina, C., et al., Hepatic lipase deficiency: clinical, biochemical, and molecular genetic characteristics. *Atherosclerosis and Thrombosis* 13:720, 1993.

Cell constituents are synthesized and degraded continuously, and lysosomes function in digesting this cellular debris. The dynamic synthesis and degradation includes proteins and nucleic acids, as well as structures such as mitochondria and endoplasmic reticulum. During this normal self-digestion process, that is, **autolysis**, cell substances are encapsulated within a membrane vesicle that fuses with a lysosome to complete the degradation. The overall process is termed **autophagy** and is also represented in Figure 1.11.

Products of lysosomal digestion diffuse across lysosomal membranes and are reutilized by the cell. Indigestible material accumulates in vesicles referred to as residual bodies, whose contents are removed from the cell by exocytosis. In some cases, residual bodies that contain a high concentration of lipid persist for long periods of time. Lipid is oxidized and a pigmented substance, which is chemically heterogeneous and contains polyunsaturated fatty acids and proteins, accumulates in the cell. This material, **lipofuscin**, is also called the "**age pigment**" or "**wear and tear pigment**" because it accumulates in cells of older individuals. It occurs in all cells but particularly in neurons and muscle cells and has been implicated in the aging process.

Under controlled conditions lysosomal enzymes are secreted from the cell for the digestion of extracellular material; an extracellular role for some lysosomal enzymes has been demonstrated in connective tissue and prostate gland and in the process of embryogenesis. Thus they have a role in programmed cell death or **apoptosis.**

Absence of specific lysosomal enzymes occurs in a number of genetic diseases in which there is accumulation in the cell of specific cellular components that cannot be digested. Lysosomes of affected cells become enlarged with undigested material, which interferes with normal cell processes. Lysosomal storage diseases are discussed in Chapter 10 (see p. 427); see Clin. Corr. 1.4 for a description of a deficiency of lysosomal lipase.

Peroxisomes Contain Oxidative Enzymes Involving Hydrogen Peroxide

Most eukaryotic cells of mammalian origin and those of protozoa and plants have organelles, designated **peroxisomes** or **microbodies**, which contain enzymes that either produce or utilize **hydrogen peroxide** (H_2O_2). They are small (0.3–1.5 mm in diameter), spherical or oval in shape, with a granular matrix and in some cases a crystalline inclusion termed a nucleoid. Peroxisomes contain enzymes that oxidize D-amino acids, uric acid, and various 2-hydroxy acids using molecular O_2 with formation of H_2O_2. Catalase, an enzyme present in peroxisomes, catalyzes the conversion of H_2O_2 to water and oxygen and oxidation by H_2O_2 of various compounds (Figure 1.12). By having both peroxide-producing and peroxide-utilizing enzymes in one compartment, cells protect themselves from the toxicity of H_2O_2.

Peroxisomes also contain enzymes involved in lipid metabolism, particularly oxidation of very long-chain fatty acids, and synthesis of glycerolipids and glycerol ether lipids (plasmalogens) (see Chapter 10). See Clin. Corr. 1.5 for a discussion of Zellweger syndrome in which there is an absence of peroxisomes.

Peroxisomes of different tissues contain different complements of enzymes, and the peroxisome content of cells can vary depending on cellular conditions.

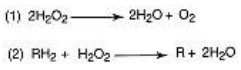


Figure 1.12
Reactions catalyzed by catalase.

Cytoskeleton Organizes the Intracellular Contents

Eukaryotic cells contain microtubules and actin filaments (microfilaments) as parts of the cytoskeletal network. The **cytoskeleton** has a role in maintenance

of cellular morphology, intracellular transport, cell motility, and cell division. **Microtubules**, multimers of the protein **tubulin**, can be rapidly assembled and disassembled depending on the cell's needs. Two very important cellular filaments, actin and myosin, occur in striated muscle and are responsible for muscular contraction (see Chapter 22). Three **mechanochemical proteins—myosin, dynein,** and **kinesin**—convert chemical energy into mechanical energy for movement of cellular components. These molecular motors are associated with the cytoskeleton; the actual mechanism for the energy conversion, however, has not been defined completely. Dynein is involved in ciliary and flagellar movement, whereas kinesin is a driving force for the movement of vesicles and organelles along microtubules.

CLINICAL CORRELATION 1.5

Zellweger Syndrome and the Absence of Functional Peroxisomes

Zellweger syndrome is a rare, autosomal recessive disease characterized by abnormalities of the liver, kidney, brain, and skeletal system. It usually results in death by age 6 months. A number of seemingly unrelated biochemical abnormalities have been described including decreased levels of glycerol-ether lipids (plasmalogens) and increased levels of very long-chain fatty acids (C-24 and C-26) and cholestanoic acid derivatives (precursors of bile acids). These abnormalities are due to the absence of functional peroxisomes in the afflicted children. Peroxisomes are responsible for synthesis of glycerol ethers, for shortening very long-chain fatty acids so that mitochondria can completely oxidize them, and for oxidation of the side chain of cholesterol needed for bile acid synthesis. Evidence indicates that there is a defect in the transport of peroxisomal enzymes between the cytosol and the interior of peroxisomes during synthesis. Cells of afflicted individuals contain empty ghosts of peroxisomes. The disease can be diagnosed prenatally by assaying amniotic fluid cells for peroxisomal enzymes or analyzing the fatty acids in the fluid.

Datta, N. S., Wilson, G. N., and Hajra, A. K. Deficiency of enzymes catalyzing the biosynthesis of glycerol-ether lipids in Zellweger syndrome. *N. Engl. J. Med.* 311:1080, 1984; Moser, A. E., Singh, I., Brown, F. R., Solish, G. I., Kelley, R. I., Benke, P. J., and Moser, H. W. The cerebrohepatorenal (Zellweger) syndrome. Increased levels and impaired degradation of very long chain fatty acids and their use for prenatal diagnosis. *N. Engl. J. Med.* 310:1141, 1984; and Wanders, R. J., Schutgens, R.B., and Barth, P. G. Peroxisomal disorders: a review. *J. Neuropathol. Exp. Neurol.* 54:726, 1995.

Cytosol Contains Soluble Cellular Components

The least complex in structure, but not in chemistry, is the organelle-free cell sap, or **cytosol**. It is here that many of the chemical reactions of metabolism occur and where substrates and cofactors interact with various enzymes. Although there is no apparent structure to the cytosol, the high protein content precludes it from being a truly homogeneous mixture of soluble components. Many reactions are localized in selected areas where substrate availability is more favorable. The actual physicochemical state of the cytosol is poorly understood. A major role of the cytosol is to support synthesis of proteins on the rough endoplasmic reticulum by supplying cofactors and energy. The cytosol also contains free ribosomes, often in a polysome form, for synthesis of intracellular proteins.

Studies with isolated cytosol suggest that many reactions are catalyzed by soluble enzymes, but in the intact cell some of these enzymes may be loosely attached to one of the many membrane structures or to cytoskeletal components and are readily released upon cell disruption.

Conclusion

A eukaryotic cell is a complex structure whose purpose is to replicate itself when necessary, maintain an intracellular environment to permit a myriad of complex reactions to occur as efficiently as possible, and to protect itself from the hazards of its surrounding environment. Cells of multicellular organisms also participate in maintaining the well-being of the whole organism by exerting influences on each other to maintain all tissue and cellular activities in balance. Thus, as we proceed to study the separate chemical components and activities of cells in subsequent chapters, it is important to keep in mind the concurrent and surrounding activities, constraints, and influences of the environment. Only by bringing together all the parts and activities of a cell, that is, reassembling the puzzle, will we appreciate the wonder of living cells.

Bibliography

Water and Electrolytes

Dick, D. A. T. *Cell Water*. Washington, DC: Butterworths, 1966.

Eisenberg, D., and Kauzmann, W. *The Structures and Properties of Water*. Fairlawn, NJ: Oxford University Press, 1969.

Morris, J. G. *A Biologist's Physical Chemistry*. Reading, MA: Addison-Wesley, 1968.

Stillinger, F. H. Water revisited. *Science* 209:451, 1980.

Westof, E., *Water and Biological Macromolecules*. Boca Raton, FL: CRC Press, 1993.

Cell Structure

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

Becker, W. M., and Deamer, D. W. *The World of the Living Cell*, 2nd. ed. Redwood City, CA: Benjamin, 1991.

DeDuve, C. *Guided Tour of the Living Cell*, Vols. 1 and 2. New York: Scientific American Books, 1984.

Dingle, J. T., Dean, R. T., and Sly, W. S. (Eds.). *Lysosomes in Biology and Pathology*. Amsterdam: Elsevier (a serial publication covering all aspects of lysosomes).

Fawcett, D. W. *The Cell*. Philadelphia: Saunders, 1981.

Holtzman, E., and Novikoff, A. B. *Cells and Organelles,* 3rd ed. New York: Holt, Rinehart & Winston, 1984.

Porter, K. R., and Bonneville, M. A. Fine Structure of Cells and Tissues. Philadelphia: Lea & Febiger, 1972.

Vale, R. D. Intracellular transport using microtubule-based motors. *Annu. Rev. Cell Biol.* 3:347, 1987.

Cell Organelles

Attardi, G., and Chomyn, A. *Mitochondrial Biogenesis and Genetics*. San Diego, CA: Academic Press, 1995.

Holtzman, E. *Lysosomes*. New York: Plenum Press, 1989.

Latruffe, N., and Bugaut, M. *Peroxisomes*. New York: Springer-Verlag, 1994.

Pavelka, M. *Functional Morphology of the Golgi Apparatus*. New York: Springer-Verlag, 1987.

Preston, T. M., King, C. A., and Hyams, J. S. *The Cytoskeleton and Cell Motility*. New York: Chapman and Hall, 1990.

Strauss, P. R., and Wilson, S. H. *The Eukaryotic Nucleus: Molecular Biochemistry and Macromolecular Assemblies*. Caldwell, NJ: Telford Press, 1990.

Tyler, D. D. *Mitochondrion in Health and Disease*. New York: VCH, 1992.

Tzagoloff, A. *Mitochondria*. New York: Plenum Press, 1982.

Questions

J. Baggot and C. N. Angstadt

1. Prokaryotic cells, but not eukaryotic cells, have:

   A. endoplasmic reticulum.

   B. histones.

   C. nucleoid.

   D. a nucleus.

   E. a plasma membrane.

2. Factors responsible for the polarity of the water molecule include:

   A. the similarity in electron affinity of hydrogen and oxygen.

   B. the tetrahedral structure of liquid water.

   C. the magnitude of the H–O–H bond angle.

   D. the ability of water to hydrogen bond to various chemical structures.

   E. the difference in bond strength between hydrogen bonds and covalent bonds.

3. Hydrogen bonds can be expected to form only between electronegative atoms such as oxygen or nitrogen and a hydrogen atom bonded to:

   A. carbon.

   B. an electronegative atom.

   C. hydrogen.

   D. iodine.

   E. sulfur.

4. Which of the following is least likely to be soluble in water?

   A. nonpolar compound

   B. weakly polar compound

   C. strongly polar compound

   D. weak electrolyte

   E. strong electrolyte

5. Which of the following is most likely to be partly associated in weak aqueous solution?

   A. alcohol

   B. lactic acid

   C. potassium sulfate (K_2SO_4)

   D. sodium chloride (NaCl)

   E. sodium lactate

6. The ion product of water:

   A. is independent of temperature.

   B. has a numerical value of 10^{-14} at 25°C.

   C. is the equilibrium constant for the reaction $HOH \rightleftharpoons H^+ + OH^-$

   D. requires that $[H^+]$ and $[OH^-]$ always be identical.

   E. is an approximation that fails to take into account the presence of the hydronium ion, H_3O^+.

7. Which of the following is both a Brønsted acid and a Brønsted base in water?

   A. H_2PO_{4-}

   B. H_2CO_3

   C. NH_3

   D. NH_4^+

   E. Cl^-

Refer to the following information for Questions 8 and 9.

A. pyruvic acid	pK = 2.50
B. acetoacetic acid	pK = 3.6
C. lactic acid	pK = 3.86
D. β-hydroxybutyric acid	pK = 4.7
E. propionic acid	pK = 4.86

8. Which weak acid will be 91% neutralized at pH 4.86?

9. Assuming that the sum of [weak acid] + [conjugate base] is identical for buffer systems based on the acids listed above, which has the greatest buffer capacity at pH 4.86?

10. All of the following subcellular structures can be isolated essentially intact EXCEPT:

   A. endoplasmic reticulum.

   B. lysosomes.

   C. mitochondria.

   D. nuclei.

   E. peroxisomes.

11. Biological membranes are associated with all of the following EXCEPT:

   A. prevent free diffusion of ionic solutes.

   B. release of proteins when damaged.

   C. contain specific systems for the transport of uncharged molecules.

   D. sites for biochemical reactions.

   E. proteins and nucleic acids cross freely.

12. Mitochondria are associated with all of the following EXCEPT:

   A. ATP synthesis.

   B. DNA synthesis.

   C. protein synthesis.

   D. hydrolysis of various macromolecules at low pH.

   E. two different membranes.

13. Analysis of the composition of the major fluid compartments of the body shows that:

A. the major blood plasma cation is K^+.

B. the major cell fluid cation is Na^+.

C. one of the major intracellular anions is Cl^-.

D. one of the major intracellular anions is phosphate.

E. plasma and the cell fluid are all very similar in ionic composition.

Refer to the following for Questions 14–17.

A. peroxisome

B. nucleus

C. cytoskeleton

D. endoplasmic reticulum

E. Golgi apparatus

14. Consists of microtubules and actin fibers.

15. Oxidizes very long-chain fatty acids.

16. Connected to the plasma membrane by a network of membranous channels.

17. Transfers carbohydrate precursors to proteins during glycoprotein synthesis.

Answers

1. C Prokaryotic DNA is organized into a structure that also contains RNA and protein, called nucleoid. A, B, and D are found in eukaryotic cells, and E is an element of both prokaryotic and eukaryotic cells (p. 2).

2. C Water is a polar molecule because the bonding electrons are attracted more strongly to oxygen than to hydrogen. The bond angle gives rise to asymmetry of the charge distribution; if water were linear, it would not be a dipole (p. 4). A: Hydrogen and oxygen have very different electron affinity. B and D are consequences of water's structure, not factors responsible for it.

3. B Only hydrogen atoms bonded to one of the electronegative elements (O, N, F) can form hydrogen bonds (p. 5). A hydrogen atom participating in hydrogen bonding must have an electronegative element on both sides of it.

4. A In general, compounds that interact with the water dipoles are more soluble than those that do not. Thus ionized compounds and polar compounds tend to be soluble. Nonpolar compounds prefer to interact with one another rather than with polar solvents such as water (p. 5).

5. B Lactic acid is a weak acid, and weak acids dissociate only partially in aqueous solution (p. 6) A: Alcohol is fully associated. C–E: These are salts and are considered to be fully dissociated under physiological conditions, although at high concentration some association occurs.

6. B The constant is a function of temperature and is numerically equal to the equilibrium constant for the dissociation of water divided by the molar concentration of water (p. 6). D: $[H^+] = [OH^-]$ in pure water, but not in solutions of solutes that contribute H^+ or OH^-.

7. A $H_2PO_4^-$ can donate a proton to become HPO_4^{2-}. It can also accept a proton to become H_3PO_4. B and D are Brønsted acids; C is a Brønsted base. The Cl^- ion in water is neither (p. 8).

8. C If weak acid is 91% neutralized, 91 parts are present as conjugate base and 9 parts remain as the weak acid. Thus the conjugate base/acid ratio is 10 : 1. Substituting into the Henderson–Hasselbalch equation, $4.86 = pK + \log (10/1)$, and solving for pH gives the answer (p. 9).

9. E The buffer capacity of any system is maximal at pH = pK (p. 10). Buffer concentration also affects buffer capacity, but in this case concentrations are equal.

10. A Gentle disruption of cells will not destroy B–E. The tube-like endoplasmic reticulum, however, is disrupted and forms small vesicles. These vesicles, not the original structure from which they were derived, may be isolated (pp. 12, 16).

11. E (p. 17).

12. D This is a lysosomal function (p. 17). Mitochondrial properties are described on p. 17.

13. D Phosphate and protein are the major intracellular anions. A, B, and E: Plasma and cell fluid are strikingly different. The Na^+ ion is the major cation of plasma. C: Most chloride is extracellular (p. 14, Figure 1.10).

14. C (p. 19).

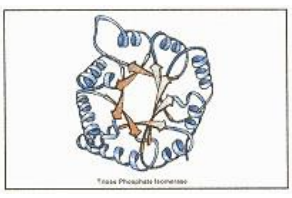
15. A Fatty acid oxidation occurs in the mitochondria, but the oxidation of very long-chain fatty acids involves the peroxisomes (p. 19).

16. B This describes only the nucleus (p. 16).

17. E Lipids, too, are attached covalently to certain proteins in the Golgi apparatus (p. 17).

Chapter 2—
Proteins I:
Composition and Structure

Richard M. Schultz and Michael N. Liebman



2.1— Functional Roles of Proteins in Humans

Proteins perform a surprising variety of essential functions in mammalian organisms. These may be grouped into dynamic and structural. Dynamic functions include transport, metabolic control, contraction, and catalysis of chemical transformations. In their structural functions, proteins provide the matrix for bone and connective tissue, giving structure and form to the human organism.

An important class of dynamic proteins are the enzymes. They catalyze chemical reactions, converting a substrate to a product at the enzyme's active site. Almost all of the thousands of chemical reactions that occur in living organisms require a specific enzyme catalyst to ensure that reactions occur at a rate compatible with life. The character of any cell is based on its particular chemistry, which is determined by its specific enzyme composition. Genetic traits are expressed through synthesis of enzymes, which catalyze reactions that establish the phenotype. Many genetic diseases result from altered levels of enzyme production or specific alterations to their amino acid sequence. Transport is another major function for proteins. Particular examples discussed in greater detail in this text are hemoglobin and myoglobin, which transport oxygen in blood and in muscle, respectively. Transferrin transports iron in blood. Transport proteins bind and carry steroid hormones in blood from their site of synthesis to their site of action. Many drugs and toxic compounds are transported bound to proteins. Proteins participate in contractile mechanisms. Myosin and actin function in muscle contraction.

Proteins have a protective role through a combination of dynamic functions. Immunoglobulins and interferon are proteins that protect the human against bacterial or viral infection. Fibrin stops the loss of blood on injury to the vascular system.

Many hormones are proteins or peptides. Protein hormones include insulin, thyrotropin, somatotropin (growth hormone), luteinizing hormone, and follicle-stimulating hormone. Many diverse polypeptide hormones have a low molecular weight (<5000) and are referred to as peptides. In general, the term **protein** is used for molecules composed of over 50 amino acids and the term **peptide** is used for molecules of less than 50 amino acids. Important peptide hormones include adrenocorticotropin hormone, antidiuretic hormone, glucagon, and calcitonin.

Proteins control and regulate gene transcription and translation. These include histones that are closely associated with DNA, repressor and enhancer transcription factors that control gene transcription, and proteins that form a part of the heteronuclear RNA particles and ribosomes.

Structural proteins function in "brick-and-mortar" roles. They include collagen and elastin, which form the matrix of bone and ligaments and provide structural strength and elasticity to organs and the vascular system. α-Keratin forms the structure of epidermal tissue.

An understanding of both the normal functioning and the pathology of the mammalian organism requires a clear understanding of the properties of the proteins.

2.2— Amino Acid Composition of Proteins

Proteins Are Polymers of α-Amino Acids

It is notable that all the different types of proteins are initially synthesized as polymers of only 20 amino acids. These **common amino acids** are defined as those for which at least one specific codon exists in the DNA genetic code. There are 20 amino acids for which DNA codons are known. Transcription and translation of the DNA code result in polymerization of amino acids into a specific linear sequence characteristic of a protein (Figure 2.1). In addition to the common amino acids, proteins may contain **derived amino acids**, which are usually formed by an enzyme-facilitated reaction on a common amino acid after that amino acid has been incorporated into a protein structure. Examples of derived amino acids are cystine (see p. 30), desmosine and isodesmosine found in elastin, hydroxyproline and hydroxylysine found in collagen, and -carboxyglutamate found in prothrombin.

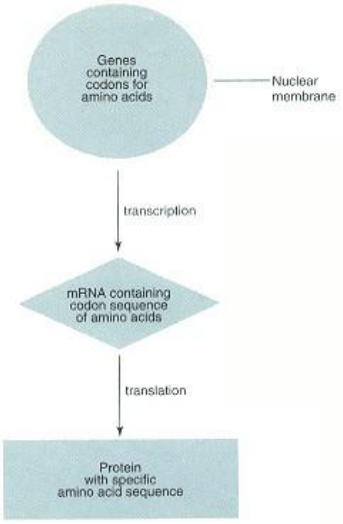


Figure 2.1
Genetic information is transcribed from a DNA sequence into mRNA and then translated to the amino acid sequence of a protein.

Common Amino Acids Have a General Structure

Common amino acids have the general structure depicted in Figure 2.2. They contain in common a central ***alpha* (α)-carbon** atom to which a carboxylic acid group, an amino group, and a hydrogen atom are covalently bonded. In addition, the α-carbon atom is bound to a specific chemical group, designated R and called the side chain, that uniquely defines each of the 20 common amino acids. Figure 2.2 depicts the ionized form of a common amino acid in solution at pH 7. The α-amino group is protonated and in its ammonium ion form; the carboxylic acid group is in its unprotonated or carboxylate ion form.

Side Chains Define Chemical Nature and Structures of Different Amino Acids

Structures of the common amino acids are shown in Figure 2.3. Alkyl amino acids have alkyl group side chains and include glycine, alanine, valine, leucine, and isoleucine. **Glycine** has the simplest structure, with R = H. **Alanine** contains a methyl (CH_3–) side chain group. **Valine** has an isopropyl R group (Figure 2.4). The leucine and isoleucine R groups are butyl groups that are structural isomers of each other. In **leucine** the branching in the isobutyl side chain occurs on the *gamma* (γ)-carbon of the amino acid. In isoleucine it is branched at the *beta* (β)-carbon.

The aromatic amino acids are phenylalanine, tyrosine, and tryptophan. The **phenylalanine** R group contains a benzene ring, **tyrosine** contains a phenol group, and the **tryptophan** R group contains the heterocyclic structure, indole.

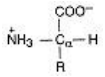


Figure 2.2
General structure of the common amino acids.

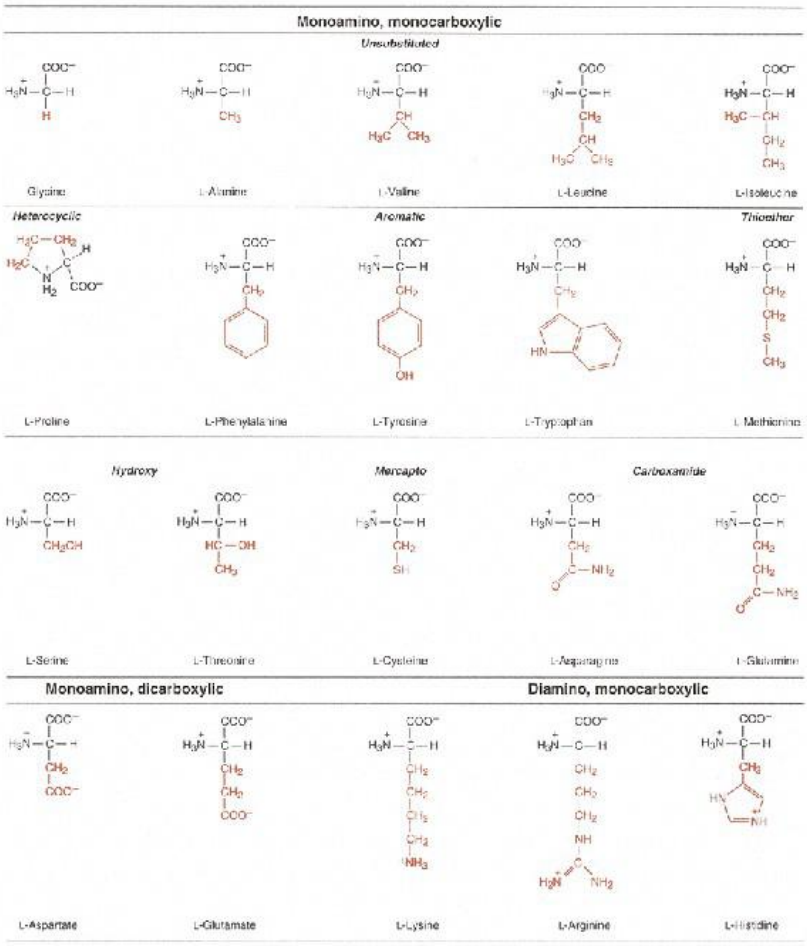


Figure 2.3
Structures of the common amino acids. Charge forms are those present at pH 7.0.

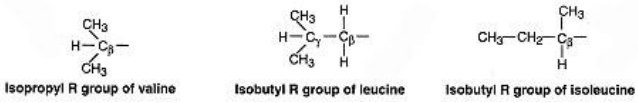


Figure 2.4
Alkyl side chains of valine, leucine, and isoleucine.

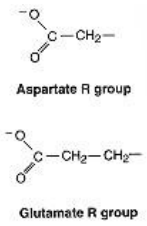


Figure 2.5
Side chains
of aspartate
and glutamate.

In each case the aromatic moiety is attached to the α-carbon through a methylene ($-CH_2-$) carbon (Figure 2.3).

Sulfur-containing common amino acids are cysteine and methionine. The **cysteine** side chain group is a thiolmethyl ($HSCH_2-$). In **methionine** the side chain is a methyl ethyl thiol ether ($CH_3SCH_2CH_2-$).

There are two hydroxy (alcohol)-containing common amino acids, serine and threonine. The **serine** side chain is a hydroxymethyl ($HOCH_2-$). In **threonine** an ethanol structure is connected to the α-carbon through the carbon containing the hydroxyl substituent, resulting in a secondary alcohol structure ($CH_3-CHOH-CH_\alpha-$).

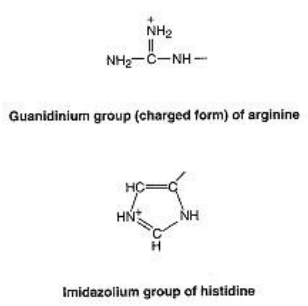


Figure 2.6
Guanidinium and imidazolium groups
of arginine and histidine.

The **proline** side chain is unique in that it incorporates the α-amino group. Thus proline is more accurately classified as an α-imino acid, since its α-amine is a secondary amine with its α-nitrogen having two covalent bonds to carbon (to the α-carbon and side chain carbon), rather than a primary amine. Incorporation of the α-amino nitrogen into a five-membered ring constrains the rotational freedom around the $-N_\alpha-C_\alpha-$ bond in proline to a specific rotational angle, which limits participation of proline in polypeptide chain conformations.

TABLE 2.1 Abbreviations for the Amino Acids

	Abbreviation	
Amino Acid	***Three Letter***	***One Letter***
Alanine	Ala	A
Arginine	Arg	R
Asparagine	Asn	N
Aspartic	Asp	D
Asparagine or aspartic	Asx	B
Cysteine	Cys	C
Glycine	Gly	G
Glutamine	Gln	Q
Glutamic	Glu	E
Glutamine or glutamic	Glx	Z
Histidine	His	H
Isoleucine	Ile	I
Leucine	Leu	L
Lysine	Lys	K
Methionine	Met	M
Phenylalanine	Phe	F
Proline	Pro	P
Serine	Ser	S
Threonine	Thr	T
Tryptophan	Trp	W
Tyrosine	Tyr	Y
Valine	Val	V

The amino acids discussed so far contain side chains that are uncharged at physiological pH. The **dicarboxylic monoamino acids** contain a carboxylic group in their side chain. **Aspartate** contains a carboxylic acid group separated by a methylene carbon ($-CH_2-$) from the α-carbon (Figure 2.5). In **glutamate** (Figure 2.5), the carboxylic acid group is separated by two methylene ($-CH_2-CH_2-$) carbon atoms from the α-carbon (Figure 2.2). At physiological pH, side chain carboxylic acid groups are unprotonated and negatively charged. **Dibasic monocarboxylic acids** include lysine, arginine, and histidine (Figure 2.3). In these structures, the R group contains one or two nitrogen atoms that act as a base by binding a proton. The **lysine** side chain is a *N*-butyl amine. In **arginine**, the side chain contains a guanidino group (Figure 2.6) separated from the α-carbon by three methylene carbon atoms. Both the guanidino group of arginine and the ε-amino group of lysine are protonated at physiological pH (pH~7) and in their charged form. In **histidine** the side chain contains a five-membered heterocyclic structure, the imidazole (Figure 2.6). The pK'_a of the imidazole group is approximately 6.0 in water; physiological solutions contain relatively high concentrations of both basic (imidazole) and acidic (imidazolium) forms of the histidine side chain (see Section 2.3).

The last two common amino acids are glutamine and asparagine. They contain an amide moiety in their side chain. **Glutamine** and **asparagine** are structural analogs of glutamic acid and aspartic acid with their side chain carboxylic acid groups amidated. Unique DNA codons exist for glutamine and asparagine separate from those for glutamic acid and aspartic acid. The amide side chains of glutamine and asparagine cannot be protonated and are uncharged at physiological pH.

In order to represent the sequence of amino acids in a protein, three-letter and one-letter abbreviations for the common amino acids have been established (Table 2.1). These abbreviations are universally accepted and will be used

throughout the book. The three-letter abbreviations of aspartic acid (Asp) and glutamic acid (Glu) should not be confused with those for asparagine (Asn) and glutamine (Gln). In experimentally determining the amino acids of a protein by chemical procedures, one cannot easily differentiate between Asn and Asp, or between Gln and Glu, because the side chain amide groups in Asn and Gln are hydrolyzed and generate Asp and Glu (see Section 2.9). In these cases, the symbols of Asx for Asp or Asn, and Glx for Glu or Gin indicate this ambiguity. A similar scheme is used with the one-letter abbreviations to symbolize Asp or Asn, and Glu or Gln.

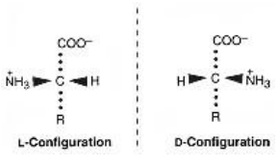


Figure 2.7
Absolute con figuration of an amino acid.

Amino Acids Have an Asymmetric Center

The common amino acids with the general structure in Figure 2.2 have four substituents (R, H, COO^-, NH_3^+) covalently bonded to the α-carbon atom in the α-amino acid structure. A carbon atom with four different substituents arranged in a tetrahedral configuration is asymmetric and exists in two enantiomeric forms. Thus each of the amino acids exhibits optical isomerism except glycine, in which R = H and thus two of the four substituents on the α-carbon atom are hydrogen. The absolute configuration for an amino acid is depicted in Figure 2.7 using the Fischer projection to show the direction in space of the tetrahedrally arranged α-carbon substituents. The α-COO^- group is directed up and behind the plane of the page, and the R group is directed down and behind the plane of the page. The α-H and α-NH_3^+ groups are directed toward the reader. An amino acid held in this way projects its α-NH_3^+ group either to the left or right of the α-carbon atom. By convention, if the α-NH_3^+ is projected to the left, the amino acid has an L absolute configuration. Its optical enantiomer, with α-NH_3^+ projected toward the right, has a D absolute configuration. In mammalian proteins only amino acids of L configuration are found. The L and D designations refer to the ability to rotate polarized light to the left (L, levo) or right (D, dextro) from its plane of polarization. As the amino acids in proteins are asymmetric, the proteins that contain them also exhibit asymmetric properties.

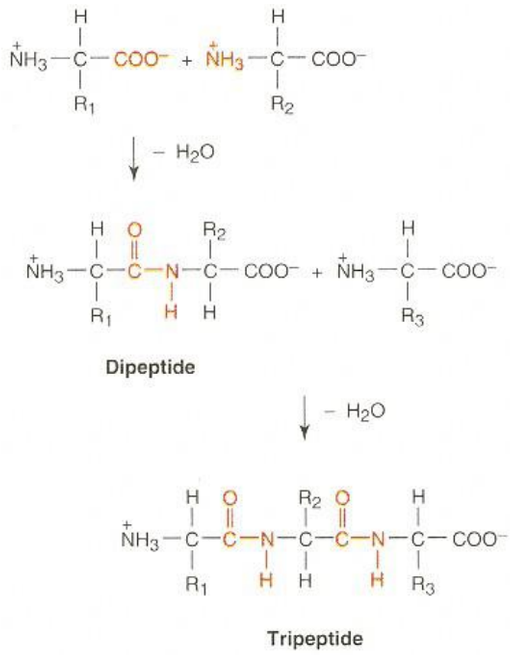


Figure 2.8
Peptide bond formation.

Amino Acids Are Polymerized into Peptides and Proteins

Polymerization of the 20 common amino acids into polypeptide chains in cells is catalyzed by enzymes and is associated with the ribosomes (Chapter 15). Chemically, this polymerization is a dehydration reaction (Figure 2.8). The α-carboxyl group of an amino acid with side chain R_1 forms a covalent **peptide bond** with the α-amino group of the amino acid with side chain R_2 by elimination of a molecule of water. The **dipeptide** (two amino acid residues joined by a single peptide bond) can then form a second peptide bond through its terminal carboxylic acid group and the α-amino of a third amino acid (R_3), to generate a tripeptide (Figure 2.8). Repetition of this process generates a **polypeptide** or protein of specific amino acid sequence (R_1-R_2-R_3-R_4-· · ·R_n). The amino acid sequence of the polypeptide chains is the **primary structure** of the protein, and it is predetermined by the DNA sequence of its gene (Chapter 14). It is the unique primary structure that enables a polypeptide chain to fold into a specific three-dimensional structure that gives the protein its chemical and physiological properties.

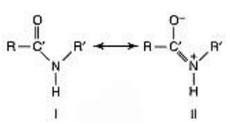
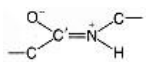


Figure 2.9
Electronic isomer
structures of a peptide bond.

A peptide bond can be represented using two **resonance isomers** (Figure 2.9). In structure I, a double bond is located between the carbonyl carbon and carbonyl oxygen (C =O), and the carbonyl carbon to nitrogen (C –N) linkage is a single bond. In structure II, the carbonyl carbon to oxygen bond (C –O^-) is a single bond and the bond located between the carbonyl carbon and nitrogen is a double bond (C =N). In structure II there is a negative charge on the oxygen and a positive charge on the nitrogen. Actual peptide bonds are a

resonance hybrid of these two electron isomer structures, the carbonyl carbon to nitrogen bond having a 50% double-bond character. The hybrid bond is supported by spectroscopic measurements and X-ray diffraction studies, the latter showing that the carbonyl carbon to nitrogen peptide bond length (1.33 Å) is approximately half-way between that found for a C–N single bond (~1.45 Å) and a C=N double bond (~1.25 Å).

A consequence of this partial double-bond character is that, as for normal double-bond structures, rotation does not occur about the carbonyl carbon to nitrogen of a peptide bond at physiological temperatures. Also, a consequence of the C =N double-bond's chemistry is that the atoms attached to C and N



all lie in a common plane. Thus a polypeptide chain is a polymer of peptide-bond planes interconnected at the α-carbon atoms. The α-carbon interconnects peptide bonds through single bonds that allow rotation of adjacent peptide planes with respect to each other. Each **amino acid residue** contributes one α-carbon (two single bonds and a peptide bond, Figure 2.10) to the polypeptide chain. The term residue refers to the atoms contributed by an amino acid to a polypeptide chain including the atoms of the side chain.

The peptide bond in Figure 2.11*a* shows a ***trans* configuration** between the oxygen (O) and the hydrogen (H) atoms of the peptide bond. This is the most stable configuration for the peptide bond with the two side chains (R and R) also in *trans*. The ***cis* configuration** (Figure 2.11*b*) brings the two side chain groups to the same side of the C =N bond, where unfavorable repulsive steric forces occur between the two side chain (R) groups. Accordingly, *trans*-peptide bonds are always found in proteins except where there are proline residues. In proline the side chain is linked to its α-amino group, and the *cis*- and *trans*-peptide bonds with the proline α-imino group have near equal energies. The configuration of the peptide bond actually found for a proline in a protein will depend on the specific forces generated by the unique folded three-dimensional structure of the protein molecule.

One of the largest natural polypeptide chains in humans is that of apolipo-protein B-100, which contains 4536 amino acid residues in one polypeptide chain. Chain length alone, however, does not determine the function of a polypeptide. Many small peptides with less than ten amino acids perform important biochemical and physiological functions in humans (Table 2.2). Primary structures are written in a standard convention and sequentially numbered from their NH_2-terminal end toward their COOH-terminal end, consistent with the order of addition of the amino acid to the chain during biosynthesis. Accordingly, for thyrotropin-releasing hormone (Table 2.2) the glutamic acid residue written on the left is the NH_2-terminal amino acid of the tripeptide and is designated amino acid residue 1 in the sequence. The proline is the COOH-terminal amino acid and is designated residue 3. The defined direction of the polypeptide chain is from Glu to Pro (NH_2-terminal amino acid to COOH-terminal amino acid).

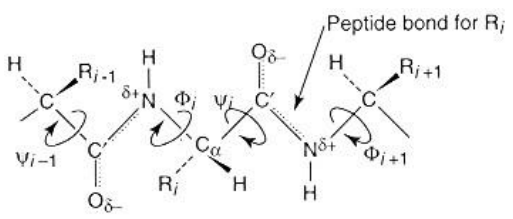


Figure 2.10
Amino acid residue.
Each amino acid residue of a polypeptide contributes two single bonds and one peptide bond to the chain. The single bonds are those between the C_α and carbonyl C′ atoms, and the C_α and N atoms. See p. 43 for definition of ϕ and ψ.

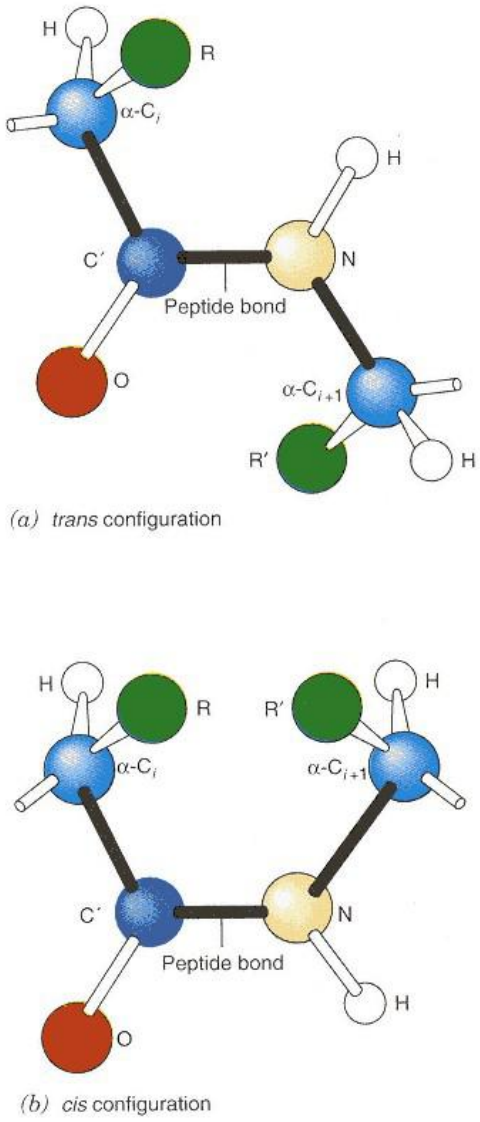


Figure 2.11
(a) *Trans*-peptide bond and
(b) the rare
***cis*-peptide bond.**
The C′–N have a partial double-bond character.

TABLE 2.2 Some Examples of Biologically Active Peptides

Amino Acid Sequence	*Name*	*Function*
1 3 pyroGlu-His-Pro(NH_2)[a]	Thyrotropin-releasing hormone	Secreted by hypothalamus; causes anterior pituitary gland to release thyrotropic hormone
1 9 H-Cys-Tyr-Phe-Gln-Asn-Cys-Pro-Arg-Gly(NH_2)[b] S——S	Vasopressin (antidiuretic hormone)	Secreted by posterior pituitary gland; causes kidney to retain water from urine
1 5 H-Tyr-Gly-Gly-Phe-Met-OH	Methionine enkephalin	Opiate-like peptide found in brain that inhibits sense of pain
1 10 pyroGlu-Gly-Pro-Trp-Leu-Glu-Glu-Glu-Glu- 11 17 Ala-Tyr-Gly-Trp-Met-Asp-Phe(NH_2)[a,c] SO_3	Little gastrin (human)	Hormone secreted by mucosal cells in stomach; causes parietal cells of stomach to secrete acid
1 10 H-His-Ser-Gln-Gly-Thr-Phe-Thr-Ser-Asp-Tyr-	Glucagon (bovine)	Pancreatic hormone involved in regulating glucose metabolism
11 20 Ser-Lys-Tyr-Leu-Asp-Ser-Arg-Arg-Ala-Gln-		
21 29 Asp-Phe-Val-Gln-Trp-Leu-Met-Asn-Thr-OH		
1 8 H-Asp-Arg-Val-Tyr-Ile-His-Pro-Phe-OH	Angiotensin II (horse)	Pressor or hypertensive peptide; also stimulates release of aldosterone from adrenal cortex
1 9 H-Arg-Pro-Pro-Gly-Phe-Ser-Pro-Phe-Arg-OH	Plasma bradykinin (bovine)	Vasodilator peptide
1 10 H-Arg-Pro-Lys-Pro-Gln-Phe-Phe-Gly-Leu-Met(NH_2)	Substance P	Neurotransmitter

[a] The NH_2 terminal Glu is in the pyro form in which its -COOH is covalently joined to its -NH_2 via amide linkage; the COOH terminal amino acid is amidated and thus also not free.

[b] Cysteine-1 and cysteine-6 are joined to form a disulfide bond structure within the nonapeptide.

[c] The Tyr 12 is sulfonated on its phenolic side chain OH.

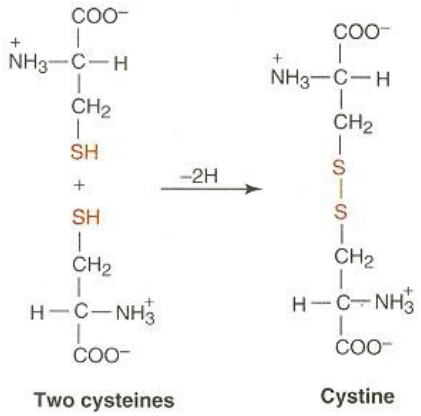


Figure 2.12
Cystine bond formation.

Cystine Is a Derived Amino Acid

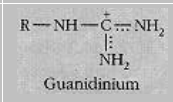
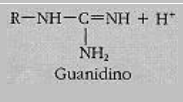
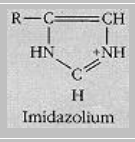
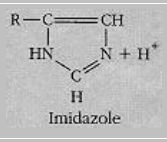
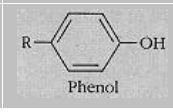
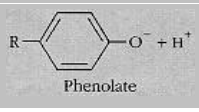
A derived amino acid found in many proteins is cystine. It is formed by the oxidation of two cysteine thiol side chains, joined to form a disulfide covalent bond (Figure 2.12). Within proteins disulfide links of cystine formed from cysteines, separated from each other in the primary structure, have an important role in stabilizing the folded conformation of proteins.

2.3— Charge and Chemical Properties of Amino Acids and Proteins

Ionizable Groups of Amino Acids and Proteins Are Critical for Biological Function

Ionizable groups common to proteins and amino acids are shown in Table 2.3. The acid forms are on the left of the equilibrium sign and the base forms on the right side. In forming its conjugate base, the acid form releases a proton. In reverse, the base form associates with a proton to form the respective acid. The proton dissociation of an acid is characterized by an acid **dissociation constant** pK'_a depends on the environment in which an acid group is placed. For example, when a

TABLE 2.3 Characteristic pK'_a Values for the Common Acid Groups in Proteins

Where Acid Group Is Found	*Acid Form*		*Base Form*	*Approximate pK_a Range for Group*
NH_2-terminal residue in peptides, lysine	$R–NH_{3+}$ Ammonium	⇌	$R–NH_2 + H^+$ Amine	7.6–10.6
COOH-terminal residue in peptides, glutamate, aspartate	R–COOH Carboxylic acid	⇌	$R–COO^- + H^+$ Carboxylate	3.0–5.5
Arginine	 Guanidinium	⇌	 Guanidino	11.5–12.5
Cysteine	R–SH Thiol	⇌	$R–S^- + H^+$ Thiolate	8.0–9.0
Histidine	 Imidazolium	⇌	 Imidazole	6.0–7.0
Tyrosine	 Phenol	⇌	 Phenolate	9.5–10.5

positive-charged ammonium group ($–NH_3^+$) is placed near a negatively charged group within a protein, the negative charge stabilizes the positively charged acid form of the amino group, making it more difficult to dissociate its proton. The pK'_a values and are called **acidic amino acids.** They are predominantly in their unprotonated forms and are negatively charged at physiological pH. Proteins in which the ratio (Lys + Arg)/(Glu + Asp) is greater than 1 are referred to as **basic proteins.** Proteins in which the above ratio is less than 1 are referred to as **acidic proteins.**

TABLE 2.4 pK'_a of Side Chain and Terminal Acid Groups in Protein Ribonuclease

	$—NH_3^+$	*—COOH*
Side chain	Lysines ≃ 4.6	
Chain end	N-terminal = 7.8	C-terminal = 3.8

Ionic Form of an Amino Acid or Protein Can Be Determined at a Given pH

$$pH = pK_a + \log \frac{[\text{conjugate base}]}{[\text{conjugate acid}]}$$

or

$$pH - pK_a = \log \frac{[\text{conjugate base}]}{[\text{conjugate acid}]}$$

Figure 2.13
Henderson–Hasselbalch equation.
For a more detailed discussion
of this equation, see p. 9.

From a knowledge of the pK'_a and the ratio of [imidazole]/[imidazolium] is 10:1 (Table 2.5). Based on this ratio, the enzyme exhibits 10/(10 + 1) × 100 = 91% of its maximum potential activity. Thus a change in pH has a dramatic effect on the enzyme's activity. Most protein activities demonstrate similar pH dependency due to their acid and base group(s).

Titration of a Monoamino Monocarboxylic Acid: Determination of the Isoelectric pH

An understanding of a protein's acid and base forms and their relation to charge is made more clear by following the titration of the ionizable groups for a simple amino acid. As presented in Figure 2.14, leucine contains an α-COOH with $pK'_a = 9.6$. At pH 1.0 the predominant ionic form (form I) has a charge of +1 and migrates toward the cathode in an

TABLE 2.5 Relationship Between the Difference of pH and Acid pK'_a and the Ratio of the Concentrations of Base to Its Conjugate Acid

pH – pK'_a (Difference Between pH and pK'_a	*Ratio of Concentration of Base to Conjugate Acid*
0	1
1	10
2	100
3	1000
–1	0.1
–2	0.01
–3	0.001

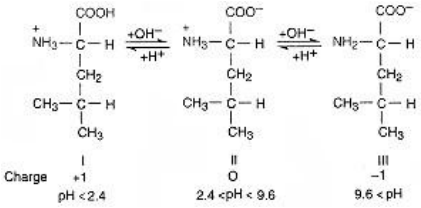


Figure 2.14
Ionic forms of leucine.

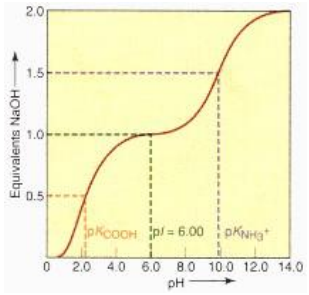


Figure 2.15
Titration curve of leucine.

electrical field. The addition of 0.5 equivalent of base half-titrates the α-COOH group of the leucine; that is, the ratio of $[COO^-]/[COOH]$ will equal 1. The Henderson–Hasselbalch equation, with the second term on the right side of the equation $\log_{10}[(\text{base})/(\text{acid})] = \log_{10}[1] = 0$ at a ratio of conjugate base to acid of 1 : 1, shows that the pH (when the α-COOH is half-titrated) is directly equal to the $pK_{a(\alpha\text{-COOH})}$ (Figure 2.15).

Addition of 1 equivalent of base completely titrates the α-COOH but leaves the α-NH_3^+ group intact. In the resulting form (II), the negatively charged α-COO^- and positively charged α-NH_3^+ cancel each other and the net charge of this ionic form is zero. Form II is thus the **zwitterion** form, that is, the ionic form in which the total of positive charges is exactly equal to the total of the negative charges. As the net charge on a zwitterion molecule is zero, it will not migrate toward either the cathode or anode in an electric field. Further addition of 0.5 equivalent of base to the zwitterion form of leucine (total base added is 1.5 equivalents) will then half-titrate the α-NH_3^+ group. At this point in the titration, the ratio of $[NH_2]/[NH_3^+] = 1$, and the pH is equal to the value of the pK'_a for the α-NH_3^+ group (Figure 2.15). Addition of a further 0.5 equivalent of base (total of 2 full equivalents of base added; Figure 2.15) completely titrates the α-NH_3^+ group to its base form (α-NH_2). The solution pH is greater than 11, and the predominant molecular species has a negative charge of –1 (form III).

It is useful to calculate the exact pH at which an amino acid is electrically neutral and in its zwitterion form. This pH is known as the **isoelectric pH** for the molecule, and the symbol is **p*I***. The p*I* value is a constant of a compound at a particular ionic strength and temperature. For simple molecules, such as leucine, p*I* is directly calculated as the average of the two pK'_a values that regulate the boundaries of the zwitterion form. Leucine has two ionizable groups that regulate the zwitterion form boundaries, and the p*I* is calculated as follows:

$$pI = \frac{pK'_a\text{COOH} + pK'_a\text{NH}_3^+}{2} = \frac{2.4 + 9.6}{2} = 6.0$$

At pH > 6.0, leucine assumes a partial negative charge that formally rises at high pH to a full negative charge of –1 (form III) (Figure 2.14). At pH < 6, leucine has a partial positive charge until at very low pH it has a charge of +1 (form I) (Figure 2.14). The partial charge at any pH can be calculated from the Henderson–Hasselbalch equation or from extrapolation from the titration curve of Figure 2.15.

Titration of a Monoamino Dicarboxylic Acid

A more complicated example of the relationship between molecular charge and pH is provided by glutamic acid. Its ionized forms and titration curve are

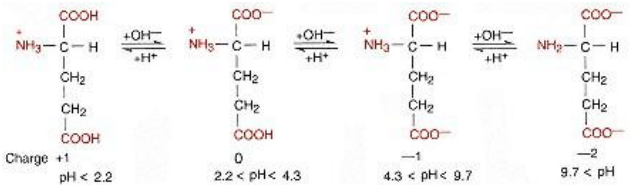


Figure 2.16
Ionic forms of glutamic acid.

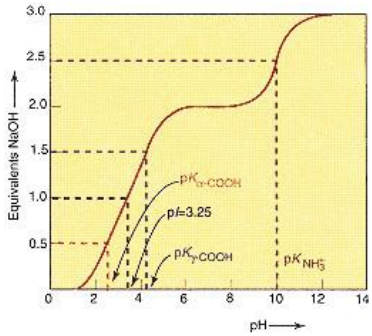


Figure 2.17
Titration curve of glutamic acid.

shown in Figures 2.16 and 2.17. In glutamic acid the α-COOH pK'_a values that control the boundaries of the zwitterion form:

$$pI = \frac{2.2 + 4.3}{2} = 3.25$$

Accordingly, at values above pH 3.25 the molecule assumes a net negative charge until at high pH the molecule has a net charge of –2. At pH < 3.25 glutamic acid is positively charged, and at extremely low pH it has a net positive charge of +1.

General Relationship between Charge Properties of Amino Acids and Proteins and pH

Analysis of charge forms present in other common amino acids shows that the relationship observed between pH and charge for leucine and glutamate is generally true. That is, at a solution pH less than p*I*, the amino acid is positively charged. At a solution pH greater than p*I*, the amino acid is negatively charged. The degree of positive or negative charge is a function of the magnitude of the difference between pH and p*I*. As a protein is a complex polyelectrolyte containing multiple ionizable acid groups that regulate the boundaries of its zwitterion form, calculation of a protein's isoelectric pH from its acid pK'_a values utilizing the Henderson–Hasselbalch relationship would be difficult. Accordingly, the p*I* values for proteins are always experimentally measured by determining the pH value at which the protein does not move in an electrical field. p*I* values for some representative proteins are given in Table 2.6.

TABLE 2.6 p*I* Values for Some Representative Proteins

Protein	*pI*
Pepsin	~1
Human serum albumin	5.9
α_1-Lipoprotein	5.5
Fibrinogen	5.8
Hemoglobin A	7.1
Ribonuclease	7.8
Cytochrome-*c*	10.0
Thymohistone	10.6

pH > p*I*, then protein charge negative
pH < p*I*, then protein charge positive

Figure 2.18
Relationship between solution pH, protein p*I*, and protein charge.

As with the amino acids, at a pH greater than the p*I*, a protein has a net negative charge. At a pH less than the p*I*, a protein has a net positive charge (Figure 2.18). The magnitude of the net charge of a protein increases as the difference between pH and p/increases. An example is human plasma albumin with 585 amino acid residues of which there are 61 glutamates, 36 aspartates, 57 lysines, 24 arginines, and 16 histidines. The titration curve for this complex molecule is shown in Figure 2.19. Albumin's p*I* = 5.9, at which pH its net charge is zero. At pH 7.5 the imidazolium groups of histidines have been partially titrated and albumin has a negative charge of –10. At pH 8.6 additional groups have been titrated to their base forms, and the net charge is approximately –20. At pH 11 the net charge is approximately –60. On the acid side of the p*I* value, at pH 3, the approximate net charge of albumin is +60.

Amino Acids and Proteins Can Be Separated Based on pI Values

The techniques of electrophoresis, isoelectric focusing, and ion-exchange chromatography separate and characterize biological molecules on the basis of differences in their p*I* (see p. 34). In clinical medicine, separation of plasma proteins by electrophoresis has led to the classification of the proteins based on their relative electrophoretic mobility. The separation is commonly carried out at pH 8.6, which is higher than the p*I* values of the major plasma proteins.

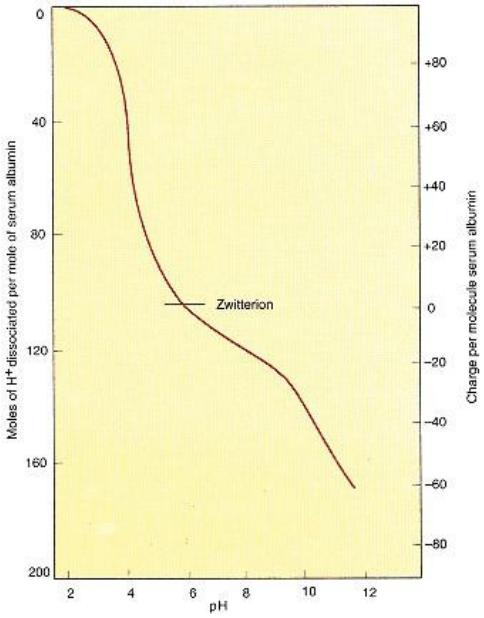


Figure 2.19
Titration curve of human serum albumin at 25°C and an ionic strength of 0.150.
Redrawn from Tanford, C. *J. Am. Chem. Soc.* 72:441, 1950.

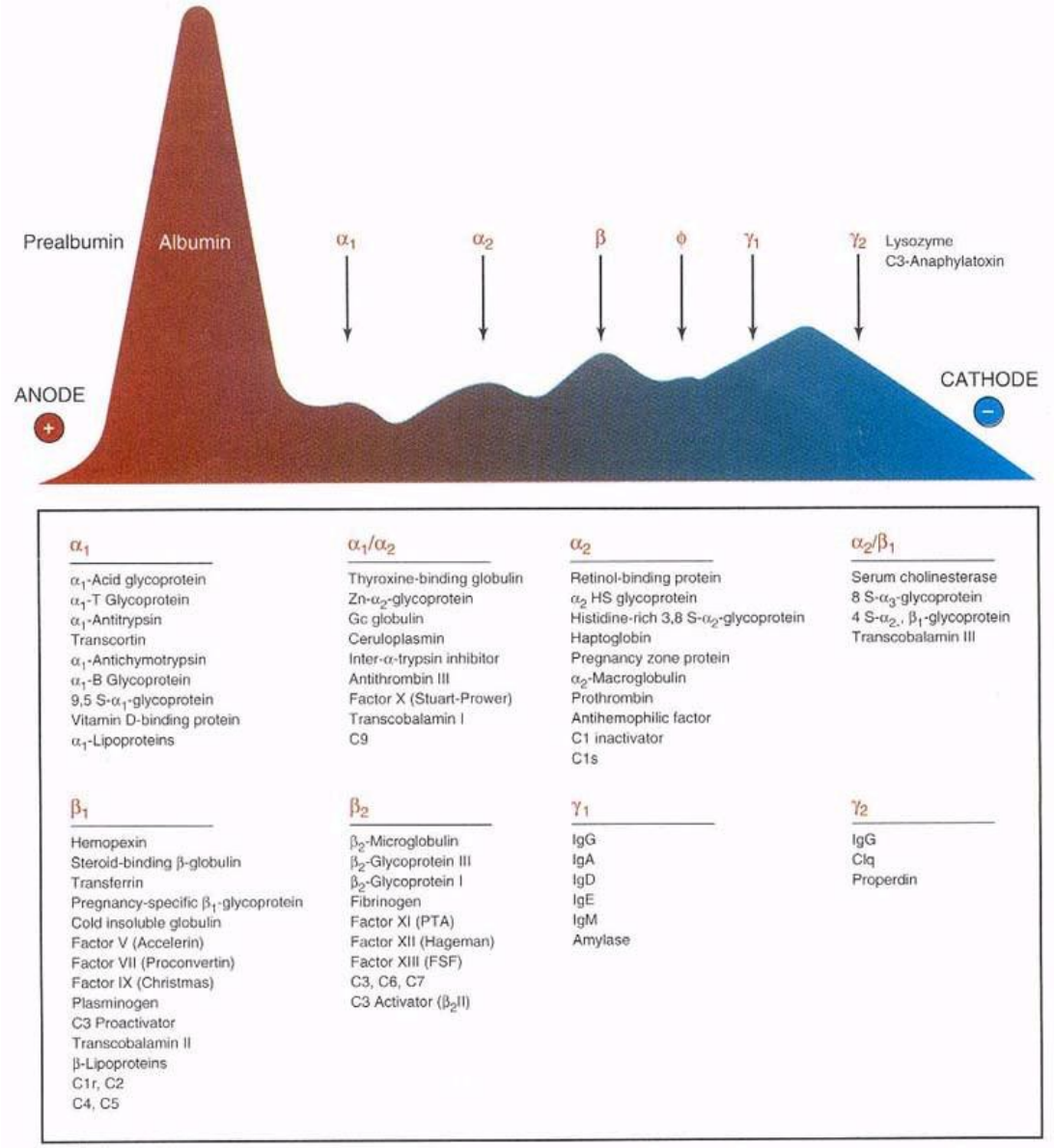


Figure 2.20
Electrophoresis pattern for plasma proteins at pH 8.6.
Plot shows the order of migration along the horizontal axis with proteins of highest mobility closest to the anode. Height of the band along the vertical axis shows the protein concentration. Different major proteins are designated underneath their electrophoretic mobility peaks.
Reprinted with permission from Heide, K., Haupt, H., and Schwick, H. G. In: F. W. Putnam (Ed.), *The Plasma Proteins*, 2nd ed., Vol. III. New York Academic Press, 1977, p. 545.

Accordingly, the proteins are negatively charged and move toward the anode at a rate dependent on their net charge. Major peaks observed in order of their migration are those of albumin, α_1^-, α_2^-, and β-globulins, fibrinogen, and the γ_1^- and γ_2^--globulins (Figure 2.20). Some of these peaks represent tens to hundreds of different plasma proteins that have a similar migration rate at pH 8.6. However, certain proteins predominate in each peak and variation in their relative amounts is characteristic of certain diseases (Figures 2.20 and 2.21; see Clin. Corr. 2.1).

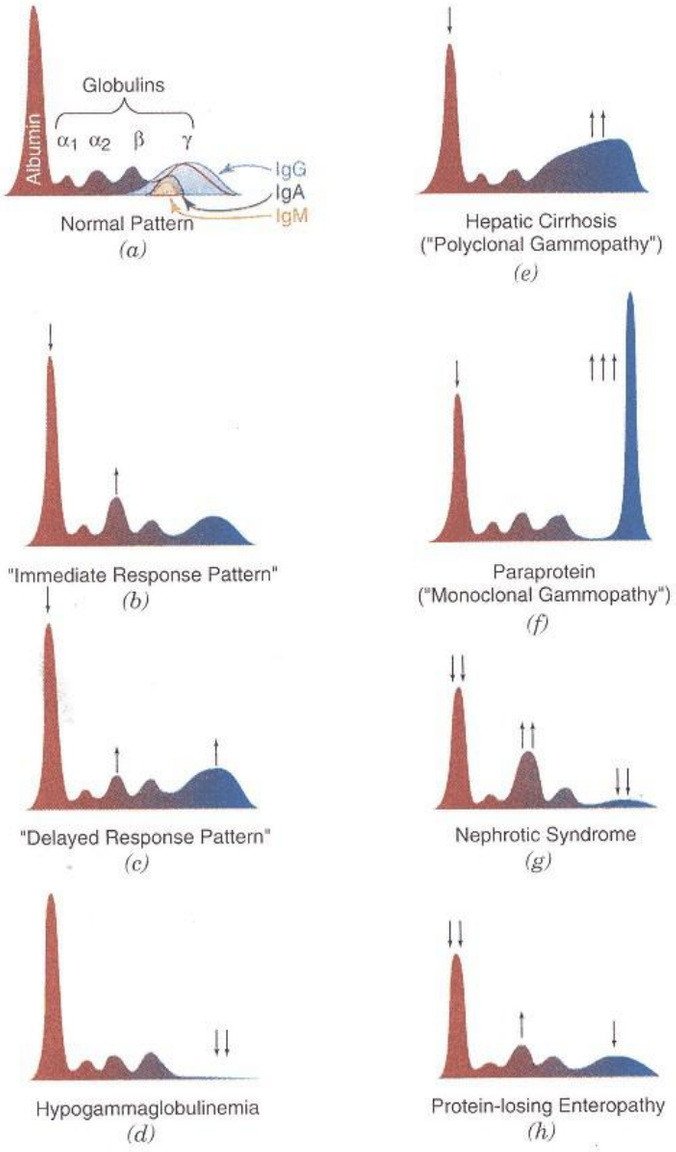


Figure 2.21
Examples of the electrophoretic mobility patterns observed for a normal individual and patients with abnormal concentrations of serum proteins, analyzed by agarose gel electrophoresis.
Redrawn from McPherson, R. A. Specific proteins. In: J. B. Henry (Ed.), *Clinical Diagnosis and Management*, 17th ed. Philadelphia: Saunders Co, 1984.

CLINICAL CORRELATION 2.1

Plasma Proteins in Diagnosis of Disease

Electrophoretic analysis of the plasma proteins is commonly used in diagnosis of disease. Electrophoresis of plasma buffered at pH 8.6 separates the major plasma proteins as they migrate to the anode in the electric field into bands or peaks, based on their charge differences (see text). Examples of abnormal electrophoresis patterns are shown in Figure 2.21. An "immediate response" that occurs with stress or inflammation caused by infection, injury, or surgical trauma is shown in pattern (*b*) in which haptoglobins in the α_2 mobility hand are selectively increased. A "late response" shown in pattern (*c*) is correlated with infection and shows an increase in the τ-globulin peaks due to an increase in immunoglobulins. An example of a hypogammaglobulinemia due to an immunosuppressive disease is shown in pattern (*d*). In hepatic cirrhosis there is a broad elevation of the τ-globulins with reduction of albumin, as in pattern (*e*). Monoclonal gammopathies are due to the clonal synthesis of a unique immunoglobulin and give rise to a sharp τ-globulin band, as in pattern (*f*). Nephrotic syndrome shows a selective loss of lower molecular weight proteins from plasma, as in pattern (*g*). The pattern shows a decrease in albumin (65 kDa), but a retention of the bands composed of the higher molecular weight proteins α_2-macroglobulin (725 kDa) and β-lipoproteins (2000 kDa) in the α_2 band. Pattern (*h*) is from a patient with a protein-losing enteropathy. The slight increase in the α_{2-} band in pattern (*h*) is due to an immediate or late response from a stressful stimulus, as previously observed in patterns (*b*) and (*c*).

Ritzmann, S. E., and Daniels, J. C. Serum protein electrophoresis and total serum proteins. In: S. E. Ritzmann and J. C. Daniels (Eds.), *Serum Protein Abnormalities, Diagnostic and Clinical Aspects*. Boston: Little, Brown and Co., 1975, pp. 3–25; and McPherson, R. A. Specific proteins. In: J. B. Henry (Ed.), *Clinical Diagnosis and Management by Laboratory Methods*, 17th ed. Philadelphia: Saunders, 1984, pp. 204–215.

Amino Acid Side Chains Have Polar or Apolar Properties

The relative **hydrophobicity** of amino acid side chains is critical for the folding of a protein to its native structure and for the stability of the folded protein. Figure 2.22 plots the values of relative hydrophobicity of the common amino acids based on the tendency of each amino acid to partition itself in a mixture of water and a nonpolar solvent. The scale is based on a value of zero for glycine. The side chains that preferentially dissolve in the nonpolar solvent relative to glycine show a positive (+) hydrophobicity value, the more positive the greater the preference for the nonpolar solvent. Most hydrophobic are those amino acids found buried in folded protein structures away from the water solvent that interacts with the surface of a soluble protein. However, the general correlation is not perfect due to the amphoteric nature of many of the hydrophobic amino acids that place the more polar portions of their side chain structure near the surface to interact with the polar solvent water on the outside. In addition, contrary to expectation, not all hydrophobic side chains are in a buried position in a folded three-dimensional structure of a globular protein. When on the surface, the hydrophobic groups are generally dispersed among the polar side chains. When clustering of nonpolar side chains occurs on the surface, it is usually associated with a function of the protein, such as to provide a site for binding of substrate molecules through hydrophobic interactions.

Most charged side chains are found on the surface of soluble globular proteins where they are stabilized by favorable energetic interactions with the water solvent. The rare positioning of a charged side chain in the interior of a globular protein usually implies an important functional role for that "buried" charge within the nonpolar interior in stabilizing conformation of the folded protein or participation in a catalytic mechanism.

Amino Acids Undergo a Variety of Chemical Reactions

Amino acids in proteins undergo a variety of chemical reactions with reagents that may be used to investigate the function of specific side chains. Some common chemical reactions are presented in Table 2.7. Reagents for amino acid side chain modification have also been synthesized that bind to specific sites in a folded protein's structure, like the substrate-binding site. The strategy

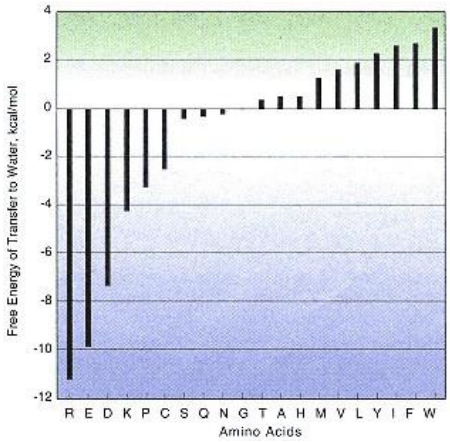


Figure 2.22
Relative hydrophobicity of the amino acid side chains.
Based on the partition of the amino acid between organic solvent and water. Negative values indicate preference for water and positive values preference for nonpolar solvent (ethanol or dioxane) relative to glycine (see text). Based on data from Von Heijne, G., and Blomberg, C. *Eur. J. Biochem*. 97:175, 1979; and from Nozaki, Y., and Tanford, C. *J. Biol. Chem.* 246:2211, 1971.

TABLE 2.7 Some Chemical Reactions of the Amino Acids

Reactive Group	*Reagent or Reaction*	*Product*
Amine ($-NH_2$) groups	Ninhydrin	Blue colored product that absorbs at 540 nm[a]
	Fluorescamine	Product that fluoresces
Carboxylic acid groups	Alcohols	Ester products
	Amines	Amide products
	Carbodiimide	Activates for reaction with nucleophiles
$-NH_2$ of Lys	2,3,6-Trinitrobenzene sulfonate	Product that absorbs at 367 nm
	Anhydrides	Acetylates amines
	Aldehydes	Forms Shiff base adducts
Guanidino group of Arg	Sakaguchi reaction	Pink-red product that can be used to assay Arg
Phenol of Tyr	I_2	Iodination of positions ortho to hydroxyl group on aromatic ring
	Acetic anhydride	Acetylation of –OH
S atom of Met side chain	CH_3I	Methyl sulfonium product
	$[O^-]$ or H_2O_2	Methionine sulfoxide or methionine sulfone
–SH of Cys	Iodoacetate	Carboxymethyl thiol ether product
	N-Ethylmaleimide	Addition product with S
	Organic mercurials	Mercurial adducts
	Performic acid	Cysteic acid ($-SO_3H$) product
	Dithionitrobenzoic acid	Yellow product that can be used to quantitate –SH groups
Imidazole of His and phenol of Tyr	Pauly's reagent	Yellow to reddish product

[a] Proline imino group reacts with ninhydrin to form product that absorbs light at 440 nm (yellow color).

is to model the structural features of the enzyme's natural substrate into the modifying reagent. The reagent binds to the active site like a natural substrate and, while within the active site, reacts with a specific side chain in the enzyme active site. This identifies the modified amino acid as being located in the substrate-binding site and helps identify its role in the catalytic mechanism.

2.4— Primary Structure of Proteins

The **primary structure** (amino acid sequence) of a protein is required for an understanding of a protein's structure, its mechanism of action at a molecular level, and its relationship to other proteins with similar physiological roles. The primary structure of **insulin** illustrates the value of this knowledge for understanding a protein's biosynthesis and physiological forms. Insulin is produced in pancreatic islet cells as a single chain precursor, **proinsulin,** with the primary structure shown in Figure 2.23. The polypeptide chain contains 86 amino acids and 3 intrachain cystine disulfide bonds. It is transformed into biologically active insulin by proteolytic modifications in its primary structure as it is secreted from the islet cells. Proinsulin is cleaved by proteases present in the islet cells that cleave two peptide bonds in proinsulin between residues 30 and 31 and 65 and 66. This releases a 35 amino acid segment (the **C-peptide**)

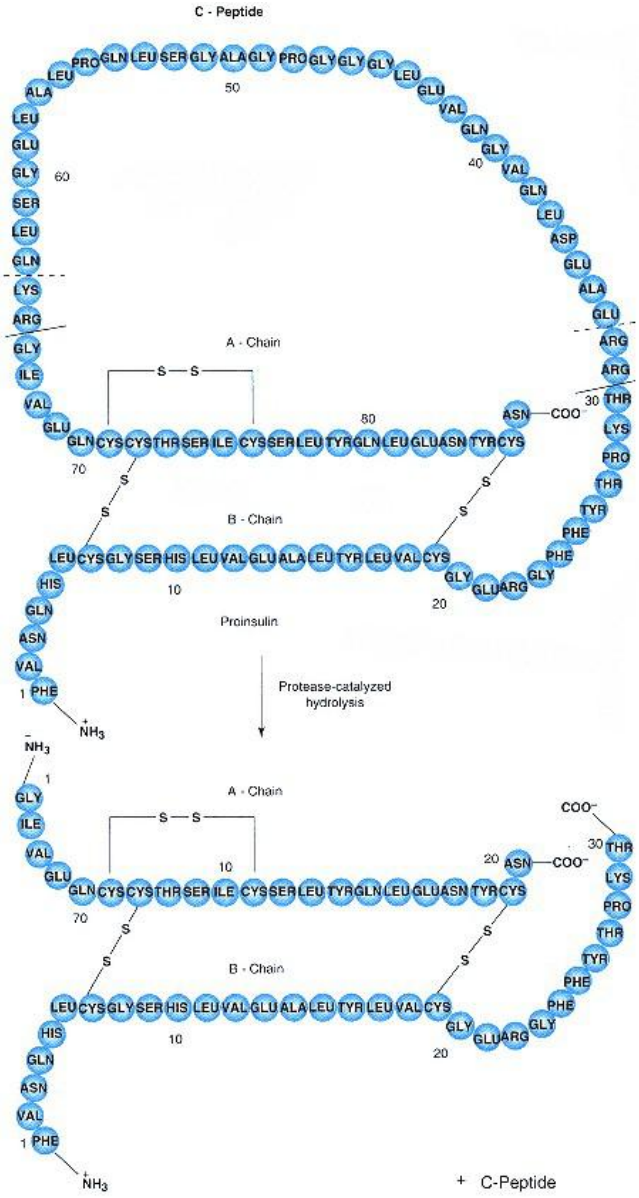


Figure 2.23
Primary structures of human proinsulin, insulin, and C-peptide.
In proinsulin, the B-chain peptide extends from Phe at position 1 to Thr at position 30, the C-peptide from Arg at position 31 to Arg at position 65, and the A-chain peptide from Gly at position 66 to Asn at position 86. Cystine bonds from positions 7 to 72, 19 to 85, and 71 to 76 are found in proinsulin.
Redrawn from Bell, G. I., Swain, W. F., Pictet, R., Cordell, B., Goodman, H. M., and Rutter, W. J. *Nature* 282:525, 1979.

and the active insulin molecule. The active insulin consists of two polypeptide chains (A and B) of 21 amino acids and 30 amino acids, respectively, covalently joined by the same disulfide bonds present in proinsulin (Figure 2.23). The C-

TABLE 2.8 Variation in Positions A8, A9, A10, and B30 of Insulin

Species	*A8*	*A9*	*A10*	*B30*
Human	Thr	Ser	Ile	Thr
Cow	Ala	Ser	Val	Ala
Pig	Thr	Ser	Ile	Ala
Sheep	Ala	Gly	Val	Ala
Horse	Thr	Gly	Ile	Ala
Dog	Thr	Ser	Ile	Ala
Chicken[a]	His	Asn	Thr	Ala
Duck[a]	Glu	Asn	Pro	Thr

[a] Positions 1 and 2 of B chain are both Ala in chicken and duck; whereas in the other species in the table, position 1 is Phe and position 2 is Val in B chain.

peptide is further processed in the pancreatic islet cells by proteases that hydrolyze a dipeptide from the COOH terminal and a second dipeptide from the NH_2 terminal of the C-peptide. The modified C-peptide is secreted into the blood with the active insulin. Besides giving information on the pathway for formation of active insulin, knowledge of primary structures shows the role of particular amino acids in the structure of insulin through comparison of the primary structures of the insulins from different animal species. The aligned primary structures show a residue identity in most amino acid positions, except for residues 8, 9, and 10 of the A chain and residue 30 of the B chain. Amino acids in these positions vary widely in different animal insulins (Table 2.8) and apparently do not affect the biological properties of the insulin molecule (see Clin. Corr. 2.2). Other amino acids of the primary structure are rarely substituted, suggesting that they have an essential role in insulin function.

Comparison of primary structures is commonly used to predict the similarity in structure and function between proteins. Sequence comparisons typically require aligning sequences to maximize the number of identical residues while minimizing the number of insertions or deletions required to achieve this alignment. Two sequences are termed **homologous** when their sequences are highly alignable. In its correct usage homology only refers to proteins that have evolved from the same gene. **Analogy** is used to describe sequences from proteins that are structurally similar but for which no evolutionary relationship has been demonstrated. Substitution of an amino acid by another amino acid of similar

CLINICAL CORRELATION 2.2

Differences in Primary Structure of Insulins Used in Treatment of Diabetes Mellitus

Both pig (porcine) and cow (bovine) insulins are commonly used in the treatment of human diabetics. Because of the differences in amino acid sequence from the human insulin, some diabetic individuals will have an initial allergic response to the injected insulin as their immunological system recognizes the insulin as foreign, or develop an insulin resistance due to a high anti-insulin antibody titer at a later stage in treatment. However, the number of diabetics who have a deleterious immunological response to pig and cow insulins is small; the great majority of human diabetics can utilize the nonhuman insulins without immunological complication. The compatibility of cow and pig insulins in humans is due to the small number and the conservative nature of the changes between the amino acid sequences of the insulins. These changes do not significantly perturb the three-dimensional structure of the insulins from that of human insulin. Pig insulin is usually more acceptable than cow insulin in insulin-reactive individuals because it is more similar in sequence to human insulin (see Table 2.8). Human insulin is now available for clinical use. It can be made using genetically engineered bacteria or by modifying pig insulin.

Brogdon, R. N., and Heel, R. C. Human insulin: a review of its biological activity, pharmacokinetics, and therapeutic use. *Drugs* 34:350, 1987.

CLINICAL CORRELATION 2.3

A Nonconservative Mutation Occurs in Sickle Cell Anemia

Hemoglobin S (HbS) is a variant form of the normal adult hemoglobin in which a nonconservative substitution occurs in the sixth position of the β-polypeptide chain of the normal hemoglobin (HbA_1). Whereas in HbA_1 this position is taken by a glutamic acid residue, in HbS the position is occupied by a valine. Consequently, in HbS a polar side chain group on the molecule's outside surface has been replaced with a nonpolar hydrophobic side chain (a nonconservative mutation). Through hydrophobic interactions with this nonpolar valine, HbS in its deoxy conformation polymerizes with other molecules of deoxy-HbS, leading to a precipitation of the hemoglobin within the red blood cell. This precipitation makes the red blood cell assume a sickle shape that results in a high rate of hemolysis and a lack of elasticity during circulation through the small capillaries, which become clogged by the abnormal shaped cells.

Only individuals homozygous for HbS exhibit the disease. Individuals heterozygous for HbS have approximately 50% HbA_1 and 50% HbS in their red blood cells and do not exhibit symptoms of the sickle cell anemia disease except under extreme conditions of hypoxia.

Individuals heterozygous for HbS have a resistance to the malaria parasite, which spends a part of its life cycle in red blood cells. This is a factor selecting for the HbS gene in malarial regions of the world and is the reason for the high frequency of this lethal gene in the human genetic pool. Approximately 10% of American blacks are heterozygous for HbS, and 0.4% of American blacks are homozygous for HbS and exhibit sickle cell anemia.

HbS is detected by gel electrophoresis. Because it lacks a glutamate, it is less acidic than HbA. HbS therefore does not migrate as rapidly toward the anode as does HbA. It is also possible to diagnose sickle cell anemia by recombinant DNA techniques.

Embury, S. H. The clinical pathophysiology of sickle-cell disease. *Annu. Rev. Med.* 37:361, 1986.

polarity (i.e., Val for Ile in position 10 of insulin) is called a **conservative** substitution and is commonly observed in amino acid sequences of the same protein from different animal species. If a particular amino acid is always found at the same position in these comparisons, then these are designated **invariant residues** and it can be assumed that these residues have an essential role in the structure or function of the protein. In contrast, a **nonconservative** substitution involves replacement of an amino acid by another of dramatically different polarity. This may produce severe changes in the properties of the resultant protein or occur in regions that are apparently unimportant functionally (see Clin. Corr. 2.3). Polarity is only one physical property of amino acids that determines whether a substitution will significantly alter the protein's function. Other physical properties of importance are the volume and surface area.

2.5— Higher Levels of Protein Organization

Primary structure of a protein refers to the covalent structure of a protein. It includes amino acid sequence and location of disulfide (cystine) bonds. Higher levels of protein organization refer to noncovalently generated conformational properties of the primary structure. These higher levels of protein conformation and organization are defined as the secondary, tertiary, and quaternary structures of a protein. **Secondary structure** refers to the local three-dimensional folding of the polypeptide chain in the protein. The polypeptide chain in this context is the covalently interconnected atoms of the peptide bonds and α-carbon linkages that sequentially link the amino acid residues of the protein. Side chains are not considered at the level of secondary structure. **Tertiary structure** refers to the three-dimensional structure of the polypeptide. It includes the conformational relationships in space of the side chains and the geometric relationship between distant regions of the polypeptide chain. **Quaternary structure** refers to the structure and interactions of the noncovalent association of discrete polypeptide subunits into a multisubunit protein. Not all proteins have a quaternary structure.

Proteins generally assume unique secondary, tertiary, and quaternary conformations as determined by their particular amino acid sequence and termed the **native conformation.** Folding of the primary structure into the native

conformation occurs, in most cases, spontaneously through noncovalent interactions. This unique conformation is the one of lowest total Gibbs free energy kinetically accessible to the polypeptide chain(s) for the particular conditions of ionic strength, pH, and temperature of the solvent in which the folding occurs. Chaperone proteins may facilitate the rate of protein folding.

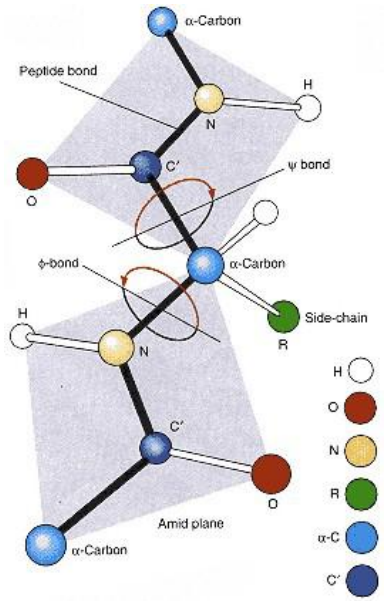


Figure 2.24
Polypeptide chain showing ϕ, ψ, and peptide bonds for residue R_i within a polypeptide chain.
Redrawn with permission from Dickerson, R. E., and Geis, I. *The Structure and Action of Proteins*. Menlo Park, CA: Benjamin, 1969, p. 25.

Proteins Have a Secondary Structure

The conformation of a polypeptide chain may be described by the rotational angles of the covalent bonds that contribute to the polypeptide chain. These are the bonds contributed by each of the amino acids between (1) the nitrogen and α-carbon and (2) the α-carbon and the carbonyl carbon. The first of these is designated the **phi (ϕ) bond** and the second is called the **psi (ψ) bond** for an amino acid residue in a polypeptide chain (Figure 2.24). The third bond contributed by each amino acid to the polypeptide chain is the peptide bond. As previously discussed, due to the partial double-bond character of the C′–N bonds, there is a barrier to free rotation about this peptide bond.

Regular secondary structure conformations in segments of a polypeptide chain occur when all ϕ bond angles in that polypeptide segment are equal, and all the ψ bond angles are equal. The rotational angles for ϕ and ψ bonds for common regular secondary structures are given in Table 2.9.

The α-helix and β-structure conformations for polypeptide chains are the most thermodynamically stable of the regular secondary structures. However, a particular sequence may form regular conformations other than α-helical or β-structure. There are also regions of unordered secondary structure, in which neither the ϕ bond angles nor the ψ bond angles are equal. Proline interrupts α-helical conformations since the pyrrolidine side chain of proline sterically interacts with the amino acid preceding it in the polypeptide sequence when in an α-helical structure. This repulsive steric interaction tends to prevent formation of α-helical structure in sections of a polypeptide chain that contain proline.

Helical structures of polypeptide chains are characterized by the number of amino acid residues per turn of helix (n) and the distance between α-carbon atoms of adjacent amino acids measured parallel to the axis of the helix (d). The **helix pitch** (p), defined as the product of $n \times d$, then measures the distance between repeating turns of the helix on a line drawn parallel to the helix axis (Figure 2.25):

$$p = n \times d$$

α-Helical Structure

An amino acid sequence in an α-helical conformation is shown in Figure 2.26.

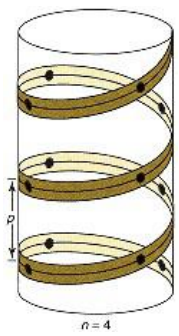


Figure 2.25
The helix pitch (p) for a helix with $n = 4$.
Each circle on a line represents an -carbon from an amino acid residue. The rise per residue would be p/n (see equation in text). From Dickerson, R. E., and Geis, I. *The Structure and Action of Proteins*. Menlo Park, CA: Benjamin, 1969, p. 26.

TABLE 2.9 Helix Parameters of Regular Secondary Structures

Structure	*Approximate Bond Angles* (°)		*Residues per turn, n*	*Helix Pitch,[a] p (A)*
	ϕ	***ψ***		
Right-handed α-helix [3.6_{13}-helix)	−57	−47	3.6	5.4
3_{10}-helix	+49	−26	3.0	6.0
Parallel β-strand	−119	+113	2.0	6.4
Antiparallel β-strand	−139	+135	2.0	6.8
Polyproline type II[b]	−78	+149	3.0	9.4

[a] Distance between repeating turns on a line drawn parallel to helix axis.

[b] Helix type found for polypeptide chains of collagen.

360° turn ($n = 3.6$). The peptide bond planes in the α-helix are parallel to the axis of the helix. In this geometry each peptide forms two hydrogen bonds, one to the peptide bond of the fourth amino acid above and the second to the peptide bond of the fourth amino acid below in the primary structure. Other α-helix parameters, such as the pitch (p), are given in Table 2.9. In the hydrogen bonds between the peptide groups of an α-helical structure, the distance between the hydrogen-donor atom and the hydrogen-acceptor atom is 2.9 Å. Also, the donor, acceptor, and hydrogen atoms are approximately collinear, in that they determine a straight line. This is an optimum geometry and distance for maximum hydrogen-bond strength (see Section 2.7).

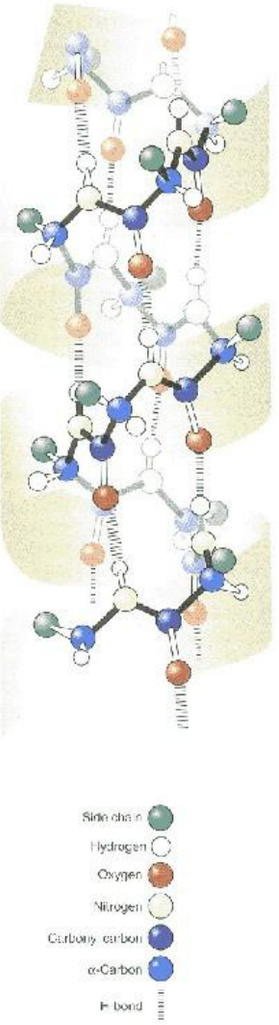


Figure 2.26
An α-helix.
Redrawn with permission, based on figure from Pauling, L. *The Nature of the Chemical Bond*, 3rd ed. Ithaca, NY: Cornell University Press, 1960.

The side chains in an α-helical conformation are on the outside of the spiral structure generated by the polypeptide chain. Due to the characteristic 3.6 residues per turn, the first and every third and fourth R group of the amino acid sequence in the helix come close to the other. Helices often present separable polar and nonpolar faces based on their amino acid sequences, which place polar or nonpolar side chains three or four amino acids apart in the sequence, which folds into the α-helix. This will give rise to unique functional characteristics of the helix. However, if every third or fourth side chain that come close together have the same charge sign or are branched at their β-carbon (valine and isoleucine), their unfavorable ionic or steric interactions destabilize the helix structure. The α-helix may theoretically form its spiral in either a left-handed or right-handed sense, giving the helix asymmetric properties and correlated optical activity. In the structure shown, a right-handed α-helix is depicted; this is more stable than the left-handed helix.

β-Structure

A polypeptide chain in a β-strand conformation (Figure 2.27) is hydrogen bonded to another similar strand aligned either in a parallel or antiparallel direction (Figure 2.28). Hydrogen-bonded β-strands appear like a pleated sheet (Figure 2.29). The side chains project above and below the pleated sheet-like structure.

Supersecondary Structures

Certain combinations of secondary structure can be observed in different folded protein structures. They are referred to as **structural motifs** and include helix-turn-helix (see p. 108), leucine zipper (see p. 110), calcium binding EF hand (see p. 209), and zinc finger (see p. 108). Even longer orderings may occur to form a domain (see below) such as the β-barrel and the immunoglobulin fold. These longer pattern lengths of secondary structure may include multiple structural motifs and when commonly observed in more than one protein are referred to as **supersecondary structures**.

Proteins Fold into a Three-Dimensional Structure Called the Tertiary Structure

The **tertiary structure** of a protein is the three-dimensional structure of a protein. It includes the geometric relationship between distant segments of primary structure and the relationship of the side chains with one another in three-dimensional space. As an example of a protein's tertiary structure, the structure for trypsin is shown in Figure 2.30. In Figure 2.30*a* the ribbon structure shows the conformation of polypeptide strands and the overall pattern of polypeptide chain folding (supersecondary structure). The tertiary structure is then further built upon in Figure 2.30*b* by showing the side chain groups and their interconnections with a stick model. Active site catalytic side chains are shown in yellow, which include the hydroxymethyl group of serine (residue 177 in the sequence), the imidazole of histidine (residue 40), and the carboxylate-containing side chain of aspartate (residue 85). Although these catalytic residues

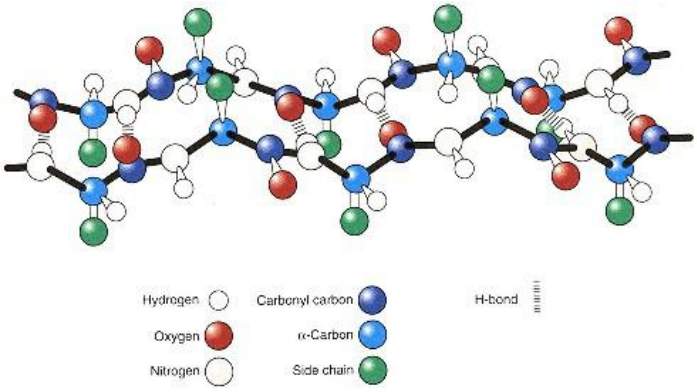


Figure 2.27
Two polypeptide chains in a β-structure conformation.
Additional polypeptide chains may be added to generate more extended structure.
Redrawn with permission from Fersht, A. *Enzyme Structure and Mechanism*, San Francisco: Freeman, 1977, p. 10.

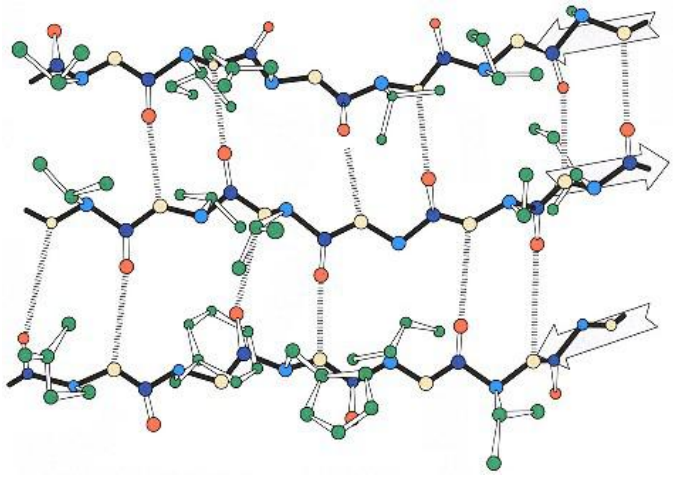


Figure 2.28
Example of antiparallel β-structure (residues 93–98, 28–33, and 16–21 of Cu,Zn superoxide dismutase).
Dashed line shows hydrogen bonds between carbonyl oxygen atoms and peptide nitrogen atoms; arrows show direction of polypeptide chains from N terminal to C terminal. In the characteristic antiparallel -structure, pairs of closely spaced interchain hydrogen bonds alternate with widely spaced hydrogen bond pairs.
Redrawn with permission from Richardson, J. S. *Adv. Protein Chem*. 34:168, 1981.

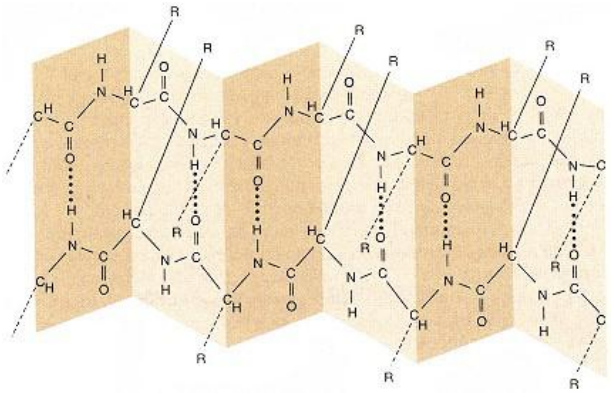


Figure 2.29
β-Pleated sheet structure between two polypeptide chains.
Additional polypeptide chains may be added above and below to generate a more extended structure.

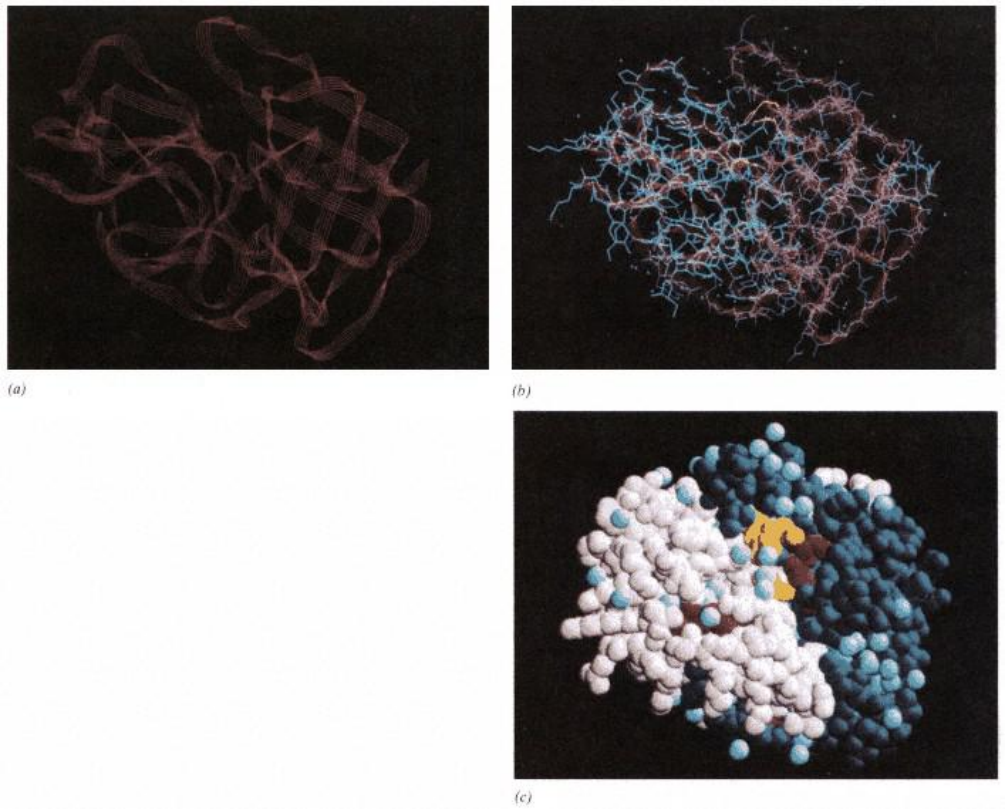


Figure 2.30
Tertiary structure of trypsin.
(a) Ribbon structure outlines the conformation of the polypeptide chain.
(b) Structure shows side chains including active site residues (in yellow) with outline of polypeptide chain (ribbon) superimposed.
(c) Space-filling structure in which each atom is depicted as the size of its van der Waals radius. Hydrogen atoms are not shown. Different domains are shown in dark blue and white. The active site residues are in yellow and intrachain disulfide bonds of cystine in red. Light blue spheres represent water molecules associated with the protein. This structure shows the density of packing within the interior of the protein.

are widely separated in the primary structure, the folded tertiary structure brings them together in space to form the catalytic site. In Figure 2.30*c* a space-filling model shows C, N, and O atoms represented by balls of radius proportional to their van der Waals radius.

The tertiary structure of trypsin conforms to the general rules of folded proteins (see Section 2.3). Hydrophobic side chains are generally in the interior of the structure, away from the water interface. Ionized side chains occur on the outside of a protein structure, where they are stabilized by water of solvation. Within the protein structure (not shown) are buried water molecules, noncovalently associated, which exhibit specific arrangements. A large number of water molecules form a solvation shell around the outside of the protein.

A long polypeptide strand often folds into multiple compact semi-independent folded regions or **domains**, each domain having a characteristic compact geometry with a hydrophobic core and polar outside. They typically contain 100–150 contiguous amino acids. The domains of a **multidomain protein** may be connected by a segment of the polypeptide chain lacking regular secondary structure. Alternatively, the dense spherical folded regions are separated by a cleft or less dense region of tertiary structure (Figure 2.31). There are two folded domains in the trypsin molecule with a cleft between the domains

that includes the substrate-binding catalytic site of the protein. An active site within an interdomain interface is an attribute of many enzymes. Different domains within a protein can move with respect to each other. Hexokinase (Figure 2.32), which catalyzes phosphorylation of a glucose molecule by adenosine triphosphate (ATP), has the glucose-binding site in a region between two domains. When the glucose binds in the active site, the surrounding domains move to enclose the substrate to trap it for phosphorylation (Figure 2.32). In enzymes with more than one substrate or allosteric effector site (see Chapter 4), the different sites may be located within different domains. In multifunctional proteins, each domain performs a different task.

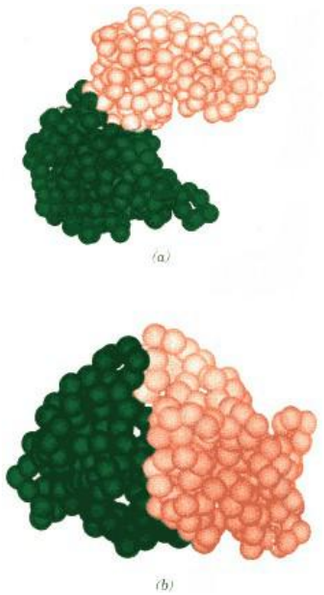


Figure 2.31
Globular domains within proteins.
(a) Phosphoglycerate kinase has two domains with a relatively narrow neck in between.
(b) Elastase has two tightly associated domains separated by a narrow cleft. Each sphere in the space-filling drawing represents the -carbon position for an amino acid within the protein structure.
Reprinted with permission from Richardson, J. S. *Adv. Protein Chem.* 34:168, 1981.

Homologous Three-Dimensional Domain Structures Are Often Formed from Common Arrangements of Secondary Structures

A protein can adopt a range of conformations for a particular amino acid sequence. Although each native structure is unique, a comparison of the tertiary structures of different proteins solved by X-ray crystallography shows similar arrangements of secondary structure motifs that form the tertiary structures of domains. Thus proteins unrelated by function, sequence, or evolution show similar patterns of arrangement of their secondary structures or supersecondary

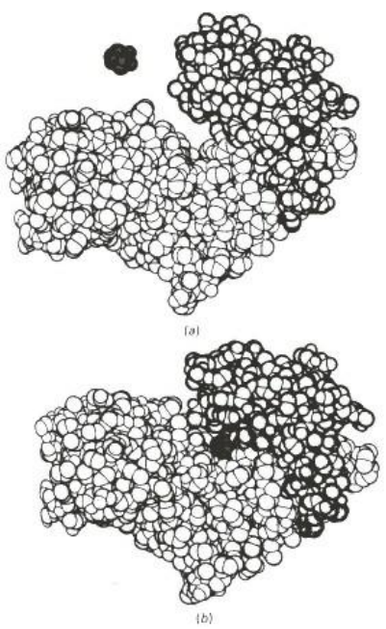


Figure 2.32
Drawings of (a) unliganded form of hexokinase and free glucose and (b) the conformation of hexokinase with glucose bound. In this space-filling drawing each circle represents the van der Waals radius of an atom in the structure. Glucose is black, and each domain is differently shaded.
Reprinted with permission from Bennett, W. S., and Huber, R. *CRC Rev. Biochem.* 15:291, 1984.

Figure 2.33
An example of an all α-folded domain.
In this drawing and those that follow (Figures 2.34–2.36), only the outline of the polypeptide chain is shown. -Structure strands are shown by arrows with the direction of the arrow showing the N C terminal direction of the chain; lightning bolts represent disulfide bonds, and circles represent metal ion cofactors (when present). Redrawn with permission from Richardson, J. S. *Adv. Protein Chem*. 34:168, 1981.

structures. A classification system for supersecondary patterns places common folding patterns for secondary structures into structural families. The key super-secondary structures are formed because of the thermodynamic stability of their folding patterns.

A common all-α structure is found in the enzyme lysozyme (Figure 2.33). Other examples of all-α structure are in myoglobin and the subunits of hemoglobin, whose structures are discussed in Chapter 3. In this supersecondary folding pattern, seven or eight sections of α-helices are joined by smaller segments of polypeptide chains that allow the helices to fold back upon themselves to form a characteristic globular shape. Another common supersecondary structure is the **α,β-domain structure** shown by triose phosphate isomerase (Figure 2.34) in which the strands (designated by arrows) are wound into a **β-barrel.** Each β-strand in the interior of the β-barrel is interconnected by α-helical regions of the polypeptide chain on the outside of the molecule. A similar supersecondary structure is found in pyruvate kinase (Figure 2.34). A different type of α,β-domain supersecondary structure is seen in lactate dehydrogenase and phosphoglycerate kinase (Figure 2.35). In these the interior polypeptide sections participate in a **twisted-sheet** β-structure. The β-structure segments are joined by α-helix regions positioned on the outside of the molecule to give a characteristic α,β-domain folding pattern. An **all-β-domain** supersecondary structure is present in Cu,Zn superoxide dismutase, in which the antiparallel β-sheet forms a **Greek key β-barrel** (Figure 2.36). A similar pattern occurs in each of the domains of the immunoglobulins, discussed in Chapter 3. Concanavalin A (Figure 2.36) shows an all-β-domain structure in which the antiparallel β-strands form a β-barrel pattern called a "jellyroll." Protein structures used to define these classes have been observed by X-ray crystallographic analysis (Section 2.9), primarily of globular proteins that are water soluble. Proteins that are not water soluble may contain different supersecondary patterns (see Section 2.6).

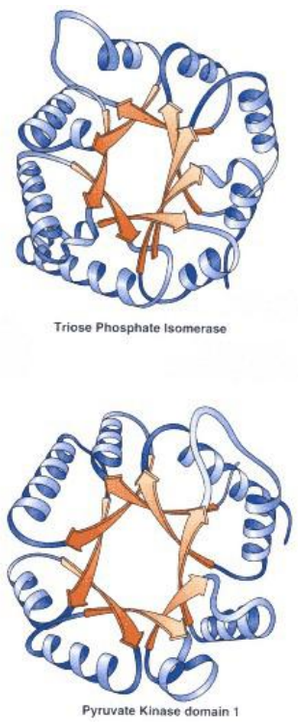


Figure 2.34
Examples of α,β-folded domains in which β-structural strands form a β-barrel in the center of the domain (see legend to Figure 2.33).
Redrawn with permission from Richardson, J. S. *Adv. Protein Chem*. 34: 168, 1981.

A Quaternary Structure Occurs When Several Polypeptide Chains Form a Specific Noncovalent Association

Quaternary structure refers to the arrangement of polypeptide chains in a multichain protein. The subunits in a quaternary structure must be in noncovalent association, α-Chymotrypsin contains three polypeptide chains covalently joined together by interchain disulfide bonds into a single covalent unit and therefore does not have a quaternary structure. Myoglobin consists of a single polypeptide chain and has no quaternary structure. However, hemoglobin A

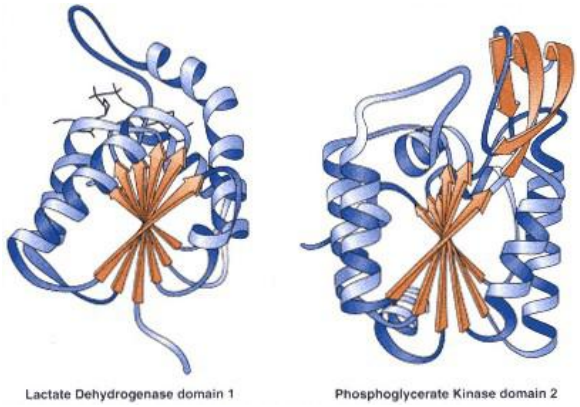


Figure 2.35
Examples of α,β-folded domains in which β-structure strands are in the form of a classical twisted β-sheet (see legend to Figure 2.33).
Redrawn with permission from Richardson, J. S. *Adv. Protein Chem.* 34:168, 1981.

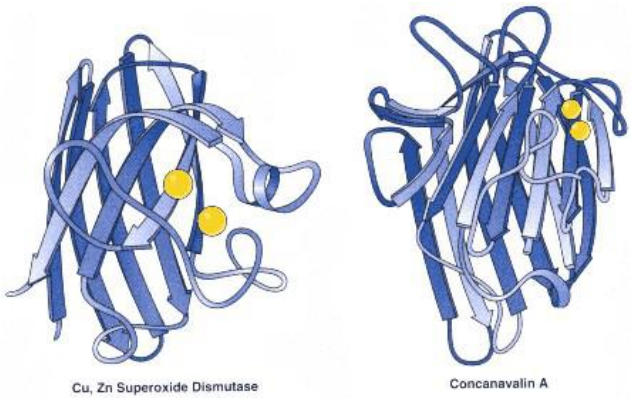


Figure 2.36
Examples of all β-folded domains (see legend to Figure 2.33).
Redrawn with permission from Richardson, J. S. *Adv. Protein Chem.* 34:168, 1981.

contains four polypeptide chains ($\alpha_2\beta_2$) held together noncovalently in a specific conformation as required for its function (see Chapter 3). Thus hemoglobin has a quaternary structure. Aspartate transcarbamylase (see Chapter 13) has a quaternary structure comprised of 12 polypeptide subunits. The poliovirus protein coat contains 60 polypeptide subunits, and the tobacco mosaic virus protein has 2120 polypeptide subunits held together noncovalently in a specific structural arrangement.

2.6— Other Types of Proteins

The characteristics of protein structure, discussed above, are based on observations on globular, water-soluble proteins. Other proteins, such as the fibrous proteins, are nonglobular and have a low water solubility; lipoproteins and

glycoproteins have a heterogeneous composition and may or may not be water soluble.

Fibrous Proteins Include Collagen, Elastin, α-Keratin, and Tropomyosin

Globular proteins have a spheroidal shape, variable molecular weights, relatively high water solubility, and a variety of functional roles as catalysts, transporters, and control proteins for the regulation of metabolic pathways and gene expression. In contrast, fibrous proteins characteristically contain larger amounts of regular secondary structure, a long cylindrical (rod-like) shape, a low solubility in water, and a structural rather than a dynamic role in the cell or organism. Examples of fibrous proteins are collagen, α-keratin, and tropomyosin.

Distribution of Collagen in Humans

Collagen is present in all tissues and organs where it provides the framework that gives the tissues their form and structural strength. Its importance is shown by its high concentration in all organs; the percentage of collagen by weight for some representative human tissues and organs is 4% of the liver, 10% of lung, 12–24% of the aorta, 50% of cartilage, 64% of the cornea, 23% of whole cortical bone, and 74% of skin (see Clin. Corr. 2.4).

Amino Acid Composition of Collagen

The amino acid composition of collagen is quite different from that of a typical globular protein. The amino acid composition of type I skin collagen and of globular proteins ribonuclease and hemoglobin are given in Table 2.10. Skin collagen is comparatively rich in glycine (33% of its amino acids), proline (13%), the derived amino acid 4-hydroxyproline (9%), and another derived amino acid 5-hydroxylysine (0.6%) (Figure 2.37). Hydroxyproline is unique to collagens being formed enzymatically from prolines within a collagen polypeptide chain. The enzyme-catalyzed hydroxylation of proline requires the presence of ascorbic acid (vitamin C); thus in vitamin C deficiency (scurvy) there is poor synthesis of new collagen. Most hydroxyprolines in a collagen have the hydroxyl group in the 4-position (γ-carbon) of the proline structure, although a small amount of 3-hydroxyproline is also formed (Table 2.10). Collagens are glycoproteins with carbohydrate covalently joined to the derived amino acid, 5-hydroxylysine, by an *O*-glycosidic bond through the δ-carbon hydroxyl group. Formation of 5-hydroxylysine from lysines and addition of the carbohydrate to the 5-hydroxylysine occur after polypeptide chain formation but prior to the folding of the collagen chains into their unique supercoiled structure.

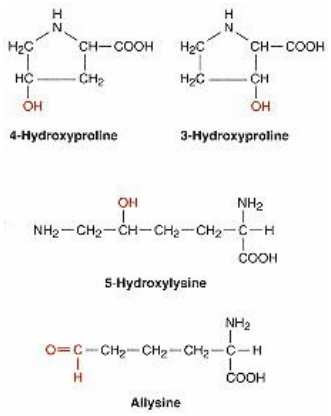


Figure 2.37
Derived amino acids found in collagen.
Carbohydrate is attached to 5-OH in 5-hydroxylysine by a type III glycosidic linkage (see Figure 2.45).

Amino Acid Sequence of Collagen

The molecular unit of mature collagen or tropocollagen contains three polypeptide chains. Various distinct collagen chains exist that make up the different

CLINICAL CORRELATION 2.4

Symptoms of Diseases of Abnormal Collagen Synthesis

Collagen is present in virtually all tissues and is the most abundant protein in the body. Certain organs depend heavily on normal collagen structure to function physiologically. Abnormal collagen synthesis or structure causes dysfunction of cardiovascular organs (aortic and arterial aneurysms and heart valve malfunction), bone (fragility and easy fracturing), skin (poor healing and unusual distensibility), joints (hypermobility and arthritis), and eyes (dislocation of the lens). Examples of diseases caused by abnormal collagen synthesis include Ehlers–Danlos syndrome, osteogenesis imperfecta, and scurvy. These diseases may result from abnormal collagen genes, abnormal posttranslational modification of collagen, or deficiency of cofactors needed by the enzymes that carry out posttranslational modification of collagen.

Byers, P. H. Disorders of collagen biosynthesis and structure. In: C. R. Scriver, A. L. Beaudet, W. S. Sly, and D. Valle (eds.), *The Metabolic and Molecular Bases of Inherited Disease*, 7th ed. McGraw-Hill, 1995, Chap. 134.

TABLE 2.10 Comparison of the Amino Acid Content of Human Skin Collagen (Type I) and Mature Elastin with That of Two Typical Globular Proteins[a]

Amino Acid	*Collagen (Human Skin)*	*Elastin (Mammalian)*	*Ribonuclease (Bovine)*	*Hemoglobin (Human)*
COMMON AMINO ACIDS		PERCENT OF TOTAL		
Ala	11	22	8	9
Arg	5	0.9	5	3
Asn			8	3
Asp	5	1	15	10
Cys	0	0	0	1
Glu	7	2	12	6
Gln			6	1
Gly	33	31	2	4
His	0.5	0.1	4	9
Ile	1	2	3	0
Leu	2	6	2	14
Lys	3	0.8	11	10
Met	0.6	0.2	4	1
Phe	1	3	4	7
Pro	13	11	4	5
Ser	4	1	11	4
Thr	2	1	9	5
Trp	2	1	9	2
Tyr	0.3	2	8	3
Val	2	12	8	10
DERIVED AMINO ACIDS				
Cystine	0	0	7	0
3-Hydroxyproline	0.1		0	0
4-Hydroxyproline	9	1		0
5-Hydroxylysine	0.6	0	0	0
Desmosine and isodesmosine	0	1	0	0

[a] Boxed numbers emphasize important differences in amino acid composition between the fibrous proteins (collagen and elastin) and typical globular proteins.

collagen types, each with their own genes. In some types, the three polypeptide chains have an identical amino acid sequence. In others such as type I (Table 2.11), two of the chains are identical while the amino acid sequence of the third chain is slightly different. In type I collagen, the identical chains are designated α1(I) chains and the third nonidentical chain, α2(I). In type V collagen all three chains are different, designated α1(V), α2(V), and α3(V). Different types of collagen differ in their physical properties due to differences in the amino acid sequence among chains, even though there are large regions of homologous sequence among the different chain types. Collagen has covalently attached carbohydrate and the collagen types differ in their carbohydrate component. Table 2.11 describes some characteristics of collagen types I–VI; additional collagen types (designated up through type XVI) have been reported.

The amino acid sequence of the chains of collagens is unusual. In long segments of all the collagen types glycine occurs as every third residue and proline or hydroxyproline also occurs three residues apart in these same regions. Accordingly, the amino acid sequences Gly-Pro-Y and Gly-X-Hyp (where X and Y are any of the amino acids) are repeated in tandem several hundred times. In type I collagen, the triplet sequences are reiterated over 200 times, encompassing over 600 amino acids within a chain of approximately 1000 amino acids.

TABLE 2.11 Classification of Collagen Types

Type	*Chain Designations*	*Tissue Found*	*Characteristics*
I	$[\alpha 1(I)]_2\alpha 2(I)$	Bone, skin, tendon, scar tissue, heart valve, intestinal, and uterine wall	Low carbohydrate; <10 hydroxylysines per chain; two types of polypeptide chains
II	$[\alpha 1(II)]_3$	Cartilage, vitreous	10% carbohydrate; >20 hydroxylysines per chain
III	$[\alpha 1(III)]_3$	Blood vessels, newborn skin, scar tissue, intestinal, and uterine wall	Low carbohydrate; high hydroxyproline and Gly; contains Cys
IV	$[\alpha 1(IV)]_3$ $[\alpha 2(IV)]_3$	Basement membrane, lens capsule	High 3-hydroxyproline; >40 hydroxylysines per chain; low Ala and Arg; contains Cys; high carbohydrate (15%)
V	$[\alpha 1(V)]_2\alpha 2(V)$ $[\alpha 1(V)]_3$ $\alpha 1(V)\alpha 2(V)\alpha 3(V)$	Cell surfaces or exocytoskeleton; widely distributed in low amounts	High carbohydrate, relatively high glycine, and hydroxylysine
VI	–	Aortic intima, placenta, kidney, and skin in low amounts	Relatively large globular domains in telopeptide region; high Cys and Tyr; molecular weight relatively low (~160,000); equimolar amounts of hydroxylysine and hydroxyproline

Structure of Collagen

Polypeptides that contain only proline can be synthesized in the laboratory. These polyproline chains assume a regular secondary structure in aqueous solution in which the chain is in a tightly twisted extended helix with three residues per turn of the helix ($n = 3$). This helix with all *trans*-peptide bonds is designated the **polyproline type II** helix (see Figure 2.11 for differences between *cis*- and *trans*-peptide bonds). The polyproline helix has the same characteristics as the helix found in collagen chains in regions of the primary structure that contain a proline or hydroxyproline at approximately every third position. Since the helix structure in collagen is the same as that of polyproline, the thermodynamic forces leading to formation of the collagen helix structure are due to the properties of proline. In proline, the ϕ angle contributed to the polypeptide chain is part of the five-member cyclic side chain. The five-member ring constrains the C_α–N bond to an angle compatible with the polyproline helix structure.

In polyproline type II helix, the plane of each peptide bond is perpendicular to the axis of the helix. In this geometry the peptide carbonyl groups are pointed toward neighboring chains and are correctly oriented to form strong interchain hydrogen bonds with other chains of the collagen molecule. This is in contrast to the α-helix, in which the plane containing the atoms of the peptide bond is

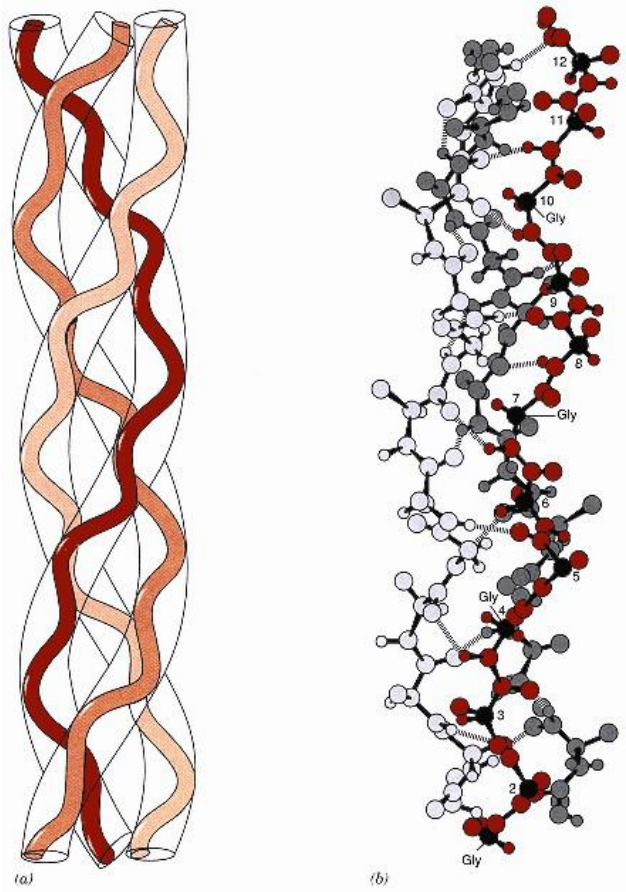


Figure 2.38
Diagram of collagen demonstrating necessity for glycine in every third residue to allow the different chains to be in close proximity in the structure.
(a) Ribbon model for supercoiled structure of collagen with each individual chain in a polyproline type II helix.
(b) More detailed model of supercoiled conformation.
All -carbon atoms are numbered and proposed hydrogen bonds are shown by dashed lines.
Redrawn with permission from Dickerson, R. E., and Geis, I. *The Structure and Actions of Proteins*, Menlo Park, CA: Benjamin, 1969, pp. 41, 42.

parallel to the α-helix axis and the peptide bonds form only intrachain hydrogen bonds with peptide bonds in the same polypeptide chain. The three chains of a collagen molecule, where each of the chains is in a polyproline type II helix conformation, are wound about each other in a defined way to form a **superhelical structure** (Figure 2.38). The three-chain superhelix has a characteristic rise (d) and pitch (p) as does the single-chain helix. The collagen superhelix forms because glycines have a side chain of low steric bulk (R = H). As the polyproline type II helix has three residues per turn (n = 3) and glycine is at every third position, the glycines in each of the polypeptide chains

are aligned along one side of the helix, forming an **apolar edge** of the chain. The glycine edges from the three polypeptide chains associate noncovalently in a close arrangement, held together by hydrophobic interactions, to form the superhelix structure of collagen. A larger side chain than that of glycine would impede the adjacent chains from coming together in the superhelix structure (Figure 2.38).

In collagen molecules the superhelix conformation may propagate for long stretches of the sequence, which is especially true for type I collagen where only the COOH-terminal and NH_2-terminal segments (known as the **telopeptides**) are not in a superhelical conformation. The type I collagen molecule has a length of 3000 Å and a width of only 15 Å, a very long cylindrical structure. In other collagen types, the superhelical regions may be broken periodically by regions of the chain that fold into globular domains.

Formation of Covalent Cross-links in Collagen

An enzyme present in extracellular space acts on the secreted collagen molecules (see p. 747) to convert some of the ε-amino groups of lysine side chains to δ-aldehydes (Figure 2.39). The resulting amino acid, containing an aldehydic R group, is the derived amino acid **allysine.** The newly formed aldehyde side chain spontaneously undergoes nucleophilic addition reactions with nonmodified lysine ε-amino groups and with the δ-carbon atoms of other allysine aldehydic groups to form linking covalent bonds (Figure 2.39). These covalent linkages can be between chains within the superhelical structure or between adjacent superhelical collagen molecules in a collagen fibril.

Elastin Is a Fibrous Protein with Allysine-Generated Cross-links

Elastin gives tissues and organs the capacity to stretch without tearing. It is classified as a fibrous protein because of its structural function and relative insolubility. It is abundant in ligaments, lungs, walls of arteries, and skin. Elastin does not contain repeating sequences of Gly-Pro-Y or Gly-X-Hyp and does not fold into either a polyproline helix or a superhelix. It appears to lack a regular secondary structure, but rather contains an unordered coiled structure in which amino acid residues within the folded structure are highly mobile. The highly mobile, kinetically free, though extensively cross-linked structure gives the

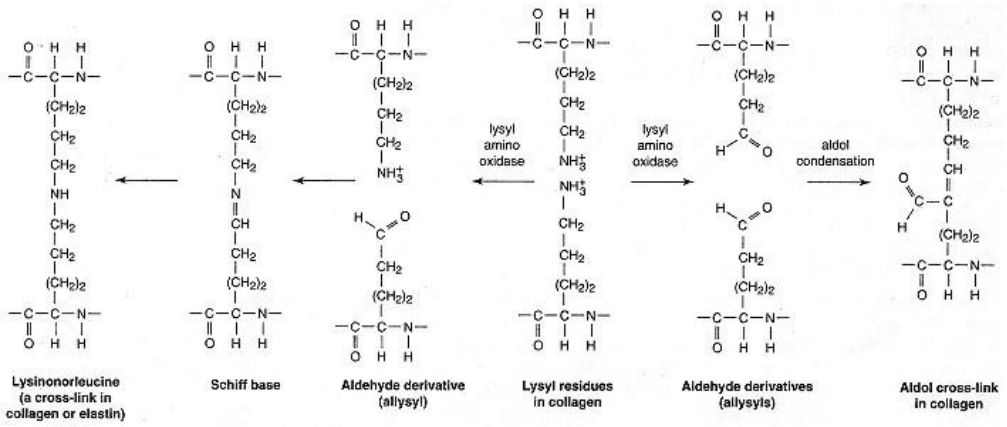


Figure 2.39
Covalent cross-links formed in collagen through allysine intermediates.
Formation of allysines is catalyzed by lysyl amino oxidase.

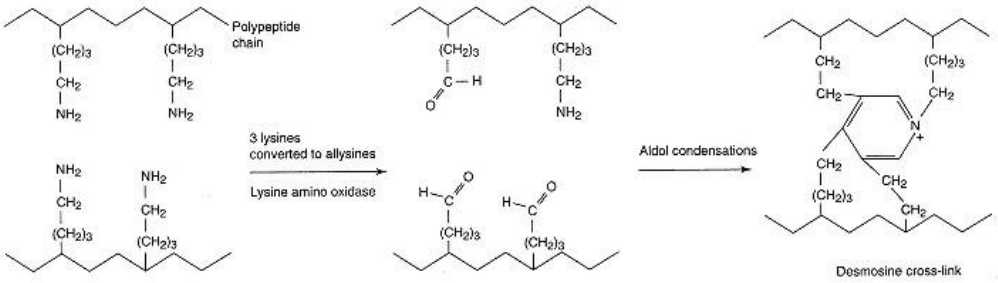


Figure 2.40
Formation of desmosine covalent cross-link in elastin from lysine and allysines.
Polypeptide chain drawn schematically with intersections of lines representing the placement of -carbons.

protein a rubber-like elasticity. As in collagen, allysines form cross-links in **elastin**. An extracellular **lysine amino oxidase** converts lysine side chains of elastin to allysines. The amino oxidase is specific for lysines in the sequence -Lys-Ala-Ala-Lys- and -Lys-Ala-Ala-Ala-Lys-. Three allysines and an unmodified lysine in these sequences, from different regions in the polypeptide chains, react to form the heterocyclic structure of **desmosine** or **isodesmosine**. The desmosines covalently cross-link the polypeptide chains in elastin fibers (Figure 2.40).

α-Keratin and Tropomyosin

α-Keratin and tropomyosin are fibrous proteins in which each chain has an α-helical conformation. α-Keratin is found in the epidermal layer of skin, in nails, and in hair. Tropomyosin is a component of the thin filament in muscle tissue. Analysis of the α-helical sequences in both these proteins shows the tandem repetition of seven amino acid segments, in which the first and fourth amino acids have hydrophobic side chains and the fifth and seventh polar side chains. The reiteration of hydrophobic and polar side chains in seven amino acid segments is symbolically represented by the formulation $(a\text{-}b\text{-}c\text{-}d\text{-}e\text{-}f\text{-}g)_i$, where residues a and d are hydrophobic amino acids, and residues e and g are polar or ionized side chain groups. Since a seven amino acid segment represents two complete turns of an α-helix ($n = 3.6$), the apolar residues at a and d align to form an apolar edge along one side of the α-helix (Figure 2.41). This apolar edge in α-keratin interacts with polypeptide apolar edges of other α-keratin chains to form a superhelical structure containing two or three polypeptide chains. Each strand also contains a polar edge, due to residues e and g, that interacts with the water solvent on the outside of the superhelix and also stabilizes the superhelical structure. Similarly, two tropomyosin polypeptide strands in α-helical conformation wind around each other to form a tropomyosin superhelical structure.

Thus collagen, α-keratin, and tropomyosin molecules are multistrand structures in which polypeptide chains with a highly regular secondary structure (polyproline type II helix in collagen, α-helix in α-keratin and tropomyosin) are wound around each other to form a rod-shaped supercoiled conformation. In turn, the supercoiled molecules are aligned into multimolecular fibrils stabilized by covalent cross-links. The amino acid sequences of the chains are repetitive, generating edges on the cylindrical surfaces of each of the chains that stabilize a hydrophobic interaction between the chains required for generation of the supercoiled conformation.

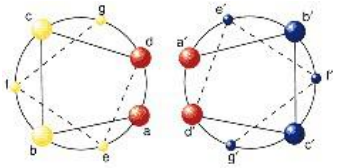


Figure 2.41
Interaction of an apolar edge of two chains in α-helical conformation as in α-keratin and tropomyosin.
Interaction of apolar a-d and d-a residues of two -helices aligned parallel in an NH_2-terminal (top) to COOH-terminal direction is presented. Redrawn from McLachlan, A. D., and Stewart, M. *J. Mol. Biol.* 98:293, 1975.

Lipoproteins Are Complexes of Lipids with Proteins

Lipoproteins are multicomponent complexes of proteins and lipids that form distinct molecular aggregates with an approximate stoichiometry between protein and lipid components within the complex. Each type of lipoprotein has a characteristic molecular mass, size, chemical composition, density, and physiological role. The protein and lipid in the complex are held together by noncovalent forces.

Plasma lipoproteins are extensively characterized and changes in their relative amounts are predictive of atherosclerosis, a major human disease (see Clin. Corr. 2.5). They have a wide variety of roles in blood including transport of lipids from tissue to tissue and participating in lipid metabolism (see Chapter 9). Four classes of plasma lipoproteins exist in normal fasting humans (Table 2.12); in the postabsorptive period a fifth type, **chylomicrons,** is also present. Lipoprotein classes are identified by their density, as determined by ultracentrifugation and by electrophoresis (Figure 2.42). The protein components of a lipoprotein particle are the **apolipoproteins**. Each type of lipoprotein has a

TABLE 2.12 Hydrated Density Classes of Plasma Lipoproteins

Lipoprotein Fraction	*Density (g mL^{-1})*	*Flotation Rate, S_f (Svedberg units)*	*Molecular Weight (daltons)*	*Particle Diameter (Å)*
HDL	1.063–1.210		HDL_2, 4×10^5	70–130
			HDL_3, 2×10^5	50–100
LDL (or LDL_2)	1.019–1.063	0–12	2×10^6	200–280
IDL (or LDL_1)	1.006–1.019	12–20	4.5×10^6	250
VLDL	0.95–1.006	20–400	5×10^6–10^7	250–750
Chylomicrons	<0.95	>400	10^9–10^{10}	10^3–10^4

Source: Data from Soutar, A. K., and Myant, N. B. In: R. E. Offord (Ed.), *Chemistry of Macromolecules*, IIB. Baltimore, MD: University Park Press, 1979.

CLINICAL CORRELATION 2.5

Hyperlipidemias

Hyperlipidemias are disorders of the rates of synthesis or clearance of lipoproteins from the bloodstream. Usually they are detected by measuring plasma triacylglycerol and cholesterol and are classified on the basis of which class of lipoproteins is elevated.

Type I hyperlipidemia is due to accumulation of chylomicrons. Two genetic forms are known: lipoprotein lipase deficiency and ApoCII deficiency. ApoCII is required by lipoprotein lipase for full activity. Patients with type I hyperlipidemia have exceedingly high plasma triacylglycerol levels (over 1000 mg dL^{-1}) and suffer from eruptive xanthomas (triacylglycerol deposits in the skin) and pancreatitis.

Type II hyperlipidemia is characterized by elevated LDL levels. Most cases are due to genetic defects in the synthesis, processing, or function of the LDL receptor. Heterozygotes have elevated LDL levels; hence the trait is dominantly expressed. Homozygous patients have very high LDL levels and may suffer myocardial infarctions before age 20.

Type III hyperlipidemia is due to abnormalities of ApoE, which interfere with the uptake of chylomicron and VLDL remnants. Hypothyroidism can produce a very similar hyperlipidemia. These patients have an increased risk of atherosclerosis.

Type IV hyperlipidemia is the commonest abnormality. The VLDL levels are increased, often due to obesity, alcohol abuse, or diabetes. Familial forms are also known but the molecular defect is unknown.

Type V hyperlipidemia is, like type I, associated with high chylomicron triacylglycerol levels, pancreatitis, and eruptive xan-thomas.

Hypercholesterolemia also occurs in certain types of liver disease in which biliary excretion of cholesterol is reduced. An abnormal lipoprotein called lipoprotein X accumulates. This disorder is not associated with increased cardiovascular disease from atherosclerosis.

Havel, R. J., and Kane, J. P. Introduction: structure and metabolism of plasma lipoproteins. In: C. R. Scriver, A. L. Beaudet, W. S. Sly, and D. Valle (Eds.), *The Metabolic and Molecular Basis of Inherited Disease*, 7th ed. New York: McGraw-Hill, 1995, Chap. 56; and Goldstein, J. L., Hobbs, H. H., and Brown, M. S. Familial hypercholesterolemia. In: C. R. Scriver, A. L. Beaudet, W. S. Sly, and D. Valle (Eds.), *The Metabolic and Molecular Bases of Inherited Disease*, 7th ed., New York: McGraw Hill, 1995, Chap. 62.

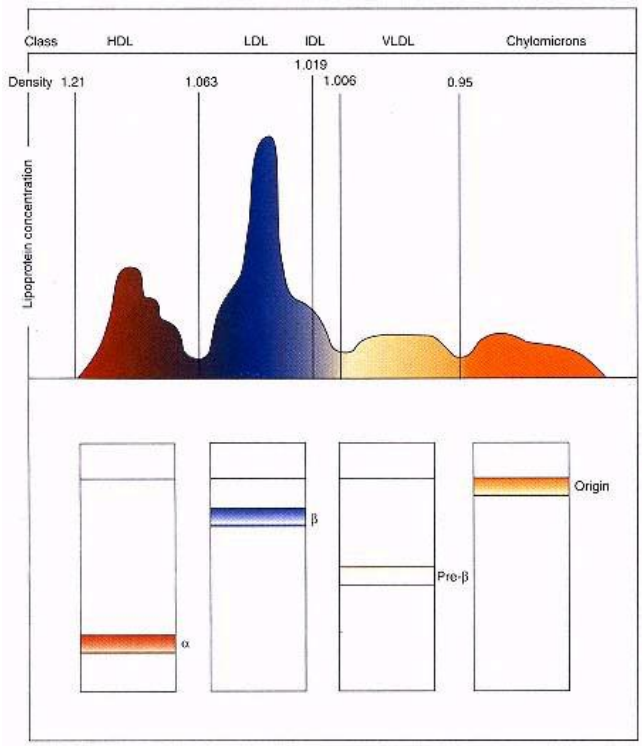


Figure 2.42
Correspondence of plasma lipoprotein density classes with electrophoretic mobility in a plasma electrophoresis.
In the upper diagram an ultracentrifugation Schlieren pattern is shown. At the bottom, electrophoresis on a paper support shows the mobilities of major plasma lipoprotein classes with respect to - and -globulin bands.
Reprinted with permission from Soutar, A. K., and Myant, N. B. In: R. E. Offord (Ed.), *Chemistry of Macromolecules*, IIB. Baltimore, MD: University Park Press, 1979.

characteristic apolipoprotein composition, the different apolipoproteins often being present in a set ratio. The most prominent apolipoprotein in **high density lipoproteins (HDLs)** is apolipoprotein A-I (ApoA-I) (Table 2.13). In **low den-**

TABLE 2.13 Apolipoproteins of Human Plasma Lipoproteins (Values in Percentage of Total Protein Present)[a]

Apolipoprotein	*HDL_2*	*HDL_3*	*LDL*	*IDL*	*VLDL*	*Chylomicrons*
ApoA-I	85	70–75	Trace	0	0–3	0–3
ApoA-II	5	20	Trace	0	0–0.5	0–1.5
ApoD	0	1–2			0	1
ApoB	0–2	0	95–100	50–60	40–50	20–22
ApoC-I	1–2	1–2	0–5	<1	5	5–10
ApoC-II	1	1	0.5	2.5	10	15
ApoC-III	2–3	2–3	0–5	17	20–25	40
ApoE	Trace	0–5	0	15–20	5–10	5
ApoF	Trace	Trace				
ApoG	Trace	Trace				

Source: Data from Soutar, A. K., and Myant, N. B. In: R. E. Offord (Ed.), *Chemistry of Macromolecules*, IIB. Baltimore, MD: University Park Press, 1979; Kostner, G. M. *Adv. Lipid Res.* **20**:1, 1983.

[a] Values show variability from different laboratories.

sity lipoproteins (LDLs) the prominent apolipoprotein is ApoB, which is also present in the **intermediate density lipoproteins (IDLs)** and **very low density lipoproteins (VLDLs).** The ApoC family is also present in high amounts in IDLs and VLDLs. Each apolipoprotein class (A, B, etc.) is distinct (see Clin. Corr. 2.6). Proteins within a class do not cross-react with antibodies to another class. The molecular weights of the apolipoproteins of the plasma lipoproteins vary from 6 kDa (ApoC-I) to 550 kDa for ApoB-100. This latter is one of the longest single-chain polypeptides known (4536 amino acids).

A model for a VLDL particle is shown in Figure 2.43. On the inside are neutral lipids such as cholesterol esters and triacylglycerols. Surrounding this inner core of neutral lipids, in a shell ~ 20 Å thick, reside the proteins and the

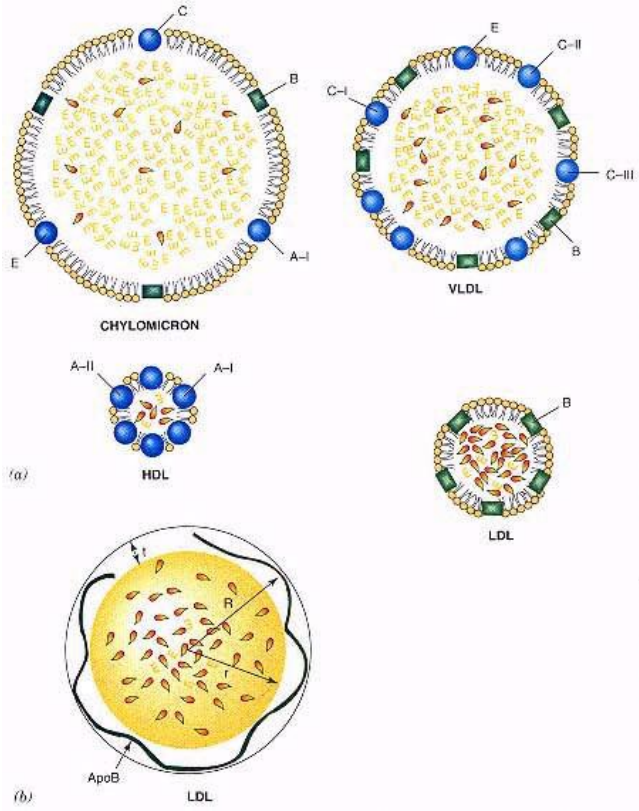


Figure 2.43
Generalized structure of plasma lipoproteins.
(a) Spherical particle model consisting of a core of triacylglycerols (yellow E's) and cholesterol esters (orange drops) with a shell ~ 20 Å thick of apolipoproteins (lettered), phospholipids, and unesterified cholesterol. Apolipoproteins are embedded with their hydrophobic edges oriented toward the core and their hydrophilic edges toward the outside. From Segrest, J. P., et al. *Adv. Protein Chem.* 45:303, 1994.
(b) LDL particle showing ApoB-100 imbedded in outer shell of particle. *From Schumaker, V. N., et al., Protein Chem. 45:205, 1994.*

CLINICAL CORRELATION 2.6

Hypolipoproteinemias

Abetalipoproteinemia is a genetic disease that is characterized by absence of chylomicrons, VLDLs, and LDLs due to an inability to synthesize apolipoprotein B-100. Patients show accumulation of lipid droplets in small intestinal cells, malabsorption of fat, acanthocytosis (spiny shaped red cells), and neurological disease (retinitis pigmentosa, ataxia, and retardation).

Tangier disease, an α-lipoprotein deficiency, is a rare autosomal recessive disease in which the HDL level is 1–5% of its normal value. Clinical features are due to the accumulation of cholesterol in the lymphoreticular system, which may lead to hepatomegaly and splenomegaly. In this disease the plasma cholesterol and phospholipids are greatly reduced.

Deficiency of the enzyme lecithin:cholesterol acyltransferase is a rare disease that results in the production of lipoprotein X (see Clin. Corr. 2.5). Also characteristic of this disease is the decrease in the α-lipoprotein and pre-β-lipoprotein bands, with the increase in the β-lipoprotein (lipoprotein X) in electrophoresis.

Kane, J. P., and Havel, R. J. Disorders of the biogenesis and secretion of lipoproteins containing the β apolipoproteins. In: C. R. Scriver, A. L. Beaudet, W. S. Sly, and D. Valle (Eds.), *The Metabolic and Molecular Basis of Inherited Disease*, 7th ed. New York: McGraw-Hill, 1995, Chap. 57; and Assmann, G., von Eckardstein, A., and Brewer, H. B. Jr. Familial high density lipoprotein deficiency: Tangier disease. In: C. R. Scriver, A. L. Beaudet, W. S. Sly, and D. Valle (Eds.), *The Metabolic and Molecular Bases of Inherited Disease*, 7th ed. New York: McGraw-Hill, 1995, Chap. 64.

charged amphoteric lipids such as unesterified cholesterol and the phosphatidylcholines (see Chapter 10). Amphoteric lipids and proteins in the outer shell place their hydrophobic apolar regions toward the inside of the particle and their charged groups toward the outside where they interact with each other and with water.

This spherical structural model with a hydrophobic inner core of neutral lipids and amphoteric lipids and proteins in the outer shell applies to all plasma lipoproteins, irrespective of their density class and particle size. The smaller lipoprotein particles, such as HDLs, have a smaller diameter. As the diameter of a spherical particle decreases, the molecules in the outer shell make up a greater percentage of the total molecules in the particle. The smaller HDL particles would therefore be theoretically predicted to have a higher percentage of surface proteins and amphoteric lipids than the larger VLDL particles. Thus the HDL particle is 45% protein and 55% lipid, while the larger VLDL particle is only 10% protein with 90% lipid (Table 2.14).

The apolipoproteins, with the exception of ApoB, have a high α-helical content when in association with lipid. The helical regions have amphipathic properties. Every third or fourth amino acid in the helix is charged and forms a polar edge along the helix that associates with the polar heads of phospholipids and the aqueous solvent on the outside. The opposite side of the helix has hydrophobic side chains that associate with the nonpolar neutral lipid core of the phospholipid particle. The α-helical structure of part of ApoC-I is shown

TABLE 2.14 Chemical Composition of the Different Plasma Lipoprotein Classes

Lipoprotein Class	*Total Protein (%)*	*Total Lipid (%)*	*Percent Composition of Lipid Fraction*			
			Phospholipids	*Esterified Cholesterol*	*Unesterified Cholesterol*	*Triacylglycerols*
HDL_2[a]	40–45	55	35	12	4	5
HDL_3[a]	50–55	50	20–25	12	3–4	3
LDL	20–25	75–80	15–20	35–40	7–10	7–10
IDL	15–20	80–85	22	22	8	30
VLDL	5–10	90–95	15–20	10–15	5–10	50–65
Chylomicrons	1.5–2.5	97–99	7–9	3–5	1–3	84–89

Source: Data from Soutar, A. K., and Myant, N. B. In R. E. Offord (Ed.), *Chemistry of Macromolecules*, IIB. Baltimore, MD: University Park Press, 1979.

[a] Subclasses of HDL.

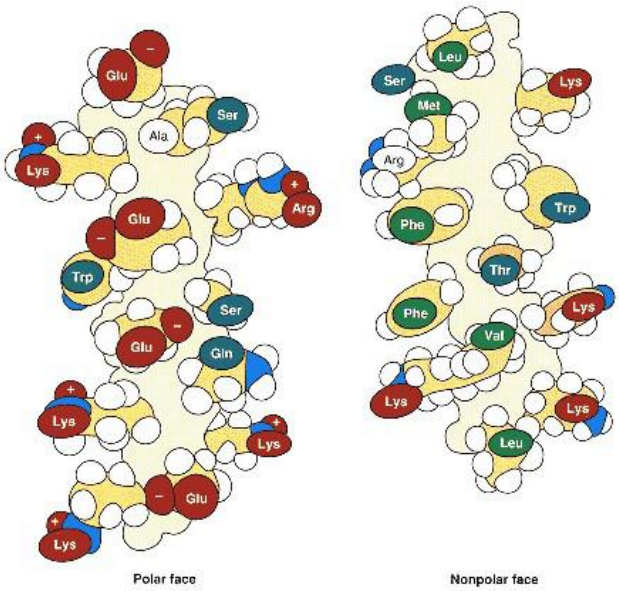


Figure 2.44
Illustration showing side chains of a helical segment of apolipoprotein C-1 between residues 32 and 53.
The polar face shows ionizable acid residues in the center and basic residues at the edge. On the other side of the helix, the hydrophobic residues form a nonpolar longitudinal face.
Redrawn with permission from Sparrow, J. T., and Gotto, A. M., Jr. *CRC Crit. Rev. Biochem*. 13:87, 1983.

in Figure 2.44. ApoB appears to have both α-helical and β-structural regions embedded in the phospholipid outer core. The long 4536 amino acid polypeptide chain of ApoB-100 surrounds the circumference of the LDL particle like a belt weaving in and out of the monolayer phospholipid outer shell (Figure 2.43). One ApoB molecule associates per LDL particle.

Glycoproteins Contain Covalently Bound Carbohydrate

Glycoproteins participate in many normal and disease-related functions of clinical relevance. Many plasma membrane proteins are glycoproteins. Some may be antigens, which determine the blood antigen system (A, B, O) and the histocompatibility and transplantation determinants of an individual. Immunoglobulin antigenic sites and viral and hormone receptor sites in plasma membranes are often glycoproteins. The carbohydrate portions of glycoproteins in membranes provide a surface code for identification by other cells and for contact inhibition in the regulation of cell growth. Changes in membrane glycoproteins can be correlated with tumorigenesis and malignant transformation in cancer. Most plasma proteins, except albumin, are glycoproteins including blood-clotting proteins, immunoglobulins, and many of the complement proteins. Some protein hormones, such as follicle-stimulating hormone (FSH) and thyroid-stimulating hormone (TSH), are glycoproteins. The structural proteins collagen, laminin, and fibronectin contain carbohydrate, as do proteins of mu-

cous secretions that perform a role in lubrication and protection of epithelial tissue.

The percentage of carbohydrate in glycoproteins is variable. IgG antibody molecules contain low amounts of carbohydrate (4%), whereas glycophorin of human red blood cell membranes is 60% carbohydrate. Human gastric glycoprotein is 82% carbohydrate. The carbohydrate can be distributed evenly along the polypeptide chain or concentrated in defined regions. For plasma membrane proteins, typically only the portion located on the outside of the cell has carbohydrate covalently attached. The carbohydrate attached at one or at multiple points along a polypeptide chain usually contains less than 15 sugar residues and in some cases only one sugar residue. Glycoproteins with the same function from different animal species often have homologous amino acid sequences but variable carbohydrate structures. Heterogeneity in carbohydrate content can occur in the same protein within a single organism. For example, pancreatic ribonuclease A and B forms have an identical primary structure but differ in their carbohydrate composition.

Functional glycoproteins are also found in different stages of completion. Addition of complex carbohydrate units occurs in a series of enzyme-catalyzed reactions as the polypeptide chain is transported through the endoplasmic reticulum and Golgi network (see Chapter 17). Immature glycoproteins are sometimes expressed with intermediate stages of carbohydrate additions.

Types of Carbohydrate–Protein Covalent Linkages

Different types of covalent linkages join the sugar moieties and protein in a glycoprotein. The two most common are the ***N*-glycosidic linkage** (type I linkage) between an asparagine amide group and a sugar, and the ***O*-glycosidic linkage** (type II linkage) between a serine or threonine hydroxyl group and a sugar (Figure 2.45). In type I linkage the bond to asparagine is within the sequence Asn-X-Thr(Ser). Another linkage found in mammalian glycoproteins is an *O*-glycosidic bond to a 5-hydroxylysine residue (type III linkage) found in collagens and in the serum complement protein C1q. Less common linkages include attachment to the hydroxyl group of 4-hydroxyproline (type IV linkage), to a cysteine thiol side chain (type V linkage), and to a NH_2-terminal α-amino group of a polypeptide chain (type VI linkage). High concentrations of type VI linkages are spontaneously formed with hemoglobin and blood glucose in uncontrolled diabetics. Assay of the concentration of glycosylated hemoglobin is used to follow changes in blood glucose concentration (see Clin. Corr. 2.7).

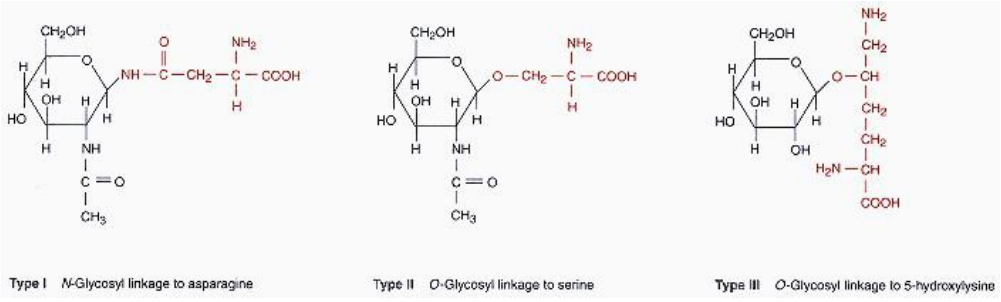
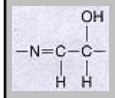


Figure 2.45
Examples of glycosidic linkages to amino acids in proteins.
Type I is an *N*-glycosidic linkage through an amide nitrogen of Asn; type II is an *O*-glycosidic linkage through the OH of Ser or Thr; and type III is an *O*-glycosidic linkage to the 5-OH of 5-hydroxylysine.

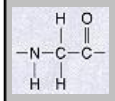
CLINICAL CORRELATION 2.7

Glycosylated Hemoglobin, HbA_{1c}

A glycosylated hemoglobin, designated HbA_{1c}, is formed spontaneously in red blood cells by combination of the NH_2-terminal amino groups of the hemoglobin β-chain and glucose. The aldehyde group of the glucose first forms a Schiff base with the NH_2-terminal amino group,



which then rearranges to a more stable amino ketone linkage,



by a spontaneous (nonenzymatic) reaction known as the Amadori rearrangement. The concentration of HbA_{1c} is dependent on the concentration of glucose in the blood and the duration of hyperglycemia. In prolonged hyperglycemia the concentration may rise to 12% or more of the total hemoglobin. Patients with diabetes mellitus have high concentrations of blood glucose and therefore high amounts of HbA_{1c}. The changes in the concentration of HbA_{1c} in diabetic patients can be used to follow the effectiveness of treatment for the diabetes.

Bunn, H. F. Evaluation of glycosylated hemoglobin in diabetic patients. *Diabetes* 30:613, 1980; and Brown, S.B., and Bowes, M. A. Glycosylated haemoglobins and their role in management of diabetes mellitus. *Biochem. Educ.* 13:2, 1985.

2.7— Folding of Proteins from Randomized to Unique Structures: Protein Stability

The Protein Folding Problem: A Possible Pathway

The ability of a primary protein structure to fold spontaneously to its native secondary or tertiary conformation, without any information other than the amino acid sequence and the noncovalent forces that act on the sequence, has been demonstrated. RNase will spontaneously refold to its native conformation after being denatured with loss of native structure but without the hydrolysis of peptide bonds. Such observations led to the hypothesis that a polypeptide sequence contains the properties sufficient to promote spontaneous protein folding to its unique active conformation under the correct solvent conditions and in the presence of prosthetic groups that may be a part of its structure. As described below chaperone proteins may facilitate the rate of protein folding. Quaternary structures also assemble spontaneously, after the tertiary structure of the individual polypeptide subunits are formed.

It may appear surprising that a protein folds into a single unique conformation given all the possible *a priori* rotational conformations available around single bonds in the primary structure. For example, the α-chain of hemoglobin contains 141 amino acids in which there are at least four single bonds per amino acid residue around which free rotation can occur. If each bond about which free rotation occurs has two or more stable rotamer conformations accessible to it, then there are a minimum of 4^{141} or 5×10^{86} possible conformations for the α-chain amino acid sequence.

The **conformation of a protein** is that conformation of the lowest Gibbs free energy accessible to the amino acid sequence within a physiological time frame. Thus folding is under thermodynamic and kinetic control. Although an exact knowledge of *de novo* folding of a polypeptide is at present an unattainable goal, certain processes appear reasonable. There is evidence that folding is initiated by short-range interactions forming secondary structures in small regions of the polypeptide. Short-range interactions are noncovalent interactions that occur between a side chain and its nearest neighbors. Particular side chains have a propensity to promote the formation of α-helices, β-structure, and sharp turns or bends (β-turns) in the polypeptide. The interaction of a side chain with its nearest neighbors in the polypeptide determines the secondary structure

into which that section of the polypeptide strand folds. Sections of polypeptide, called **initiation sites**, thus spontaneously fold into small regions of secondary structure. The partially folded structures then condense with each other to form a **molten-globular** state. This is a condensed intermediate on the folding pathway that contains much of the secondary structure elements of the native structure, but a large number of incorrect tertiary structure interactions. Segments of secondary structure in the molten-globular state are highly mobile relative to one another, and the molten-globular structure is in rapid equilibrium with the fully unfolded denatured state. The correct medium- and long-range interactions between different initiation sites are found by rearrangements within the molten-globule and the low free energy, native tertiary structure for the polypeptide chain is formed. With formation of the native tertiary structure, the correct disulfide bonds (cystine) are formed. The rate-determining step for folding and unfolding of the native conformation lies between the molten-globular state and the native structure.

Chaperone Proteins May Assist the Protein Folding Process

Cells contain proteins that facilitate the folding process. These include *cis-trans*-prolyl isomerases, protein disulfide isomerases, and chaperone proteins. ***cis-trans*-Protyl isomerases** increase the rate of folding by catalyzing interconversion of *cis*- and *trans*-peptide bonds of proline residues within the polypeptide chain. This allows the correct proline peptide bond conformation to form for each proline as required by the folded native structure. **Protein disulfide isomerases** catalyze the breakage and formation of disulfide cystine linkages so incorrect linkages are not stabilized and the correct arrangement of cystine linkages for the folded conformation is rapidly achieved.

Chaperone proteins were discovered as **heat shock proteins (hsps),** a family of proteins whose synthesis is increased at elevated temperatures. The chaperones do not change the final outcome of the folding process but act to prevent protein aggregation prior to the completion of folding and to prevent formation of metastable dead-end or nonproductive intermediates during folding. They increase the rate of the folding process by limiting the number of unproductive folding pathways available to a polypeptide. Chaperones of the hsp 70-kDa family bind to polypeptide chains as they are synthesized on the ribosomes, shielding the hydrophobic surfaces that would normally be exposed to solvent. This protects the protein from aggregation until the full chain is synthesized and folding can occur. Some proteins, however, cannot complete their folding process while in the presence of hsp70 chaperones and are delivered to the hsp60 family (GroEL in *Escherichia coli*) of chaperone proteins, also called **chaperonins.** The chaperonins form long cylindrical multisubunit quaternary structures that bind unfolded proteins in their molten-globular state within their central hydrophobic cavity. Chaperonins have an ATPase activity, hydrolyzing ATP as they facilitate folding. The folding process in *E. coli* is presented in Figure 2.46. Chaperone proteins are also required for refolding of proteins after they cross cellular membranes. A system of chaperones exists to facilitate protein transport into mitochondria and into and through the endoplasmic reticulum. Proteins cross the lipid bilayer of the mitochondrial and endoplasmic reticulum membranes in an unfolded conformation, and local chaperones are required to facilitate their folding.

Noncovalent Forces Lead to Protein Folding and Contribute to a Protein's Stability

Noncovalent forces cause a polypeptide to fold into a unique conformation and then stabilize the native structure against denaturation. Noncovalent forces are weak bonding forces with strengths of 1–7 kcal mol^{-1} (4–29 kJ mol^{-1}). This may be compared to the strength of covalent bonds that have a bonding strength

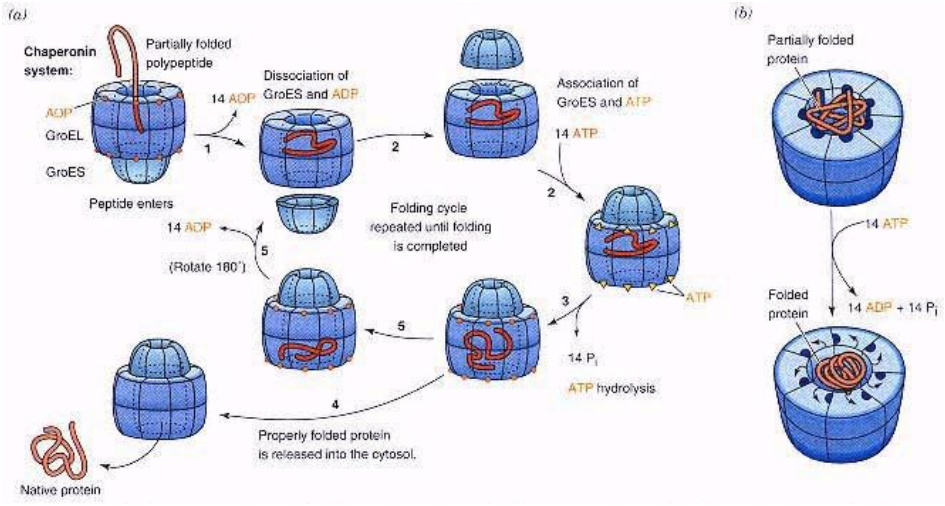


Figure 2.46
Chaperonin directed protein folding in *E. coli*.
(a) A proposed reaction cycle of the *E. coli* chaperonins GroEL and GroES in protein folding. (1) GroEL is a complex of 14 subunits, each with ADP attached. An associated ring of 8 GroES subunits binds an unfolded polypeptide in its central cavity and ADP and GroES are released. (2) Each GroEL subunit binds an ATP, weakening the interaction between unfolded polypeptide and GroEL. GroES is rebound on the opposite face of GroEL. (3) The 14 ATP are simultaneously hydrolyzed, releasing the bound polypeptide inside GroEL. The polypeptide, which is probably in its molten-globular state, folds in a protected microenvironment, preventing aggregation with other partially folded polypeptides. (4) The polypeptide is released from GroEL after folding into its native conformation. (5) If the polypeptide fails to attain its native fold, it remains bound to GroEL and reenters the reaction cycle at step 2. In the diagram GroEL turns over by 180°. GroES binds but does not hydrolyze ATP and facilitates the binding of ATP to GroEL. It coordinates simultaneous hydrolysis of ATP and prevents escape of a partially folded polypeptide from the GroEL cavity.
(b) A model for the ATP-dependent release of an unfolded polypeptide from its multiple attachment sites in GroEL. ATP binding and hydrolysis mask the hydrophobic sites of GroEL (darker areas) that bind to the unfolded polypeptide, thus permitting it to fold in an isolated environment.
Adapted from Hartl, R.-U, Hlodan, R., and Langer, T. *Trends Biochem. Sci*. 19:23, 1994. Figure reproduced with permission from Voet, D., and Voet, J. *Biochemistry*, 2nd ed., New York: John Wiley, 1995.

of at least 50 kcal mol^{-1} (Table 2.15). Even though individually weak, the large number of individually weak noncovalent contacts within a protein add up to a large energy factor that promotes protein folding.

Hydrophobic Interaction Forces

The most important noncovalent forces that cause a randomized polypeptide conformation to lose rotational freedom and fold into its native structure are **hydrophobic interaction forces.** The strength of a hydrophobic interaction is not due to an intrinsic attraction between nonpolar groups, but rather to the properties of the water solvent in which the nonpolar groups are dissolved. A nonpolar molecule or a region of a protein molecule dissolved in water induces a solvation shell of water in which water molecules are highly ordered. When two nonpolar side chains come together on folding of a polypeptide, the surface area exposed to solvent is reduced and some of the highly ordered water

TABLE 2.15 Bond Strength of Typical Bonds Found in Protein Structures

Bond Type	*Bond Strength* (*kcal mol*$^{-1}$)
Covalent bonds	>50
Noncovalent bonds	0.6–7
Hydrophobic bond	2–3
(i.e., two benzyl side chain groups of Phe)	
Hydrogen bond	1–7
Ionic bond	1–6
(low dielectric environment)	
van der Waals	<1
Average energy of kinetic motion (37°C)	0.6

molecules in the solvation shell are released to bulk solvent. Accordingly, the entropy of the system (i.e., net disorder of the water molecules in the system) is increased. The increase in entropy is thermodynamically favorable and is the driving force causing nonpolar moieties to come together in aqueous solvent. A favorable free energy change of –2 kcal mol^{-1} for association of two phenylal-anine side chain groups in water is due to this favorable water solvent entropy gain (Figure 2.47).

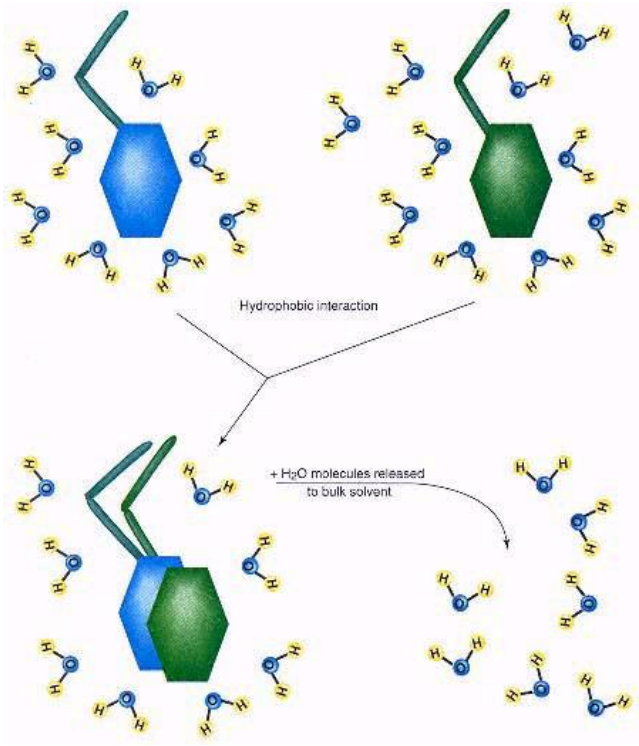


Figure 2.47
Formation of hydrophobic interaction between two phenylalanine side chain groups.

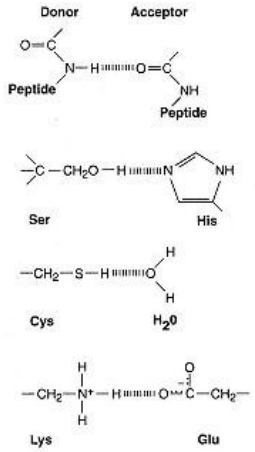


Figure 2.48
Some common hydrogen bonds found in proteins.

In transition from a random into a regular secondary conformation such as an α-helix or β-structure, approximately one-third of the ordered water of solvation about the unfolded polypeptide is lost to bulk solvent. This approximates 0.5–0.9 kcal mol^{-1} for each peptide residue. An additional one-third of the original solvation shell is lost when a protein that has folded into a secondary structure folds into a tertiary structure. The tertiary folding brings different segments of folded polypeptide chains into close proximity with the release of water of solvation between the polypeptide chains.

Hydrogen Bonds

Another noncovalent force in proteins is hydrogen bonding. **Hydrogen bonds** are formed when a hydrogen atom covalently bonded to an electronegative atom is shared with a second electronegative atom. The atom to which the hydrogen atom is covalently bonded is designated the hydrogen-**donor atom.** The atom with which the hydrogen atom is shared is the hydrogen-**acceptor atom.** Typical hydrogen bonds found in proteins are shown in Figure 2.48. α-Helical and β-structure conformations are extensively hydrogen bonded.

The strength of a hydrogen bond is dependent on the distance between the donor and acceptor atoms. High bonding energies occur when the distance is between 2.7 and 3.1 A. Of lesser importance, but not negligible, to bonding strength is the dependence of hydrogen-bond strength on geometry. Bonds of higher energy are geometrically collinear, with donor, hydrogen, and acceptor atoms lying in a straight line. The dielectric constant of the medium around the hydrogen bond may also be reflected in the bonding strength. Typical hydrogen-bond strengths in proteins are 1–7 kcal mol^{-1}. Although hydrogen bonds contribute to thermodynamic stability of a protein's conformation, their formation may not be as major a driving force for folding as we might at first believe. This is because peptide bonds and other hydrogen-bonding groups in proteins form hydrogen bonds to the water solvent in the denatured state, and these bonds must be broken before the protein folds. The energy required to break the hydrogen bonds to water must be subtracted from the energy gained from formation of new hydrogen bonds between atoms in the folded protein in calculating the net contribution of hydrogen-bonding forces to the folding.

Electrostatic Interactions

Electrostatic interactions (also referred to as **ionic** or **salt linkages**) between charged groups are important in the stabilization of protein structure and in binding of charged ligands and substrates to proteins. Electrostatic forces are repulsive or attractive depending on whether the interacting charges are of the same or opposite sign. The strength of an electrostatic force (E_{el}) is directly dependent on the charge (Z) of each ion and is inversely dependent on the dielectric constant (D) of the solvent and the distance between the charges (r_{ab}) (Figure 2.49).

$$E_{el} \approx \frac{Z_A \cdot Z_B \cdot \varepsilon^2}{D \cdot r_{ab}}$$

Figure 2.49
Strength of electrostatic interactions.

Water has a high dielectric constant ($D = 80$), and interactions in water are relatively weak in comparison to the strength of charge interactions in the interior of a protein where the dielectric constant is low. However, most charged groups of proteins remain on the surface of the protein where they do not interact with other charged groups from the protein because of the high dielectric constant of the water solvent, but are stabilized by hydrogen bonding and polar interactions to the water. These water interactions generate the dominant forces that lead to placement of most charged groups of a protein on the outside of the folded structures.

Van der Waals–London Dispersion Forces

Van der Waals and **London dispersion forces** are the weakest of the noncovalent forces. They have an attractive term (A) inversely dependent on the 6th

power of the distance between two interacting atoms (r_{ab}), and a repulsive term (B) inversely dependent on the 12th power of r_{ab} (Figure 2.50). The A term contributes at its optimum distance an attractive force of less than 1 kcal mol^{-1} per atomic interaction due to the induction of complementary partial charges or dipoles in the electron density of adjacent atoms when the electron orbitals of the two atoms approach to a close distance. As the atoms come even closer, however, the repulsive component (term B) of the van der Waals force predominates as the electron orbitals of the adjacent atoms begin to overlap. The repulsive force is commonly called **steric hindrance.**

$$E_{VDW} = -\frac{A}{r_{ab}^{6}} + \frac{B}{r_{ab}^{12}}$$

Figure 2.50
Strength of van der Waals interactions.

The distance of maximum favorable interaction between two atoms is the **van der Waals contact distance,** which is the sum of the van der Waals radii for the two atoms (Figure 2.51). The van der Waals radii for atoms found in proteins are given in Table 2.16.

The van der Waals repulsive forces between atoms attached to a peptide bond are weakest at the specific ϕ and ψ angles compatible with the α-helix and β-strand structures. Thus van der Waals forces are critical for secondary structure formation in proteins. In folding into a tertiary structure, the number of weak van der Waals interactions that occur are in the thousands. Thus the total contribution of van der Waals–London dispersion forces to the stability of a folded structure is substantial, even though a single interaction between any two atoms is less than 1 kcal mol^{-1}.

A special type of interaction (π-electron–π-electron) occurs when two aromatic rings approach each other with electrons of their aromatic rings favorably interacting (Figure 2.52). This interaction can result in attractive forces of up to 6 kcal mol^{-1}. A number of π-π aromatic interactions occur in a typical folded protein, contributing to the stability of the folded structure.

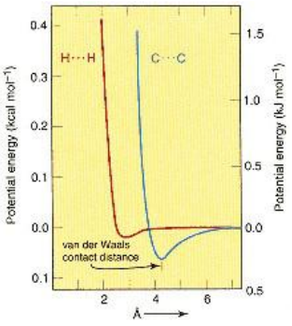


Figure 2.51
Van der Waals–London dispersion interaction energies between two hydrogen atoms and two (tetrahedral) carbon atoms.
Negative energies are favorable and positive energies unfavorable. Redrawn from Fersht, A. *Enzyme Structure and Mechanism*. San Francisco: Freeman, 1977, p. 228.

Denaturation of Proteins Leads to Loss of Native Structure

Denaturation occurs when a protein loses its native secondary, tertiary, and/or quaternary structure. The primary structure is not necessarily broken by denaturation. The **denatured state** is always correlated with the loss of a protein's function. Loss of a protein's function is not necessarily synonymous with denaturation, however, because small conformational changes can lead

TABLE 2.16 Covalent Bond Radii and van der Waals Radii for Selected Atoms

Atom	*Covalent Radius (Å)*	*van der Waals Radius (Å)*[a]
Carbon (tetrahedral)	0.77	2.0
Carbon (aromatic)	0.69 along=bond	1.70
	0.73 along–bond	
Carbon (amide)	0.72 to amide N	1.50
	0.67 to oxygen	
	0.75 to chain C	
Hydrogen	0.33	1.0
Oxygen (–O–)	0.66	1.35
Oxygen (=O)	0.57	1.35
Nitrogen (amide)	0.60 to amide C	1.45
	0.70 to hydrogen bond H	
	0.70 to chain C	
Sulfur, diagonal	1.04	1.70

Source: Fasman, G. D. (Ed.), *CRC Handbook of Biochemistry and Molecular Biology*, 3rd ed., Sect. D, Vol. II, Boca Raton, FL: CRC Press, 1976, p. 221.

[a] The van der Waals contact distance is the sum of the two van der Waals radii for the two atoms in proximity.

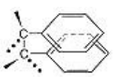


Figure 2.52
-Electron– -electron interactions between two aromatic rings.

to loss of function. A change in conformation of a single side chain in the active site of an enzyme or a change in protonation of a side chain can result in loss of activity, but does not lead to a complete loss of the native protein structure.

Even though conformational differences between denatured and native structures may be substantial, the free energy difference between such structures can in some cases be as low as the free energy of three or four noncovalent bonds. Thus the loss of a single hydrogen bond or electrostatic or hydrophobic interaction can lead to destabilization of a folded structure. A change in stability of a noncovalent bond can, in turn, be caused by a change in pH, ionic strength, or temperature. Binding of prosthetic groups, cofactors, and substrates also affects stability of the native conformation.

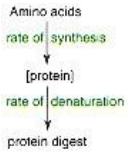


Figure 2.53
Steady-state concentration of a protein is due to its rates of synthesis and denaturation.

The statement that the breaking of a single noncovalent bond in a protein can cause denaturation apparently conflicts with the observation that the amino acid sequence can often be extensively varied without loss of a protein's structure. The key to the resolution of this apparent conflict is the word "essential." Many noncovalent interactions are not essential to the structural stability of the native conformation of a protein. However, substitution or modification of an essential amino acid that provides a critical noncovalent interaction dramatically affects the stability of a native protein structure.

The concentration of a protein in a cell is controlled by its rate of synthesis and degradation (Figure 2.53). Understanding the processes that control protein degradation is therefore as equally important as an understanding of the processes that regulate protein synthesis. Under many circumstances the denaturation of a protein is the rate-controlling step in its degradation. Cellular enzymes and organelles that digest proteins "recognize" denatured protein conformations and eliminate them rapidly. In experimental situations, protein denaturation occurs on addition of urea or detergents (sodium dodecyl sulfate or guanidine hydrochloride) that weaken hydrophobic bonding in proteins. These reagents stabilize the denatured state and shift the equilibrium toward the denatured form of the protein. Addition of strong base, acid, or organic solvent, or heating to temperatures above 60°C are also common ways to denature a protein.

2.8— Dynamic Aspects of Protein Structure

While high-resolution X-ray diffraction experiments yield atomic coordinates for each atom in a protein structure, experimental evidence from NMR, fluorescence spectroscopy, and the temperature dependence of the X-ray diffraction reveals that the atoms in a folded protein molecule have a fluid-like dynamic motion and do not exist in a single static position. Rather than an exact location, the atomic coordinates obtained by X-ray diffraction represent the time-averaged position for each atom. The time frame for position averaging is the length of time for data collection, which may be several days. Thus the active conformation may differ from the average conformation. An X-ray structure also shows small "defects" in packing of the folded structure, indicating the existence of "holes" in the structure that will allow the protein space for flexibility. The concept that each atom in a protein is in constant motion such as molecules within a fluid, although constrained by its covalent bonds and the secondary and tertiary structure, is an important aspect of protein structure.

Theoretical **molecular dynamics** calculations describe the changes in coordinates of atoms in a folded protein structure, with corresponding changes in position of regions of the structure due to summation of the movements of atoms in that region. The dynamic motion computation is based on the solving of Newton's equations of motion simultaneously for all the atoms of the protein and the solvent that interacts with the protein. The energy functions used in the equation include representations of covalent and noncovalent bonding energies due to electrostatic forces, hydrogen bonding, and van der Waals

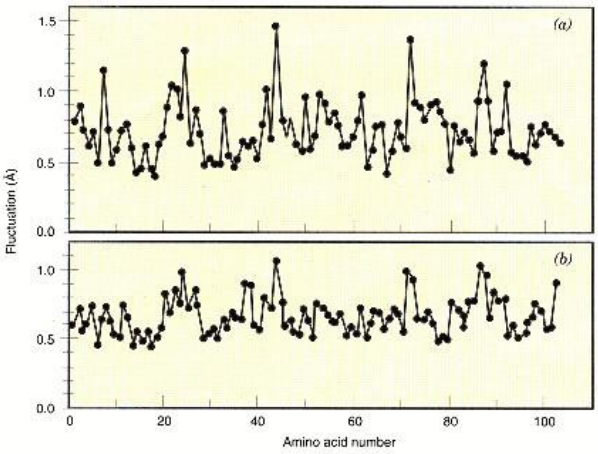


Figure 2.54
Fluctuation of structure of cytochrome c.
(a) Calculated fluctuation on a picosecond time scale of -carbons within each amino acid residue in the folded structure of cytochrome-c and (b) experimentally observed fluctuation of each -carbon of the amino acid residues determined from the temperature dependency of the X-ray diffraction pattern for the protein. Cytochrome-*c* has 103 amino acid residues. The *x*-axis plots the amino acid residues in cytochrome-*c* from 1 to 103, and the *y*-axis the fluctuation distances in angstroms.
Redrawn from Karplus, M., and McCammon, J. A. *Annu. Rev. Biochem*. 53:263, 1983.

forces. Individual atoms are randomly assigned a velocity from a theoretical distribution and Newton's equations are used to "relax" the system at a given "temperature." The calculation is a computationally intensive activity, even when limited to less than several hundred picoseconds (1 ps = 10^{-12} s) of protein dynamic time, and frequently requires supercomputers. These calculations indicate that the average atom within a typical protein is oscillating over a distance of 0.7 Å on the picosecond scale. Some atoms or groups of atoms move smaller or larger distances than this calculated average (Figure 2.54).

Net movement of any segment of a polypeptide over time represents the sum of forces due to rapid atomic oscillations and the local jiggling and elastic movements of covalently attached groups of atoms. These movements within the closely packed interior of a protein molecule are frequently large enough to allow the planar aromatic rings of buried tyrosines to flip. Furthermore, the small amplitude fluctuations provide the "lubricant" for large motions in proteins such as domain motions and quaternary structure changes, like those observed in hemoglobin on O_2 binding (see Chapter 3). The dynamic behavior of proteins is implicated in conformational changes induced by substrate, inhibitor, or drug binding to enzymes and receptors, generation of allosteric effects in hemoglobins, electron transfer in cytochromes, and in the formation of supramolecular assemblies such as viruses. The movements may also have a functional role in the protein's mechanism of action.

2.9—
Methods for Characterization, Purification, and Study of Protein Structure and Organization

Separation of Proteins on Basis of Charge

In **electrophoresis,** the protein dissolved in a solution buffered at a particular pH is placed in an electric field. Depending on the relationship of the buffer pH to the p*I* of the protein, the molecule moves toward the cathode (–) or the anode (+) or remains stationary (pH = p*I*). Procedures for electrophoresis use supports such as polymer gels (e.g., polyacrylamide), starch, or paper. The inert supports are saturated with buffer solution, a sample of protein is placed

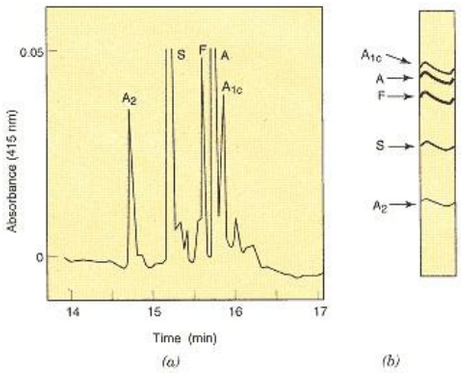


Figure 2.55
Isoelectric focusing of hemoglobins from patient heterozygous for HbS and β-thalassemia.
Figure shows separation by isoelectric focusing of HbA_{1c} (HbA glycosylated on NH_2 end, see Clin. Corr. 2.7), normal adult HbA, fetal HbF, sickle cell HbS (see Clin. Corr. 2.3), and the minor adult HbA_2.
(a) Isoelectric focusing carried out by capillary electrophoresis with ampholyte pH range between 6.7 and 7.7 and detection of bands at 415 nm.
(b) Isoelectric focusing carried out on gel with Pharmacia PhastSystem; ampholyte pH range is between 6.7 and 7.7.
From Molteni, S., Frischknecht, H., and Thormann, W. *Electrophoresis* 15:22, 1994 (Figure 4, parts A and B).

on the support, an electric field is applied across the support, and the charged proteins migrate in the support toward the oppositely charged pole.

An electrophoresis technique with extremely high resolution is **isoelectric focusing,** in which mixtures of polyamino–polycarboxylic acid ampholytes with a defined range of p*I* values are used to establish a pH gradient across the applied electric field. A charged protein migrates through the pH gradient in the electric field until it reaches a pH region in the gradient equal to its p*I* value. At this point the protein becomes stationary and may be visualized (Figure 2.55). Proteins that differ by as little as 0.0025 in p*I* values are separated on the appropriate pH gradient.

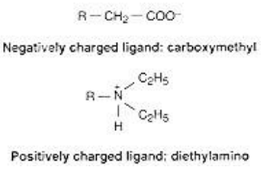


Figure 2.56
Two examples of charged ligands used in ion-exchange chromatography.

Ion-exchange column chromatography is used for preparative separation of proteins by charge. Ion-exchange resins consist of insoluble materials (agarose, polyacrylamide, cellulose, and glass) that contain negatively or positively charged groups (Figure 2.56). Negatively charged resins bind cations strongly and are **cation-exchange resins.** Positively charged resins bind anions strongly and are **anion-exchange resins.** The degree of retardation of a protein (or an amino acid) by a resin depends on the magnitude of the charge on the protein at the particular pH of the experiment. Molecules of the same charge as the resin are eluted first in a single band, followed by proteins with an opposite charge to that of the resin, in an order based on the protein's charge density (Figure 2.57). When it is difficult to remove a molecule from the resin because of the strength of the attractive interaction between the bound molecule and resin, systematic changes in pH or in ionic strength are used to weaken the interaction. For example, an increasing pH gradient through a cation-exchange resin reduces the difference between the solution pH and the p*I* of the bound protein. This decrease between pH and p*I* reduces the magnitude of the net charge on the protein and decreases the strength of the charge interaction between the protein and the resin. An increasing gradient of ionic strength also decreases the strength of charge interactions and elutes tightly bound electrolytes from the resin.

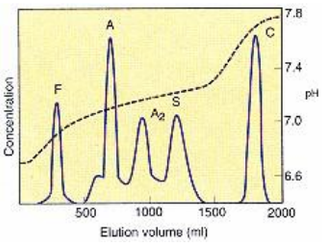


Figure 2.57
Example of ion-exchange chromatography.
Elution diagram of an artificial mixture of hemoglobins F, A, A_2, S, and C on carboxymethyl– Sephadex C-50.
From Dozy, A. M., and Juisman, T. H. J. J. *Chromatog.* 40:62, 1969.

Capillary Electrophoresis

Electrophoresis within a fused silica capillary tube has a high separation efficiency, utilizes very small samples, and requires only several minutes for an assay. A long capillary tube is filled with the electrophoretic medium, the sample is injected in a narrow band near the anode end of the tube, and the molecules of the sample are separated by their mobility toward the negatively charged pole. The fused silica wall of the capillary has a negatively charged surface to

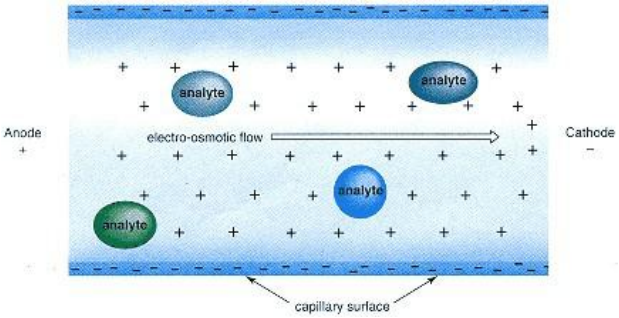


Figure 2.58
Generation of electro-osmotic flow toward cathode in capillary electrophoresis.

which an immobile cationic layer is fixed. An adjacent diffuse layer of cations moves toward the cathode in the applied electric field and causes a flow of solvent toward the cathode. This electro-osmotic flow creates a "current" that carries analyte molecules toward the cathode, irrespective of the analyte's charge (Figure 2.58). Molecules with a high positive charge to mass ratio "swim" with the current and have the highest mobility, followed by neutral molecules. Anionic molecules will be repelled by the cathode and will "swim" against the electro-osmotic flow. However, the electro-osmotic current toward the cathode overcomes any negative migration, and anions also migrate toward the cathode but at a slower rate than the cationic or neutral molecules.

In addition to zone electrophoresis, in which the separations are run in the presence of a single buffer, capillary electrophoresis may be performed in the presence of ampholytes to separate proteins by isoelectric focusing, in the presence of a porous gel to separate proteins by molecular weight, or in the presence of a micellar component to separate by hydrophobicity. Detectors that utilize UV light, fluorescence, Raman spectroscopy, electrochemical detection, or mass spectroscopy make the capillary method sensitive and versatile.

Separation of Proteins Based on Molecular Mass or Size

Ultracentrifugation:
Definition of Svedberg Coefficient

A protein subjected to centrifugal force moves in the direction of the force at a velocity dependent on its mass. The rate of movement is measured with an appropriate optical detection system, and from the rate the sedimentation coefficient is calculated in **Svedberg units** (units of 10^{-13} s). In the equation (Figure 2.59), v is the measured velocity of protein movement, ω the angular velocity of the centrifuge rotor, and r the distance from the center of the tube in which the protein is placed to the center of rotation. Sedimentation coefficients between 1 and 200 Svedberg units (S) have been found for proteins (Table 2.17). Equations have been derived to relate the sedimentation coefficient to the molecular mass of a protein. One of the more simple equations is shown in Figure 2.60, in which R is the gas constant, T the temperature, s the sedimentation coefficient, D the diffusion coefficient of the protein, $\bar{\nu}$ are difficult, usually only the sedimentation coefficient for a molecule is reported. A protein's sedimentation coefficient is a qualitative measurement of molecular mass.

$$s = \frac{v}{\omega^2 r}$$

Figure 2.59
Equation for calculation of the Svedberg coefficient.

$$\text{Molecular weight} = \frac{RTs}{D(1 - \bar{\nu}\rho)}$$

Figure 2.60
An equation relating the Svedberg coefficient to molecular weight.

TABLE 2.17 Svedberg Coefficients for Some Plasma Proteins of Different Molecular Weights

Protein	$s_{20,} \times 10^{-13}$ *(cm s^{-1} dyn^{-1})*[a]	*Molecular Weight*
Lysozyme	2.19	15,000–16,000
Albumin	4.6	69,000
Immunoglobulin G	6.6–7.2	153,000
Fibrinogen	7.63	341,000
Clq (factor of complement)	11.1	410,000
α_2-Macroglobulin	19.6	820,000
Immunoglobulin M	18–20	1,000,000
Factor VIII of blood coagulation	23.7	1,120,000

Source: Fasman, G.D. (Ed.), *CRC Handbook of Biochemistry and Molecular Biology*, 3rd ed., Sect. A, Vol. II. Boca Raton, FL: CRC Press, 1976, p. 242.

[a] $s_{20,} \times 10^{-13}$ is sedimentation coefficient in Svedberg units, referred to water at 20°C, and extrapolated to zero concentration of protein.

Molecular Exclusion Chromatography

A porous gel in the form of small insoluble beads is commonly used to separate proteins by size in column chromatography. Small protein molecules penetrate the pores of the gel and have a larger solvent volume through which to travel in the column than large proteins, which are sterically excluded from the pores. Accordingly, a protein mixture is separated by size. The larger proteins are eluted first, followed by the smaller proteins, which are retarded by their accessibility to a larger solvent volume (Figure 2.61). As with ultracentrifugation, an assumption is made as to the geometry of an unknown protein in the determination of molecular mass. Elongated nonspheroidal proteins give anomalous molecular masses when analyzed using a standard curve determined with proteins of spheroidal geometry.

Polyacrylamide Gel Electrophoresis in the Presence of a Detergent

If a charged detergent is added to a protein electrophoresis assay and electrophoresis occurs through a sieving support, the separation of proteins is based on protein size and not charge. A common detergent is **sodium dodecyl sulfate (SDS)** and a common sieving support is **cross-linked polyacrylamide.** The dodecyl sulfates are amphophilic C12 alkyl sulfate molecules, which stabilize a denatured protein by forming a charged micellar SDS solvation shell around its polypeptide chain. The inherent charge of the polypeptide chain is obliterated by the negatively charged micelle of SDS molecules, and each protein–SDS solubilized aggregate has an identical charge per unit volume. Negatively charged micelle particles move through the polyacrylamide gel toward the anode. Polyacrylamide acts as a molecular sieve and the protein–micelle complexes are separated by size; proteins of larger mass are retarded. A single band in an SDS polyacrylamide electrophoresis experiment is often used to demonstrate the purity of a protein. The conformation of the native structure is not a factor in the calculation of molecular mass, which is determined by comparison to known standards that are similarly denatured. The detergent dissociates quaternary structure into its constituent subunits. Only the molecular masses of the covalent polypeptide subunits within a protein are determined by this method.

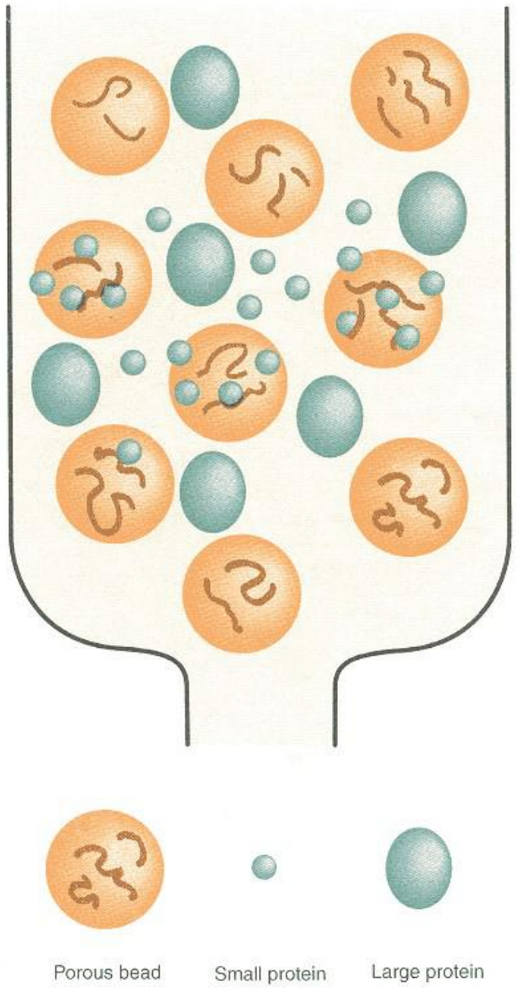


Figure 2.61
Molecular exclusion chromatography.
A small protein can enter the porous gel particles and will be retarded on the column with respect to a larger protein that cannot enter the porous gel particles.

HPLC Chromatographic Techniques Separate Amino Acids, Peptides, and Proteins

In high-performance liquid chromatography (HPLC), a liquid solvent containing a mixture of components to be identified is passed through a column densely

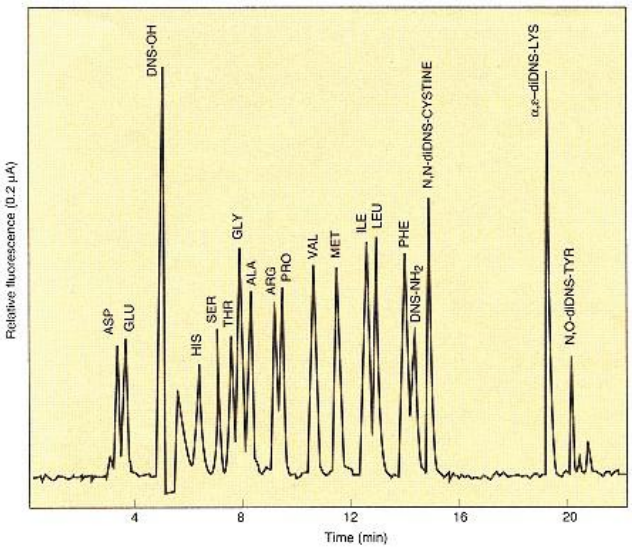


Figure 2.62
Separation of amino acids utilizing reverse-phase HPLC.
The x-axis is time of elution from column. Amino acids are derivatized by reaction with dansyl chloride (DNS) so that they emit a fluorescence that is used to assay them as they are eluted from the column.
Reprinted with permission from Hunkapiller, M. W., Strickler, J. E., and Wilson, K. J. *Science* 226:304, 1984.

packed with a small-diameter insoluble bead-like resin. In column chromatography, the smaller and more tightly packed the resin beads, the greater the resolution of the separation technique. In HPLC, the resin is so tightly packed that in order to overcome the resistance the liquid must be pumped through the column at high pressure. Therefore HPLC uses precise high-pressure pumps with metal plumbing and columns rather than glass and plastics as used in gravity chromatography. Resin beads are coated with charged chemical groups to separate compounds by ion exchange or with hydrophobic groups to retard hydrophobic nonpolar molecules. In hydrophobic chromatography, tightly associated nonpolar compounds are eluted from the hydrophobic beads in aqueous solvents containing various percentages of an organic reagent. The higher the percentage of organic solvent in the eluent, the faster the nonpolar component is eluted from the hydrophobic resin. This latter type of chromatography over nonpolar resin beads is called **reverse-phase HPLC** (Figure 2.62). The HPLC separations have extremely high resolution and reproducibility.

Affinity Chromatography

Proteins have a high affinity for their substrates, prosthetic groups, membrane receptors, specific noncovalent inhibitors, and antibodies made against them. These high-affinity compounds can be covalently attached to an insoluble resin and the modified resin used to purify its conjugate protein in column chromatography. In a mixture of proteins eluted through the resin, the one of interest is selectively retarded.

General Approach to Protein Purification

A protein must be purified prior to a meaningful characterization of its chemical composition, structure, and function. As living cells contain thousands of genetically distinct proteins, the purification of a single protein from the other cellular constituents may be difficult. The first task in the purification of a protein is

the development of a simple assay for the protein. Whether it utilizes the rate of a substrate transformation to a product, an antibody–antigen reaction, or a physiological response in an animal assay system, a protein assay must give a quantitative measurement of activity per unit of protein concentration. This quantity is known as the sample's **specific activity.** The purpose of a purification procedure is to increase a sample's specific activity to the value expected for the pure protein. A typical protocol for purification of a soluble cellular protein involves the disruption of the cellular membrane, followed by a differential centrifugation in a density gradient to isolate the protein from subcellular particles and high molecular weight aggregates. Further purification may utilize selective precipitation by addition of inorganic salts (salting out) or addition of miscible organic solvent to the solution containing the protein. Final purification will include a combination of techniques that separate based on molecular charge, molecular size, and affinity.

CLINICAL CORRELATION 2.8

Use of Amino Acid Analysis in Diagnosis of Disease

There are a number of clinical disorders in which a high concentration of amino acids is found in plasma and urine. An abnormally high concentration of an amino acid in urine is called an aminoaciduria. In phenylketonuria patients lack sufficient amounts of the enzyme phenylalanine hydroxylase, which catalyzes the transformation of phenylalanine to tyrosine. As a result, large concentrations of phenylalanine, phenylpyruvate, and phenyllactate accumulate in the plasma and urine. Phenylketonuria occurs clinically in the first few weeks after birth, and if the infant is not placed on a special diet, severe mental retardation will occur (see Clin. Corr. 11.5). Cystinuria is a genetically transmitted defect in the membrane transport system for cystine and the basic amino acids (lysine, arginine, and the derived amino acid ornithine) in epithelial cells. Large amounts of these amino acids are excreted in urine. Other symptoms of this disease may arise from the formation of renal stones composed of cystine precipitated within the kidney (see Clin. Corr. 11.9). Hartnup disease is a genetically transmitted defect in epithelial cell transport of neutral amino acids (mono-amino monocarboxylic acids), and high concentrations of these amino acids are found in the urine. The physical symptoms of the disease are primarily caused by a deficiency of tryptophan. These symptoms may include a pellagra-like rash (nicotinamide is partly derived from tryptophan) and cerebellar ataxia (irregular and jerky muscular movements) due to the toxic effects of indole derived from the bacterial degradation of unabsorbed tryptophan present in large amounts in the gut. Fanconi's syndrome is a generalized aminoaciduria associated with hypophosphatemia and a high excretion of glucose. Abnormal reabsorption of amino acids, phosphate, and glucose by the tubular cells is the underlying defect.

Determination of Amino Acid Composition of a Protein

Determination of the amino acid composition is an essential component in the study of a protein's structure and physiological properties. Analysis of the amino acid composition of physiological fluids (i.e., blood and urine) is utilized in diagnosis of disease (see Clin. Corr. 2.8). A protein is hydrolyzed to its constituent amino acids by heating the protein at 110°C in 6 N HCl for 18–36 h, in a sealed tube under vacuum to prevent degradation of oxidation-sensitive amino acid side chains by oxygen in air. Tryptophan is destroyed in this method and alternative procedures are used for its analysis. Asparagine and glutamine side chain amides are hydrolyzed to unsubstituted carboxylic acid side chains and free ammonia; thus they are counted within the glutamic acid and aspartic acid content in the analysis.

Common procedures for amino acid identification use cation-exchange chromatography or reverse-phase HPLC to separate the amino acids, which are then reacted with ninhydrin, fluorescamine, dansyl chloride, or similar chromophoric or fluorophoric reagents to quantitate the separated amino acids (Figure 2.62). With some types of derivatization, amino acids are identified at concentrations as low as 0.5×10^{-12} mol (pmol).

Techniques to Determine Amino Acid Sequence of a Protein

The ability to clone genes for proteins has led to the determination of the amino acid sequence of a protein as derived from the DNA or RNA sequences (see Chapter 18). This is a much faster method for obtaining an amino acid sequence. Sequencing of a protein, however, is required for the determination of modifications to the protein structure that occur after its biosynthesis, to identify a part of the protein sequence in order that its gene can be cloned, and to identify a protein as the product of a particular gene (see Chapter 17). Determination of the primary structure of a protein requires a purified protein. Many proteins contain several polypeptide chains and it is necessary to determine the number of chains in the protein. Individual chains are purified by the same techniques used in purification of the whole protein. If disulfide bonds covalently join the chains, these bonds have to be broken (Figure 2.63).

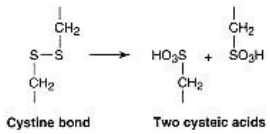


Figure 2.63
Breaking of disulfide bonds by oxidation to cysteic acids.

Polypeptide chains are most commonly sequenced by the **Edman reaction** (Figure 2.64) in which the polypeptide chain is reacted with phenylisothiocyanate, which forms a covalent bond to the NH_2-terminal amino acid. In this derivative, acidic conditions catalyze an intramolecular cyclization that results in cleavage of the NH_2-terminal amino acid from the polypeptide chain as a phenylthiohydantoin derivative. This NH_2-terminal amino acid derivative may be separated chromatographically and identified against standards. The polypeptide chain minus the NH_2-terminal amino acid is then isolated, and the

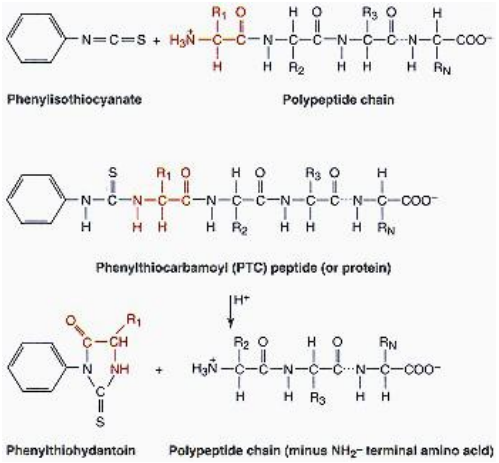


Figure 2.64
Edman reaction.

Edman reaction is repeated to identify the next NH_2-terminal amino acid. This series of reactions can theoretically be repeated until the sequence of the entire polypeptide chain is determined but under favorable conditions can only be carried out for 30 or 40 amino acids into the polypeptide chain. At this point in the analysis, impurities generated from incomplete reactions in the reaction series make further **Edman cycles** unfeasible. Since most polypeptide chains in proteins contain many more than 30 or 40 amino acids, they have to be hydrolyzed into smaller fragments and sequenced in sections.

Enzymatic and chemical methods are used to break polypeptide chains into smaller polypeptide fragments for sequencing. The enzyme **trypsin** preferentially catalyzes hydrolysis of the peptide bond on the COOH-terminal side of the basic amino acid residues of lysine and arginine within polypeptide chains. **Chymotrypsin** hydrolyzes peptide bonds on the COOH-terminal side of residues with large apolar side chains. Other proteolytic enzymes cleave polypeptide chains on the COOH-terminal side of glutamic and aspartic acid (Figure 2.65). **Cyanogen bromide** specifically cleaves peptide bonds on the COOH-terminal side of methionine residues within polypeptide chains (Figure

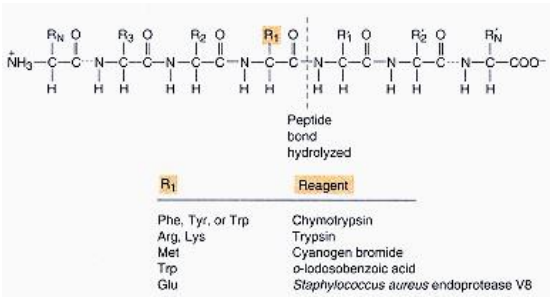


Figure 2.65
Specificity of some polypeptide cleaving reagents.

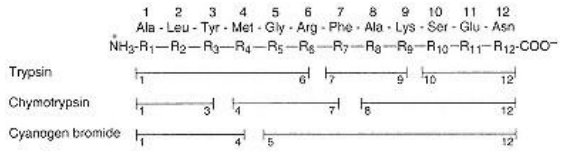


Figure 2.66
Ordering of peptide fragments from overlapping sequences produced by specific proteolysis of a peptide.

2.65). To establish the amino acid sequence of a large polypeptide, it is subjected to **partial hydrolysis** by one of the specific cleaving reagents, the polypeptide segments are separated, and the amino acid sequence of each of the small segments is determined by the Edman reaction. To place the sequenced peptides correctly into the complete sequence of the original polypeptide, a sample of the original polypeptide is subjected to a second partial hydrolysis by a different hydrolytic reagent from that used initially. This generates overlapping sequences to the first set of sequences, leading to the complete sequence (Figure 2.66).

X-Ray Diffraction Techniques Are Used to Determine the Three-Dimensional Structure of Proteins

X-ray diffraction enables determination of the three-dimensional structure of proteins at near atomic resolution. The approach requires formation of a protein crystal, which contains solvent and is thus a concentrated solution, for use as the target. Our present understanding of the detailed components of protein structure derived from experiments performed in this crystalline state correlate well with other physical measurements of protein structure in solution such as those made using NMR spectroscopy (see p. 81).

Generation of the protein crystal can be the most time-consuming aspect of the process. A significant factor in both experimental and computational handling of protein crystals, in contrast with most small molecule crystals, stems from the content of the protein crystalline material. Proteins exhibit molecular dimensions an order of magnitude greater than small molecules, and the packing of large protein molecules into the crystal lattice generates a crystal with large "holes" or solvent channels. A protein crystal typically contains 40–60% solvent and may be considered a concentrated solution rather than the hard crystalline solid associated with most small molecules. This attribute proves both beneficial and detrimental. The presence of solvent and unoccupied volume in the crystal allows the infusion of inhibitors and substrates into the protein molecules in the "crystalline state" but also permits a **dynamic flexibility** within regions of the protein structure. The flexibility may be seen as "disorder" in the X-ray diffraction experiment. Disorder is used to describe the situation in which the observed electron density can be fitted by more than a single local conformation. Two explanations for the disorder exist and must be distinguished. The first involves the presence of two or more static molecular conformations, which are present in a stoichiometric relationship. The second involves the actual dynamic range of motion exhibited by atoms or groups of atoms in localized regions of the molecule. An experimental distinction between the two explanations can be made by lowering the environmental temperature of the crystal to a point where dynamic disorder is "frozen out"; in contrast, the static disorder is not temperature dependent and persists. Analysis of dynamic disorder by its temperature dependency using X-ray diffraction determinations is an important method for studying protein dynamics (see Section 2.8). Crystallization techniques have advanced so that crystals are now obtainable from less abundant proteins. Interesting structures are reported for proteins in which specific amino

acid residues have been substituted, of antibody–antigen complexes, and of viral products such as the protease required for the infection of the human immunodeficiency virus (HIV) that causes acquired immunodeficiency syndrome (AIDS). Many structures have been solved by X-ray diffraction and the details are stored in a database called the **Protein Data Bank,** which is readily accessible.

Diffraction of X-ray radiation by a crystal occurs with incident radiation of a characteristic wavelength (e.g., copper, K_α = 1.54 Å). The X-ray beam is diffracted by the electrons surrounding the atomic nuclei in the crystal, with an intensity proportional to the number of electrons around the nucleus. Thus the technique establishes the **electron distribution** of the molecule and infers the nuclear distribution. The actual positions of atomic nuclei can be determined directly by diffraction with **neutron beam radiation,** an interesting but very expensive technique as it requires a source of neutrons (nuclear reactor or particle accelerator). With the highest resolution now available for X-ray diffraction determinations of protein structures, the electron diffraction from C, N, O, and S atoms can be observed. The diffraction from hydrogen atoms is not observed due to the low density of electrons—that is, a single electron—around a hydrogen nucleus. Detectors of the diffracted beam, typically photographic film or electronic area detectors, permit the recording of the amplitude (intensity) of radiation diffracted in a defined orientation. However, the data do not give information about phases of the radiation, which are essential to the solution of a protein's structure. Determination of the **phase angles** historically required the placement of heavy atoms (such as iodine, mercury, or lead) in the protein structure. Modern procedures, however, can often solve the phase problem without use of a heavy atom.

It is convenient to consider an analogy between X-ray crystallography and light microscopy to understand the processes involved in carrying out the three-dimensional structure determination. In light microscopy, incident radiation is reflected by an object under study and the reflected beam is recondensed by the objective lens to form an image of the object. The analogy is appropriate to incident X rays with the notable exception that no known material exists that can serve as an objective lens for X-ray radiation. To replace the objective lens, amplitude and phase angle measurements of the diffracted radiation are mathematically reconstructed by **Fourier synthesis** to yield a three-dimensional **electron-density map** of the diffracted object. Initially a few hundred reflections are obtained to construct a low-resolution electron-density map. For example, in one of the first protein crystallographic structures, 400 reflections were utilized to obtain a 6-Å map of myoglobin. At this level of resolution it is possible to locate clearly the molecule within the unit cell of the crystal and study the overall packing of the subunits in a protein with a quaternary structure. A trace of the polypeptide chain of an individual protein molecule is made with difficulty. Utilizing the low-resolution structure as a base, further reflections may be used to obtain higher-resolution maps. For myoglobin, where 400 reflections were utilized to obtain the 6-Å map, 10,000 reflections were needed for a 2-Å map, and 17,000 reflections for an extremely high-resolution 1.4-Å map. Many of these steps are now partially automated using computer graphics. A two-dimensional slice through a three-dimensional electron-density map of trypsinogen is shown in Figure 2.67. The known primary structure of the protein is fitted to the electron-density pattern by refinement. **Refinement** is the computer-intensive process of aligning a protein's amino acid sequence to the electron-density map until the best fit is obtained.

Whereas X-ray diffraction has provided extensive knowledge on protein structure, such a structure provides incomplete evidence for a protein's mechanism of action. The X-ray determined structure is an average structure of a molecule in which atoms are normally undergoing rapid fluctuations in solution (see Section 2.8). Thus the average crystalline structure determined by X-ray

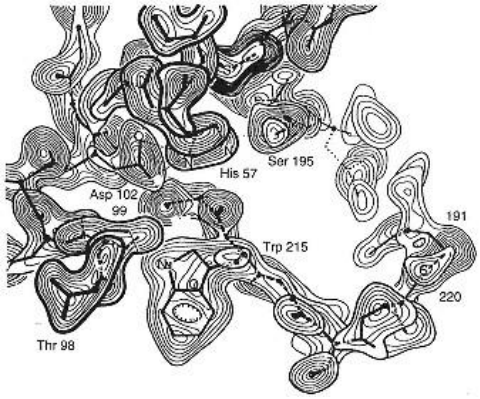


Figure 2.67
Electron-density map at 1.9-Å resolution of active site region of proenzyme form of trypsin.
Active site amino acid residues are fitted onto density map.
Reprinted with permission from Kossiakoff, A. A., Chambers, J. L., Kay, L. M., and Stroud, R. M. *Biochemistry* 16:654, 1977.

diffraction may not be the active structure of a particular protein in solution. A second important consideration is that it currently takes at least a day to collect data in order to determine a structure. On this time scale, the structures of reactive enzyme–substrate complexes, intermediates, and reaction transition states of an enzyme can not be observed. Rather, these structures must be inferred from the static pictures of an inactive form of the protein or from complexes with inactive analogs of the substrates of the protein (Figure 2.68). Newer methods for X-ray diffraction using synchrotron radiation to generate a X-ray beam at least 10,000 times brighter than that of standard X-ray generators will enable collection of diffraction data to solve a protein structure on a millisecond time scale. Application of the later X-ray techniques will enable scientists to determine short-lived structures and solve mechanistic and dynamic structural questions not addressable by current technology.

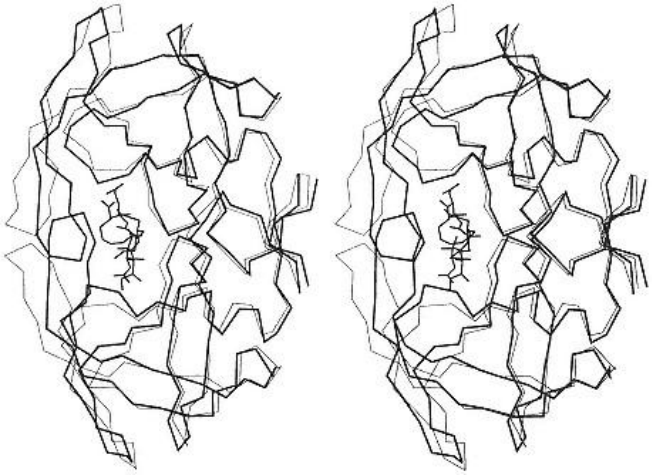


Figure 2.68
Stereo tracing of superimposed α-carbon backbone structure of HIV protease with inhibitor bound (thick lines) and the native structure of HIV protease without inhibitor bound (thin lines).
Redrawn with permission from Miller, M., Schneider, J., Sathyanarayana, B. K., Toth, M. V., Marshall, G. R., Clawson, L., Selk, L., Kent, S. B. H., and Wlodawer, A. *Science* 246: 1149, 1989.

Various Spectroscopic Methods Are Employed in Evaluating Protein Structure and Function

Ultraviolet Light Spectroscopy

The side chains of tyrosine, phenylalanine, and tryptophan, as well as peptide bonds in proteins, absorb **ultraviolet (UV) light**. The efficiency of light energy absorption for each **chromophore** is related to its **molar extinction coefficient** (ε). A typical protein spectrum is shown in Figure 2.69. The absorbance between 260 and 300 nm is primarily due to phenylalanine, tyrosine, and tryptophan side chain groups (Figure 2.70). When the tyrosine side chain is ionized at high pH (the tyrosine R group has a $pK'_a \approx 10$), the absorbance for tyrosine is shifted to a higher wavelength (red shifted) and its molar absorptivity is increased (Figure 2.70). Peptide bonds absorb in the far-UV (180–230 nm). A peptide bond in α-helix conformation interacts with the electrons of other peptide bonds above and below it in the spiral conformation to create an **exciton system** in which electrons are delocalized. The result is a shift of the absorption maximum from that of an isolated peptide bond to either a lower or higher wavelength (Figure 2.71). Thus UV spectroscopy can be used to study changes in a protein's secondary and tertiary structure. As a protein is denatured (helix unfolded), differences are observed in the absorption characteristics of the peptide bonds due to the disruption of the exciton system. In addition, the absorption maximum for an aromatic chromophore appears at a lower wavelength in an aqueous environment than in a nonpolar environment.

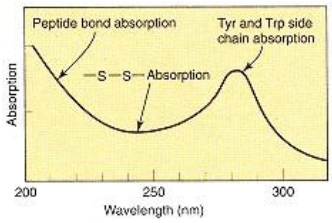


Figure 2.69
Ultraviolet absorption spectrum of the globular protein -chymotrypsin.

The **molar absorbency** of a chromophoric substrate often changes on binding to a protein. This change in the binding molecule's extinction coefficient can be used to measure its binding constant. Changes in chromophore extinction coefficients during the enzyme catalysis of a chemical reaction are used to obtain the kinetic parameters for the reaction.

Fluorescence Spectroscopy

The energy of an excited electron produced by light absorption is lost by various mechanisms and most commonly as thermal energy in a collision process. In some chromophores the excitation energy is dissipated by fluorescence. The **fluorescent emission** is always at a longer wavelength of light (lower energy) than the absorption wavelength of the fluorophore. Higher vibrational energy

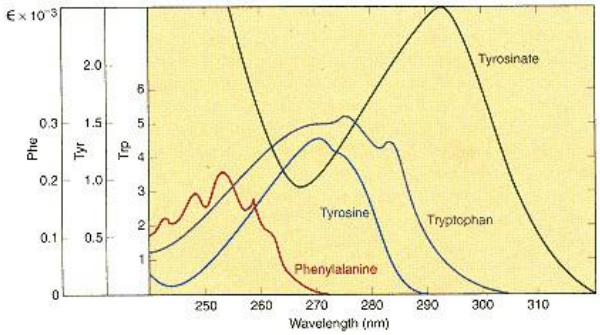


Figure 2.70
Ultraviolet absorption for the aromatic chromophores in Phe, Tyr,Trp, and tyrosinate.
Note differences in extinction coefficients on left axis for the different chromophores.
Redrawn from d'Albis *, A., and Gratzer, W. B. In: A. T. Bull, J. R. Lagmado, J.O. Thomas, and K.F. Tipton (Eds.), *Companion to Biochemistry*. London: Longmans, 1974, p. 170.

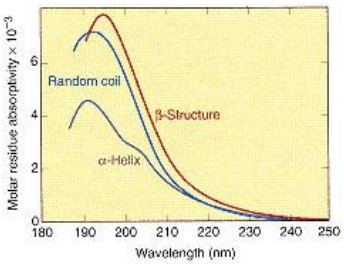


Figure 2.71
Ultraviolet absorption of the peptide bonds of a polypeptide chain in α-helix, random coil, and antiparallel β-structure conformations.
Redrawn from d'Albis, A., and Gratzer, W.B. In: A.T. Bull, J.R. Lagmado, J.O. Thomas, and K.F. Tipton (Eds.), *Companion to Biochemistry*. London: Longmans, 1970, p. 175.

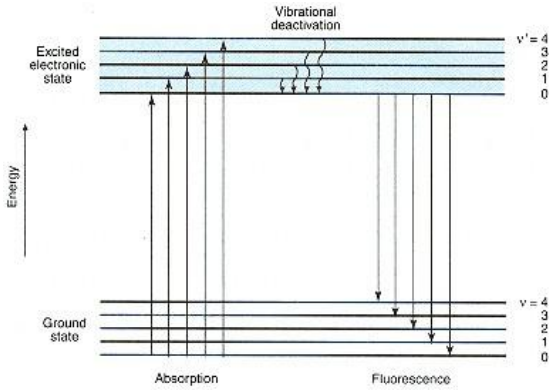


Figure 2.72
Absorption and fluorescence electronic transitions.
Excitation is from zero vibrational level in ground state to various higher vibrational levels in the excited state. Fluorescence is from zero vibrational level in excited electronic state to various vibrational levels in the ground state.
Redrawn from d'Albis *, A., and Gratzer, W. B. In: A. T. Bull, J. R. Lagmado, J. O. Thomas, and K. F. Tipton (Eds.), *Companion to Biochemistry*. London: Longmans, 1970, p. 166.

levels, formed in the excited electron state during the excitation event, are lost prior to the fluorescent event (Figure 2.72). If a chromophoric molecule is present that absorbs light energy emitted by the **fluorophore**, the emitted fluorescence is not observed. Rather, the fluorescence energy is transferred to the absorbing molecule. The **acceptor molecule**, in turn, either emits its own characteristic fluorescence or loses its excitation energy by an alternative process. If the acceptor molecule loses its excitation energy by a nonfluorescent process, it is acting as a **quencher** of the **donor molecule's** fluorescence. The efficiency of the **excitation transfer** is dependent on the distance and orientation between donor and acceptor molecules.

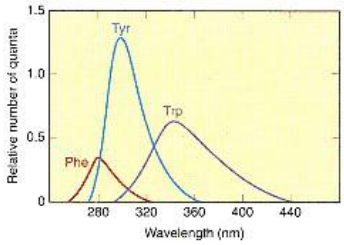


Figure 2.73
Characteristic fluorescence of aromatic groups in proteins.
Redrawn from d'Albis, A., and Gratzer, W. B. In: A. T. Bull, J. R. Lagmado, J. O. Thomas, and K. F. Tipton (Eds.), *Companion to Biochemistry*. London: Longmans, 1970, p. 478.

Fluorescence emission spectra for phenylalanine, tyrosine, and tryptophan side chains are shown in Figure 2.73. The emission wavelengths for phenylalanine overlap with the absorption wavelengths for tyrosine. In turn, the emission wavelengths for tyrosine overlap with the absorption wavelengths for tryptophan. Because of these overlaps in emission and absorption wavelengths, primarily only tryptophan fluorescence is observed in proteins that contain all of these amino acids. **Excitation energy transfers** occur over distances up to 80 Å, which are typical diameter distances in folded globular proteins. On protein denaturation, the distances between donor and acceptor groups become greater and decrease the efficiency of energy transfer to tryptophan. Accordingly, an increase in fluorescence due to tyrosines and/or phenylalanines is observed on denaturation of a protein. Since excitation transfer processes in proteins are distance and orientation dependent, the fluorescence yield is dependent on the conformation of the protein. The greatest sensitivity of this analysis occurs in its ability to detect changes due to solvent or binding interactions rather than establish absolute structure.

Optical Rotatory Dispersion and Circular Dichroism Spectroscopy

Optical rotation is caused by differences in the refractive index and **circular dichroism** (CD) is caused by differences in the light absorption between the clockwise and counterclockwise component vectors of a beam of polarized light as it travels through a solution containing an optically active molecule

such as an L-amino acid. In proteins the aromatic amino acids and the polypeptide chain generate an optical rotation and CD spectrum (Figure 2.74). Because of the differences between α-helical, β-structure, and random polypeptide spectra, circular dichroism has been a sensitive assay for the amount and type of secondary structure in a protein. Newer developments in vibrational circular dichroism examine the CD in regions of the spectrum more sensitive to protein backbone conformation.

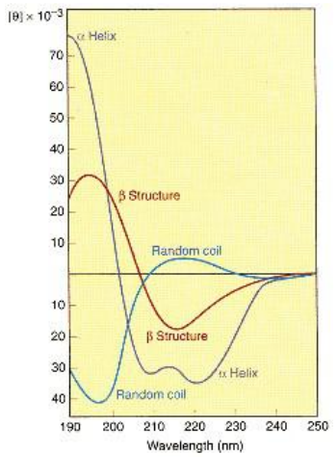


Figure 2.74
Circular dichroism spectra for polypeptide chains in α-helix, β-structure, and random coil conformations.
Redrawn from d'Albis *, A., and Gratzer, W. B. In: A. T. Bull, J. R. Lagmado, J. O. Thomas, and K. F. Tipton (Eds.), *Companion to Biochemistry*. London: Longmans, 1970, p. 190.

Nuclear Magnetic Resonance

With **two-dimensional** (2-D) **NMR** and powerful NMR spectrometers it is possible to obtain the solution conformation of small proteins of approximately 150 amino acids or less. Multidimensional NMR and triple resonance can extend the NMR to solve protein structures with up to 250 amino acids.

Conventional NMR techniques involve use of radiofrequency (rf) radiation to study the environment of atomic nuclei that are magnetic. The requirement for magnetic nuclei is absolute and is based on an unpaired spin state in the nucleus. Thus the naturally abundant carbon (^{12}C), nitrogen (^{14}N), and oxygen (^{16}O) do not absorb, while ^{13}C, ^{15}N, and ^{17}O do absorb. The absorption bands in an NMR spectrum are characterized by (1) a position or chemical shift value, reported as the frequency difference between that observed for a specific absorption band and that for a standard reference material; (2) the intensity of the peak or integrated area, which is proportional to the total number of absorbing nuclei; (3) the half-height peak width, which reflects the degree of motion in solution of the absorbing species; and (4) the coupling constant, which measures the extent of direct interaction or influence of neighboring nuclei on the absorbing nuclei. These four measurements enable the determination of the identity and number of nearest-neighbor groups that can affect the response of absorbing species through bonded interactions. They give no information on through-space (nonbonded) interaction due to the three-dimensional structure of the protein. To determine through-space interactions and protein tertiary structure requires the use of **nuclear Overhauser effects** (NOEs) and the application of the two-dimensional technique.

The major difference in two-dimensional versus one-dimensional (1-D) NMR is the addition of a second time delay rf pulse. The technique first requires the identification in the spectrum of a proton absorbance from a particular position in the protein structure. A maximum distance of approximately 5 Å is the limit for which these through-space interactions can be observed. Upon the generation of distance information for interresidue pairs through the protein structure, three-dimensional protein conformations consistent with the spectra are generated. In this calculation, a distance matrix is constructed containing ranges of distances (minimum and maximum) for as many interresidue interactions as may be measured. Possible structures are generated from the data consistent with the constraints imposed by the NMR spectra. Computational refinements of the initially calculated structures can be made to optimize covalent bond distances and bond angles. The method generates a family of structures, the variability showing either the imprecision of the technique or the dynamic "disorder" of the folded structure (Figure 2.75). Such computations based on NMR experiments have yielded structures for proteins that do not significantly differ from the time-averaged structure observed with X-ray diffraction methods.

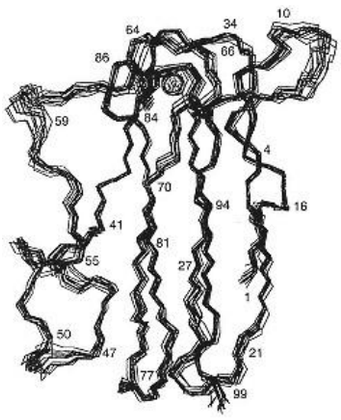


Figure 2.75
NMR structure of the protein plastocyanin from the French bean.
The structure shows superposition of eight structures of the polypeptide backbone for the protein, calculated from the constraints of the NMR spectrum.
From Moore, J. M., Lepre, C. A., Gippert, G. P., Chazin, W. J., Case, D. A., and Wright, P. E. *J. Mol. Biol.* 221:533, 1991. Figure generously supplied by P. E. Wright.

Other enhancements to NMR, which are applicable to the determination of protein structure, include the ability to synthesize proteins containing isotopically enriched (e.g., containing ^{13}C or ^{15}N) amino acids, and development of paramagnetic shift reagents to study localized environments on paramagnetic resonances, such as the lanthanide ion reporting groups.

Bibliography

Physical and Structural Properties of Proteins

Bowie, J. U., Luethy, R., and Eisenberg, D. A method to identify protein sequences that fold into a known three-dimensional structure. *Science* 253:164, 1991.

Bryson, J. W., Betz, S. F., Lu, H. S., Suich, D. H., Zhou, H. X., O'Neil, K. T., and DeGrado, W. F. Protein design: a hierarchic approach. *Science* 270:935, 1995.

Chothia, C. Principles that determine the structure of proteins. *Annu. Rev. Biochem*. 53:537, 1984.

Doolittle, R. F. The multiplicity of domains in proteins. *Annu. Rev. Biochem*. 64:287, 1995.

Eisenberg, D., and McLachlan, A. D. Solvation energy in protein folding and binding. *Nature* 319:199, 1986.

Fasman, G. D. Protein conformational prediction. *Trends Biochem. Sci*. 14:295, 1989.

Finkelstein, A. V., Gutun, A. M., and Badretdinov, A. Ya. Why are the same protein folds used to perform different functions? *FEBS Lett*. 325:23, 1993.

Flöckner, H., Braxenthaler, M., Lackner, P., Jaritz, M., Ortner, M., and Sippl, M. J. Progress in fold recognition. *Proteins 23*:376, 1995.

Mosimann, S., Meleshko, R., and James, M. N. G. A critical assessment of comparative molecular modeling of tertiary structures of proteins. *Proteins* 23:301, 1995.

Orengo, C. A., Jones, D. T., and Thornton, J. M. Protein superfamilies and domain superfolds. *Nature* 372:631, 1994.

Richardson, J. S. The anatomy and taxonomy of protein structure. *Adv. Protein Chem*. 34:168, 1981.

Rose, G. D., and Wolfenden, R. Hydrogen bonding, hydrophobicity, packing, and protein folding. *Annu. Rev. Biophys. Biomol. Struct*. 22:381, 1993.

Protein Folding

Anfinsen, C. B., and Scheraga, H. Experimental and theoretical aspects of protein folding. *Adv. Protein Chem*. 29:205, 1975.

Baldwin, R. L. How does protein folding get started? *Trends Biochem. Sci*. 14:291, 1989.

Bychkova, V. E., and Ptitsyn, O. B. The molten globule state is involved in genetic disease? *FEBS Lett*. 359:6, 1995.

Chen, S., Roseman, A. M., Hunter, A. S., Wood, S. P. Burston, S. G., Ranson, N. A., Clarke, A. R., and Saibil, H. R. Location of a folding protein and shape changes in GroEL–GroES complexes imaged by cryo-electron microscopy. *Nature* 371:261, 1994.

Dobson, C. M., Evans, P. A., and Radford, S. E. Understanding how proteins fold: the lysozyme story so far. *Trends Biochem. Sci*. 19:31, 1994.

Hendrick, J. P., and Hartl, F. U. The role of molecular chaperones in protein folding. *FASEB J*. 9:1559, 1995.

Jaenicke, R. Protein folding: local structures, domains, subunits and assemblies. *Biochemistry* 30:3147, 1991.

Lins, L., and Brasseur, R. The hydrophobic effect in protein folding. *FASEB J*. 9:535, 1995.

Ptitsyn, O. B. How the molten globule became. *Trends Biochem. Sci*. 20:376, 1995.

Shakhnovich, E., Abkevich, V., and Ptitsyn, O. Conserved residues and the mechanism of protein folding. *Nature* 379:96, 1996.

Sosnick, T. R., Mayne, L., Hiller, R., and Englander, S. W. The barriers in protein folding. *Nature Struct. Biol*. 1:149, 1994.

Wolynes, P. G., Onuchic, J. N., and Thirumalai, D. Navigating the folding routes. *Science* 267:1619, 1995.

Techniques for the Study of Proteins

Bax, A., and Grzesiek, S. Methodological advances in protein NMR. *Acc. Chem. Res*. 26:131, 1993.

Clauser, K. R., Hall, S. C., Smith, D. M., Webb, J. W., Andrews, L. E., Tran, H. M., Epstein, L. B., and Burlingame, A. L. Rapid mass spectrometric peptide sequencing and mass matching for characterization of human melanoma proteins isolated by two dimensional PAGE. *Proc. Natl. Acad. Sci. U.S.A*. 92:4801, 1995.

Farrow, N. A., Muhandiram, D. R., Singer, A. U., Pascal, S. M., Kay, C. M., Gish, G., Shoelson, S. E., Pawson, T., and Forman-Kay, L. E. Back-bone dynamics of a free and phosphopeptide-complexed Src homology 2 domain studied by ^{15}N NMR relaxation. *Biochemistry* 33:5984, 1994.

Nikolov, C. B., Chen, H., Halay, E. D., Usheva, A. A., Hisatake, K., Lee, D. K., Roeder, R. G., and Burley, S. K. Crystal structure of a TFIIB-TBP-TATA-element ternary complex. *Nature* 377:119, 1995.

Reif, O. W., Lausch, R., and Fritag, R. High-performance capillary electrophoresis of human serum and plasma proteins. *Adv. Chromatogr*. 34:1, 1994.

Shively, J. E., Paxton, R. J., and Lee, T. D. Highlights of protein structural analysis. *Trends Biochem. Sci*. 14:246, 1989.

Dynamics in Folded Proteins

Daggett, V., and Levitt, M. Realistic simulations of native-protein dynamics in solution and beyond. *Annu. Rev. Biophys. Biomol. Struct*. 22:353, 1993.

Joseph, D., Petsko, G. A., and Karplus, M. Anatomy of a conformational change: hinged lid motion of the triosephosphate isomerase loop. *Science* 249:1425, 1990.

Karplus, M., and McCammon, J. A. Dynamics of proteins: elements and function. *Annu. Rev. Biochem*. 53:263, 1983.

Karplus, M., and McCammon, J. A. The dynamics of proteins. *Sci. Am*. 254:42, 1986.

Smith, P. E., van Schaik, R. C., Szyperski, T., Würthrick, K., and van Gunsteren, W. F. Internal mobility of the basic pancreatic trypsin inhibitor in solution: a comparison of NMR spin relaxation measurements and molecular dynamics simulations. *J. Mol. Biol*. 246:356, 1995.

Glycoproteins

Brockhausen, I. Clinical aspects of glycoprotein biosynthesis. *Crit. Rev. Clin. Lab. Sci*. 30:65, 1993.

Lis., H., and Sharon, N. Protein glycosylation. Structural and functional aspects. *Eur. J. Biochem*. 218:1, 1993.

Paulson, J. C. Glycoproteins: what are the sugar chains for? *Trends Biochem. Sci*. 14:272, 1989.

Lipoproteins

Gotto, A. M. Jr. Plasma lipoproteins. In: A. Neuberger and L. L. M. van Deenen (Eds.), *New Comprehensive Biochemistry*. Amsterdam: Elsevier, 1987.

Myers, G. L., Cooper, G. R., and Sampson, E. J. Traditional lipoprotein profile: clinical utility, performance requirement, and standardization. *Atherosclerosis* 108:S157, 1994.

Schumaker, V. N., Phillips, M. L., and Chatterton, J. E. Apolipoprotein B and low-density lipoprotein structure: implications for biosynthesis of triglyceride-rich lipoproteins. *Adv. Protein Chem*. 45:205, 1994.

Schuster, B., Prassl, R., Nigon, F., Chapman, M. J., and Laggner, P. Core lipid structure is a major determinant of the oxidative resistance of low density lipoprotein. *Proc. Natl. Acad. Sci. U.S.A*. 92:2509, 1995.

Segrest, J. P., Garber, D. W., Brouillette, C. G., Harvey, S. C., and Ananthara-maiah, G. M. The amphipathic α helix: a multifunctional structural motif in plasma apolipoproteins. *Adv. Protein Chem.* 45:303, 1994.

Collagen

Brodsky, B., and Shah, N. K. The triple-helix motif in proteins. *FASEB J.* 9:1537, 1995.

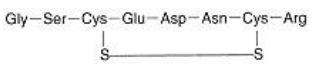
Kuivaniemi, H., Tromp, G., and Prockop, D. J. Mutations in collagen genes: causes of rare and some common diseases in humans. *FASEB J.* 5:2052, 1991.

Prockop, D. J., and Kivirikko, K. I. Collagens: molecular biology, diseases, and potentials for therapy. *Annu. Rev. Biochem.* 64:403, 1995.

Questions

J. Baggott and C. N. Angstadt

Refer to the following structure for Questions 1 and 2.



1. The peptide shown above:

   A. has arginine in position 1 of the sequence.

   B. contains a derived amino acid.

   C. is basic.

   D. consists entirely of amino acids with charged or nonpolar side chains.

   E. contains more amino acids with side chains that are charged than ones with electrically neutral side chains at pH 7.0.

2. The charge on the peptide shown above is about:

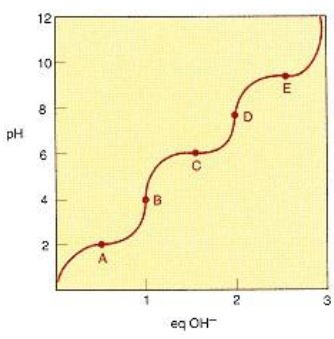
   A. –2 at pH > 13.5.

   B. –1 at pH ~ 11.5.

   C. +1 at pH ~ 6.5.

   D. +2 at pH ~ 5.5.

   E. 0 at pH ~ 4.5.



3. The figure above shows the titration curve of one of the common amino acids. From this curve we can conclude:

   A. the amino acid contains two carboxyl groups.

   B. at point B the amino acid is zwitterionic.

   C. the amino acid contains an aromatic hydroxyl group.

   D. point D corresponds to the pK'_a of an ionizable group.

   E. at point E the amino acid has a net negative charge.

4. Which of the following can be used for a quantitative determination of amino acids in general?

   A. acetic anhydride

   B. iodoacetate

   C. ninhydrin

   D. Pauly's reagent

   E. the Sakaguchi reaction

5. Which of the following is a conservative substitution?

   A. Val for Ile

   B. His for Pro

   C. Asp for Ala

   D. Lys for Leu

   E. Ser for Ala

6. Which of the following has quaternary structure?

   A. α-chymotrypsin

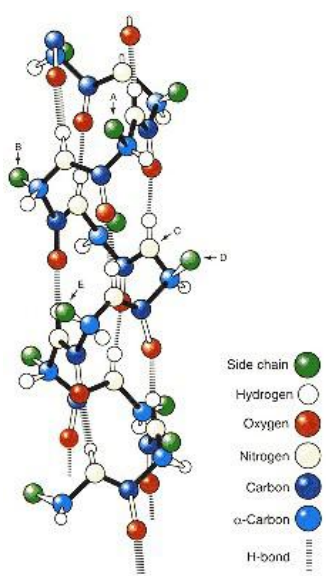
   B. hemoglobin

   C. insulin

   D. myoglobin

   E. trypsin

Refer to the drawing for Questions 7 and 8.



7. When group E contains a negatively charged carboxyl function, the structure is destabilized by:

   A. alanine at position A.

   B. arginine at position B.

   C. serine at position B.

   D. proline at position D.

   E. aspartate at position D.

8. The properties of atom C are essential to which stabilizing force in the structure?

A. hydrogen bonding

B. steric effects

C. ionic attraction

D. disulfide bridge

E. none of the above

Refer to the following for Questions 9–11.

A. disulfide bond formation

B. hydrogen bonding

C. hydrophobic interaction

D. electrostatic interactions

E. van der Waals forces

9. Which interaction is NOT formed when protein subunits combine to form a quaternary structure?

10. Driven by the entropy of water.

11. Repulsive forces between atoms attached to a peptide bond are weakest when the ϕ and φ angles are compatible with the α-helix and β-structures.

12. In collagen:

A. intrachain hydrogen bonding stabilizes the native structure.

B. three chains with polyproline type II helical conformation can wind about one another to form a superhelix because of the structure of glycine.

C. the ϕ angles contributed by proline are free to rotate, but the φ angles are constrained by the ring.

D. regions of superhelicity comprise the entire structure except for the N and C termini.

E. cross-links between triple helices form after an intracellular enzyme converts some of the lysine to allysine.

13. Chaperone proteins:

A. all require ATP to exert their effect.

B. cleave incorrect disulfide bonds, allowing correct ones to subsequently form.

C. guide the folding of polypeptide chains into patterns that would be thermodynamically unstable without the presence of chaperones.

D. are involved in transport of proteins across mitochondrial and endoplasmic reticulum membranes.

E. act primarily on fully synthesized polypeptide chains, unfolding incorrect structures so they can refold correctly.

14. Proteins may be separated according to size by:

A. isoelectric focusing.

B. polyacrylamide gel electrophoresis.

C. ion-exchange chromatography.

D. molecular exclusion chromatography.

E. reverse-phase HPLC.

Refer to the following for Questions 15–18.

A. primary structure

B. secondary structure

C. tertiary structure

D. quaternary structure

E. random conformation

15. All ϕ angles are equal and all φ angles are equal.

16. May bring distant segments of a single polypeptide chain into close juxtaposition.

17. Unaffected by binding of a charged detergent, such as SDS.

18. Exemplified by the β-structure (pleated sheet).

19. Changes in protein conformation can be detected rapidly by:

A. ultraviolet absorbance spectroscopy.

B. fluorescence emission spectroscopy.

C. optical rotatory dispersion.

D. circular dichroism.

E. all of the above.

Answers

1. B Cystine, formed by joining two cysteine residues through a disulfide bridge, is a derived amino acid (p. 30). A: The convention is to write the N terminal to the left. Numbering begins at the N terminal, so glycine is in position 1 (p. 29). C: The peptide contains two acidic amino acids, glutamate and aspartate, and only one basic amino acid, arginine, so it is acidic (p. 31). D and E: Cysteine is nonpolar and uncharged, and glutamate, aspartate, and arginine are charged at physiological pH; serine and asparagine are polar but are not charged (p. 26).

2. E At pH 4.5 the peptide is in the following ionic state: the N-terminal amino group is +1, the side chain carboxyls of glutamate and aspartate each average about –0.5 (since this pH is about at their pK values), the side chain of arginine is +1, and its terminal carboxyl group is –1. The sum is zero (pp. 30–33).

3. E The axes of this titration curve are reversed from the presentation in the text. The abscissa shows that three ionizable groups are present. The pK values, where the groups are 50% titrated, are at points A (pH ~ 2), C (pH ~ 6.5), and E (pH ~ 9.5). Histidine is the only common amino acid with these pK values. At point B, its net charge is –1 + 1 + 1 = +1. At point E, the net charge is –1 + 0 + 0.5 = –0.5 (pp. 31–34).

4. C A: Acetic anhydride acetylates tyrosyl residues. B: Iodoacetate reacts with the –SH of cysteine. D: Pauly's reagent reacts only with histidine and tyrosine. E: The Sakaguchi reaction is for arginine. (See Table 2.7.)

5. A B: His and Pro both have rings in their side chains, but the rings are very different. C–E: In each of these pairs the first amino acid is charged or polar, and the second has an aliphatic R group (p. 26).

6. B Quaternary structure consists of a specific noncovalent association of subunits having their own tertiary structures (p. 44). Myoglobin is a single polypeptide chain (p. 48). Insulin, trypsin (p. 40), and α-chymotrypsin (p. 46) are multichain proteins covalently joined by cystine bonds, and, each chain having arisen from a single polypeptide chain by proteolytic cleavage.

7. D Like charges in the third or fourth position in either direction from the designated position destabilize the helix due to charge repulsion. Thus aspartate at position D is harmless, whereas glutamate at position A or B would destabilize. Alanine has a small side chain. Proline destabilizes the α-helix conformation and is usually not found in either α-helix or β-structure (p. 44).

8. A Atom C is an amide nitrogen. The attached hydrogen atom participates in hydrogen bonding (pp. 43–44). Hydrogen bonds contribute to the stability of the structure (p. 43).

9. A Quaternary structure is stabilized exclusively by noncovalent interactions. Disulfide bonds are covalent (p. 48).

10. C Hydrophobic groups in contact with water result in formation of a relatively highly ordered solvation shell of water around the group. If the hydrophobic groups come together, eliminating the bound water, the water becomes more random, a favorable process (pp. 64–66).

11. E Van der Waals repulsive forces (as opposed to the attractive element of van der Waals forces) are weakest at these angles (p. 66).

12. B The close contacts in the interior of the triple helix are possible only when the R group of the amino acid at that position is very small, that is, hydrogen. A: The hydrogen bonding in collagen is interchain. C: The ϕ angle is part of the proline ring and is not free to rotate. D: Although the statement is true of type I collagen, the superhelical regions in other collagen types may be broken by regions of globular domains (p. 52). E: The conversion and cross-linking are extracellular.

13. D A: The hsp60 family of chaperones is ATP-linked, but the hsp70 family is not. B: Disulfide isomerases catalyze this reaction. C: The final product is thermodynamically stable; chaperones merely prevent unfavorable intermediate interactions. E: Hsp70 chaperones react with nascent polypeptide chains as they are synthesized by the ribosome. The protein may then be delivered to a hsp60 chaperone for facilitation of final folding (pp. 62–63).

14. D A–C separate molecules on the basis of charge (p. 72). E: Reverse-phase HPLC effects separations on the basis of polarity (p. 73).

15. B This statement is a definition of secondary structure (p. 43).

16. C This is a consequence of folding into a compact structure (p. 44).

17. A SDS binding produces an extended conformation of a polypeptide chain due to charge repulsion, but no peptide bonds are broken (p. 72).

18. B β-Structure is an important type of secondary structure (p. 44).

19. E A: Peptide bond absorption (180–230 nm) in the α-helical conformation differs from that in other conformations (p. 79). B: Excitation transfers become less efficient as donor and acceptor groups become further apart, as in denaturation. C and D: These effects of optically active chromophors upon polarized light are sensitive to environment; in addition, the peptide bond itself becomes part of an optically active system when it forms an asymmetric structure like the α-helix (p. 80).