

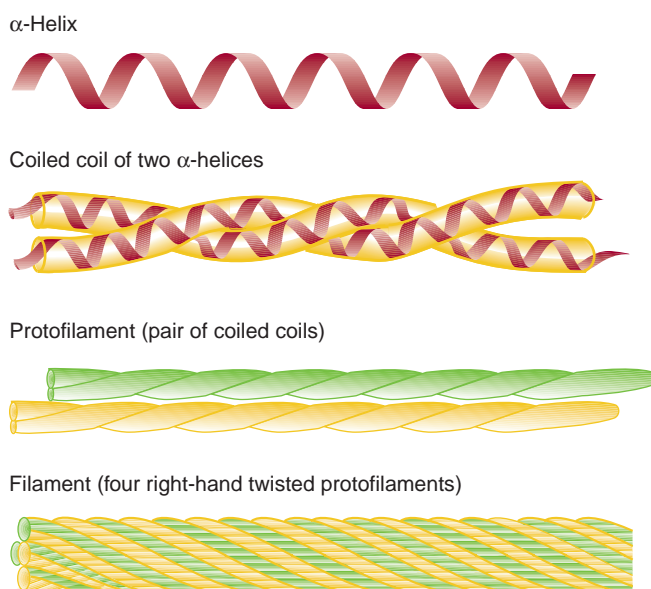
roles. Keratin (Figure 5.33) is a fibrous protein composed of bundles of  $\alpha$ -helices, whereas the polypeptide chains of silk fibroin (Figure 5.34) are arranged in antiparallel  $\beta$ -pleated sheets. The structural features of collagen, the most abundant protein in vertebrates, are described in some detail.

**COLLAGEN** Collagen is synthesized by connective tissue cells and then secreted into the extracellular space to become part of the connective tissue matrix. The 20 major families of collagen molecules include many closely related proteins that have diverse functions. The genetically distinct collagen molecules in skin, bones, tendons, blood vessels, and corneas impart to these structures many of their special properties (e.g., the tensile strength of tendons and the transparency of corneas).

Collagen is composed of three left-handed polypeptide helices that are twisted around each other to form a right-handed triple helix (Figure 5.35). Type I collagen molecules, found in teeth, bone, skin, and tendons, are about 300 nm long and approximately 1.5 nm wide. Approximately 90% of the collagen found in humans is type I.

The amino acid composition of collagen is distinctive. Glycine constitutes approximately one-third of the amino acid residues. Proline and 4-hydroxyproline may account for as much as 30% of a collagen molecule's amino acid composition. Small amounts of 3-hydroxyproline and 5-hydroxylysine also occur. (Specific proline and lysine residues in collagen's primary sequence are hydroxylated within the rough ER after the polypeptides have been synthesized. These reactions, which are discussed in Chapter 19, require ascorbic acid (p. 763).

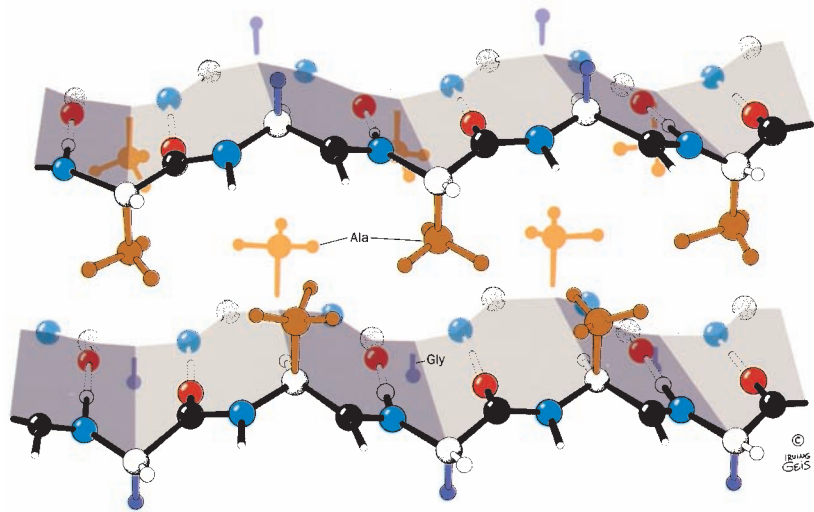
Collagen's amino acid sequence primarily consists of large numbers of repeating triplets with the sequence of Gly—X—Y, in which X and Y are often proline and hydroxyproline. Hydroxylysine is also found in the Y position. Simple carbohydrate groups are often attached to the hydroxyl group of hydroxylysine residues. It has been suggested that collagen's carbohydrate components are



**FIGURE 5.33**

**$\alpha$ -Keratin**

The  $\alpha$ -helical rodlike domains of two keratin polypeptides form a coiled coil. Two staggered antiparallel rows of these dimers form a supercoiled protofilament. Hydrogen bonds and disulfide bridges are the principal interactions between subunits. Hundreds of filaments, each containing four protofilaments, form a macrofibril. Each hair cell, also called a fiber, contains several macrofibrils. Each strand of hair consists of numerous dead cells packed with keratin molecules. In addition to hair, the keratins are also found in wool, skin, horns, and fingernails.

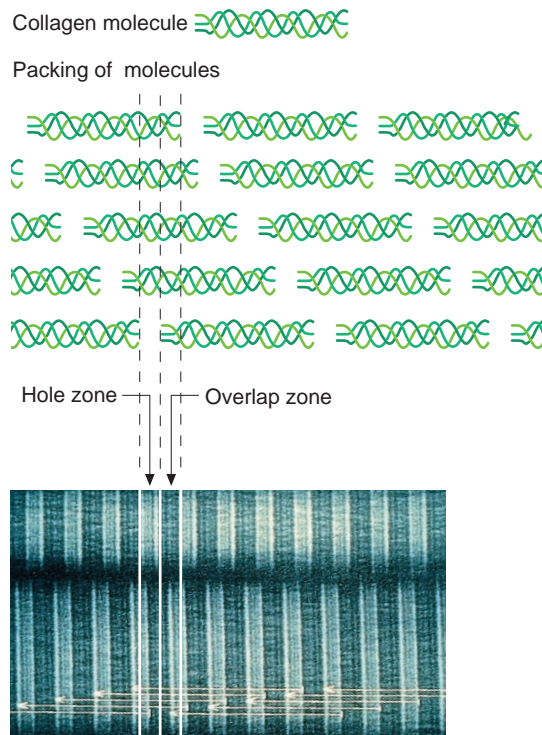


**FIGURE 5.34**  
Molecular Model of Silk Fibroin

In fibroin, the silk fibrous protein, the polypeptide chains are arranged in fully extended antiparallel  $\beta$ -pleated sheet conformations. Note that the R groups of alanine on one side of each  $\beta$ -pleated sheet interdigitate with similar residues on the adjacent sheet. Silk fibers (fibroin embedded in an amorphous matrix) are flexible because the pleated sheets are loosely bonded to each other (primarily with weak van der Waals forces) and slide over each other easily.

**FIGURE 5.35**  
Collagen Fibrils

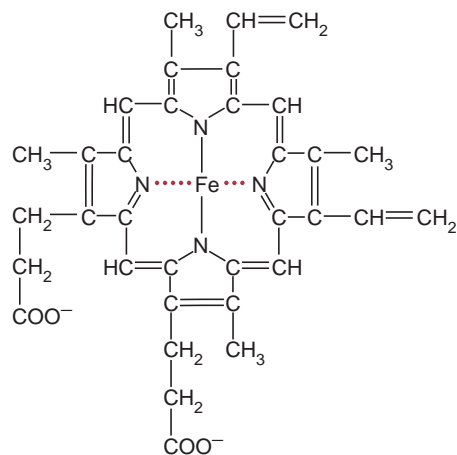
The bands are formed by staggered collagen molecules. Cross-striations are about 680 Å apart. Each collagen molecule is about 3000 Å long.



required for *fibrilogenesis*, the assembly of collagen fibers in their extracellular locations, such as tendons and bone.

The enzyme lysyl oxidase converts some of the lysine and hydroxylysine side groups to aldehydes through oxidative deamination, and this facilitates the spontaneous nonenzymatic formation of strengthening aldimine, and aldol cross-links. (An aldol cross-link is formed in a reaction, called an **aldol condensation**, in



**FIGURE 5.36****Heme**

Heme consists of a porphyrin ring (composed of four pyrroles) with Fe<sup>2+</sup> in the center.

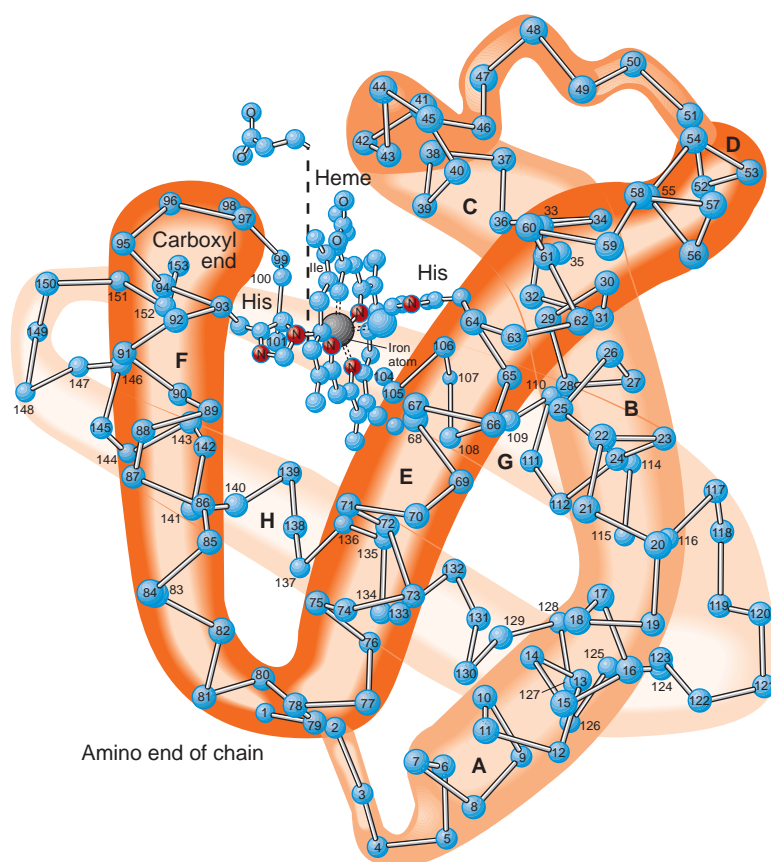
proteins. Each protein possesses one or more unique cavities or clefts whose structure is complementary to a specific ligand. After ligand binding, a conformational change occurs in the protein that is linked to a biochemical event. For example, the binding of ATP to myosin in muscle cells is a critical event in muscle contraction.

The oxygen-binding proteins myoglobin and hemoglobin are interesting and well-researched examples of globular proteins. They are both members of the hemoproteins, a specialized group of proteins that contain the prosthetic group heme. Although the heme group (Figure 5.36) in both proteins is responsible for the reversible binding of molecular oxygen, the physiological roles of myoglobin and hemoglobin are significantly different. The chemical properties of heme are dependent on the Fe<sup>2+</sup> ion in the center of the prosthetic group. Fe<sup>2+</sup>, which forms six coordinate bonds, is bound to the four nitrogens in the center of the protoporphyrin ring. Two other coordinate bonds are available, one on each side of the planar heme structure. In myoglobin and hemoglobin, the fifth coordination bond is to the nitrogen atom in a histidine residue, and the sixth coordination bond is available for binding oxygen. In addition to serving as a reservoir for oxygen within muscle cells, myoglobin facilitates the diffusion of oxygen in metabolically active cells. The role of hemoglobin, the primary protein of red blood cells, is to deliver oxygen to cells throughout the body. A comparison of the structures of these two proteins illustrates several important principles of protein structure, function, and regulation.

**MYOGLOBIN** Myoglobin, found in high concentration in skeletal and cardiac muscle, gives these tissues their characteristic red color. The muscles of diving mammals such as whales, which remain submerged for long periods, have high myoglobin concentrations. Because of the extremely high concentrations of myoglobin, such muscles are typically brown. The protein component of myoglobin, called globin, is a single polypeptide chain that contains eight segments of  $\alpha$ -helix (Figure 5.37). The folded globin chain forms a crevice that almost completely encloses a heme group. Free heme [Fe<sup>2+</sup>] has a high affinity for O<sub>2</sub> and is irreversibly oxidized to form hematin [Fe<sup>3+</sup>]. Hematin cannot bind O<sub>2</sub>. Noncovalent interactions between amino acid side chains and the nonpolar porphyrin ring within the oxygen-binding crevice decrease heme's affinity for O<sub>2</sub>. The decreased affinity protects Fe<sup>2+</sup> from oxidation and allows for the reversible binding of O<sub>2</sub>. All of the heme-interacting amino acids are nonpolar except for two histidines, one of which (the proximal histidine) binds directly to the heme iron atom (Figure 5.38). The other (the distal histidine) stabilizes the oxygen-binding site.

**HEMOGLOBIN** Hemoglobin is a roughly spherical molecule found in red blood cells, where its primary function is to transport oxygen from the lungs to every tissue in the body. Recall that HbA is composed of two  $\alpha$ -chains and two  $\beta$ -chains (Figure 5.39). The HbA molecule is commonly designated  $\alpha_2\beta_2$ . [There is another type of adult hemoglobin, however: approximately 2% of human hemoglobin is HbA<sub>2</sub>, which contains  $\delta$  (delta)-chains instead of  $\beta$ -chains.] Before birth, several additional hemoglobin polypeptides are synthesized. The  $\epsilon$  (epsilon)-chain, which appears in early embryonic life, and the  $\gamma$ -chain, found in the fetus, closely resemble the  $\beta$ -chain. Because both  $\alpha_2\epsilon_2$  and  $\alpha_2\gamma_2$  hemoglobins have a greater affinity for oxygen than does  $\alpha_2\beta_2$  (HbA), the fetus can preferentially absorb oxygen from the maternal bloodstream.

Although the three-dimensional configurations of myoglobin and the  $\alpha$ - and  $\beta$ -chains of hemoglobin are very similar, their amino acid sequences have many differences. Comparison of these molecules from dozens of species has revealed nine invariant amino acid residues. Several invariant residues directly affect the oxygen-binding site, whereas others stabilize the  $\alpha$ -helical peptide segments. The remaining residues may vary considerably. However, most



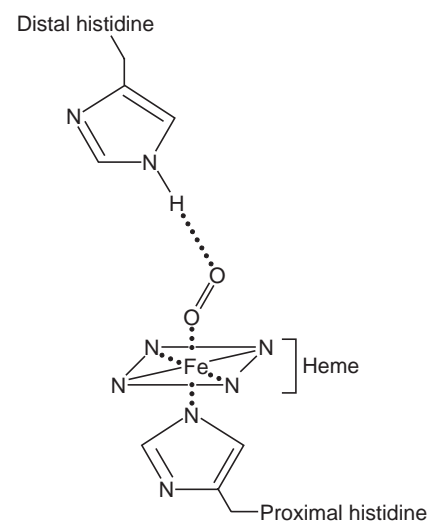
**FIGURE 5.37**  
**Myoglobin**

With the exception of the side chain groups of two histidine residues, only the  $\alpha$ -carbon atoms of the globin polypeptide are shown. Myoglobin's eight helices are designated A through H. The heme group has an iron atom that binds reversibly with oxygen. To improve clarity one of heme's propionic acid side chains has been displaced.

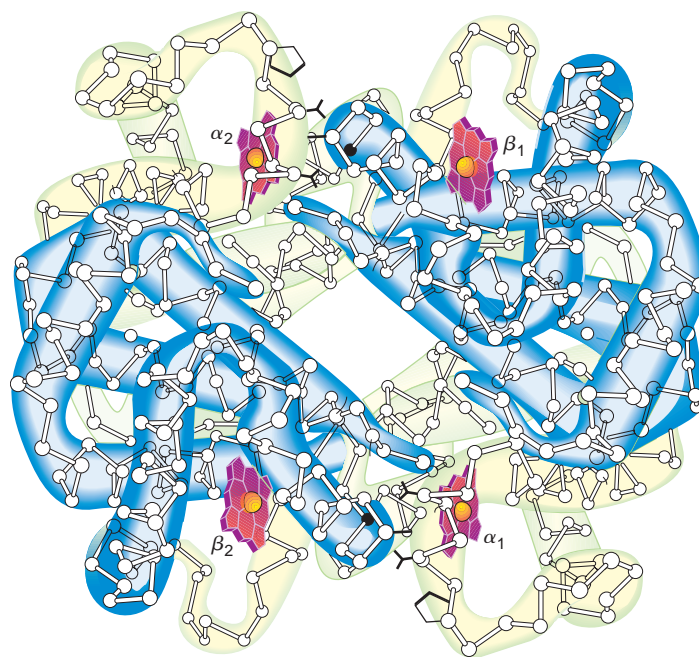
substitutions are conservative. For example, each polypeptide's interior remains nonpolar.

The four chains of hemoglobin are arranged in two identical dimers, designated as  $\alpha_1\beta_1$ , and  $\alpha_2\beta_2$ . Each globin polypeptide has a heme-binding unit similar to that described for myoglobin. Although both myoglobin and hemoglobin bind oxygen reversibly, the latter molecule has a complex structure and more complicated binding properties. The numerous noncovalent interactions (mostly hydrophobic) between the subunits in each  $\alpha\beta$ -dimer remain largely unchanged when hemoglobin interconverts between its oxygenated and deoxygenated forms. In contrast, the relatively small number of interactions between the two dimers change substantially during this transition. When hemoglobin is oxygenated, the salt bridges and certain hydrogen bonds are ruptured as the  $\alpha_1\beta_1$  and  $\alpha_2\beta_2$  dimers slide by each other and rotate  $15^\circ$  (Figure 5.40). The deoxygenated conformation of hemoglobin (deoxyHb) is often referred to as the T (taut) state and oxygenated hemoglobin (oxyHb) is said to be in the R (relaxed) state. The oxygen-induced readjustments in the interdimer contacts are almost simultaneous. In other words, a conformational change in one subunit is rapidly propagated to the other subunits. Consequently, hemoglobin alternates between two stable conformations, the T and R states.

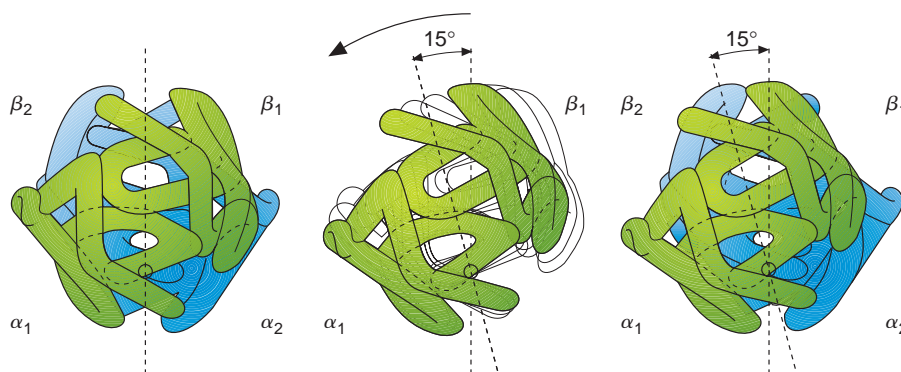
Because of subunit interactions, the oxygen dissociation curve of hemoglobin has a sigmoidal shape (Figure 5.41). As the first  $O_2$  binds to hemoglobin, the binding of additional  $O_2$  to the same molecule is enhanced. This binding pattern, called **cooperative binding**, results from changes in hemoglobin's three-dimensional structure that are initiated when the first  $O_2$  binds. The binding of the first  $O_2$  facilitates the binding of the remaining three  $O_2$  molecules to the tetrameric hemoglobin molecules. In the lungs, where  $O_2$  tension is high, hemoglobin is quickly saturated (converted to the R state). In tissues depleted of  $O_2$ , hemoglobin gives up about half its oxygen. In contrast to hemoglobin, myoglobin's oxygen dissociation curve is hyperbolic. This simpler binding pattern,



**FIGURE 5.38**  
**The Oxygen-Binding Site of Heme Created by a Folded Globin Chain**

**FIGURE 5.39****Hemoglobin**

The protein contains four subunits, designated  $\alpha$  and  $\beta$ . Each subunit contains a heme group that binds reversibly with oxygen.

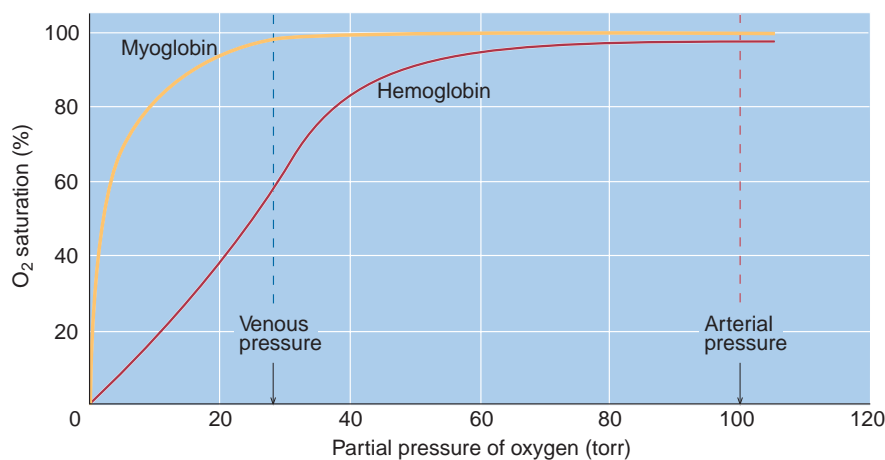
**FIGURE 5.40****The Hemoglobin Allosteric Transition**

When hemoglobin is oxygenated, the  $\alpha_1\beta_1$  and  $\alpha_2\beta_2$  dimers slide by each other and rotate  $15^\circ$ .

**(a) Deoxyhemoglobin****(b) Oxyhemoglobin**

a consequence of myoglobin's simpler structure, reflects several aspects of this protein's role in oxygen storage. Because its dissociation curve is well to the left of the hemoglobin curve, myoglobin gives up oxygen only when the muscle cell's oxygen concentration is very low (i.e., during strenuous exercise). In addition, because myoglobin has a greater affinity for oxygen than does hemoglobin, oxygen moves from blood to muscle.

The binding of ligands other than oxygen affects hemoglobin's oxygen-binding properties. For example, the dissociation of oxygen from hemoglobin is enhanced if pH decreases. By this mechanism, called the *Bohr effect*, oxygen is delivered to cells in proportion to their needs. Metabolically active cells, which require large amounts of oxygen for energy generation, also produce large amounts of the waste product  $\text{CO}_2$ . As  $\text{CO}_2$  diffuses into blood, it reacts with water to form  $\text{HCO}_3^-$  and  $\text{H}^+$ . (The bicarbonate buffer was discussed on p. 96.) The subsequent binding of  $\text{H}^+$  to several ionizable groups on hemoglobin molecules enhances the dissociation of  $\text{O}_2$  by converting hemoglobin to its T state. (Hydrogen ions bind preferentially to deoxyHb. Any increase in  $\text{H}^+$  concentration stabilizes the deoxy conformation of the protein and therefore shifts the equilibrium distribution between the T and R states.) When a small number of  $\text{CO}_2$  molecules bind to terminal amino groups on hemoglobin (forming carbamate or  $-\text{NHCOO}^-$  groups) the deoxy form (T state) of the protein is additionally stabilized.



**FIGURE 5.41**  
Equilibrium Curves Measure the Affinity of Hemoglobin and Myoglobin for Oxygen

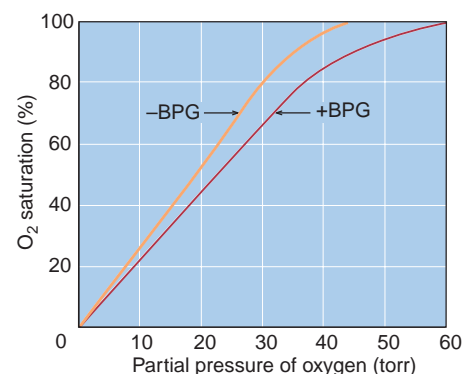
2,3-Bisphosphoglycerate (BPG) (also called glycerate-2,3-bisphosphate) is also an important regulator of hemoglobin function. Although most cells contain only trace amounts of BPG, red blood cells contain a considerable amount. BPG is derived from glycerate-1,3-bisphosphate, an intermediate in the breakdown of the energy-rich compound glucose. In the absence of BPG, hemoglobin has a very high affinity for oxygen (Figure 5.42). As with  $H^+$  and  $CO_2$ , binding BPG stabilizes deoxyHb. A negatively charged BPG molecule binds in a central cavity within hemoglobin that is lined with positively charged amino acids.

In the lungs the process is reversed. A high oxygen concentration drives the conversion from the deoxyHb configuration to that of oxyHb. The change in the protein's three-dimensional structure initiated by the binding of the first oxygen molecule releases bound  $CO_2$ ,  $H^+$ , and BPG. The  $H^+$  recombines with  $HCO_3^-$  to form carbonic acid, which then dissociates to form  $CO_2$  and  $H_2O$ . Afterward,  $CO_2$  diffuses from the blood into the alveoli.

### KEY CONCEPTS



- Globular protein function usually involves binding to small ligands or to other macromolecules.
- The oxygen-binding properties of myoglobin and hemoglobin are determined in part by the number of subunits they contain.



**FIGURE 5.42**  
The Effect of 2,3-Bisphosphoglycerate (BPG) on the Affinity Between Oxygen and Hemoglobin

In the absence of BPG ( $-BPG$ ), hemoglobin has a high affinity for  $O_2$ ; where BPG is present and binds to hemoglobin ( $+BPG$ ), its affinity for  $O_2$  decreases.

### QUESTION 5.11

Fetal hemoglobin (HbF) binds to BPG to a lesser extent than does HbA. Why do you think HbF has a greater affinity for oxygen than does maternal hemoglobin?

Myoglobin stores  $O_2$  in muscle tissue to be used by the mitochondria only when the cell is in oxygen debt, while hemoglobin can effectively transport  $O_2$  from the lungs and deliver it discriminately to cells in need of  $O_2$ . Describe the structural features that allow these two proteins to accomplish separate functions.

# BIOCHEMISTRY IN THE LAB

## Protein Technology

Living organisms produce a stunning variety of proteins. Consequently, it is not surprising that considerable time, effort, and funding have been devoted to investigating their properties. Since the amino acid sequence of bovine insulin was determined by Frederick Sanger in 1953, the structures of several thousand proteins have been elucidated.

In contrast to the 10 years required for insulin, current technologies allow protein sequence determination within a few days. In addition to the Edman degradation method and mass spectrometry, the amino acid sequence of a protein can be generated from its DNA or mRNA sequence if this information is available. After a brief review of protein purification methods, the Edman degradation method and mass spectrometry are described. Note that all the techniques for isolating, purifying, and characterizing proteins exploit differences in charge, molecular weight, and binding affinities. Many of these technologies apply to the investigation of other biomolecules.

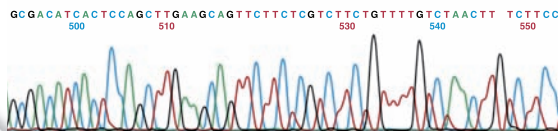
### Purification

Protein analysis begins with isolation and purification. Extraction of a protein requires cell disruption and homogenization (see Biochemistry in the Lab, Cell Technology, Chapter 2). This process is often followed by differential centrifugation and, if the protein is a component of an organelle, by density gradient centrifugation. After the protein-containing fraction has been obtained, several relatively crude methods may be used to enhance purification. **Salting out** is a technique in which high concentrations of salts such as ammonium sulfate  $[(\text{NH}_4)_2\text{SO}_4]$  are used to precipitate proteins. Because each protein has a characteristic salting-out point, this technique removes many impurities. (Unwanted proteins that remain in solution are discarded when the liquid is decanted.) When proteins are tightly bound to membrane, organic solvents or detergents often aid in their extraction. Dialysis is routinely used to remove low-molecular-weight impurities such as salts, solvents, and detergents.

As a protein sample becomes progressively more pure, more sophisticated methods are used to achieve further purification. Among the most commonly used techniques are chromatography and electrophoresis.

### Chromatography

Originally devised to separate low-molecular-weight substances such as sugars and amino acids, chromatography has become an invaluable tool in protein purification. A wide variety of chromatographic techniques are used to separate protein mixtures on the basis of molecular properties such as size, shape, and weight, or certain binding affinities. Often several techniques must be used sequentially to obtain a demonstrably pure protein.



In all chromatographic methods the protein mixture is dissolved in a liquid known as the **mobile phase**. As the protein molecules pass across the **stationary phase** (a solid matrix), they separate from each other because they are differently distributed between the two phases. The relative movement of each molecule results from its capacity to remain associated with the stationary phase while the mobile phase continues to flow.

Three chromatographic methods commonly used in protein purification are gel-filtration chromatography, ion-exchange chromatography, and affinity chromatography. **Gel-filtration chromatography** (Figure 5D) is a form of size-exclusion chromatography in which particles in an aqueous solution flow through a column (a hollow tube) filled with gel and are separated according to size. Molecules that are larger than the gel pores are excluded and therefore move through the column quickly. Molecules that are smaller than the gel pores diffuse in and out of the pores, so their movement through the column is retarded. Differences in the rates of particle movement separate the protein mixture into bands, which are then collected separately.

**Ion-exchange chromatography** separates proteins on the basis of their charge. Anion-exchange resins, which consist of positively charged materials, bind reversibly with a protein's negatively charged groups. Similarly, cation-exchange resins bind positively charged groups. After proteins that do not bind to the resin have been removed, the protein of interest is recovered by an appropriate change in the solvent pH and/or salt concentration. (A change in pH alters the protein's net charge.)

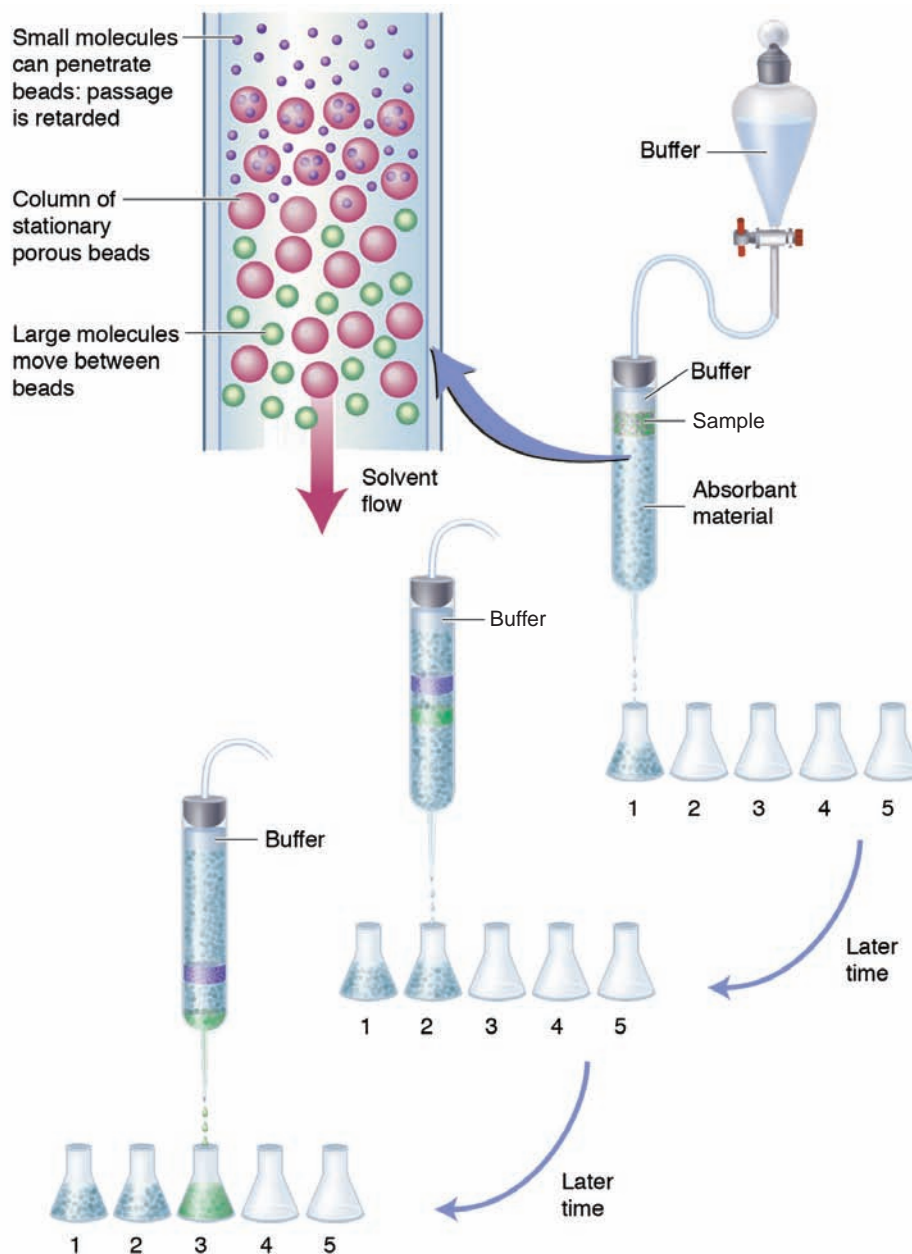
**Affinity chromatography** takes advantage of the unique biological properties of proteins. That is, it uses a special noncovalent binding affinity between the protein and a special molecule (the ligand). The ligand is covalently bound to an insoluble matrix, which is placed in a column. After nonbinding protein molecules have passed through the column, the protein of interest is removed by altering the conditions that affect binding (i.e., pH or salt concentration).

### Electrophoresis

Because proteins are electrically charged, they move in an electric field. In this process, called **electrophoresis**, molecules separate from each other because of differences in their net charge. For example, molecules with a positive net charge migrate toward the negatively charged electrode (cathode). Molecules with a net negative charge will move toward the positively charged electrode (anode). Molecules with no net charge will not move at all.

Electrophoresis, one of the most widely used techniques in biochemistry, is nearly always carried out by using gels such as polyacrylamide or agarose. The gel, functioning much as it does in gel-filtration chromatography, also acts to separate proteins on the basis of their molecular weight and shape. Consequently, gel



**BIOCHEMISTRY** IN THE LAB cont


**FIGURE 5D**  
**Gel-Filtration Chromatography**

In gel-filtration chromatography the stationary phase is a gelatinous polymer with pore sizes selected by the experimenter to separate molecules according to their sizes. The sample is applied to the top of the column and is eluted with buffer (the mobile phase). As elution proceeds, larger molecules travel faster through the gel than smaller molecules, whose progress is slowed because they can enter the pores. If fractions are collected, the larger molecules appear in the earlier fractions and later fractions contain smaller molecules.



## BIOCHEMISTRY IN THE LAB cont

electrophoresis is highly effective at separating complex mixtures of proteins or other molecules.

Bands resulting from a gel electrophoretic separation may be treated in several ways. Specific bands may be excised from the gel after visualization with ultraviolet light. Each protein-containing slice is then eluted with buffer and prepared for further analysis. Because of its high resolving power, gel electrophoresis is also used to assess the purity of protein samples. Staining gels with a dye such as Coomassie Brilliant Blue is a common method for quickly assessing the success of a purification step.

SDS–polyacrylamide gel electrophoresis (SDS-PAGE) is a widely used variation of electrophoresis that can be used to determine molecular weight (Figure 5E). SDS, a negatively charged detergent, binds to the hydrophobic regions of protein molecules, causing the proteins to denature and assume rodlike shapes. Because most molecules bind SDS in a ratio roughly proportional to their molecular weights, during electrophoresis SDS-treated proteins migrate toward the anode (+ pole) only in relation to their molecular weight.

### Protein Sequence Analysis

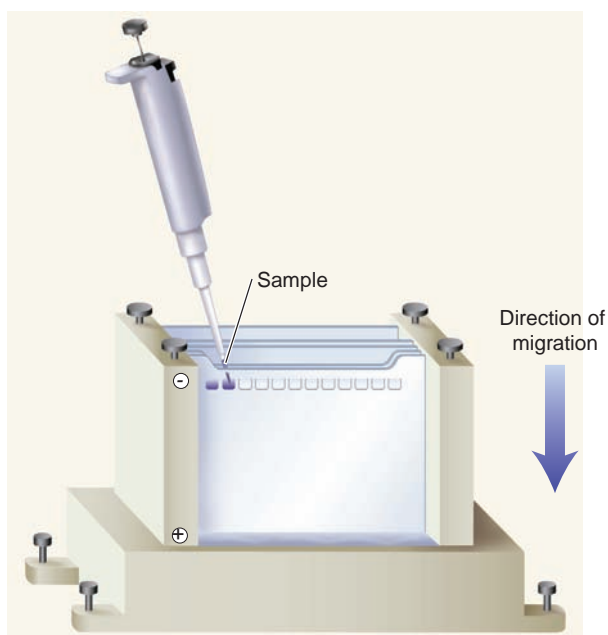
The first step in protein sequence analysis is to determine how many of each type of amino acid residue are present in the molecule. The process for obtaining this information, referred to as the *amino acid composition*, begins with the complete hydrolysis of all peptide bonds. Hydrolysis is typically accomplished

with 6 N HCl for 10 to 100 hours. Long reaction times are required because of difficulties in the hydrolysis of three aliphatic amino acids (Leu, Ile, and Val). Hydrolysis is followed by analysis of the resulting amino acid mixture, referred to as the *hydrolysate*. Because of the vigorous conditions of acid hydrolysis, several amino acids (Trp, Gln, Ser, Thr, Tyr, and cystine) are degraded. The concentrations of these molecules in the protein are determined by alternate means.

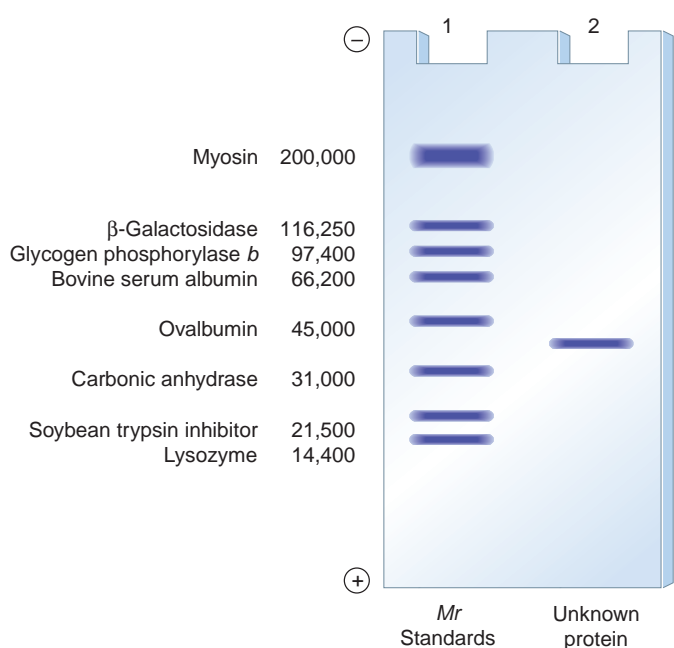
Currently, most protein hydrolysates are analyzed by automated high-pressure liquid chromatography (HPLC). In HPLC, after the hydrolysate has been treated with compounds such as Edman's reagent (described shortly), the products are forced at high pressure through a stainless steel column packed with a stationary phase. Each amino acid derivative is identified according to its retention time on the column. Amino acid analysis by HPLC takes about 1 hour.

Determining a protein's primary structure is similar to solving a complex puzzle. Several steps are involved in solving the amino acid sequence of any protein.

- 1. Cleavage of all disulfide bonds.** Oxidation with performic acid is commonly used.
- 2. Determination of the N-terminal and C-terminal amino acids.** Several methods are available to determine the N-terminal amino acid. In Sanger's method, the polypeptide chain is reacted with 1-fluoro-2,4-dinitrobenzene. The



(a)



(b)

**FIGURE 5E**  
Gel Electrophoresis

(a) Gel apparatus. The samples are loaded into wells. After an electric field is applied, the proteins move into the gel. (b) Molecules separate and move in the gel as a function of molecular weight and shape.



## BIOCHEMISTRY IN THE LAB cont

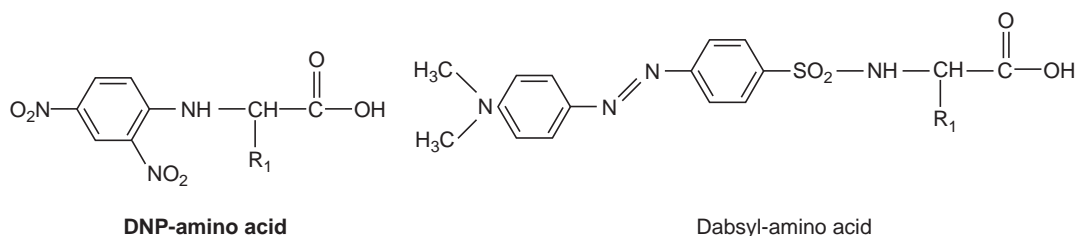
dinitrophenyl (DNP) derivative of the N-terminal amino acid can then be isolated. Alternatively, dabsyl chloride is used to determine N-terminal amino acid residues.

Dabsylation is a very sensitive technique because the dabsyl group, a fluorescent marker, is easily detected in small amounts. The N-terminal derivatives of both of these reagents can then be identified by HPLC analysis. A group of enzymes called the carboxypeptidases are used to identify the C-terminal residue. Because these enzymes sequentially cleave peptide bonds starting at the C-terminal residue, the first amino acid liberated is the C-terminal residue.

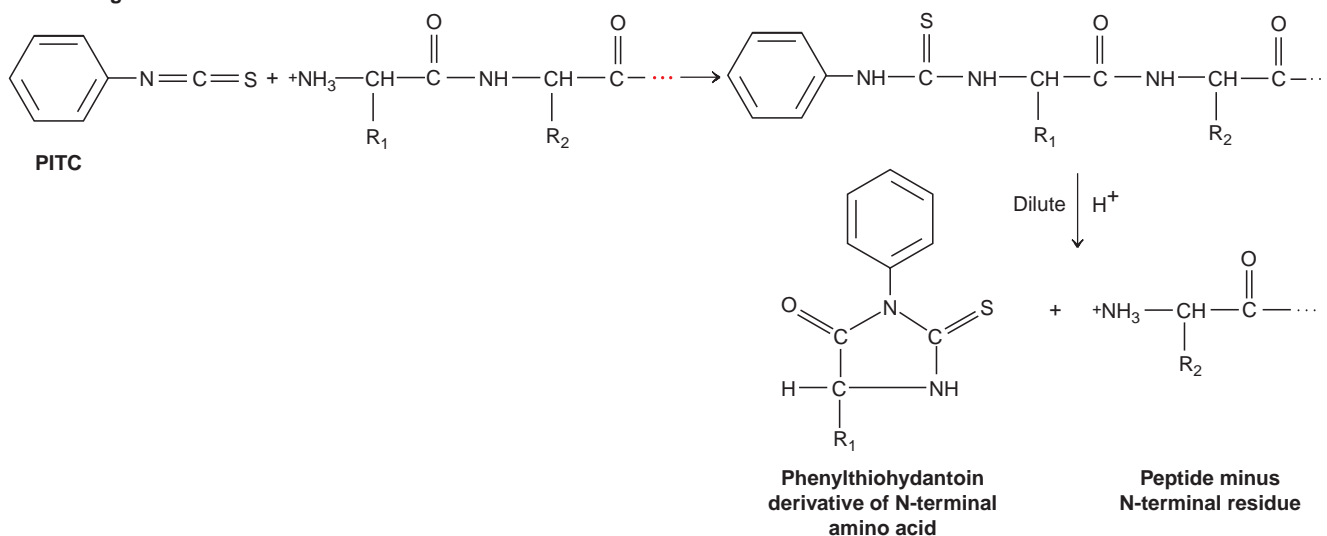
- 3. Cleavage of the polypeptide into fragments.** The polypeptide is broken into smaller peptides because technical problems prevent the direct sequencing of long polypeptides. The use of several reagents, each of which cuts the chain at a different site, creates overlapping sets of fragments. After the amino acid sequence of each fragment has been determined, the investigator uses this information to work out the entire sequence of the polypeptide. Of all the enzymes commonly used, the pancreatic enzyme trypsin is the most reliable. It cleaves peptide bonds on the carboxy side of either lysine or arginine residues. The peptide fragments, referred to as *tryptic*

*peptides*, have lysine or arginine carboxy terminal residues. Chymotrypsin, another pancreatic enzyme, is also often used. It breaks peptide bonds on the carboxyl side of phenylalanine, tyrosine, leucine, methionine, or tryptophan. Treating the polypeptide with the reagent cyanogen bromide also generates peptide fragments. Cyanogen bromide specifically cleaves peptide bonds on the carboxyl side of methionine residues.

- 4. Determination of the sequences of the peptide fragments.** Each fragment is sequenced through repeated cycles of a procedure called the *Edman degradation*. In this method phenylisothiocyanate (PITC), often referred to as Edman's reagent, reacts with the N-terminal residue of each fragment. Treatment of the product of this reaction with acid cleaves the N-terminal residue as a phenylthiohydantoin derivative. The derivative is then identified by comparing it with known standards, using electrophoresis or various chromatographic methods (most commonly HPLC). Because of the large number of steps involved in sequencing peptide fragments, Edman degradation is usually carried out by using a computer-programmed machine called a sequenator.
- 5. Ordering the peptide fragments.** The amino acid sequence information derived from two or more sets of



### Edman degradation



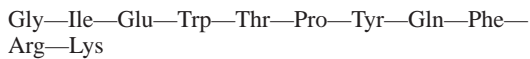
## BIOCHEMISTRY IN THE LAB cont

polypeptide fragments is next examined for overlapping segments. Such segments make it possible to piece together the overall sequence.

A typical primary sequence determination problem is given, along with its solution.

### Problem

Consider the following peptide:



What amino acids and peptides are produced when this peptide is treated with each of the following reagents?

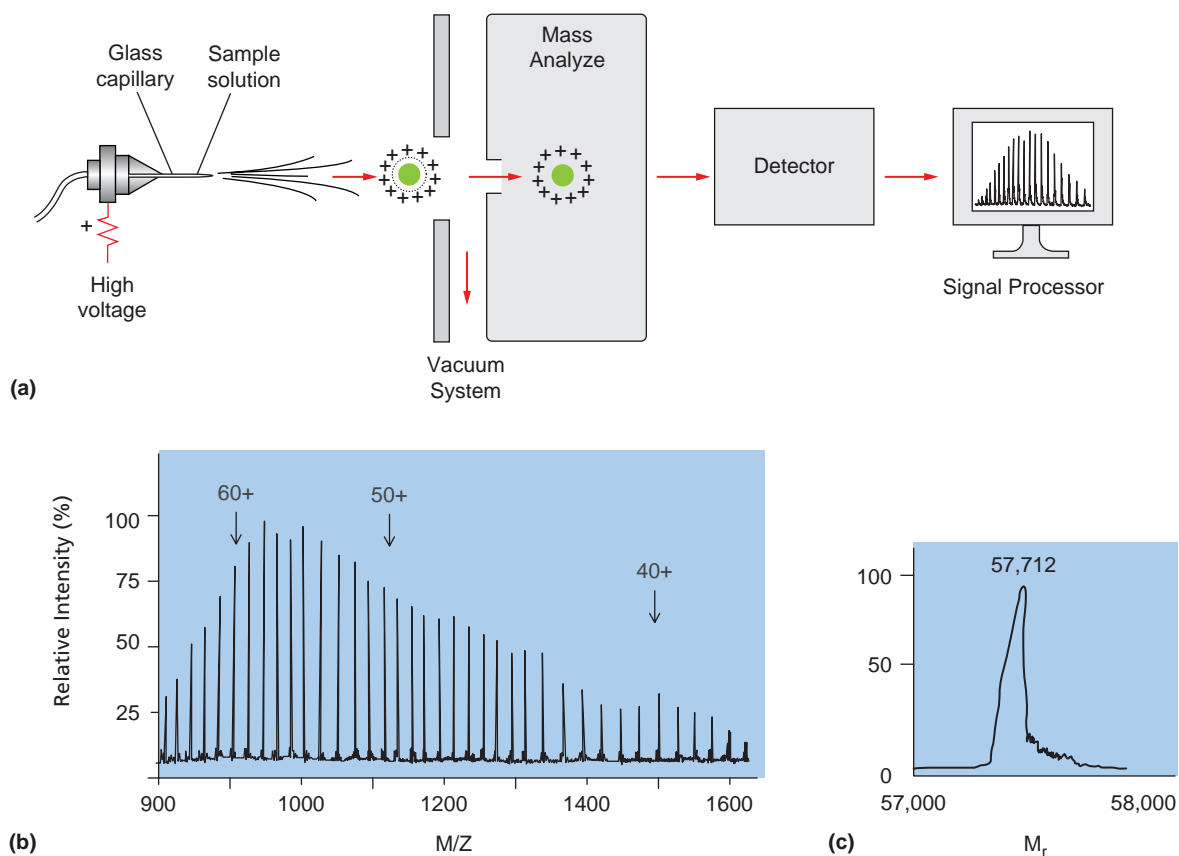
- a. Carboxypeptidase    b. Chymotrypsin

c. Trypsin

d. DNFB (dinitrofluorobenzene)

### Solution

- a. Because carboxypeptidase cleaves at the carboxyl end of peptides, the products are  
Gly—Ile—Glu—Trp—Thr—Pro—Tyr—Gln—Phe—Arg  
and Lys
- b. Because chymotrypsin cleaves peptide bonds in which aromatic amino acids (i.e., Phe, Tyr, and Trp) contribute a carboxyl group, the products are  
Gly—Ile—Glu—Trp, Thr—Pro—Tyr,  
Gln—Phe, and Arg—Lys



**FIGURE 5F**

### Mass Spectrometry

(a) The principal steps in electrospray ionization. The sample (a protein dissolved in a solvent) is injected via a glass capillary into the ionization chamber. The voltage difference between the electrospray needle and the injection port results in the creation of protein ions. The solvent evaporates during this phase. The ions enter the mass spectrometer, which then measures their  $m/z$  ratios. (b) An electrospray mass spectrum showing the  $m/z$  ratios for several peaks. (c) A computer analysis of the data showing the molecular mass of the sample protein ( $M_r$  = molecular weight).



## BIOCHEMISTRY IN THE LAB cont

- c. Trypsin cleaves at the carboxyl end of lysine and arginine. The products are

Gly—Ile—Glu—Trp—Thr—Pro—Tyr—Gln—Phe—Arg  
and Lys

- d. DNFB tags the amino-terminal amino acid. The product is

DNP—Gly—Ile—Glu—Trp—Thr—Pro—  
Tyr—Gln—Phe—Arg—Lys

Hydrolysis then cleaves all the peptide bonds, and DNP—Gly can be identified by a chromatographic method.

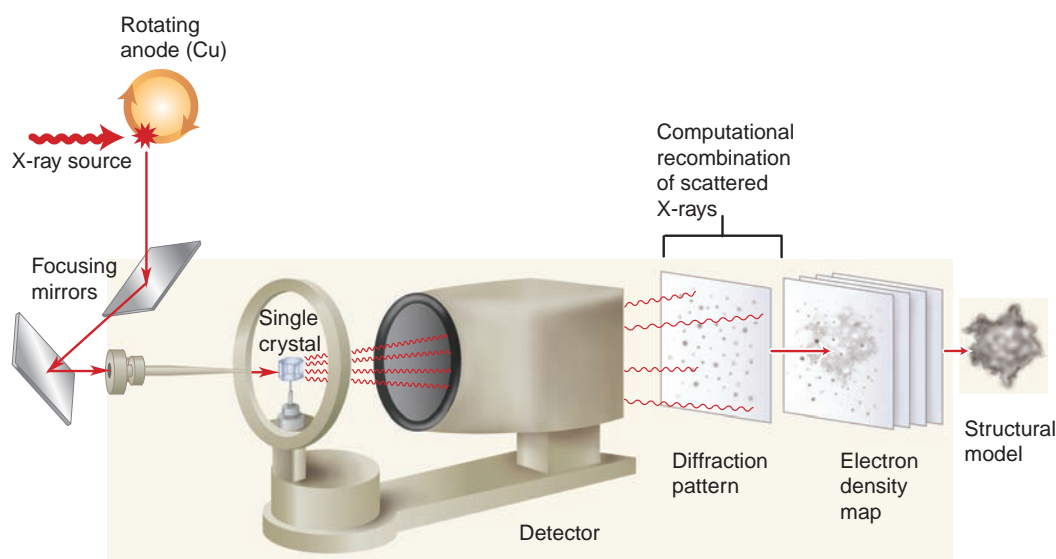
Automation of the Edman degradation method has increased the speed and accuracy of the sequencing process. Computer-assisted devices called sequencers can determine the amino acid sequence of vanishingly small samples in a fraction of the time needed for the manual method.

### Mass Spectrometry

In recent years biochemists have moved away from the Edman degradation method for sequencing proteins. Instead, many use **mass spectrometry** (MS), a powerful and sensitive technique for separating, identifying, and determining the mass of molecules by exploiting differences in their mass-to-charge ( $m/z$ ) ratios. In a mass spectrometer, ionized molecules flow through a magnetic field (Figure 5F). The magnetic field force deflects the ions depending on their  $m/z$  ratios with lighter ions being more deflected from a straight-line path than heavier ions. A detector measures the deflection of each ion. In addition to protein identity and mass determinations, MS is also used to detect bound cofactors and protein modifications. Because MS analysis

involves the ionization and vaporization of the substances to be investigated, its use in the analysis of thermally unstable macromolecules such as proteins and nucleic acids did not become feasible until methods such as electrospray ionization and matrix-assisted laser desorption ionization (MALDI) had been developed. In electrospray ionization a solution containing the protein of interest is sprayed in the presence of a strong electrical field into a port in the spectrometer. As the protein droplets exit the injection device, typically an ultrafine glass tube, the protein molecules become charged. In MALDI, a laser pulse vaporizes the protein, which is embedded in a solid matrix. Once the sample has been ionized, its molecules, now in the gas phase, are separated according to their individual  $m/z$  ratios. A detector within the mass spectrometer produces a peak for each ion. In a computer-assisted process, information concerning each ion's mass is compared against data for ions of known structure and used to determine the sample's molecular identity.

Protein sequencing analysis makes use of tandem MS (two mass spectrometers linked in series, MS/MS). A protein of interest, often extracted from a band in a gel, is then digested by a proteolytic enzyme. Subsequently, the enzyme digest is injected into the first mass spectrometer, which separates the oligopeptides according to their  $m/z$  ratios. One by one, each oligopeptide ion is directed into a collision chamber, where it is fragmented by collisions with hot inert gas molecules. Product ions, peptides that differ from each other in size by one amino acid residue, are then sequentially directed into the second mass spectrometer. A computer identifies each peak and automatically determines the amino acid sequence of the peptides. The process is then repeated



**FIGURE 5G**  
**Schematic Diagram of X-Ray Crystallography**

X-rays are useful in the analysis of biomolecules because their wavelength range is quite similar to the magnitude of chemical bonds. Consequently, the resolving power of X-ray crystallography is equivalent to interatomic distances.



## BIOCHEMISTRY IN THE LAB cont

with oligopeptides derived from digestion with another enzyme. The computer uses the sequence information derived from both digests to determine the amino acid sequence of the original polypeptide.

### X-Ray Crystallography

Much of the three-dimensional structural information about proteins was obtained by X-ray crystallography. Because the bond distances in proteins are approximately 0.15 nm, the electromagnetic radiation used to resolve protein structure must have a short wavelength. Visible light wavelengths [ $\lambda$  =

400–700 nm] clearly does not have sufficient resolving power for biomolecules. X-rays, however, have very short wavelengths (0.07–0.25 nm).

In X-ray crystallography, highly ordered crystalline specimens are exposed to an X-ray beam (Figure 5G). As the X-rays hit the crystal, they are scattered by the atoms in the crystal. The diffraction pattern that results is recorded on charge-coupled device (CCD) detectors. The diffraction patterns are used to construct an electron density map. Because there is no objective lens to recombine the scattered X-rays, the three-dimensional image is reconstructed mathematically. Computer programs now perform these extremely complex and laborious computations. ■

## Chapter Summary

1. Polypeptides are amino acid polymers. Proteins may consist of one or more polypeptide chains.
2. Each amino acid contains a central carbon atom (the  $\alpha$ -carbon) to which an amino group, a carboxylate group, a hydrogen atom, and an R group are attached. In addition to comprising protein, amino acids have several other biological roles. According to their capacity to interact with water, amino acids may be separated into four classes: nonpolar, polar, acidic, and basic.
3. Titration of amino acids and peptides illustrates the effect of pH on their structures. The pH at which a molecule has no net charge is called its isoelectric point.
4. Amino acids undergo several chemical reactions. Two reactions are especially important: peptide bond formation and cysteine oxidation.
5. Proteins have a vast array of functions in living organisms. In addition to serving as structural materials, proteins are involved in metabolic regulation, transport, defense, and catalysis. Some proteins are multifunctional; that is, they have two or more seemingly unrelated functions. Proteins can also be classified into families and superfamilies, according to their sequence similarities as well as their shapes and composition. Fibrous proteins (e.g., collagen) are long, rod-shaped molecules that are insoluble in water and physically tough. Globular proteins (e.g., hemoglobin) are compact, spherical molecules that are usually soluble in water.
6. Biochemists have distinguished four levels of protein structure. Primary structure, the amino acid sequence, is specified by genetic information. As the polypeptide chain folds, local folding patterns constitute the protein's secondary structure. The overall three-dimensional shape that a polypeptide assumes is called the tertiary structure. Proteins that consist of two or more polypeptides have quaternary structure. The functions of numerous proteins, especially molecules that participate in eukaryotic regulatory processes, are partially or completely unstructured. Many physical and chemical conditions disrupt protein structure. Denaturing agents include strong acids or bases, reducing agents, organic solvents, detergents, high salt concentrations, heavy metals, temperature changes, and mechanical stress.
7. One of the most important aspects of protein synthesis is the folding of polypeptides into their biologically active conformations. Despite decades of investigation into the physical and chemical properties of polypeptide chains, the mechanism by which a primary sequence dictates the molecule's final conformation is unresolved. Many proteins require molecular chaperones to fold into their final three-dimensional conformations. Protein misfolding is now known to be an important feature of several human diseases, including Alzheimer's disease and Huntington's disease.
8. Fibrous proteins (e.g.,  $\alpha$ -keratin and collagen), which contain high proportions of  $\alpha$ -helices or  $\beta$ -pleated sheets, have structural rather than dynamic roles. Despite their varied functions, most globular proteins have features that allow them to bind to specific ligands or sites on certain macromolecules. These binding events involve conformational changes in the globular protein's structure.
9. The biological activity of complex multisubunit proteins is often regulated by allosteric interactions in which small ligands bind to the protein. Any change in the protein's activity is due to changes in the interactions among the protein's subunits. Effectors can increase or decrease the function of a protein.



Take your learning further by visiting the **companion website** for Biochemistry at [www.oup.com/us/mckee](http://www.oup.com/us/mckee) where you can complete a multiple-choice quiz on amino acids, peptides, and proteins to help you prepare for exams.

## Suggested Readings

- Branden, C., and Tooze, J., *Introduction to Protein Structure*, 2nd ed., Garland, New York, 1999.
- Bustamonte, C., Of Torques, Forces, and Protein Machines, *Protein Sci.* 13:3061–65, 2004.
- Chothia, C., Gough, J., Vogel, C., and Teichmann, S. A., Evolution of the Protein Repertoire, *Science* 300:1701–1703, 2003.
- Dyson, H. J., and Wright, P. E., Intrinsically Unstructured Proteins and Their Functions, *Nat. Rev. Mol. Cell Biol.* 6(3):197–208, 2005.
- Fink, A. L., Natively Unfolded Proteins, *Curr. Opin. Struct. Biol.* 15:35–41, 2005.
- Lesk, A. M., *Introduction to Protein Science: Architecture, Function, and Genomics*, Oxford University Press, Oxford, 2004.
- Lindorff, K., Rogen, P., Poci, E., Vendruscolo, M., and Dobson, M., Protein Folding and the Organization of the Protein Topology Universe, *Trends, Biochem. Sci.* 30(1):13–19, 2005.
- Mattos, C., Protein-Water Interactions in a Dynamic World, *Trends Biochem. Sci.* 27(4):203–208, 2002.
- Ponting, C. P., and Russell, R. R., The Natural History of Protein Domains, *Annu. Rev. Biophys. Biomed. Struct.* 31:45–71, 2002.
- Tompa, P., Szasz, C., and Buday, L., Structural Disorder Throws New Light on Moonlighting, *Trends Biochem. Sci.* 30(9):484–489, 2005.

## Key Words

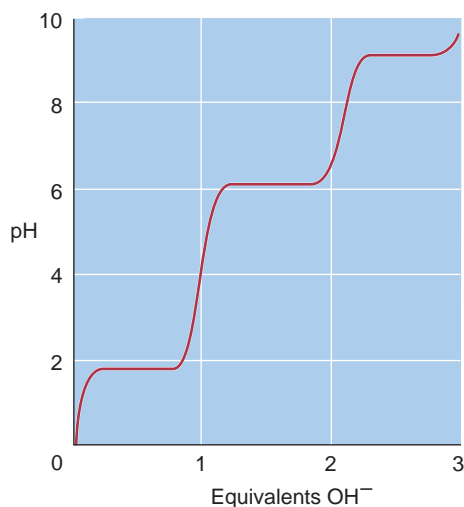
- |                              |   |                                |   |
|------------------------------|---|--------------------------------|---|
| affinity chromatography, 172 | electrophoresis, 172                    | lipoprotein, 142               | protein, 124                                |
| aldimine, 138                | enantiomer, 131                         | metalloprotein, 142            | protein family, 141                         |
| aldol condensation, 166      | fibrous protein, 142                    | mobile phase, 172              | protein folding, 148                        |
| aliphatic hydrocarbon, 128   | fold, 136                               | modular protein, 148           | protein superfamily, 141                    |
| allosteric transition, 153   | gel-filtration chromatography, 172      | modulator, 153                 | protomer, 153                               |
| allostery, 153               | globular protein, 142                   | molecular chaperone, 161       | quaternary structure, 142                   |
| Alzheimer's disease, 164     | glycoprotein, 142                       | molecular disease, 145         | response element, 155                       |
| amino acid residue, 124      | heat shock protein, 141                 | molten globule, 161            | salt bridge, 172                            |
| amphipathic molecule, 156    | hemoprotein, 142                        | motif, 147                     | salting out, 172                            |
| amphoteric molecule, 127     | holoprotein, 142                        | motor protein, 157             | Schiff base, 138                            |
| antigen, 141                 | homologous polypeptide, 142             | multifunction protein, 141     | SDS-polyacrylamide gel electrophoresis, 174 |
| apoprotein, 142              | hormone, 130                            | natively unfolded protein, 155 | secondary structure, 142                    |
| aromatic hydrocarbon, 128    | hsp60, 162                              | neurotransmitter, 130          | site-directed mutagenesis, 159              |
| asymmetric carbon, 131       | hsp70, 162                              | oligomer, 153                  | stationary phase, 172                       |
| chaperonins, 162             | Huntington's disease, 164               | optical isomer, 131            | stereoisomer, 131                           |
| chiral carbon, 131           | intrinsically unstructured protein, 155 | peptide, 124                   | subunit, 150                                |
| circular dichroism, 160      | ion-exchange chromatography, 172        | peptide bond, 136              | supersecondary structure, 147               |
| conjugated protein, 142      | isoelectric point, 133                  | phosphoprotein, 142            | tertiary structure, 142                     |
| cooperative binding, 169     | ligand, 148                             | polypeptide, 124               | zwitterion, 127                             |
| denaturation, 155            |   | primary structure, 142         |   |
| disulfide bridge, 136        |   | prosthetic group, 142          |   |
| effector, 153                |   |                                |   |

## Review Questions

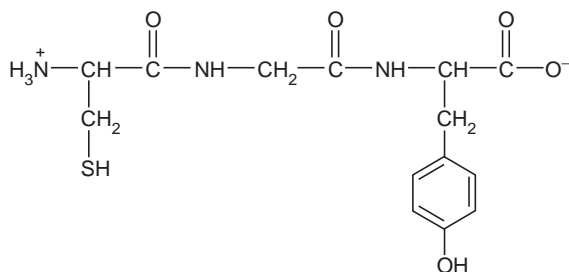
These questions are designed to test your knowledge of the key concepts discussed in this chapter, before moving on to the next chapter. You may like to compare your answers to the solutions provided in the back of the book and in the accompanying Study Guide.

- Distinguish between proteins, peptides, and polypeptides.
- Indicate whether each of the following amino acids is polar, nonpolar, acidic, or basic:
  - glycine
  - tyrosine
  - glutamic acid
  - histidine
  - proline
  - lysine
  - cysteine
  - asparagine
  - valine
  - leucine
- Arginine has the following  $pK_a$  values:  $pK_1 = 2.17$ ,  $pK_2 = 9.04$ ,  $pK_R = 12.48$ . Give the structure and net charge of arginine at the following pH values: 1, 4, 7, 10, 12

4. Shown is the titration curve for histidine.
- What species are present at each plateau?
  - Using the titration curve, determine the  $pK_a$  of each ionization of histidine.
  - What is the isoelectric point of histidine?



5. Consider the following molecule.



- Name it.
  - Use the three-letter symbols for the amino acid to represent this molecule.
6. Rotation about the peptide bond in glycylglycine is hindered. Draw the resonance forms of the peptide bond and explain why.
7. List six functions of proteins in the body.
8. Differentiate the terms in each pair:
- globular and fibrous proteins
  - simple and conjugated proteins
  - apoprotein and holoprotein
9. Define the following terms:
- asymmetric carbon
  - motor protein
  - prosthetic group
  - primary structure
  - molten globule
10. Indicate the level(s) of protein structure to which each of the following contributes:
- amino acid sequence
  - $\beta$ -pleated sheet
  - hydrogen bond
  - disulfide bond
11. What type of secondary structure would the following amino acid sequence be *most* likely to have?
- polyproline
  - polyglycine
  - Ala—Val—Ala—Val—Ala—Val—
  - Gly—Ser—Gly—Ala—Gly—Ala
12. List three factors that do not foster  $\alpha$ -helix formation.
13. Denaturation is the loss of protein function from structural change or chemical reaction. At what level of protein structure or through what chemical reaction does each of the following denaturation agents act?
- heat
  - strong acid
  - saturated salt solution
  - organic solvents (e.g., alcohol or chloroform)
14. A polypeptide has a high pI value. Suggest which amino acids might comprise it.
15. Outline the steps to isolate a typical protein. What is achieved at each step?
16. Outline the steps to purify a protein. What criteria are used to evaluate purity?
17. List the types of chromatography used to purify proteins. Describe how each separation method works.
18. In using carboxypeptidase to sequence a protein, the protein is first broken down into smaller fragments, which are separated from one another. Each fragment is then individually sequenced. If this initial fragmentation were not carried out, amino acid residues would build up in the reaction medium. How would their presence inhibit sequencing?
19. Define the following terms:
- mosaic protein
  - homologous polypeptide
  - cooperative binding
  - aldol condensation
  - globular protein
20. In an amino acid analysis, a large protein is broken down into overlapping fragments by using specific enzymes. Why must the sequences be overlapping?
21. Define the following terms:
- electrophoresis
  - molecular disease
  - $\alpha$ -carbon
  - isoelectric point
  - peptide bond
22. The following amino acid sequence represents bradykinin, a peptide released by certain organisms in response to wasp stings.
- Arg—Pro—Pro—Gly—Phe—Ser—Pro—Phe—Arg
- What amino acids or peptides are produced when bradykinin is treated with each of the following reagents?
- carboxypeptidase
  - chymotrypsin
  - trypsin
  - DNFB
23. Describe the problems associated with using a polypeptide's primary sequence to determine its final three-dimensional shape.
24. Describe the forces involved in protein folding.
25. What are the characteristics of motor proteins? How do organisms use them?

26. Briefly outline the roles of molecular chaperones in protein folding.
27. Define the following terms:
- hydrophobic amino acid
  - salt bridge
  - dabsyl-amino acid
  - site-directed mutagenesis
  - proteomics

## ThoughtQuestions

*These questions are designed to reinforce your understanding of all of the key concepts discussed in the book so far, including this chapter and all of the chapters before it. They may not have one right answer! The authors have provided possible solutions to these questions in the back of the book and in the accompanying Study Guide, for your reference.*

28. Residues such as valine, leucine, isoleucine, methionine, and phenylalanine are often found in the interior of proteins, whereas arginine, lysine, aspartic acid, and glutamic acid are often found on the surface of proteins. Suggest a reason for this observation. Where would you expect to find glutamine, glycine, and alanine?
29. Proteins that are synthesized by living organisms adopt a biologically active conformation. Yet when such molecules are prepared in the laboratory, they usually fail to spontaneously adopt their active conformations. Can you suggest why?
30. The active site of an enzyme contains sequences that are conserved because they participate in the protein's catalytic activity. The bulk of an enzyme, however, is not part of the active site. Because a substantial amount of energy is required to assemble enzymes, why are they usually so large?
31. A structural protein may incorporate large amounts of immobilized water as part of its structure. Can you suggest how protein molecules "freeze" the water in place and make it part of the protein structure?
32. The peptide bond is a stronger bond than that of esters. What structural feature of the peptide bond gives it additional bond strength?
33. Because of their tendency to avoid water, nonpolar amino acids play an important role in forming and maintaining the three-dimensional structure of proteins. Can you suggest how these molecules accomplish this feat?
34. Hydrolysis of  $\beta$ -endorphin (a peptide containing 31 amino acid residues) produces the following amino acids: Tyr, Gly (3), Phe (2), Met, Thr (3), Ser (2), Lys (5), Gln (2), Pro, Leu (2), Val (2), Asn (2), Ala (2), Ile, His, and Glu. Treatment with carboxypeptidase liberates Gln. Treatment with DNFB liberates DNP-Tyr. Treatment with trypsin produces the following peptides:  
Lys, Gly—Gln, Asn—Ala—Ile—Val—Lys,  
Tyr—Gly—Gly—Phe—Met—Thr—Ser—Glu—Lys,  
Asn—Ala—His—Lys, Ser—Gln—Thr—Pro—Leu—  
Val—Thr—Leu—Phe—Lys  
Treatment with chymotrypsin produces the following peptides:  
Lys—Asn—Ala—Ile—Val—Lys—Asn—Ala—  
His—Lys—Lys—Gly—Gln  
Tyr—Gly—Gly—Phe  
Met—Thr—Ser—Glu—Lys—Ser—Gln—Thr—Pro—  
Leu—Val—Thr—Leu—Phe  
What is the primary sequence of  $\beta$ -endorphin?.
35. Consider the following tripeptide:  
Gly—Ala—Val
- What is the approximate isoelectric point?
  - In which direction will the tripeptide move if placed in an electric field at pH 1, 5, 10, and 12?
36. Chymotrypsin is an enzyme that cleaves other enzymes during sequencing. Why don't chymotrypsin molecules attack each other?
37. Most amino acids appear bluish purple when treated with ninhydrin reagent. Proline and hydroxyproline appear yellow. Suggest a reason for the difference.
38. When the multifunction protein glyceraldehyde-3-phosphate dehydrogenase (GAPD) catalyzes a key reaction in glycolysis (a metabolic pathway in cytoplasm), it does so as a homotetramer (four identical subunits). The GAPD monomer is a nuclear DNA repair enzyme. Describe in general terms what structural properties of multifunction proteins allow this phenomenon.
39. From the following analytical results, deduce the structure of a peptide isolated from the Alantian orchid, which contains 14 amino acids. Complete hydrolysis produces the following amino acids: Gly (3), Leu (3), Glu (2), Pro, Met, Lys (2), Thr, Phe. Treatment with carboxypeptidase releases glycine. Treatment with DNFB releases DNP-glycine. Treatment with a nonspecific proteolytic enzyme produces the following fragments:  
Gly—Leu—Glu, Gly—Pro—Met—Lys,  
Lys—Glu, Thr—Phe—Leu—Leu—Gly,  
Lys—Glu—Thr—Phe—Leu,  
Leu—Leu—Gly,  
Glu—Thr—Phe, Glu—Gly—Pro,  
Pro—Met—Lys—Lys,  
and Gly—Leu
40. Many proteins have several functions. Provide examples. What natural forces are responsible for this phenomenon?
41. Why are multifunctional proteins necessary and/or desirable?
42. Given the following decapeptide sequence, which amino acids would you expect to be on the surface of this molecule once it folds into its native conformation?  
Gly—Phe—Tyr—Asn—Tyr—Met—Ser—His—Val—Leu
43. What amino acid residues of the decapeptide in Question 42 would tend to be found on the interior of the molecule?
44. What would be the products of the acid hydrolysis for 3 hours of the decapeptide in Question 42?
45. A mutational change alters a polypeptide by substituting 3 adjacent prolines for 3 glycines. What possible effect will this event have on the protein's structure?

46. As a genetic engineer, you have been given the following task: alter a protein's structure by converting a specific amino acid sequence that forms an extended  $\alpha$ -helix to one that forms a  $\beta$ -barrel. What types of amino acid are probably in the  $\alpha$ -helix, and which ones would you need to substitute?
47. Of the naturally occurring amino acids Gly, Val, Phe, His, and Ser, which would be likely to form coordination compounds with metals?
48.  $\beta$ -Endorphin, an opiate peptide, is released by the anterior pituitary gland at the base of the mammalian brain in response to stress or pain. As with other signal molecules, the effects of  $\beta$ -endorphin on its target tissue (neurons) is triggered when it binds to its receptor. In general terms, outline the process by which the  $\beta$ -endorphin receptor would be isolated and its structure characterized.
49. Amino acids are the precursors of a vast number of biologically active nitrogen-containing molecules. Which amino acids are the precursors of the following molecules?
50. You are a materials engineer who decides that a synthetic fibrous protein might possess desirable properties for a new product. If technical constraints dictate that only two or three amino acids can be used, which ones would you choose?
51. Suggest a protocol for separating oxytocin and vasopressin from an extract of the posterior pituitary gland.
52. The caging of water sequesters nonpolar amino acid residues to the interior of the folding protein. Can you explain how this phenomenon can have a significant impact on the formation and maintenance of protein native structure?

