26 Biomolecules: Amino Acids, Peptides, and Proteins

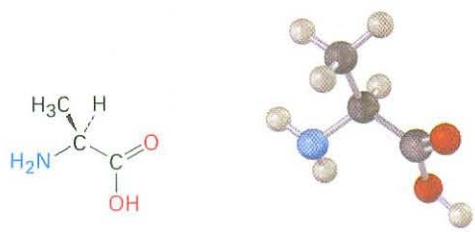
Organic KNOWLEDGE TOOLS

ThomsonNOW™ Throughout this chapter, sign in at **www.thomsonedu.com** for online self-study and interactive tutorials based on your level of understanding.

Online homework for this chapter may be assigned in Organic OWL.

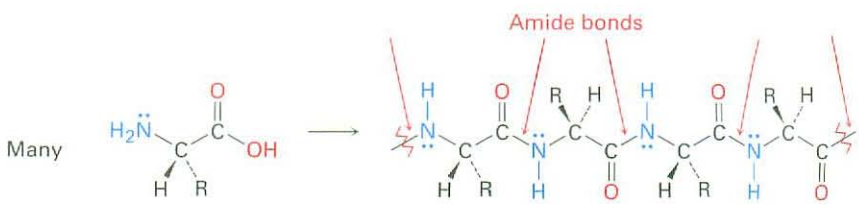
Proteins occur in every living organism, are of many different types, and have many different biological functions. The keratin of skin and fingernails, the fibroin of silk and spider webs, and the estimated 50,000 to 70,000 enzymes that catalyze the biological reactions in our bodies are all proteins. Regardless of their function, all proteins are made up of many *amino acids* linked together in a long chain.

Amino acids, as their name implies, are difunctional. They contain both a basic amino group and an acidic carboxyl group.



Alanine, an amino acid

Their value as building blocks to make proteins stems from the fact that amino acids can join together into long chains by forming amide bonds between the $-NH_2$ of one amino acid and the $-CO_2H$ of another. For classification purposes, chains with fewer than 50 amino acids are often called **peptides**, while the term **protein** is used for larger chains.



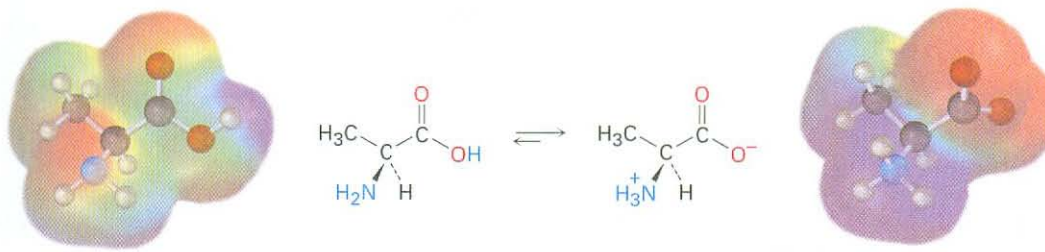
Sean Duggan

WHY THIS CHAPTER?

Continuing our look at the four main classes of biomolecules, we'll focus in this chapter on amino acids, the fundamental building blocks from which the 100,000 or so proteins in our bodies are made. We'll then see how amino acids are incorporated into proteins and the structures of those proteins. Any understanding of biological chemistry would be impossible without this study.

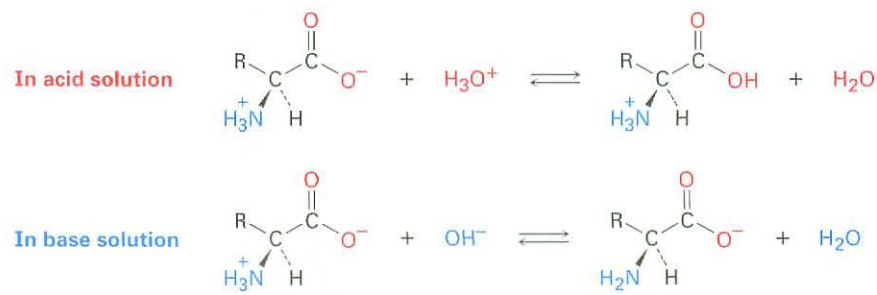
26.1 Structures of Amino Acids

We saw in Sections 20.3 and 24.5 that a carboxyl group is deprotonated and exists as the carboxylate anion at a physiological pH of 7.3, while an amino group is protonated and exists as the ammonium cation. Thus, amino acids exist in aqueous solution primarily in the form of a dipolar ion, or **zwitterion** (German *zwitter*, meaning "hybrid").



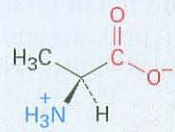
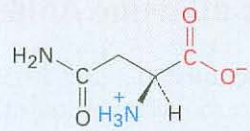
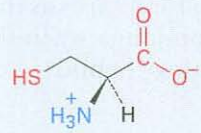
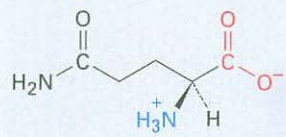
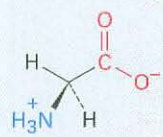
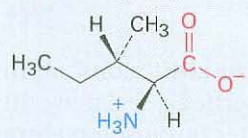
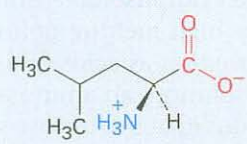
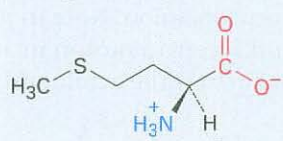
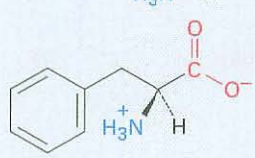
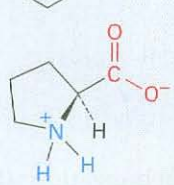
Alanine

Amino acid zwitterions are internal salts and therefore have many of the physical properties associated with salts. They have large dipole moments, are soluble in water but insoluble in hydrocarbons, and are crystalline substances with relatively high melting points. In addition, amino acids are *amphiprotic;* they can react either as acids or as bases, depending on the circumstances. In aqueous acid solution, an amino acid zwitterion is a base that *accepts* a proton to yield a cation; in aqueous base solution, the zwitterion is an acid that *loses* a proton to form an anion. Note that it is the carboxylate, $-CO_2^-$, that acts as the basic site and accepts a proton in acid solution, and it is the ammonium cation, $-NH_3^+$, that acts as the acidic site and donates a proton in base solution.



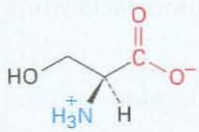
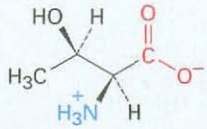
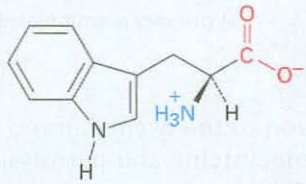
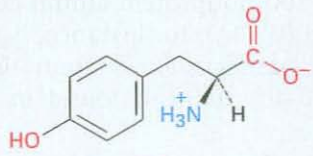
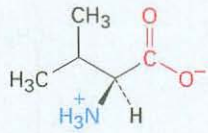
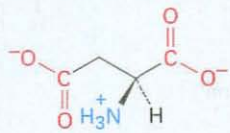
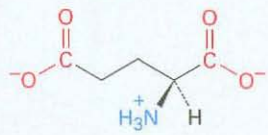
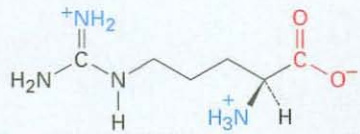
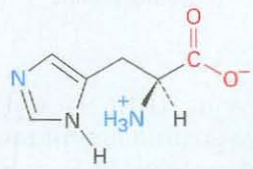
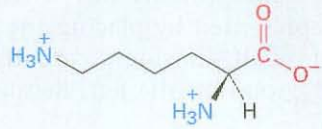
The structures, abbreviations (both three- and one-letter), and pK_a values of the 20 amino acids commonly found in proteins are shown in Table 26.1. All are

Table 26.1 The 20 Common Amino Acids in Proteins

Name	Abbreviations		MW	Stucture	pK_a α-CO_2H	pK_a α-NH_3^+	pK_a side chain	p*I*
Neutral Amino Acids								
Alanine	Ala	A	89		2.34	9.69	—	6.01
Asparagine	Asn	N	132		2.02	8.80	—	5.41
Cysteine	Cys	C	121		1.96	10.28	8.18	5.07
Glutamine	Gln	Q	146		2.17	9.13	—	5.65
Glycine	Gly	G	75		2.34	9.60	—	5.97
Isoleucine	Ile	I	131		2.36	9.60	—	6.02
Leucine	Leu	L	131		2.36	9.60	—	5.98
Methionine	Met	M	149		2.28	9.21	—	5.74
Phenylalanine	Phe	F	165		1.83	9.13	—	5.48
Proline	Pro	P	115		1.99	10.60	—	6.30

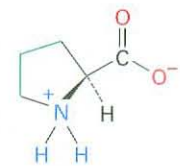
(continued)

Table 26.1 The 20 Common Amino Acids in Proteins *(continued)*

Name	Abbreviations		MW	Stucture	pK_a α-CO_2H	pK_a α-NH_3^+	pK_a side chain	p*I*
Neutral Amino Acids *continued*								
Serine	Ser	S	105		2.21	9.15	—	5.68
Threonine	Thr	T	119		2.09	9.10	—	5.60
Tryptophan	Trp	W	204		2.83	9.39	—	5.89
Tyrosine	Tyr	Y	181		2.20	9.11	10.07	5.66
Valine	Val	V	117		2.32	9.62	—	5.96
Acidic Amino Acids								
Aspartic acid	Asp	D	133		1.88	9.60	3.65	2.77
Glutamic acid	Glu	E	147		2.19	9.67	4.25	3.22
Basic Amino Acids								
Arginine	Arg	R	174		2.17	9.04	12.48	10.76
Histidine	His	H	155		1.82	9.17	6.00	7.59
Lysine	Lys	K	146		2.18	8.95	10.53	9.74

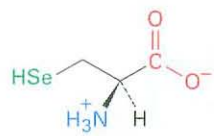
α-amino acids, meaning that the amino group in each is a substituent on the α carbon atom—the one next to the carbonyl group. Nineteen of the twenty amino acids are primary amines, RNH_2, and differ only in the nature of the substituent attached to the α carbon, called the **side chain**. Proline is a secondary amine and the only amino acid whose nitrogen and α carbon atoms are part of a ring.

A primary α-amino acid

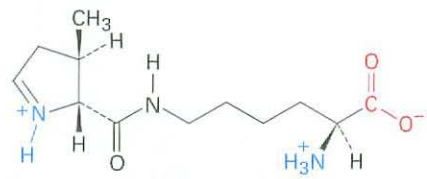


Proline, a secondary α-amino acid

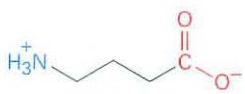
In addition to the twenty amino acids commonly found in proteins, two others—selenocysteine and pyrrolysine—are found in some organisms, and more than 700 nonprotein amino acids are also found in nature. γ-Aminobutyric acid (GABA), for instance, is found in the brain and acts as a neurotransmitter; homocysteine is found in blood and is linked to coronary heart disease; and thyroxine is found in the thyroid gland, where it acts as a hormone.



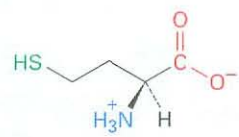
Selenocysteine



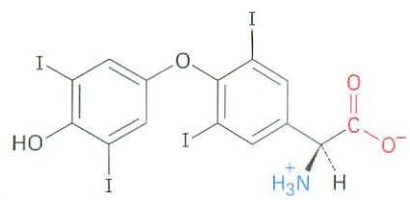
Pyrrolysine



γ-Aminobutyric acid



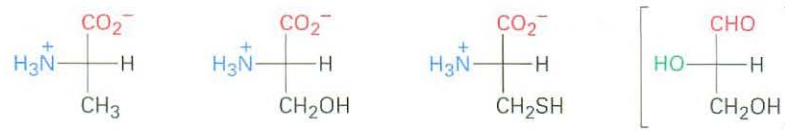
Homocysteine



Thyroxine

Except for glycine, $H_2NCH_2CO_2H$, the α carbons of amino acids are chirality centers. Two enantiomers of each are therefore possible, but nature uses only one to build proteins. In Fischer projections, naturally occurring amino acids are represented by placing the $-CO_2^-$ group at the top and the side chain down, as if drawing a carbohydrate (Section 25.2) and then placing the $-NH_3^+$ group on the left. Because of their stereochemical similarity to

L sugars (Section 25.3), the naturally occurring α-amino acids are often referred to as L amino acids.



L-Alanine
(*S*)-Alanine

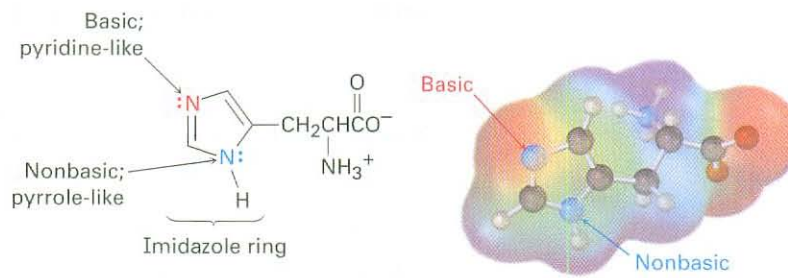
L-Serine
(*S*)-Serine

L-Cysteine
(*R*)-Cysteine

L-Glyceraldehyde

The 20 common amino acids can be further classified as neutral, acidic, or basic, depending on the structure of their side chains. Fifteen of the twenty have neutral side chains, two (aspartic acid and glutamic acid) have an extra carboxylic acid function in their side chains, and three (lysine, arginine, and histidine) have basic amino groups in their side chains. Note that both cysteine (a thiol) and tyrosine (a phenol), although usually classified as neutral amino acids, nevertheless have weakly acidic side chains that can be deprotonated in strongly basic solution.

At the physiological pH of 7.3 within cells, the side-chain carboxyl groups of aspartic acid and glutamic acid are deprotonated and the basic side-chain nitrogens of lysine and arginine are protonated. Histidine, however, which contains a heterocyclic imidazole ring in its side chain, is not quite basic enough to be protonated at pH 7.3. Note that only the pyridine-like, doubly bonded nitrogen in histidine is basic. The pyrrole-like singly bonded nitrogen is nonbasic because its lone pair of electrons is part of the 6 π electron aromatic imidazole ring (Section 24.9).



Histidine

Humans are able to synthesize only 11 of the 20 amino acids in proteins, called *nonessential amino acids*. The other 9, called *essential amino acids*, are biosynthesized only in plants and microorganisms and must be obtained in our diet. The division between essential and nonessential amino acids is not clearcut, however: tyrosine, for instance, is sometimes considered nonessential because humans can produce it from phenylalanine, but phenylalanine itself is essential and must be obtained in the diet. Arginine can be synthesized by humans, but much of the arginine we need also comes from our diet.

Problem 26.1 How many of the α-amino acids shown in Table 26.1 contain aromatic rings? How many contain sulfur? How many contain alcohols? How many contain hydrocarbon side chains?

Problem 26.2 Of the 19 L amino acids, 18 have the *S* configuration at the α carbon. Cysteine is the only L amino acid that has an *R* configuration. Explain.

Problem 26.3 The amino acid threonine, (2*S*,3*R*)-2-amino-3-hydroxybutanoic acid, has two chirality centers.

(a) Draw a Fischer projection of threonine.

(b) Draw a Fischer projection of a threonine diastereomer, and label its chirality centers as *R* or *S*.

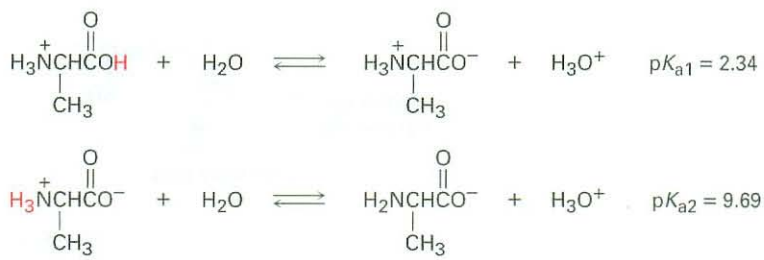
26.2 Amino Acids, the Henderson–Hasselbalch Equation, and Isoelectric Points

ThomsonNOW Click *Organic Interactive* to **learn to estimate isoelectric points for simple amino acids and peptides.**

According to the Henderson–Hasselbalch equation (Sections 20.3 and 24.5), if we know both the pH of a solution and the pK_a of an acid HA, we can calculate the ratio of $[A^-]$ to [HA] in the solution. Furthermore, when pH = pK_a, the two forms A^- and HA are present in equal amounts because log 1 = 0.

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

To apply the Henderson–Hasselbalch equation to an amino acid, let's find out what species are present in a 1.00 M solution of alanine at pH = 9.00. According to Table 26.1, protonated alanine [$^+H_3NCH(CH_3)CO_2H$] has pK_{a1} = 2.34, and neutral zwitterionic alanine [$^+H_3NCH(CH_3)CO_2^-$] has pK_{a2} = 9.69:



Since the pH of the solution is much closer to pK_{a2} than to pK_{a1}, we need to use pK_{a2} for the calculation. From the Henderson–Hasselbalch equation, we have:

$$\log\frac{[A^-]}{[HA]} = pH - pK_a = 9.00 - 9.69 = -0.69$$

so

$$\frac{[A^-]}{[HA]} = \text{antilog}(-0.69) = 0.20 \quad \text{and} \quad [A^-] = 0.20\,[HA]$$

In addition, we know that

$$[A^-] + [HA] = 1.00\ M$$

Solving the two simultaneous equations gives [HA] = 0.83 and [A^-] = 0.17. In other words, at pH = 9.00, 83% of alanine molecules in a 1.00 M solution are neutral (zwitterionic) and 17% are deprotonated. Similar calculations can be done at any other pH and the results plotted to give the *titration curve* shown in Figure 26.1.

Each leg of the titration curve is calculated separately. The first leg, from pH 1 to 6, corresponds to the dissociation of protonated alanine, H_2A^+. The second leg, from pH 6 to 11, corresponds to the dissociation of zwitterionic alanine, HA. It's as if we started with H_2A^+ at low pH and then titrated with NaOH. When 0.5 equivalent of NaOH is added, the deprotonation of H_2A^+ is 50% done; when 1.0 equivalent of NaOH is added, the deprotonation of H_2A^+ is complete and HA predominates; when 1.5 equivalent of NaOH is added, the deprotonation of HA is 50% done; and when 2.0 equivalents of NaOH is added, the deprotonation of HA is complete.

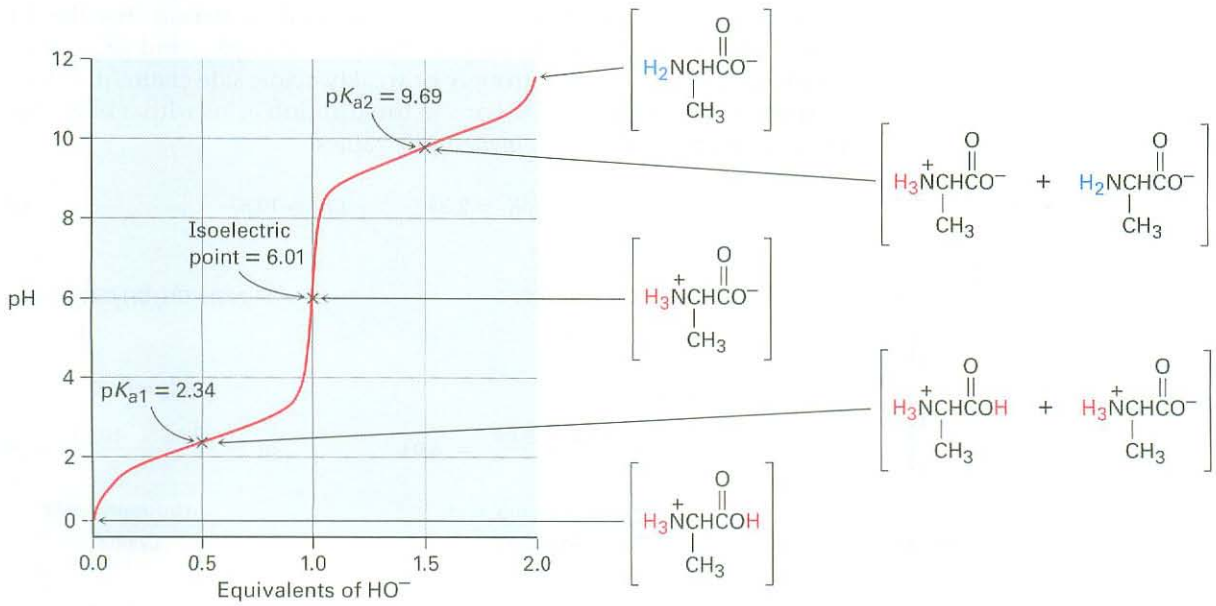
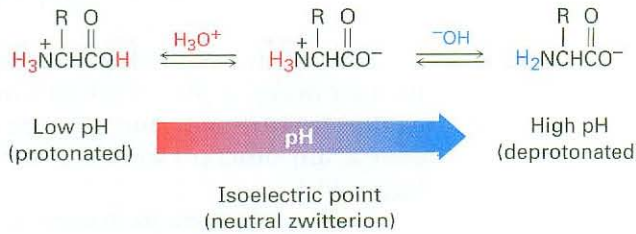


Figure 26.1 A titration curve for alanine, plotted using the Henderson–Hasselbalch equation. Each of the two legs is plotted separately. At pH < 1, alanine is entirely protonated; at pH = 2.34, alanine is a 50:50 mix of protonated and neutral forms; at pH 6.01, alanine is entirely neutral; at pH = 9.69, alanine is a 50:50 mix of neutral and deprotonated forms; at pH > 11.5, alanine is entirely deprotonated.

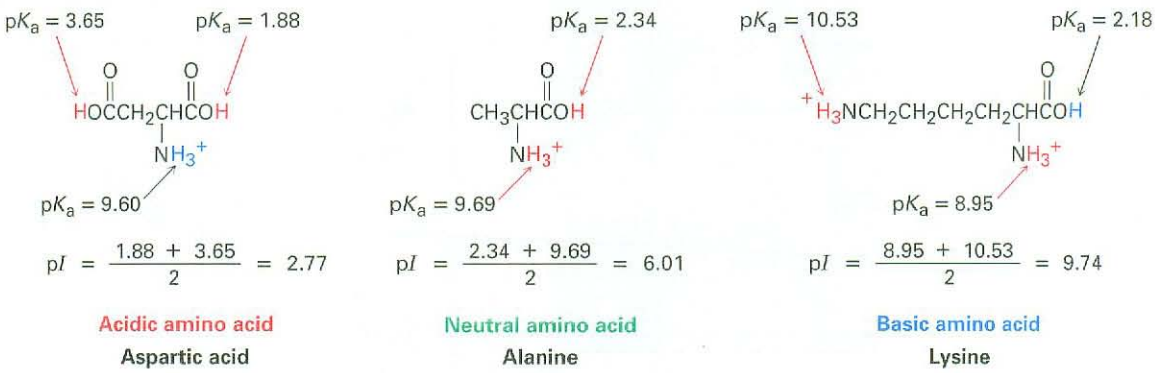
Look carefully at the titration curve in Figure 26.1. In acid solution, the amino acid is protonated and exists primarily as a cation. In basic solution, the amino acid is deprotonated and exists primarily as an anion. In between the two is an intermediate pH at which the amino acid is exactly balanced between anionic and cationic forms and exists primarily as the neutral,

dipolar zwitterion. This pH is called the amino acid's **isoelectric point (p*I*)** and has a value of 6.01 for alanine.



The isoelectric point of an amino acid depends on its structure, with values for the 20 common amino acids given in Table 26.1. The 15 neutral amino acids have isoelectric points near neutrality, in the pH range 5.0 to 6.5. The two acidic amino acids have isoelectric points at lower pH so that deprotonation of the side-chain $-CO_2H$ does not occur at their p*I*, and the three basic amino acids have isoelectric points at higher pH so that protonation of the side-chain amino group does not occur at their p*I*.

More specifically, the p*I* of any amino acid is the average of the two acid-dissociation constants that involve the neutral zwitterion. For the 13 amino acids with a neutral side chain, p*I* is the average of pK_{a1} and pK_{a2}. For the four amino acids with either a strongly or weakly acidic side chain, p*I* is the average of the two *lowest* pK_a values. For the three amino acids with a basic side chain, p*I* is the average of the two *highest* pK_a values.



Just as individual amino acids have isoelectric points, proteins have an overall p*I* because of the acidic or basic amino acids they may contain. The enzyme lysozyme, for instance, has a preponderance of basic amino acids and thus has a high isoelectric point (p*I* = 11.0). Pepsin, however, has a preponderance of acidic amino acids and a low isoelectric point (p*I* ~ 1.0). Not surprisingly, the solubilities and properties of proteins with different p*I*'s are strongly affected by the pH of the medium. Solubility is usually lowest at the isoelectric point, where the protein has no net charge, and is higher both above and below the p*I*, where the protein is charged.

We can take advantage of the differences in isoelectric points to separate a mixture of proteins into its pure constituents. Using a technique known as

electrophoresis, a mixture of proteins is placed near the center of a strip of paper or gel. The paper or gel is moistened with an aqueous buffer of a given pH, and electrodes are connected to the ends of the strip. When an electric potential is applied, those proteins with negative charges (those that are deprotonated because the pH of the buffer is above their isoelectric point) migrate slowly toward the positive electrode. At the same time, those amino acids with positive charges (those that are protonated because the pH of the buffer is below their isoelectric point) migrate toward the negative electrode.

Different proteins migrate at different rates, depending on their isoelectric points and on the pH of the aqueous buffer, thereby separating the mixture into its pure components. Figure 26.2 illustrates this separation for a mixture containing basic, neutral, and acidic components.

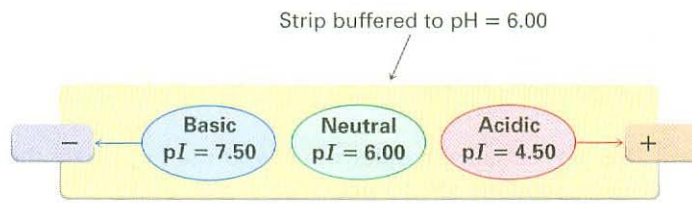
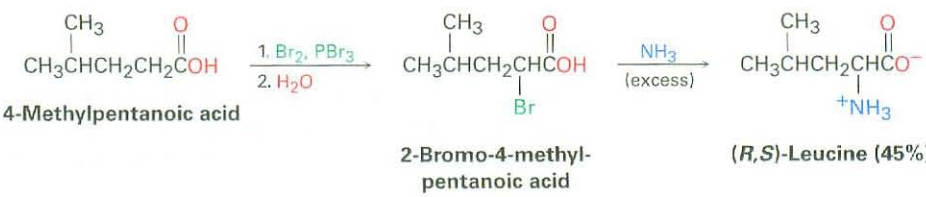


Figure 26.2 Separation of a protein mixture by electrophoresis. At pH = 6.00, a neutral protein does not migrate, a basic protein is protonated and migrates toward the negative electrode, and an acidic protein is deprotonated and migrates toward the positive electrode.

Problem 26.4 Hemoglobin has p*I* = 6.8. Does hemoglobin have a net negative charge or net positive charge at pH = 5.3? At pH = 7.3?

26.3 Synthesis of Amino Acids

α-Amino acids can be synthesized in the laboratory using some of the reactions discussed in previous chapters. One of the oldest methods of α-amino acid synthesis begins with α bromination of a carboxylic acid by treatment with Br_2 and PBr_3 (the Hell–Volhard–Zelinskii reaction; Section 22.4). S_N2 substitution of the α-bromo acid with ammonia then yields an α-amino acid.



4-Methylpentanoic acid

2-Bromo-4-methylpentanoic acid

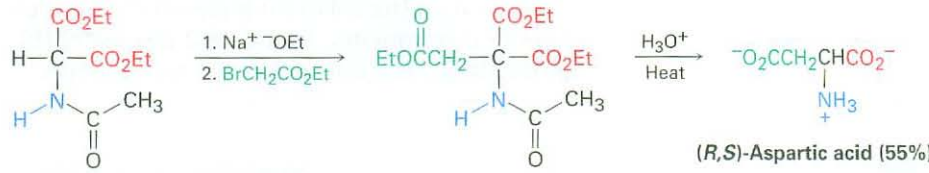
(*R,S*)-Leucine (45%)

Problem 26.5 Show how you could prepare the following α-amino acids from the appropriate carboxylic acids:

(a) Phenylalanine (b) Valine

The Amidomalonate Synthesis

A more general method for preparation of α-amino acids is the *amidomalonate synthesis*, a straightforward extension of the malonic ester synthesis (Section 22.7). The reaction begins with conversion of diethyl acetamidomalonate into an enolate ion by treatment with base, followed by S_N2 alkylation with a primary alkyl halide. Hydrolysis of both the amide protecting group and the esters occurs when the alkylated product is warmed with aqueous acid, and decarboxylation then takes place to yield an α-amino acid. For example, aspartic acid can be prepared from ethyl bromoacetate, $BrCH_2CO_2Et$:



Diethyl acetamidomalonate

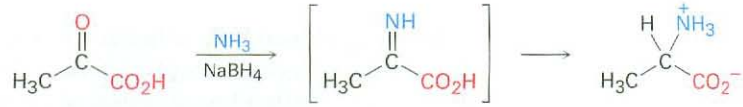
(*R,S*)-Aspartic acid (55%)

Problem 26.6 What alkyl halides would you use to prepare the following α-amino acids by the amidomalonate method?

(a) Leucine **(b)** Histidine **(c)** Tryptophan **(d)** Methionine

Reductive Amination of α-Keto Acids

Yet a third method for the synthesis of α-amino acids is by reductive amination of an α-keto acid with ammonia and a reducing agent. Alanine, for instance, is prepared by treatment of pyruvic acid with ammonia in the presence of $NaBH_4$. As described in Section 24.6, the reaction proceeds through formation of an intermediate imine that is then reduced.



Pyruvic acid **Imine intermediate** **(*R,S*)-Alanine**

Enantioselective Synthesis

The synthesis of an α-amino acid from an achiral precursor by any of the methods described in the previous section yields a racemic mixture, with equal amounts of *S* and *R* enantiomers. To use an amino acid in the laboratory synthesis of a naturally occurring protein, however, the pure *S* enantiomer must be obtained.

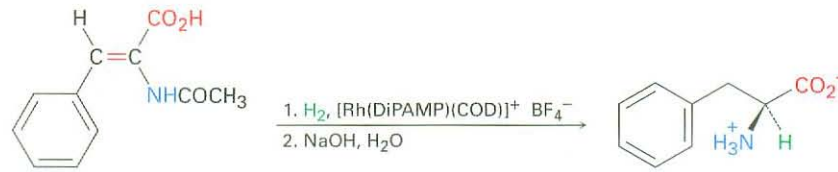
Two methods are used in practice to obtain enantiomerically pure amino acids. One way is to resolve the racemic mixture into its pure enantiomers (Section 9.8). A more direct approach, however, is to use an *enantioselective synthesis* to prepare only the desired *S* enantiomer directly. As discussed in the Chapter 19 *Focus On*, the idea behind enantioselective synthesis is to find a chiral reaction catalyst that will temporarily hold a substrate molecule in an unsymmetrical environment. While in that chiral environment, the substrate may be more

William S. Knowles

William S. Knowles (1917–) was born in Taunton, Massachusetts, and received his Ph.D. from Columbia University in 1942. Following his graduate studies, he began work at the Monsanto Company in St. Louis, Missouri, where he remained until his retirement in 1986. He received the 2001 Nobel Prize in chemistry for his work on enantioselective synthesis, one of the few non-academic scientists to be thus honored.

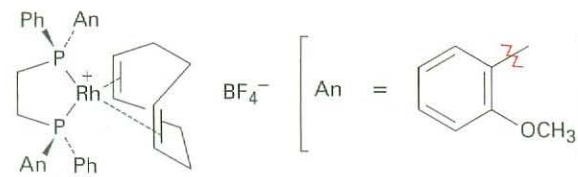
open to reaction on one side than on another, leading to an excess of one enantiomeric product over another.

William Knowles at the Monsanto Company discovered some years ago that α-amino acids can be prepared enantioselectively by hydrogenation of a *Z* enamido acid with a chiral hydrogenation catalyst. (*S*)-Phenylalanine, for instance, is prepared in 98.7% purity contaminated by only 1.3% of the (*R*) enantiomer when a chiral rhodium catalyst is used. For this discovery, Knowles shared the 2001 Nobel Prize in chemistry.



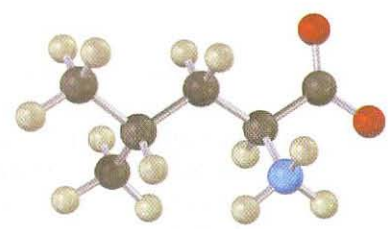
A (*Z*) enamido acid **(*S*)-Phenylalanine**

The most effective catalysts for enantioselective amino acid synthesis are coordination complexes of rhodium(I) with 1,5-cyclooctadiene (COD) and a chiral diphosphine such as (*R*,*R*)-1,2-bis(*o*-anisylphenylphosphino)ethane, the so-called DiPAMP ligand. The complex owes its chirality to the presence of the trisubstituted phosphorus atoms (Section 9.12).



[Rh(*R*,*R*-DiPAMP)(COD)]$^+$ BF_4^-

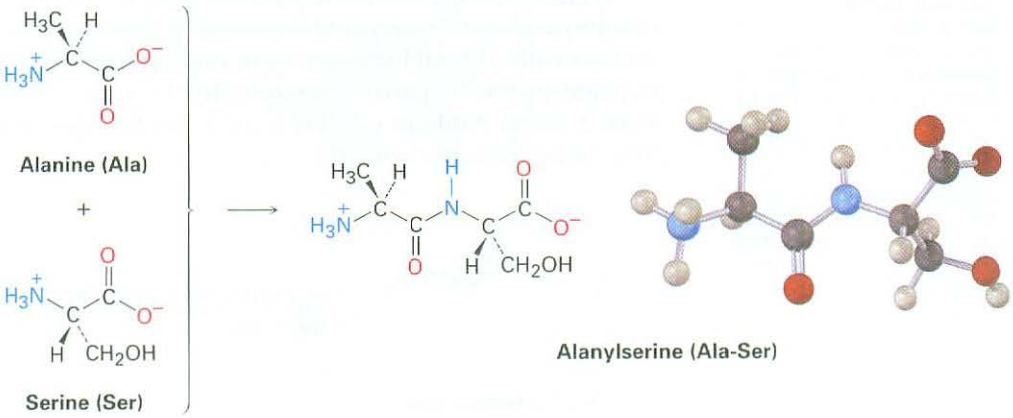
Problem 26.7 Show how you could prepare the following amino acid enantioselectively:



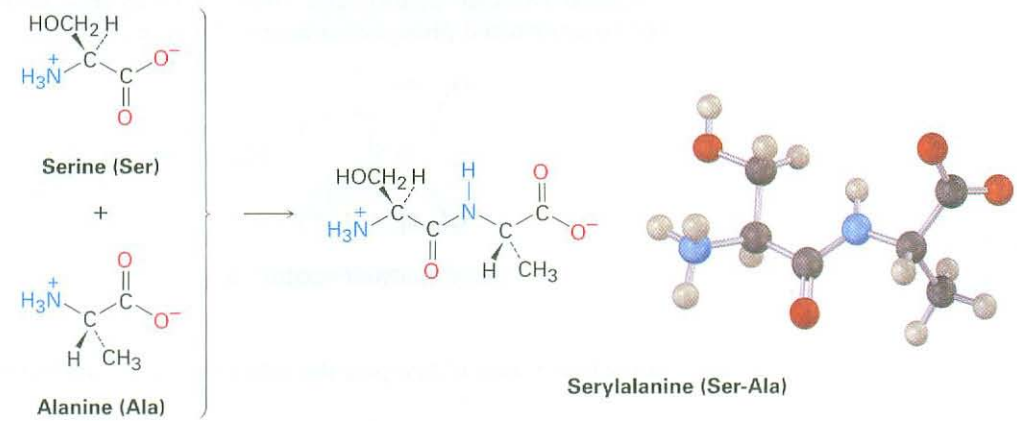
26.4 Peptides and Proteins

Proteins and peptides are amino acid polymers in which the individual amino acids, called **residues**, are linked together by amide bonds, or *peptide bonds*. An amino group from one residue forms an amide bond with the carboxyl of a second residue, the amino group of the second forms an amide bond with the carboxyl of a third, and so on. For example, alanylserine is the dipeptide that

results when an amide bond is formed between the alanine carboxyl and the serine amino group.



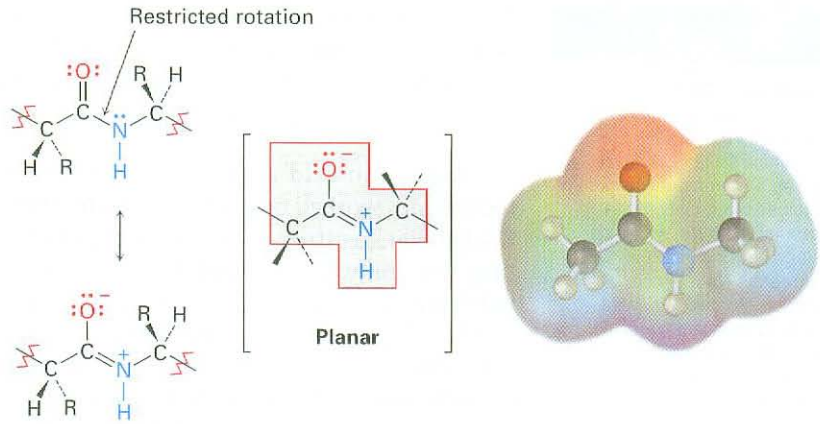
Note that two dipeptides can result from reaction between alanine and serine, depending on which carboxyl group reacts with which amino group. If the alanine amino group reacts with the serine carboxyl, serylalanine results.



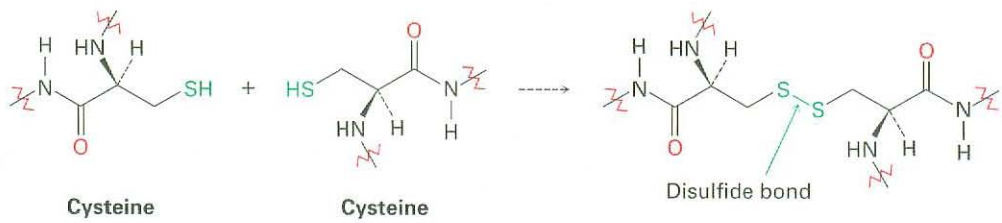
The long, repetitive sequence of $-N-CH-CO-$ atoms that make up a continuous chain is called the protein's **backbone**. By convention, peptides are written with the **N-terminal amino acid** (the one with the free $-NH_3^+$ group) on the left and the **C-terminal amino acid** (the one with the free $-CO_2^-$ group) on the right. The name of the peptide is indicated by using the abbreviations listed in Table 26.1 for each amino acid. Thus, alanylserine is abbreviated Ala-Ser or A-S, and serylalanine is abbreviated Ser-Ala or S-A. Needless to say, the one-letter abbreviations are more convenient than the older three-letter abbreviations.

The amide bond that links different amino acids together in peptides is no different from any other amide bond (Section 24.3). Amide nitrogens are nonbasic because their unshared electron pair is delocalized by interaction with the carbonyl group. This overlap of the nitrogen p orbital with the p orbitals of the carbonyl group imparts a certain amount of double-bond character to the

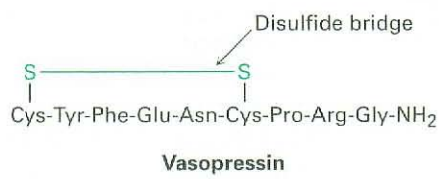
C–N bond and restricts rotation around it. The amide bond is therefore planar, and the N–H is oriented 180° to the C=O.



A second kind of covalent bonding in peptides occurs when a disulfide linkage, RS–SR, is formed between two cysteine residues. As we saw in Section 18.8, a disulfide is formed by mild oxidation of a thiol, RSH, and is cleaved by mild reduction.



A disulfide bond between cysteine residues in different peptide chains links the otherwise separate chains together, while a disulfide bond between cysteine residues in the same chain forms a loop. Such is the case, for instance, with vasopressin, an antidiuretic hormone found in the pituitary gland. Note that the C-terminal end of vasopressin occurs as the primary amide, $-CONH_2$, rather than as the free acid.



Problem 26.8 Six isomeric tripeptides contain valine, tyrosine, and glycine. Name them using both three- and one-letter abbreviations.

Problem 26.9 Draw the structure of M-P-V-G, and indicate the amide bonds.

26.5 Amino Acid Analysis of Peptides

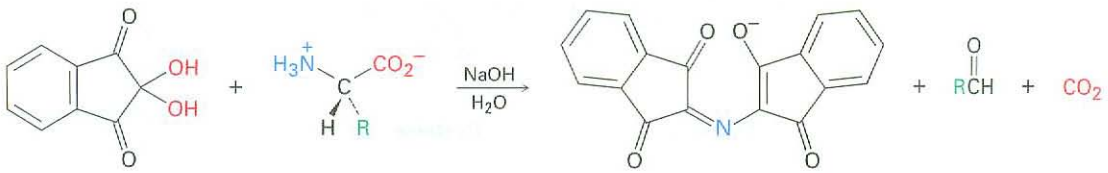
William Howard Stein

William Howard Stein (1911–1980) was born in New York City and received his Ph.D. in 1938 from the Columbia College of Physicians and Surgeons. He immediately joined the faculty of the Rockefeller Institute, where he remained until his death. In 1972, he shared the Nobel Prize in chemistry for his work with Stanford Moore on developing methods of amino acid analysis and for determining the structure of ribonuclease.

To determine the structure of a protein or peptide, we need to answer three questions: What amino acids are present? How much of each is present? In what sequence do the amino acids occur in the peptide chain? The answers to the first two questions are provided by an automated instrument called an *amino acid analyzer.*

An amino acid analyzer is an automated instrument based on analytical techniques worked out in the 1950s by William Stein and Stanford Moore at the Rockefeller Institute, now the Rockefeller University, in New York. In preparation for analysis, the peptide is broken into its constituent amino acids by reducing all disulfide bonds, capping the −SH groups of cysteine residues by S_N2 reaction with iodoacetic acid, and hydrolyzing the amide bonds by heating with aqueous 6 M HCl at 110 °C for 24 hours. The resultant amino acid mixture is then analyzed, either by high-pressure liquid chromatography (HPLC) as described in the Chapter 12 *Focus On,* or by a related technique called ion-exchange chromatography.

In the ion-exchange technique, separated amino acids exiting *(eluting)* from the end of the chromatography column mix with a solution of *ninhydrin* and undergo a rapid reaction that produces an intense purple color. The color is detected by a spectrometer, and a plot of elution time versus spectrometer absorbance is obtained.



Ninhydrin + **α-Amino acid** → **(purple color)** + RCHO + CO_2

Stanford Moore

Stanford Moore (1913–1982) was born in Chicago, Illinois, and received his Ph.D. from the University of Wisconsin in 1938. He was a professor at the Rockefeller Institute and shared the 1972 Nobel Prize in chemistry with his colleague and collaborator, William Stein.

Because the amount of time required for a given amino acid to elute from a standard column is reproducible, the identities of the amino acids in a peptide can be determined. The amount of each amino acid in the sample is determined by measuring the intensity of the purple color resulting from its reaction with ninhydrin. Figure 26.3 shows the results of amino acid analysis of a standard equimolar mixture of 17 α-amino acids. Typically, amino acid analysis requires about 100 picomoles (2–3 μg) of sample for a protein containing about 200 residues.

Problem 26.10 Show the structure of the product you would expect to obtain by S_N2 reaction of a cysteine residue with iodoacetic acid.

Problem 26.11 Show the structures of the products obtained on reaction of valine with ninhydrin.

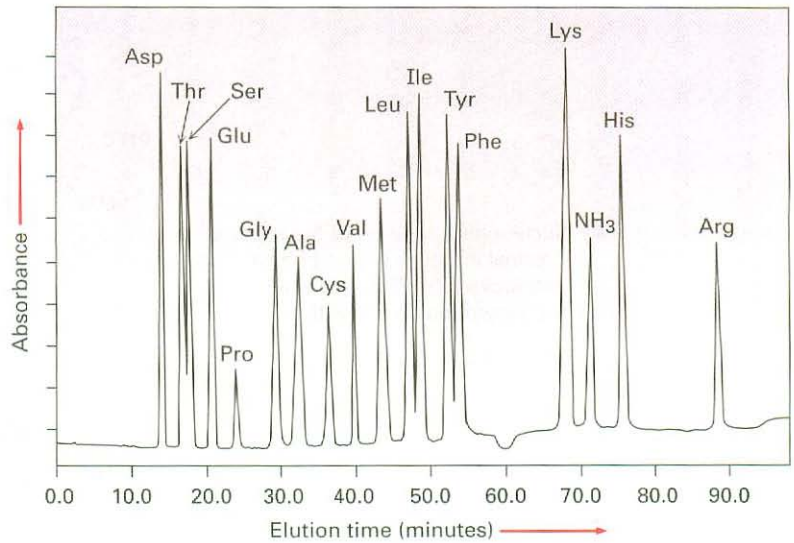


Figure 26.3 Amino acid analysis of an equimolar mixture of 17 amino acids.

26.6 Peptide Sequencing: The Edman Degradation

ThomsonNOW Click *Organic Interactive* to **predict products from degradation and modification reactions of simple peptides.**

With the identities and amounts of amino acids known, the peptide is *sequenced* to find out in what order the amino acids are linked together. Much peptide sequencing is now done by mass spectrometry, using either electrospray ionization (ESI) or matrix-assisted laser desorption ionization (MALDI) linked to a time-of-flight (TOF) mass analyzer, as described in Section 12.4. Also in common use is a chemical method of peptide sequencing called the *Edman degradation*.

The general idea of peptide sequencing by Edman degradation is to cleave one amino acid at a time from an end of the peptide chain. That terminal amino acid is then separated and identified, and the cleavage reactions are repeated on the chain-shortened peptide until the entire peptide sequence is known. Automated protein sequencers are available that allow as many as 50 repetitive sequencing cycles to be carried out before a buildup of unwanted by products interferes with the results. So efficient are these instruments that sequence information can be obtained from as little as 1 to 5 picomoles of sample—less than 0.1 μg.

Edman degradation involves treatment of a peptide with phenyl isothiocyanate (PITC), $C_6H_5{-}N{=}C{=}S$, followed by treatment with trifluoroacetic acid, as shown in Figure 26.4. The first step attaches the PITC to the $-NH_2$ group of the N-terminal amino acid, and the second step splits the N-terminal residue from the peptide chain, yielding an anilinothiazolinone (ATZ) derivative plus the chain-shortened peptide. Further acid-catalyzed rearrangement of the ATZ derivative with aqueous acid converts it into a phenylthiohydantoin (PTH), which is identified chromatographically by comparison of its elution time with the known elution times of PTH derivatives of the 20 common amino acids. The chain-shortened peptide is then automatically resubmitted to another round of Edman degradation.

Complete sequencing of large proteins by Edman degradation is impractical because of the buildup of unwanted by-products. To get around the problem, a large peptide chain is first cleaved by partial hydrolysis into a number of smaller fragments, the sequence of each fragment is determined, and the individual fragments are fitted together by matching the overlapping ends. In this way, protein chains with more than 400 amino acids have been sequenced.

Pehr Victor Edman

Pehr Victor Edman (1916–1977) was born in Stockholm, Sweden, and received an M.D. in 1946 at the Karolinska Institute. After a year in the United States at the Rockefeller Institute, he returned to Sweden as professor at the University of Lund. In 1957, he moved to St. Vincent's School of Medical Research in Melbourne, Australia, where he developed and automated the method of peptide sequencing that bears his name. A reclusive man, he never received the prizes or recognition merited by the importance of his work.

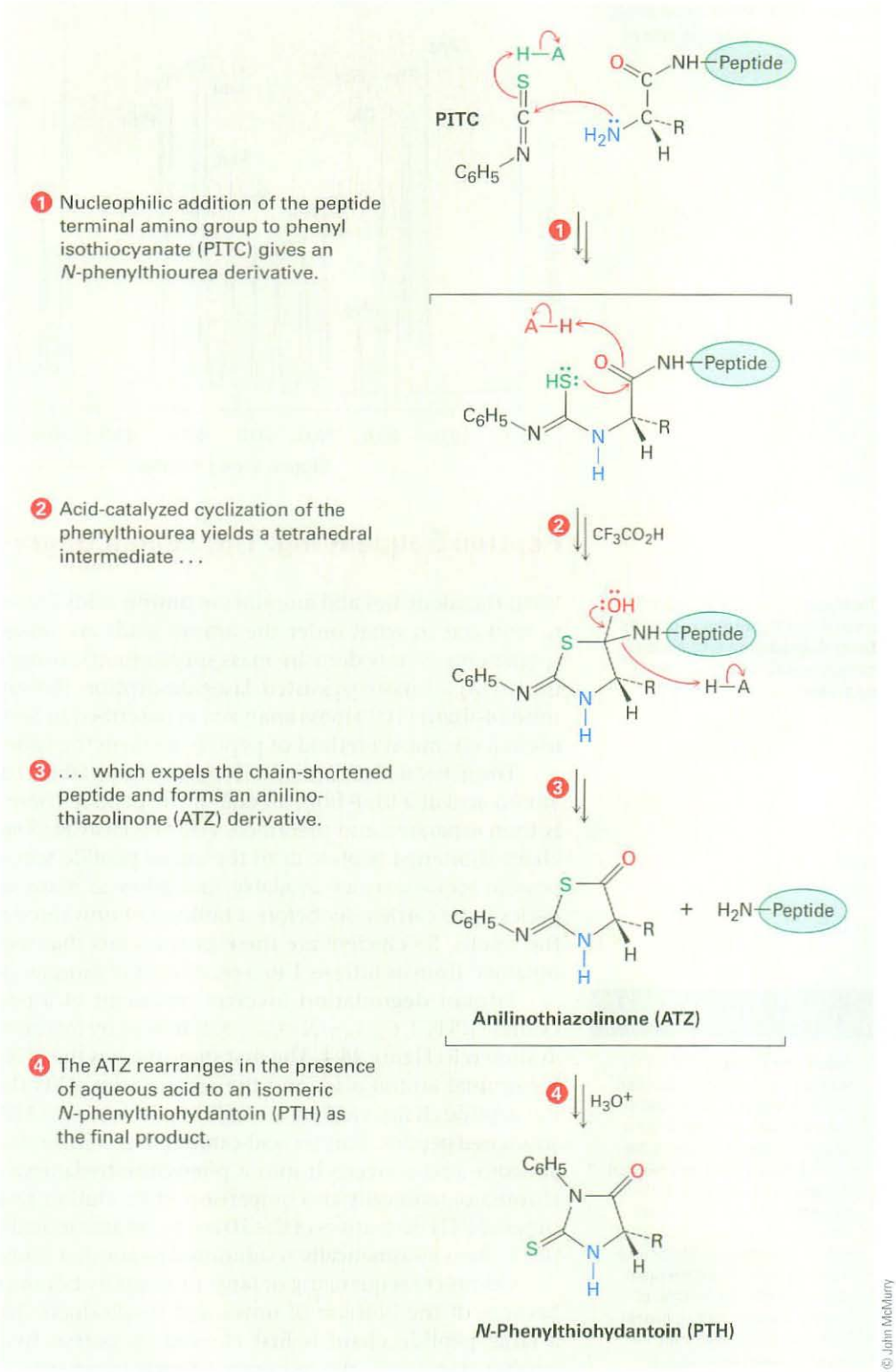
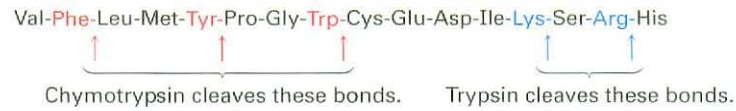


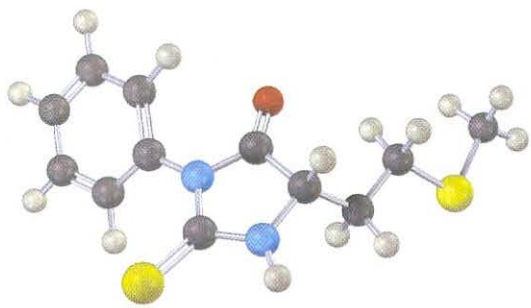
Figure 26.4 MECHANISM: Mechanism of the Edman degradation for N-terminal analysis of peptides.

Partial hydrolysis of a peptide can be carried out either chemically with aqueous acid or enzymatically. Acidic hydrolysis is unselective and leads to a more or less random mixture of small fragments, but enzymatic hydrolysis is quite specific. The enzyme trypsin, for instance, catalyzes hydrolysis of peptides only at the carboxyl side of the basic amino acids arginine and lysine; chymotrypsin cleaves only at the carboxyl side of the aryl-substituted amino acids phenylalanine, tyrosine, and tryptophan.



Problem 26.12 The octapeptide angiotensin II has the sequence Asp-Arg-Val-Tyr-Ile-His-Pro-Phe. What fragments would result if angiotensin II were cleaved with trypsin? With chymotrypsin?

Problem 26.13 What is the N-terminal residue on a peptide that gives the following PTH derivative on Edman degradation?



Problem 26.14 Draw the structure of the PTH derivative that would be formed on Edman degradation of angiotensin II (Problem 26.12).

Problem 26.15 Give the amino acid sequence of hexapeptides that produce the following sets of fragments on partial acid hydrolysis:

(a) Arg, Gly, Ile, Leu, Pro, Val gives Pro-Leu-Gly, Arg-Pro, Gly-Ile-Val

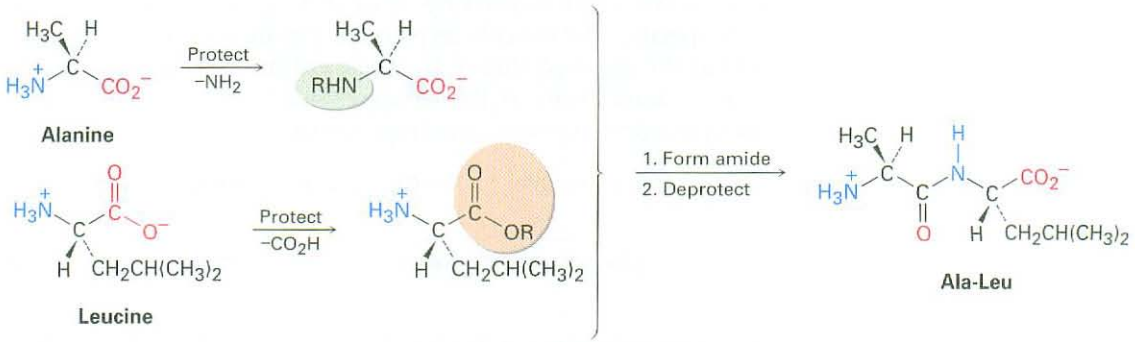
(b) N, L, M, W, V_2 gives V-L, V-M-W, W-N-V

26.7 Peptide Synthesis

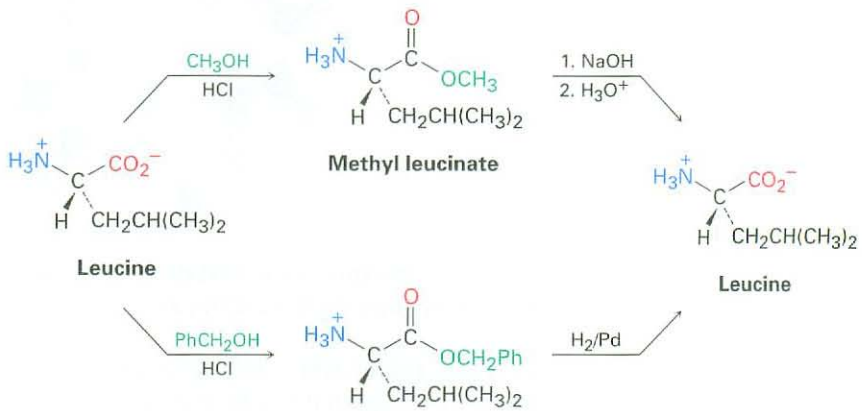
With its structure known, the synthesis of a peptide can then be undertaken—perhaps to obtain a larger amount for biological evaluation. A simple amide might be formed by treating an amine and a carboxylic acid with dicyclohexylcarbodiimide (DCC; Section 21.7), but peptide synthesis is a more difficult problem because many different amide bonds must be formed in a specific order rather than at random.

The solution to the specificity problem is to *protect* those functional groups we want to render unreactive while leaving exposed only those functional groups we want to react. For example, if we wanted to couple alanine with leucine to synthesize Ala-Leu, we could protect the $-NH_2$ group of alanine and

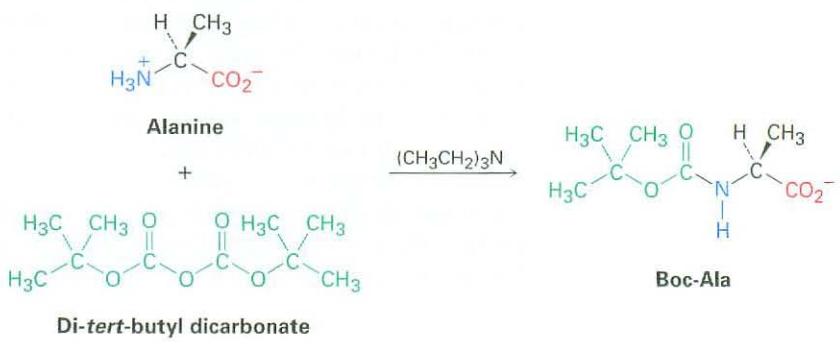
the $-CO_2H$ group of leucine to render them unreactive, then form the desired amide bond, and then remove the protecting groups.



Many different amino- and carboxyl-protecting groups have been devised, but only a few are widely used. Carboxyl groups are often protected simply by converting them into methyl or benzyl esters. Both groups are easily introduced by standard methods of ester formation (Section 21.6) and are easily removed by mild hydrolysis with aqueous NaOH. Benzyl esters can also be cleaved by catalytic *hydrogenolysis* of the weak benzylic C—O bond ($RCO_2—CH_2Ph + H_2 \rightarrow RCO_2H + PhCH_3$).

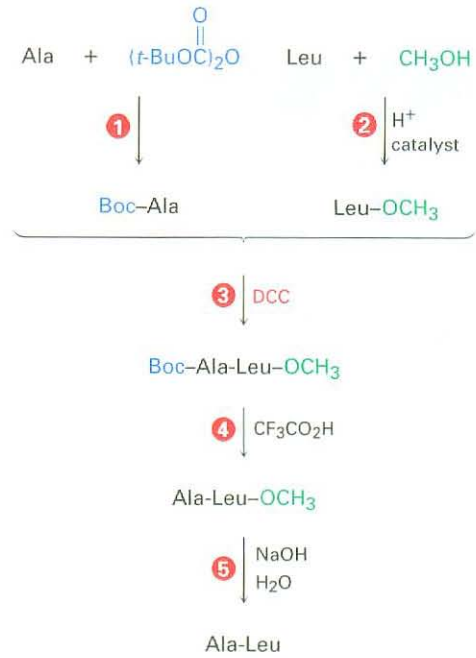


Amino groups are often protected as their *tert*-butoxycarbonyl amide, or Boc, derivatives. The Boc protecting group is introduced by reaction of the amino acid with di-*tert*-butyl dicarbonate in a nucleophilic acyl substitution reaction and is removed by brief treatment with a strong organic acid such as trifluoroacetic acid, CF_3CO_2H.



Thus, five steps are needed to synthesize a dipeptide such as Ala-Leu:

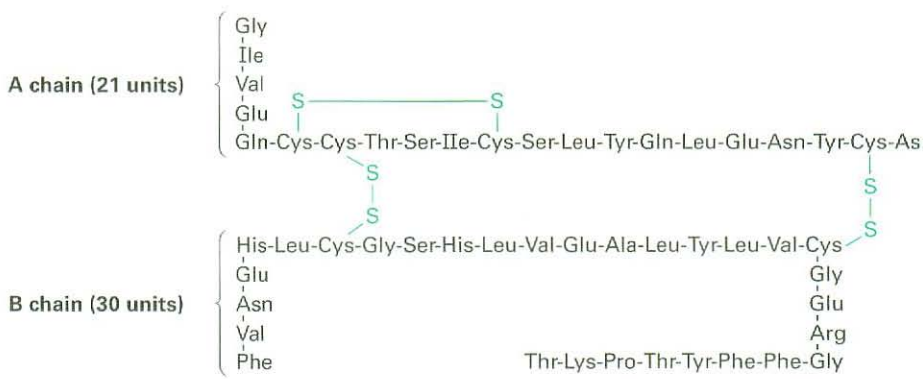
1. The amino group of alanine is protected as the Boc derivative, and
2. the carboxyl group of leucine is protected as the methyl ester.
3. The two protected amino acids are coupled using DCC.
4. The Boc protecting group is removed by acid treatment.
5. The methyl ester is removed by basic hydrolysis.



Frederick Sanger

Frederick Sanger (1918–) was born in Gloucestershire, England, and received his Ph.D. at the University of Cambridge in 1943. After 10 years on the faculty at Cambridge, he joined the Medical Research Council in 1951, where he has remained. In 1958, he was awarded the Nobel Prize in chemistry for his determination of the structure of insulin, and in 1980 he became only the fourth person ever to win a second Nobel Prize. This second prize was awarded for his development of a method for determining the sequence of nucleotides in DNA.

These steps can be repeated to add one amino acid at a time to the growing chain or to link two peptide chains together. Many remarkable achievements in peptide synthesis have been reported, including a complete synthesis of human insulin. Insulin is composed of two chains totaling 51 amino acids linked by two disulfide bridges. Its structure was determined by Frederick Sanger, who received the 1958 Nobel Prize in chemistry for his work.



Insulin

Problem 26.16 Show the mechanism for formation of a Boc derivative by reaction of an amino acid with di-*tert*-butyl dicarbonate.

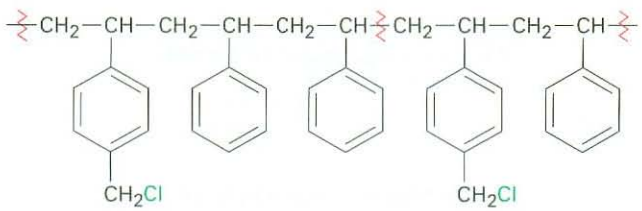
Problem 26.17 Write all five steps required for the synthesis of Leu-Ala from alanine and leucine.

26.8 Automated Peptide Synthesis: The Merrifield Solid-Phase Method

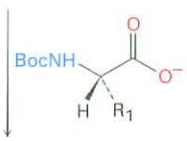
Robert Bruce Merrifield

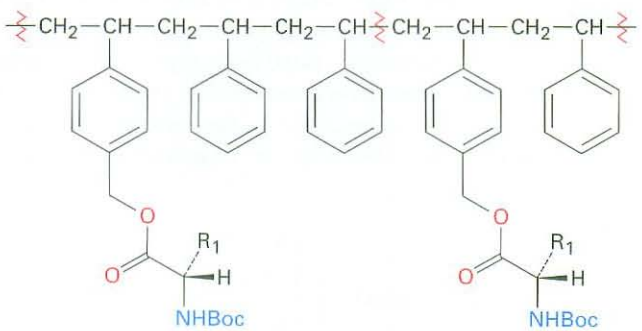
Robert Bruce Merrifield (1921–2006) was born in Fort Worth, Texas, and received his Ph.D. at the University of California, Los Angeles, in 1949. He then joined the faculty at the Rockefeller Institute, where he remained until his death. In 1984, he was awarded the Nobel Prize in chemistry for his development of methods for the automated synthesis of peptides.

The synthesis of large peptide chains by sequential addition of one amino acid at a time is long and arduous, but an immense simplification is possible using the *solid-phase* method introduced by R. Bruce Merrifield at the Rockefeller University. In the Merrifield method, peptide synthesis is carried out with the growing amino acid chain covalently bonded to small beads of a polymer resin rather than in solution. In the standard Merrifield procedure, polystyrene resin is used, prepared so that 1 of every 100 or so benzene rings contained a chloromethyl ($-CH_2Cl$) group, and a Boc-protected C-terminal amino acid is then bonded to the resin through an ester bond formed by S_N2 reaction.



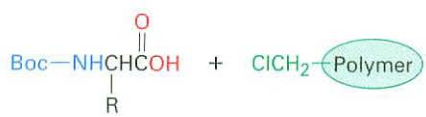
Chloromethylated polystyrene resin



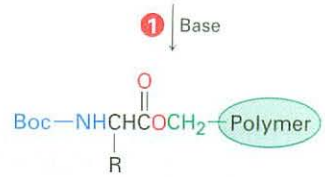


Resin-bound amino acid

With the first amino bonded to the resin, a repeating series of four steps is then carried out to build a peptide.



1 A Boc-protected amino acid is covalently linked to the polystyrene polymer by formation of an ester bond (S_N2 reaction).

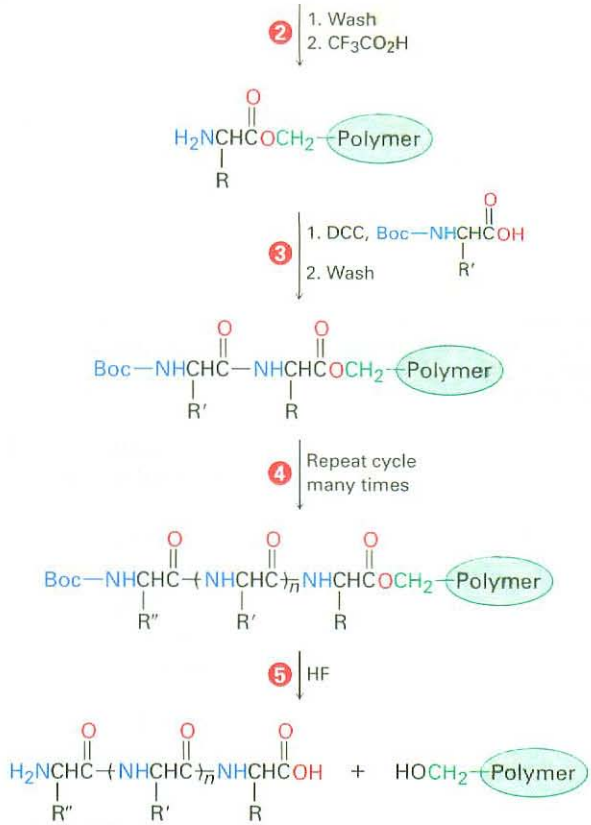


2 The polymer-bonded amino acid is washed free of excess reagent and then treated with trifluoroacetic acid to remove the Boc group.

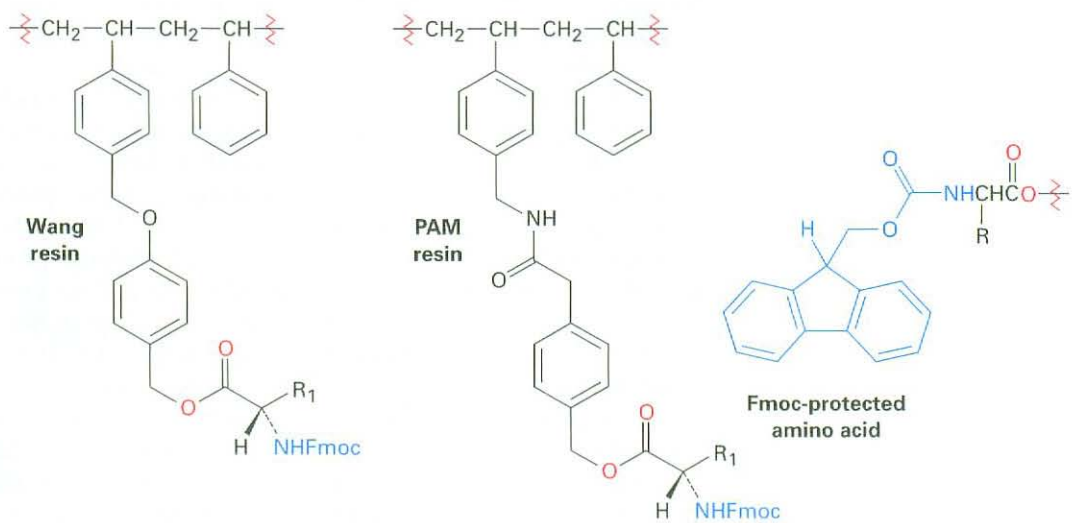
3 A second Boc-protected amino acid is coupled to the first by reaction with DCC. Excess reagents are removed by washing them from the insoluble polymer.

4 The cycle of deprotection, coupling, and washing is repeated as many times as desired to add amino acid units to the growing chain.

5 After the desired peptide has been made, treatment with anhydrous HF removes the final Boc group and cleaves the ester bond to the polymer, yielding the free peptide.



The details of the solid-phase technique have been improved substantially over the years, but the fundamental idea remains the same. The most commonly used resins at present are either the Wang resin or the PAM (phenylacetamidomethyl) resin, and the most commonly used N-protecting group is the fluorenylmethyloxycarbonyl, or Fmoc group, rather than Boc.



Robotic peptide synthesizers are now used to automatically repeat the coupling, washing, and deprotection steps with different amino acids. Each step occurs in high yield, and mechanical losses are minimized because the peptide intermediates are never removed from the insoluble polymer until the final step. Using this procedure, up to 25 to 30 mg of a peptide with 20 amino acids can be routinely prepared.

26.9 Protein Structure

ThomsonNOW Click *Organic Interactive* to **use interactive animations to view aspects of protein structure.**

Proteins are usually classified as either *fibrous* or *globular*, according to their three-dimensional shape. **Fibrous proteins**, such as the collagen in tendons and connective tissue and the myosin in muscle tissue, consist of polypeptide chains arranged side by side in long filaments. Because these proteins are tough and insoluble in water, they are used in nature for structural materials. **Globular proteins**, by contrast, are usually coiled into compact, roughly spherical shapes. These proteins are generally soluble in water and are mobile within cells. Most of the 3000 or so enzymes that have been characterized to date are globular proteins.

Proteins are so large that the word *structure* takes on a broader meaning than it does with simpler organic compounds. In fact, chemists speak of four different levels of structure when describing proteins.

- The **primary structure** of a protein is simply the amino acid sequence.
- The **secondary structure** of a protein describes how *segments* of the peptide backbone orient into a regular pattern.
- The **tertiary structure** describes how the *entire* protein molecule coils into an overall three-dimensional shape.
- The **quaternary structure** describes how different protein molecules come together to yield large aggregate structures.

Primary structure is determined, as we've seen, by sequencing the protein. Secondary, tertiary, and quaternary structures are determined by X-ray crystallography (Chapter 22 *Focus On*) because it's not yet possible to predict computationally how a given protein sequence will fold.

The most common secondary structures are the α helix and the β-pleated sheet. An **α helix** is a right-handed coil of the protein backbone, much like the coil of a telephone cord (Figure 26.5a). Each coil of the helix contains 3.6 amino acid residues, with a distance between coils of 540 pm, or 5.4 Å. The structure is stabilized by hydrogen bonds between amide N–H groups and C=O groups four residues away, with an N–H···O distance of 2.8 Å. The α helix is an extremely common secondary structure, and almost all globular proteins contain many helical segments. Myoglobin, a small globular protein containing 153 amino acid residues in a single chain, is an example (Figure 26.5b).

A **β-pleated sheet** differs from an α helix in that the peptide chain is extended rather than coiled and the hydrogen bonds occur between residues in adjacent chains (Figure 26.6a). The neighboring chains can run either in the same direction (parallel) or in opposite directions (antiparallel), although the antiparallel arrangement is more common and energetically somewhat more favorable. Concanavalin A, for instance, consists of two identical chains of 237 residues, each with extensive regions of antiparallel β sheets (Figure 26.6b).

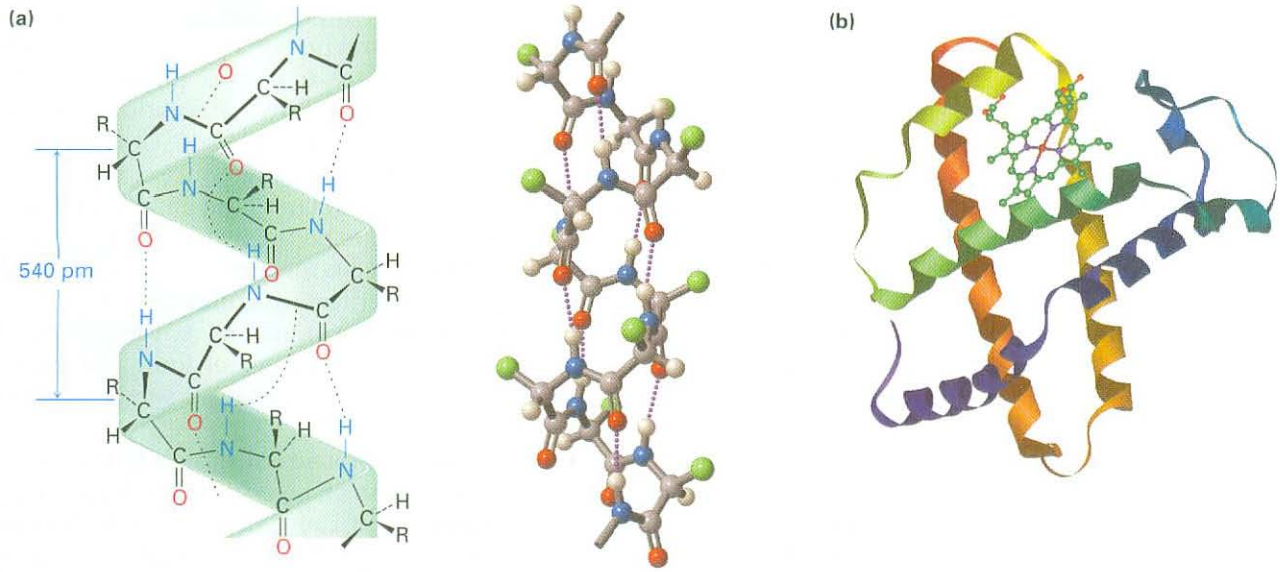


Figure 26.5 **(a)** The α-helical secondary structure of proteins is stabilized by hydrogen bonds between the N–H group of one residue and the C=O group four residues away. **(b)** The structure of myoglobin, a globular protein with extensive helical regions that are shown as coiled ribbons in this representation.

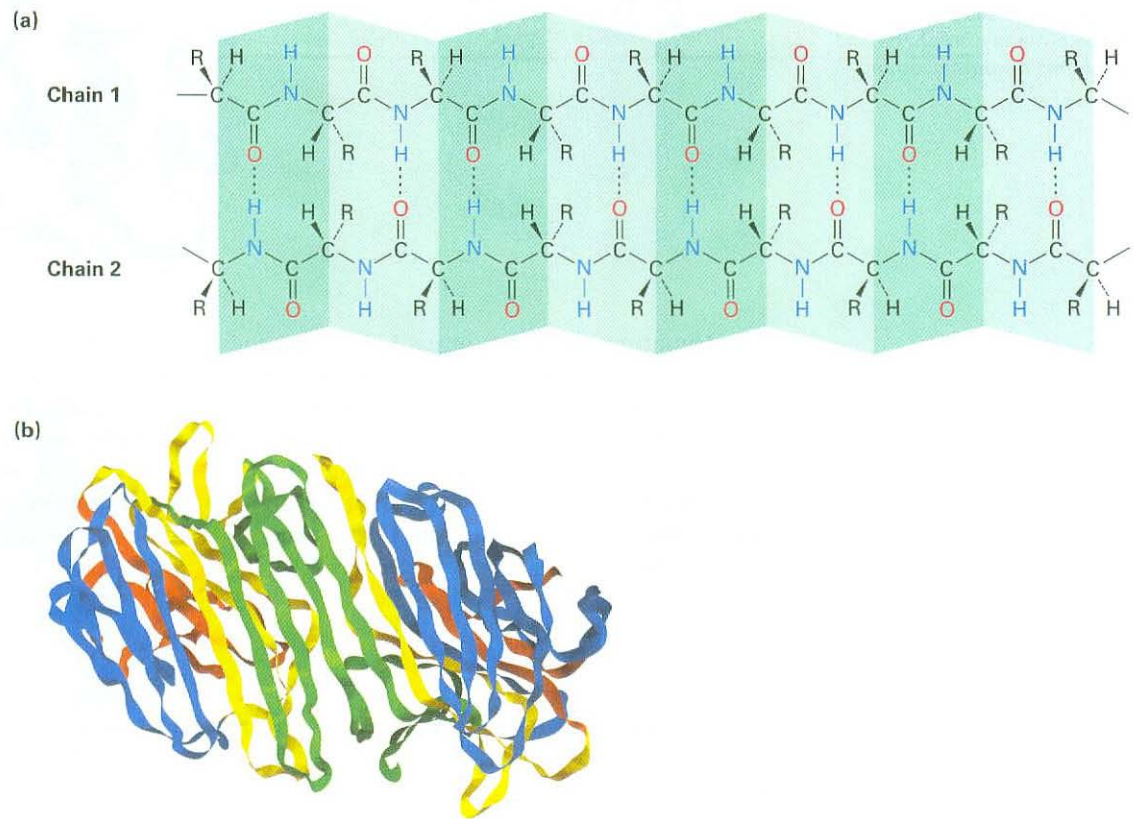


Figure 26.6 **(a)** The β-pleated sheet secondary structure of proteins is stabilized by hydrogen bonds between parallel or antiparallel chains. **(b)** The structure of concanavalin A, a protein with extensive regions of antiparallel β sheets, shown as flat ribbons.

What about tertiary structure? Why does any protein adopt the shape it does? The forces that determine the tertiary structure of a protein are the same forces that act on all molecules, regardless of size, to provide maximum stability. Particularly important are the hydrophilic (water-loving; Section 2.13) interactions of the polar side chains on acidic or basic amino acids. Those acidic or basic amino acids with charged side chains tend to congregate on the exterior of the protein, where they can be solvated by water. Those amino acids with neutral, nonpolar side chains tend to congregate on the hydrocarbon-like interior of a protein molecule, away from the aqueous medium.

Also important for stabilizing a protein's tertiary structure are the formation of disulfide bridges between cysteine residues, the formation of hydrogen bonds between nearby amino acid residues, and the presence of ionic attractions, called *salt bridges*, between positively and negatively charged sites on various amino acid side chains within the protein.

Because the tertiary structure of a globular protein is delicately held together by weak intramolecular attractions, a modest change in temperature or pH is often enough to disrupt that structure and cause the protein to become **denatured**. Denaturation occurs under such mild conditions that the primary structure remains intact but the tertiary structure unfolds from a specific globular shape to a randomly looped chain (Figure 26.7).

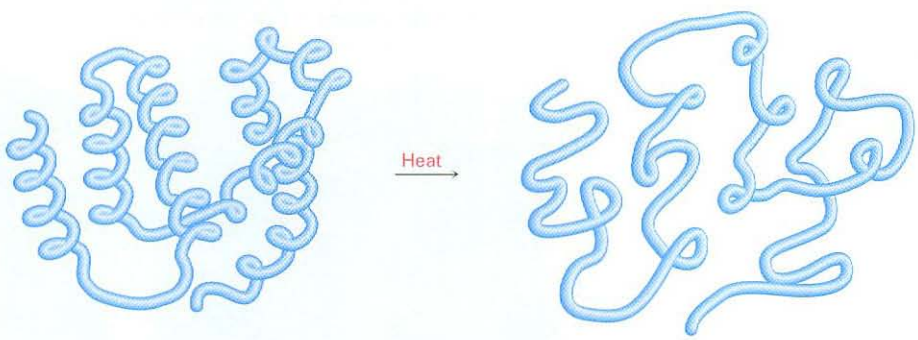


Figure 26.7 A representation of protein denaturation. A globular protein loses its specific three-dimensional shape and becomes randomly looped.

Denaturation is accompanied by changes in both physical and biological properties. Solubility is drastically decreased, as occurs when egg white is cooked and the albumins unfold and coagulate. Most enzymes also lose all catalytic activity when denatured, since a precisely defined tertiary structure is required for their action. Although most denaturation is irreversible, some cases are known where spontaneous *renaturation* of an unfolded protein to its stable tertiary structure occurs. Renaturation is accompanied by a full recovery of biological activity.

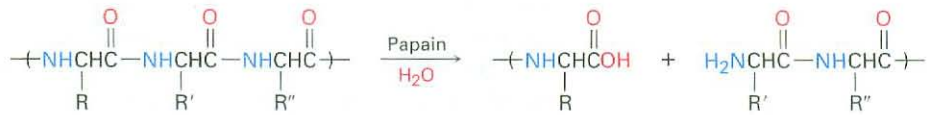
26.10 Enzymes and Coenzymes

An **enzyme**—usually a large protein—is a substance that acts as a catalyst for a biological reaction. Like all catalysts, an enzyme doesn't affect the equilibrium constant of a reaction and can't bring about a chemical change that is otherwise unfavorable. An enzyme acts only to lower the activation energy for a reaction,

thereby making the reaction take place more rapidly. Sometimes, in fact, the rate acceleration brought about by enzymes is extraordinary. Millionfold rate increases are common, and the glycosidase enzymes that hydrolyze polysaccharides increase the reaction rate by a factor of more than 10^{17}, changing the time required for the reaction from millions of years to milliseconds.

Unlike many of the catalysts that chemists use in the laboratory, enzymes are usually specific in their action. Often, in fact, an enzyme will catalyze only a single reaction of a single compound, called the enzyme's *substrate*. For example, the enzyme amylase, found in the human digestive tract, catalyzes only the hydrolysis of starch to yield glucose; cellulose and other polysaccharides are untouched by amylase.

Different enzymes have different specificities. Some, such as amylase, are specific for a single substrate, but others operate on a range of substrates. Papain, for instance, a globular protein of 212 amino acids isolated from papaya fruit, catalyzes the hydrolysis of many kinds of peptide bonds. In fact, it's this ability to hydrolyze peptide bonds that makes papain useful as a meat tenderizer and a cleaner for contact lenses.



Enzymes function through a pathway that involves initial formation of an enzyme–substrate complex E · S, a multistep chemical conversion of the enzyme-bound substrate into enzyme-bound product E · P, and final release of product from the complex.

$$E + S \rightleftharpoons E \cdot S \rightleftharpoons E \cdot P \rightleftharpoons E + P$$

The overall rate constant for conversion of the E · S complex to products E + P is called the **turnover number** because it represents the number of substrate molecules the enzyme turns over into product per unit time. A value of about 10^3 per second is typical.

The rate acceleration achieved by enzymes is due to several factors. Particularly important is the ability of the enzyme to stabilize and thus lower the energy of the transition state(s). That is, it's not the ability of the enzyme to bind the *substrate* that matters but rather its ability to bind and thereby stabilize the *transition state*. Often, in fact, the enzyme binds the transition structure as much as 10^{12} times more tightly than it binds the substrate or products. As a result, the transition state is substantially lowered in energy. An energy diagram for an enzyme-catalyzed process might look like that in Figure 26.8.

Enzymes are classified into six categories depending on the kind of reaction they catalyze, as shown in Table 26.2. *Oxidoreductases* catalyze oxidations and reductions; *transferases* catalyze the transfer of a group from one substrate to another; *hydrolases* catalyze hydrolysis reactions of esters, amides, and related substrates; *lyases* catalyze the elimination or addition of a small molecule such as H_2O from or to a substrate; *isomerases* catalyze isomerizations; and *ligases* catalyze the bonding together of two molecules, often coupled with the hydrolysis

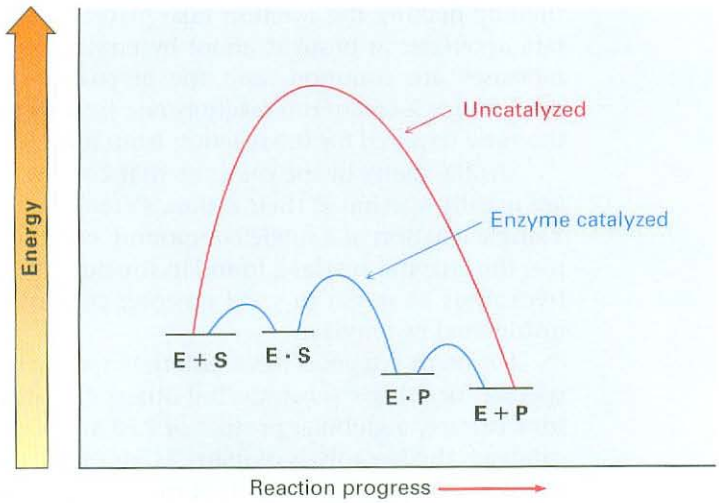


Figure 26.8 Energy diagrams for uncatalyzed (red) and enzyme-catalyzed (blue) processes. The enzyme makes available an alternative, lower-energy pathway. Rate enhancement is due to the ability of the enzyme to bind to the transition state for product formation, thereby lowering its energy.

of ATP. The systematic name of an enzyme has two parts, ending with *-ase*. The first part identifies the enzyme's substrate, and the second part identifies its class. For example, hexose kinase is a transferase that catalyzes the transfer of a phosphate group from ATP to a hexose sugar.

Table 26.2 Classification of Enzymes

Class	Some subclasses	Function
Oxidoreductases	Dehydrogenases	Introduction of double bond
	Oxidases	Oxidation
	Reductases	Reduction
Transferases	Kinases	Transfer of phosphate group
	Transaminases	Transfer of amino group
Hydrolases	Lipases	Hydrolysis of ester
	Nucleases	Hydrolysis of phosphate
	Proteases	Hydrolysis of amide
Lyases	Decarboxylases	Loss of CO_2
	Dehydrases	Loss of H_2O
Isomerases	Epimerases	Isomerization of chirality center
Ligases	Carboxylases	Addition of CO_2
	Synthetases	Formation of new bond

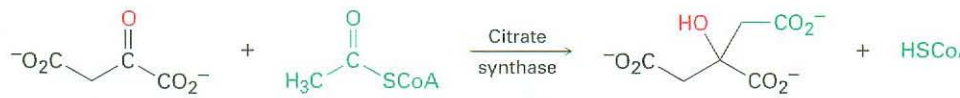
In addition to their protein part, most enzymes also contain a small non-protein part called a *cofactor*. A **cofactor** can be either an inorganic ion, such as Zn^{2+}, or a small organic molecule, called a **coenzyme**. A coenzyme is not a catalyst but is a reactant that undergoes chemical change during the reaction and

requires an additional step to return to its initial state. Many, although not all, coenzymes are derived from *vitamins*—substances that an organism requires for growth but is unable to synthesize and must receive in its diet. Coenzyme A from pantothenate (vitamin B_3), NAD^+ from niacin, FAD from riboflavin (vitamin B_2), tetrahydrofolate from folic acid, pyridoxal phosphate from pyridoxine (vitamin B_6), and thiamin diphosphate from thiamin (vitamin B_1) are examples (Table 26.3 on pages 1044–1045). We'll discuss the chemistry and mechanisms of coenzyme reactions at appropriate points later in the text.

Problem 26.18 To what classes do the following enzymes belong?
(a) Pyruvate decarboxylase (b) Chymotrypsin (c) Alcohol dehydrogenase

26.11 How Do Enzymes Work? Citrate Synthase

Enzymes work by bringing reactant molecules together, holding them in the orientation necessary for reaction, and providing any necessary acidic or basic sites to catalyze specific steps. As an example, let's look at citrate synthase, an enzyme that catalyzes the aldol-like addition of acetyl CoA to oxaloacetate to give citrate. The reaction is the first step in the *citric acid cycle*, in which acetyl groups produced by degradation of food molecules are metabolized to yield CO_2 and H_2O. We'll look at the details of the citric acid cycle in Section 29.7.



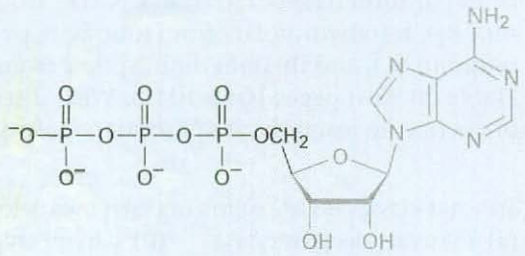
Oxaloacetate **Acetyl CoA** **Citrate**

Citrate synthase is a globular protein of 433 amino acids with a deep cleft lined by an array of functional groups that can bind to oxaloacetate. On binding oxaloacetate, the original cleft closes and another opens up to bind acetyl CoA. This second cleft is also lined by appropriate functional groups, including a histidine at position 274 and an aspartic acid at position 375. The two reactants are now held by the enzyme in close proximity and with a suitable orientation for reaction. Figure 26.9 on page 1046 shows the structure of citrate synthase as determined by X-ray crystallography, along with a close-up of the active site.

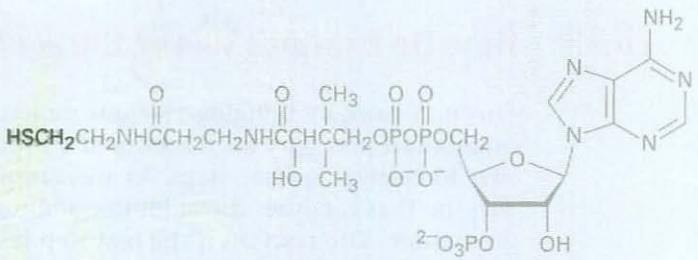
As shown in Figure 26.10 on page 1047, the first step in the aldol reaction is generation of the enol of acetyl CoA. The side-chain carboxyl of an aspartate residue acts as base to abstract an acidic α proton, while at the same time the side-chain imidazole ring of a histidine donates H^+ to the carbonyl oxygen. The enol thus produced then does a nucleophilic addition to the ketone carbonyl group of oxaloacetate. The first histidine acts as a base to remove the $-OH$ hydrogen from the enol, while a second histidine residue simultaneously donates a proton to the oxaloacetate carbonyl group, giving citryl CoA. Water then hydrolyzes the thiol ester group in citryl CoA in a nucleophilic acyl substitution reaction, releasing citrate and coenzyme A as the final products.

Table 26.3 Structures of Some Common Coenzymes

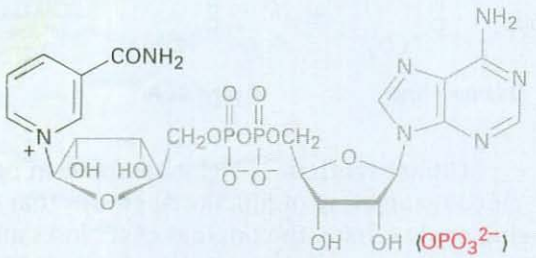
Adenosine triphosphate—ATP (phosphorylation)



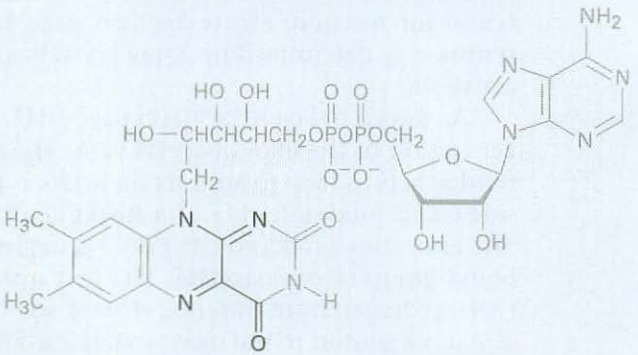
Coenzyme A (acyl transfer)



Nicotinamide adenine dinucleotide—NAD^+ (oxidation/reduction)
($NADP^+$)



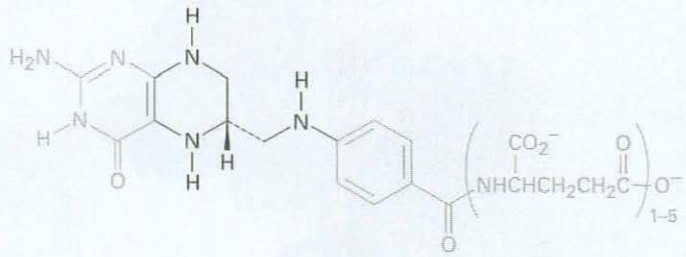
Flavin adenine dinucleotide—FAD (oxidation/reduction)



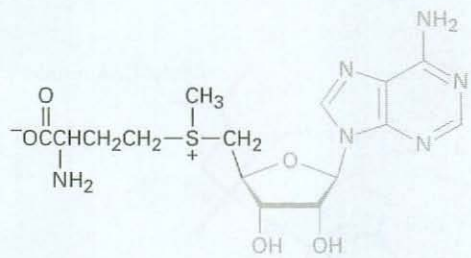
(continued)

Table 26.3 Structures of Some Common Coenzymes *(continued)*

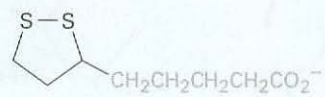
Tetrahydrofolate (transfer of C_1 units)



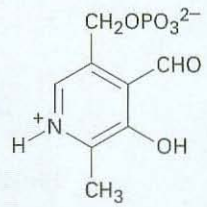
***S*-Adenosylmethionine (methyl transfer)**



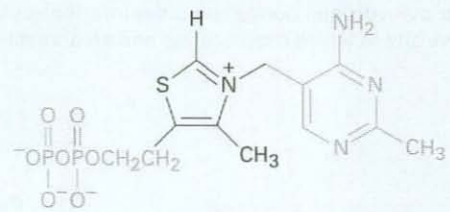
Lipoic acid (acyl transfer)



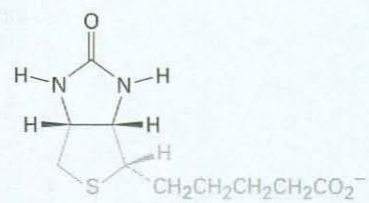
Pyridoxal phosphate (amino acid metabolism)



Thiamin diphosphate (decarboxylation)



Biotin (carboxylation)



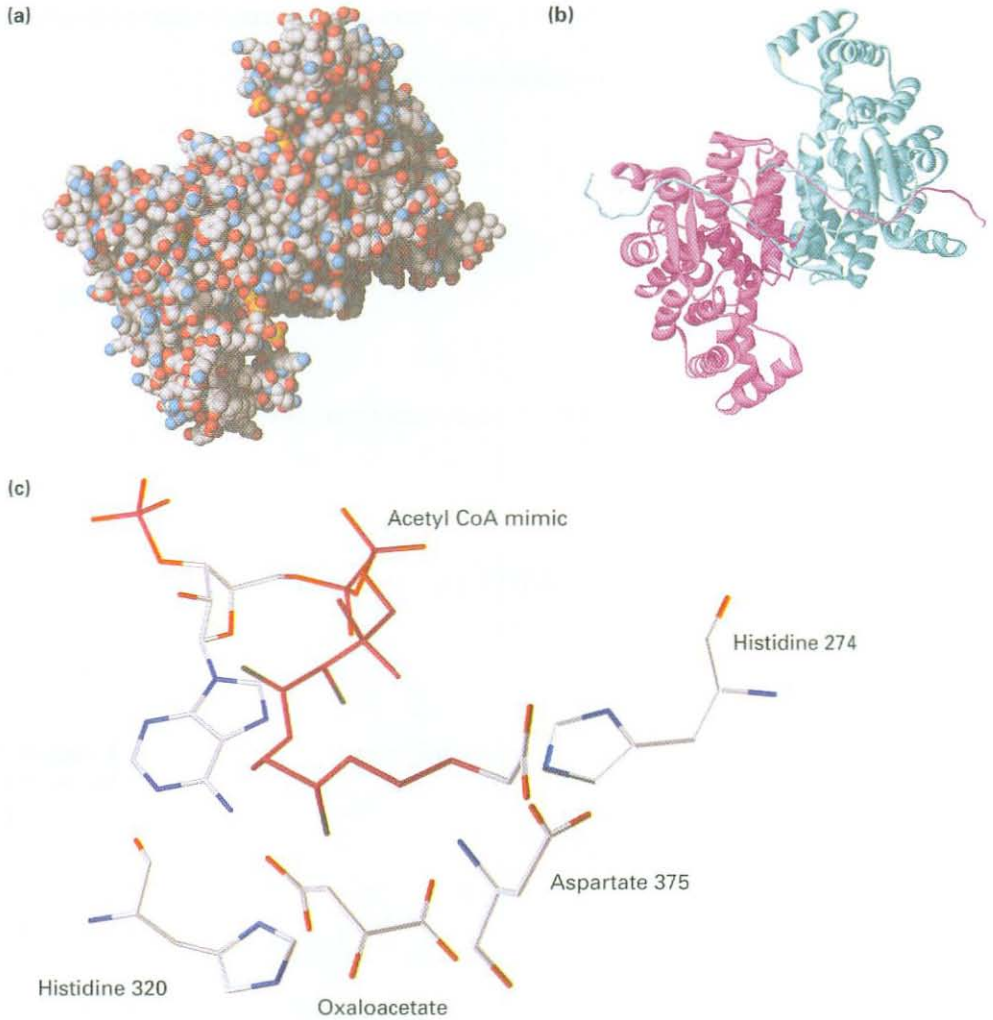
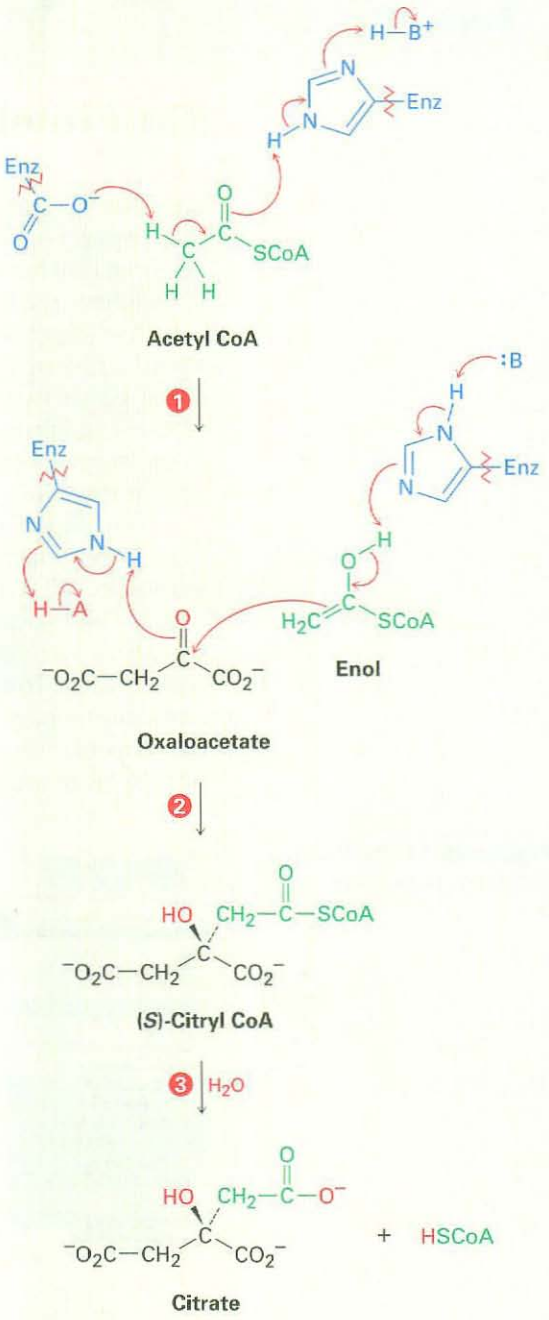


Figure 26.9 X-ray crystal structure of citrate synthase. Part **(a)** is a space-filling model and part **(b)** is a ribbon model, which emphasizes the α-helical segments of the protein chain and indicates that the enzyme is dimeric; that is, it consists of two identical chains held together by hydrogen bonds and other intermolecular attractions. Part **(c)** is a close-up of the active site in which oxaloacetate and an unreactive acetyl CoA mimic are bound.



1. The side-chain carboxylate group of an aspartic acid acts as a base and removes an acidic α proton from acetyl CoA, while the N–H group on the side chain of a histidine acts as an acid and donates a proton to the carbonyl oxygen, giving an enol.

2. A histidine deprotonates the acetyl-CoA enol, which adds to the ketone carbonyl group of oxaloacetate in an aldol-like reaction. Simultaneously, an acid N–H proton of another histidine protonates the carbonyl oxygen, producing (*S*)-citryl CoA.

3. The thioester group of citryl CoA is hydrolyzed by a typical nucleophilic acyl substitution reaction to produce citrate plus coenzyme A.

Figure 26.10 MECHANISM: Mechanism of the addition of acetyl CoA to oxaloacetate to give (*S*)-citryl CoA, catalyzed by citrate synthase.

Focus On . . .

The Protein Data Bank

Enzymes are so large, so structurally complex, and so numerous that the use of computer databases and molecular visualization programs has become an essential tool for studying biological chemistry. Of the various databases available online, the Kyoto Encyclopedia of Genes and Genomes (KEGG) database (http://www.genome.ad.jp/kegg), maintained by the Kanehisa Laboratory of Kyoto University Bioinformatics Center, is useful for obtaining information on biosynthetic pathways of the sort we'll be describing in the next few chapters. For obtaining information on a specific enzyme, the BRENDA database (http://www.brenda.uni-koeln.de), maintained by the Institute of Biochemistry at the University of Cologne, Germany, is particularly valuable.

Perhaps the most useful of all biological databases is the Protein Data Bank (PDB), operated by the Research Collaboratory for Structural Bioinformatics (RCSB). The PDB is a worldwide repository of X-ray and NMR structural data for biological macromolecules. In early 2007, data for more than 40,000 structures were available, and more than 6000 new structures were being added yearly. To access the Protein Data Bank, go to http://www.rcsb.org/pdb/ and a home page like that shown in Figure 26.11 will appear. As with much that is available online, however, the PDB site is changing rapidly, so you may not see quite the same thing.

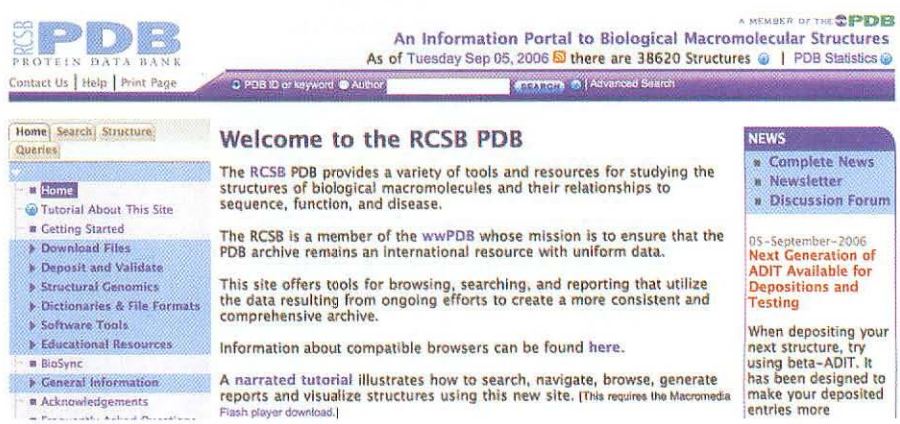


Figure 26.11 The Protein Data Bank home page.

To learn how to use the PDB, begin by running the short tutorial listed near the top of the blue sidebar on the left of the screen. After that introduction, start exploring. Let's say you want to view citrate synthase, the enzyme shown previously in Figure 26.9 that catalyzes the addition of acetyl CoA to oxaloacetate to give citrate. Type "citrate synthase" into the small

(continued)

search window on the top line, click on "Search," and a list of 30 or so structures will appear. Scroll down near the end of the list until you find the entry with a PDB code of 5CTS and the title "Proposed Mechanism for the Condensation Reaction of Citrate Synthase: 1.9 Å Structure of the Ternary Complex with Oxaloacetate and Carboxymethyl Coenzyme A." Alternatively, if you know the code of the enzyme you want, you can enter it directly into the search window. Click on the PDB code of entry 5CTS, and a new page containing information about the enzyme will open.

If you choose, you can download the structure file to your computer and open it with any of numerous molecular graphics programs to see an image like that in Figure 26.12. The biologically active molecule is a dimer of two identical subunits consisting primarily of α-helical regions displayed as coiled ribbons. For now, just click on "Display Molecule," followed by "Image Gallery," to see some of the tools for visualizing and further exploring the enzyme.

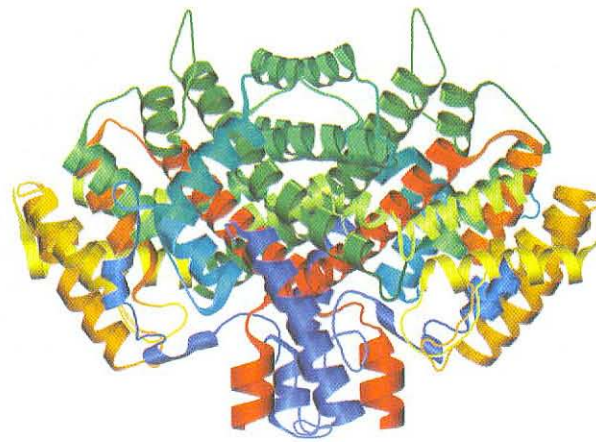


Figure 26.12 An image of citrate synthase, downloaded from the Protein Data Bank.

SUMMARY AND KEY WORDS

α-amino acid, 1020
α helix, 1038
backbone, 1028
β-pleated sheet, 1038
C-terminal amino acid, 1028
coenzyme, 1042
cofactor, 1042
denatured, 1040
Edman degradation, 1031
enzyme, 1040
fibrous protein, 1038

Proteins are large biomolecules made up of **α-amino acid residues** linked together by amide, or *peptide*, bonds. Chains with fewer than 50 amino acids are often called **peptides**, while the term **protein** is reserved for larger chains. Twenty amino acids are commonly found in proteins; all are α-amino acids, and all except glycine have stereochemistry similar to that of L sugars. In neutral solution, amino acids exist as dipolar **zwitterions.**

Amino acids can be synthesized in racemic form by several methods, including ammonolysis of an α-bromo acid, alkylation of diethyl acetamidomalonate, and reductive amination of an α-keto acid. Alternatively, an enantioselective synthesis of amino acids can be carried out using a chiral hydrogenation catalyst.

To determine the structure of a peptide or protein, the identity and amount of each amino acid present is first found by amino acid analysis. The peptide is

globular protein, 1038
isoelectric point, (p*I*), 1024
N-terminal amino acid, 1028
peptide, 1016
primary structure, 1038
protein, 1016
quaternary structure, 1038
residue, 1027
secondary structure, 1038
side chain, 1020
tertiary structure, 1038
turnover number, 1041
zwitterion, 1017

hydrolyzed to its constituent α-amino acids, which are separated and identified. Next, the peptide is sequenced. **Edman degradation** by treatment with phenyl isothiocyanate (PITC) cleaves one residue from the N terminus of the peptide and forms an easily identifiable phenylthiohydantoin (PTH) derivative of the **N-terminal amino acid**. A series of sequential Edman degradations allows the sequencing of a peptide chain up to 50 residues in length.

Peptide synthesis requires the use of selective protecting groups. An N-protected amino acid with a free carboxyl group is coupled to an O-protected amino acid with a free amino group in the presence of dicyclohexylcarbodiimide (DCC). Amide formation occurs, the protecting groups are removed, and the sequence is repeated. Amines are usually protected as their *tert*-butoxycarbonyl (Boc) derivatives, and acids are protected as esters. This synthetic sequence is often carried out by the Merrifield solid-phase method, in which the peptide is esterified to an insoluble polymeric support.

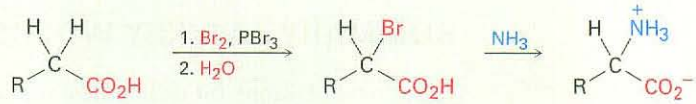
Proteins have four levels of structure. **Primary structure** describes a protein's amino acid sequence; **secondary structure** describes how segments of the protein chain orient into regular patterns—either **α-helix** or **β-pleated sheet**; **tertiary structure** describes how the entire protein molecule coils into an overall three-dimensional shape; and **quaternary structure** describes how individual protein molecules aggregate into larger structures.

Proteins are classified as either globular or fibrous. **Fibrous proteins** such as α-keratin are tough, rigid, and water-insoluble; **globular proteins** such as myoglobin are water-soluble and roughly spherical in shape. Many globular proteins are **enzymes**—substances that act as catalysts for biological reactions. Enzymes are grouped into six classes according to the kind of reaction they catalyze. They function by bringing reactant molecules together, holding them in the orientation necessary for reaction, and providing any necessary acidic or basic sites to catalyze specific steps.

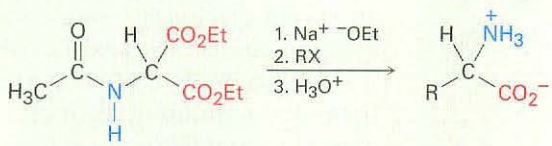
SUMMARY OF REACTIONS

1. Amino acid synthesis (Section 26.3)

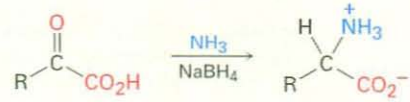
(a) From α-bromo acids



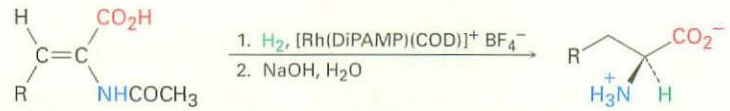
(b) Diethyl acetamidomalonate synthesis



(c) Reductive amination of an α-keto acid

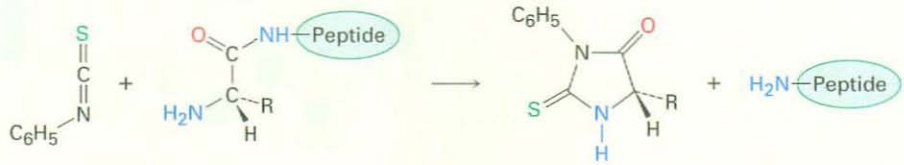


(d) Enantioselective synthesis



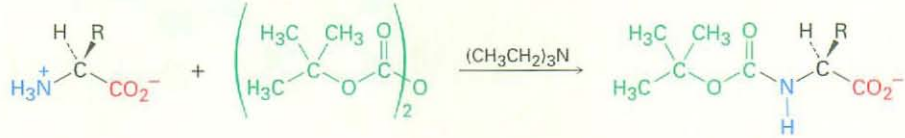
A (*Z*) enamido acid **An (*S*)-amino acid**

2. Peptide sequencing by Edman degradation (Section 26.6)



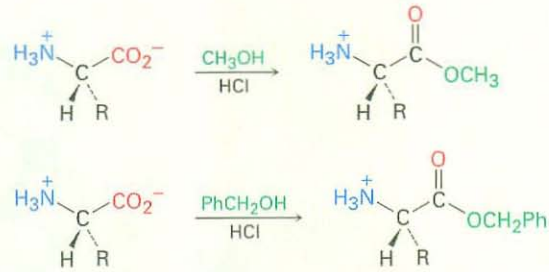
3. Peptide synthesis (Section 26.7)

(a) Amine protection



Boc-protected amino acid

(b) Carboxyl protection



EXERCISES

Organic KNOWLEDGE TOOLS

ThomsonNOW™ Sign in at **www.thomsonedu.com** to assess your knowledge of this chapter's topics by taking a pre-test. The pre-test will link you to interactive organic chemistry resources based on your score in each concept area.

Online homework for this chapter may be assigned in Organic OWL.

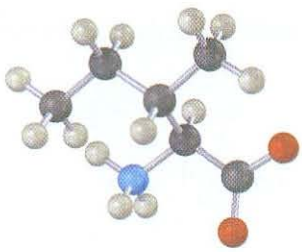
■ indicates problems assignable in Organic OWL.

VISUALIZING CHEMISTRY

(Problems 26.1–26.18 appear within the chapter.)

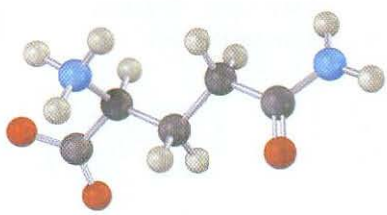
26.19 ■ Identify the following amino acids:

(a)

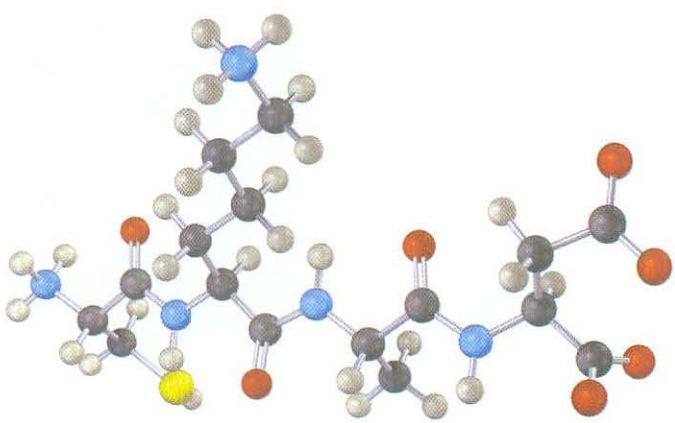


(b)

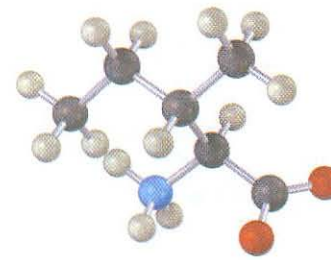
(c)



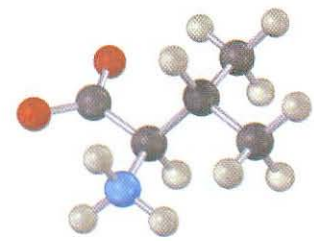
26.20 ■ Give the sequence of the following tetrapeptide (yellow = S):



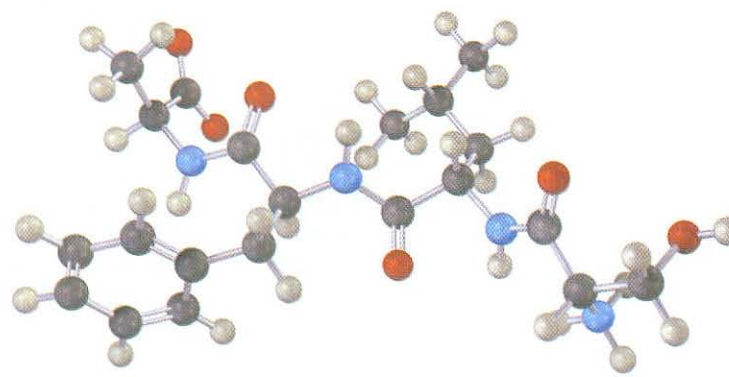
26.21 Isoleucine and threonine (Problem 26.3) are the only two amino acids with two chirality centers. Assign *R* or *S* configuration to the methyl-bearing carbon atom of isoleucine.



26.22 ■ Is the following structure a D amino acid or an L amino acid? Identify it.



26.23 Give the sequence of the following tetrapeptide:



ADDITIONAL PROBLEMS

26.24 Except for cysteine, only *S* amino acids occur in proteins. Several *R* amino acids are also found in nature, however. (*R*)-Serine is found in earthworms, and (*R*)-alanine is found in insect larvae. Draw Fischer projections of (*R*)-serine and (*R*)-alanine. Are these D or L amino acids?

26.25 Cysteine is the only amino acid that has L stereochemistry but an *R* configuration. Make up a structure for another L amino acid of your own creation that also has an *R* configuration.

26.26 Draw a Fischer projection of (*S*)-proline.

26.27 ■ Show the structures of the following amino acids in their zwitterionic forms:
(a) Trp (b) Ile (c) Cys (d) His

26.28 ■ Proline has $pK_{a1} = 1.99$ and $pK_{a2} = 10.60$. Use the Henderson–Hasselbalch equation to calculate the ratio of protonated and neutral forms at pH = 2.50. Calculate the ratio of neutral and deprotonated forms at pH = 9.70.

26.29 Using both three- and one-letter codes for amino acids, write the structures of all possible peptides containing the following amino acids:
(a) Val, Ser, Leu (b) Ser, Leu_2, Pro

26.30 ■ Predict the product of the reaction of valine with the following reagents:
(a) CH_3CH_2OH, acid (b) Di-*tert*-butyl dicarbonate
(c) KOH, H_2O (d) CH_3COCl, pyridine; then H_2O

26.31 ■ Show how you could use the acetamidomalonate method to prepare the following amino acids:
(a) Leucine (b) Tryptophan

26.32 Show how you could prepare the following amino acids using a reductive amination:
(a) Methionine (b) Isoleucine

26.33 Show how you could prepare the following amino acids enantioselectively:
(a) Pro (b) Val

26.34 Serine can be synthesized by a simple variation of the amidomalonate method using formaldehyde rather than an alkyl halide. How might this be done?

26.35 ■ Write full structures for the following peptides:
(a) C-H-E-M (b) P-E-P-T-I-D-E

26.36 ■ Propose two structures for a tripeptide that gives Leu, Ala, and Phe on hydrolysis but does not react with phenyl isothiocyanate.

26.37 Show the steps involved in a synthesis of Phe-Ala-Val using the Merrifield procedure.

26.38 ■ Draw the structure of the PTH derivative product you would obtain by Edman degradation of the following peptides:
(a) I-L-P-F (b) D-T-S-G-A

26.39 Look at the side chains of the 20 amino acids in Table 26.1, and then think about what is *not* present. None of the 20 contain either an aldehyde or a ketone carbonyl group, for instance. Is this just one of nature's oversights, or is there a likely chemical reason? What complications might an aldehyde or ketone carbonyl group cause?

26.40 The α-helical parts of myoglobin and other proteins stop whenever a proline residue is encountered in the chain. Why is proline never present in a protein α-helix?

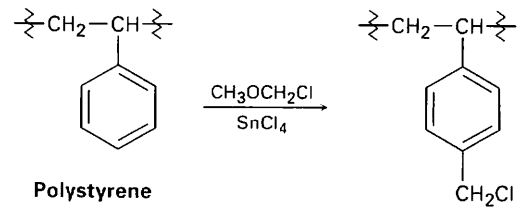
26.41 ■ Which amide bonds in the following polypeptide are cleaved by trypsin? By chymotrypsin?

Phe-Leu-Met-Lys-Tyr-Asp-Gly-Gly-Arg-Val-Ile-Pro-Tyr

26.42 What kinds of reactions do the following classes of enzymes catalyze?
(a) Hydrolases (b) Lyases (c) Transferases

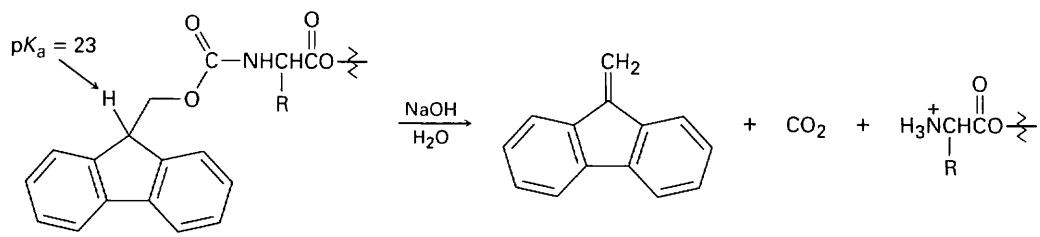
26.43 ■ Which of the following amino acids are more likely to be found on the outside of a globular protein, and which on the inside? Explain.
(a) Valine (b) Aspartic acid (c) Phenylalanine (d) Lysine

26.44 The chloromethylated polystyrene resin used for Merrifield solid-phase peptide synthesis is prepared by treatment of polystyrene with chloromethyl methyl ether and a Lewis acid catalyst. Propose a mechanism for the reaction.



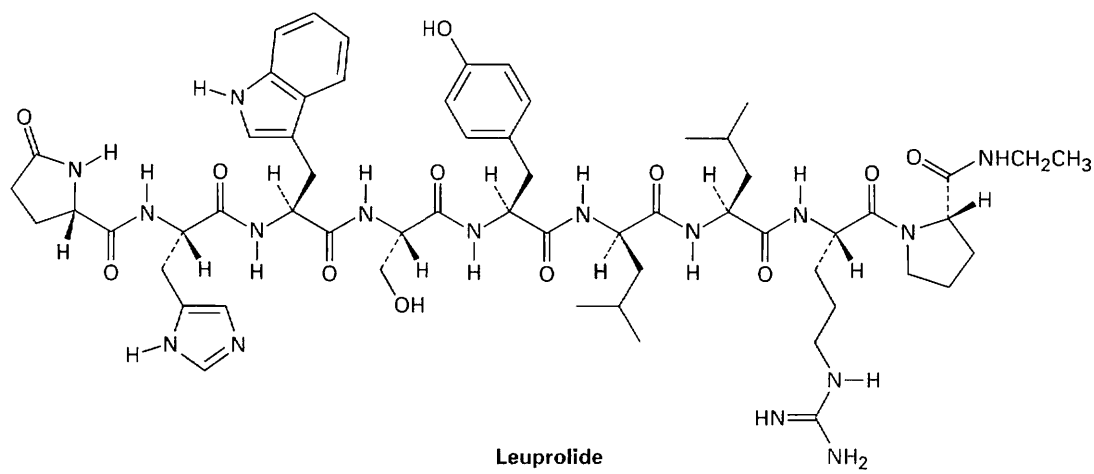
Polystyrene

26.45 An Fmoc protecting group can be removed from an amino acid by treatment with the amine base piperidine. Propose a mechanism.



Fmoc-protected amino acid

26.46 Leuprolide is a synthetic nonapeptide used to treat both endometriosis in women and prostate cancer in men.

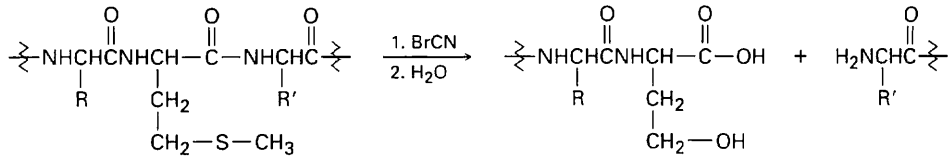


Leuprolide

(a) Both C-terminal and N-terminal amino acids in leuprolide have been structurally modified. Identify the modifications.
(b) One of the nine amino acids in leuprolide has D stereochemistry rather than the usual L. Which one?
(c) Write the structure of leuprolide using both one- and three-letter abbreviations.
(d) What charge would you expect leuprolide to have at neutral pH?

■ Assignable in OWL

26.47 Proteins can be cleaved specifically at the amide bond on the carboxyl side of methionine residues by reaction with cyanogen bromide, BrC≡N.



The reaction occurs in several steps:

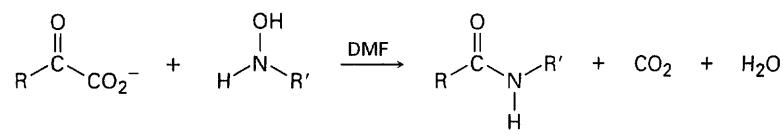
(a) The first step is a nucleophilic substitution reaction of the sulfur on the methionine side chain with BrCN to give a cyanosulfonium ion, $[R_2SCN]^+$. Show the structure of the product, and propose a mechanism for the reaction.

(b) The second step is an internal S_N2 reaction, with the carbonyl oxygen of the methionine residue displacing the positively charged sulfur leaving group and forming a five-membered ring product. Show the structure of the product and the mechanism of its formation.

(c) The third step is a hydrolysis reaction to split the peptide chain. The carboxyl group of the former methionine residue is now part of a lactone (cyclic ester) ring. Show the structure of the lactone product and the mechanism of its formation.

(d) The final step is a hydrolysis of the lactone to give the product shown. Show the mechanism of the reaction.

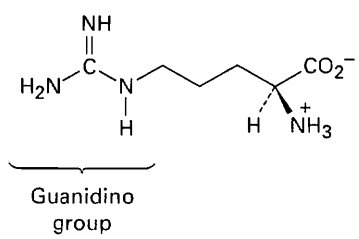
26.48 A clever new method of peptide synthesis involves formation of an amide bond by reaction of an α-keto acid with an *N*-alkylhydroxylamine:



An α-keto acid **A hydroxylamine** **An amide**

The reaction is thought to occur by nucleophilic addition of the *N*-alkylhydroxylamine to the keto acid as if forming an oxime (Section 19.8), followed by decarboxylation and elimination of water. Show the mechanism.

26.49 Arginine, the most basic of the 20 common amino acids, contains a *guanidino* functional group in its side chain. Explain, using resonance structures to show how the protonated guanidino group is stabilized.

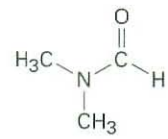


Arginine

Guanidino group

26.50 Cytochrome *c* is an enzyme found in the cells of all aerobic organisms. Elemental analysis of cytochrome *c* shows that it contains 0.43% iron. What is the minimum molecular weight of this enzyme?

26.51 Evidence for restricted rotation around amide CO–N bonds comes from NMR studies. At room temperature, the ^{1}H NMR spectrum of *N,N*-dimethylformamide shows three peaks: 2.9 δ (singlet, 3 H), 3.0 δ (singlet, 3 H), 8.0 δ (singlet, 1 H). As the temperature is raised, however, the two singlets at 2.9 δ and 3.0 δ slowly merge. At 180 °C, the ^{1}H NMR spectrum shows only two peaks: 2.95 δ (singlet, 6 H) and 8.0 δ (singlet, 1 H). Explain this temperature-dependent behavior.



***N,N*-Dimethylformamide**

26.52 ■ Propose a structure for an octapeptide that shows the composition Asp, Gly_2, Leu, Phe, Pro_2, Val on amino acid analysis. Edman analysis shows a glycine N-terminal group, and leucine is the C-terminal group. Acidic hydrolysis gives the following fragments:

Val-Pro-Leu, Gly, Gly-Asp-Phe-Pro, Phe-Pro-Val

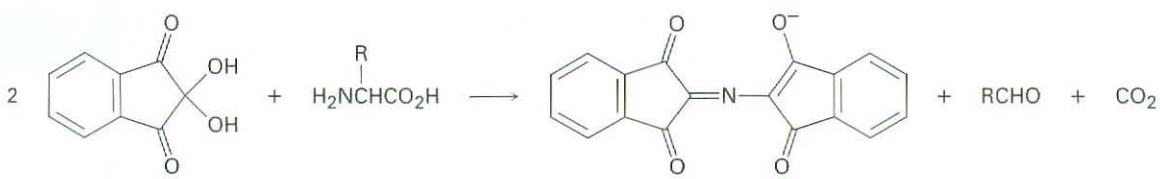
26.53 The reaction of ninhydrin with an α-amino acid occurs in several steps.

(a) The first step is formation of an imine by reaction of the amino acid with ninhydrin. Show its structure and the mechanism of its formation.

(b) The second step is a decarboxylation. Show the structure of the product and the mechanism of the decarboxylation reaction.

(c) The third step is hydrolysis of an imine to yield an amine and an aldehyde. Show the structures of both products and the mechanism of the hydrolysis reaction.

(d) The final step is formation of the purple anion. Show the mechanism of the reaction.



Ninhydrin

26.54 Draw resonance forms for the purple anion obtained by reaction of ninhydrin with an α-amino acid (Problem 26.53).

26.55 Look up the structure of human insulin (Section 26.7), and indicate where in each chain the molecule is cleaved by trypsin and chymotrypsin.

26.56 ■ What is the structure of a nonapeptide that gives the following fragments when cleaved?

Trypsin cleavage: Val-Val-Pro-Tyr-Leu-Arg, Ser-Ile-Arg

Chymotrypsin cleavage: Leu-Arg, Ser-Ile-Arg-Val-Val-Pro-Tyr

26.57 Oxytocin, a nonapeptide hormone secreted by the pituitary gland, functions by stimulating uterine contraction and lactation during childbirth. Its sequence was determined from the following evidence:

1. Oxytocin is a cyclic compound containing a disulfide bridge between two cysteine residues.
2. When the disulfide bridge is reduced, oxytocin has the constitution Asn, Cys_2, Gln, Gly, Ile, Leu, Pro, Tyr.
3. Partial hydrolysis of reduced oxytocin yields seven fragments: Asp-Cys, Ile-Glu, Cys-Tyr, Leu-Gly, Tyr-Ile-Glu, Glu-Asp-Cys, Cys-Pro-Leu.
4. Gly is the C-terminal group.
5. Both Glu and Asp are present as their side-chain amides (Gln and Asn) rather than as free side-chain acids.

What is the amino acid sequence of reduced oxytocin? What is the structure of oxytocin itself?

26.58 *Aspartame,* a nonnutritive sweetener marketed under the trade name Nutra-Sweet (among others), is the methyl ester of a simple dipeptide, Asp-Phe-OCH_3.

(a) Draw the structure of aspartame.

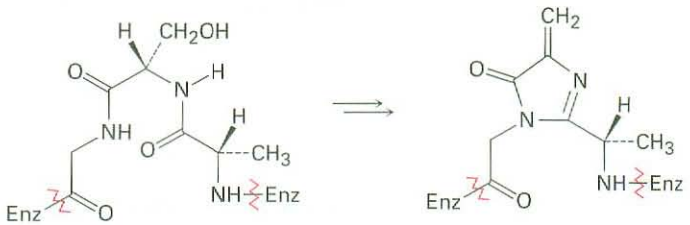
(b) The isoelectric point of aspartame is 5.9. Draw the principal structure present in aqueous solution at this pH.

(c) Draw the principal form of aspartame present at physiological pH = 7.3.

26.59 Refer to Figure 26.2 and propose a mechanism for the final step in the Edman degradation—the acid-catalyzed rearrangement of the ATZ derivative to the PTH derivative.

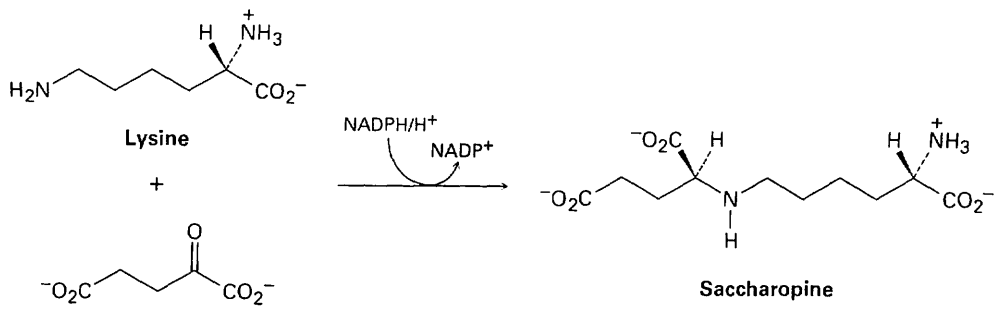
26.60 Amino acids are metabolized by a transamination reaction in which the $-NH_2$ group of the amino acid changes places with the keto group of an α-keto acid. The products are a new amino acid and a new α-keto acid. Show the product from transamination of isoleucine.

26.61 ■ The first step in the biological degradation of histidine is formation of a 4-methylideneimidazol-5-one (MIO) by cyclization of a segment of the peptide chain in the histidine ammonia lyase enzyme. Propose a mechanism.



4-Methylidene-5-imidazolone (MIO)

26.62 The first step in the biological degradation of lysine is reductive amination with α-ketoglutarate to give saccharopine. Nicotinamide adenine dinucleotide phosphate (NADPH), a relative of NADH, is the reducing agent. Show the mechanism.



Lysine + **α-Ketoglutarate**

Saccharopine