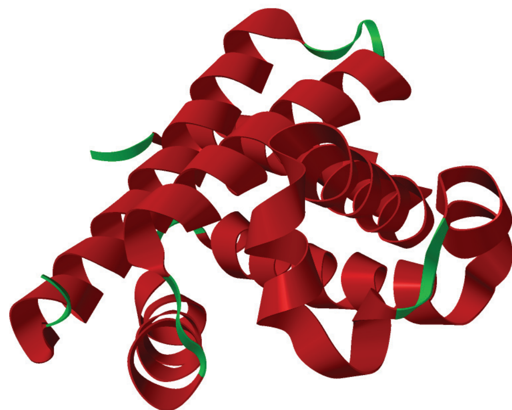
Organic Chemistry: Structure, Mechanism, and Synthesis. Robert J. Ouellette and J. David Rawn. 2014. Elsevier. 1203 pages.
27 - Amino Acids, Peptides, and Proteins. Pages 953-991.

27 Amino Acids, Peptides, and Proteins



Ribbon model of human myoglobin

The study of proteins goes back a long way. In the first half of the nineteenth century, the Dutch chemist Gerardus Johannes Mulder was studying proteins isolated from milk and eggs called albumins. He determined what he thought was the empirical formula for all albumins: $C_{40}H_{62}O_{12}N_{10}$. The Swedish chemist Jons Jacob Berzelius, one of the founders of modern chemistry, suggested that the albumins should be called *proteins*, (Greek, *proteios*, primary) because he thought they might be the most important biological substances. Although Mulder's formula was very wide of the mark—proteins contain thousands of atoms—Berzelius' guess was prophetic. Proteins participate in virtually every cellular process. We will focus narrowly in this chapter upon the structure and properties of the molecules from which proteins are made, the α-amino acids, and upon protein structure.

27.1 THE STRUCTURES OF α-AMINO ACIDS

Proteins are linear polymers of amino acids linked by secondary amide bonds. Careful hydrolysis of a protein releases up to 20 α-amino acids. All proteins from all sources, from Archaebacteria to mammals, are made from the same set of amino acids. Their sequence depends upon an underlying genetic code, which is the same for all organisms with very minor variations. Figure 27.1 shows the general structure of 19 of the 20 amino acids. In this structure, the α-carbon atom is bonded to four different groups every amino acid but one. Thus, the α-carbon is a sterogenic center. The other three groups are a hydrogen atom, a carboxyl group, an amino group, and a fourth R-group commonly called the "side chain." Based upon the configuration of D-glyceraldehyde, all of the chiral amino acids isolated from proteins have an L-configuration.

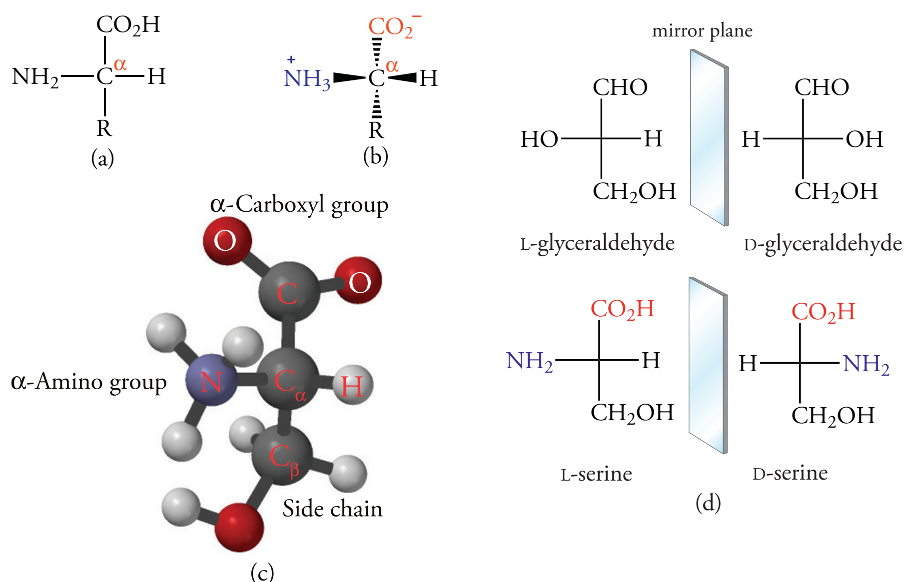


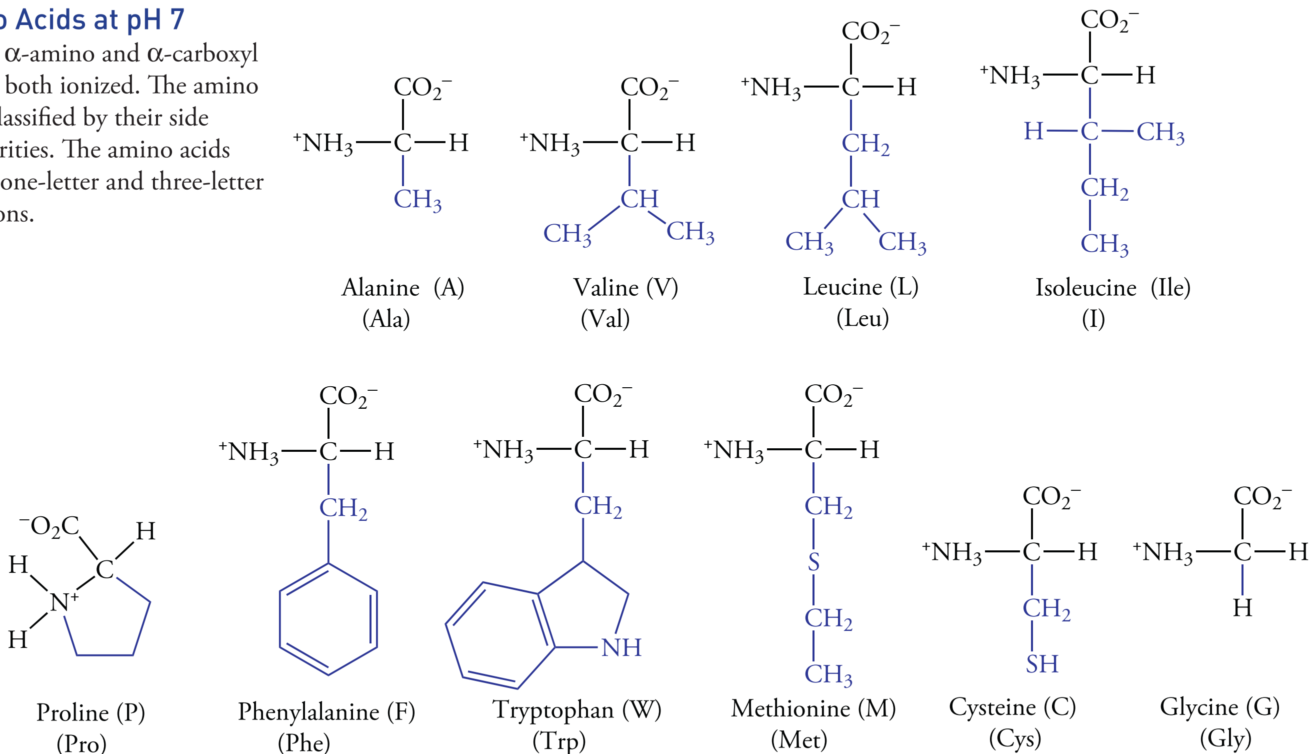
Figure 27.1 Chirality of the α-Amino Acids
(a) Planar projection of an L-amino acid in unionized form. (b) The α-carboxyl group and the α-amino group are ionized in aqueous solution at pH 7. (c) The configuration of the α-amino acids isolated from proteins is opposite to the configuration of the reference compound D-glyceraldehyde. (d) Molecular model of L-serine, whose side chain is a CH_2OH group.

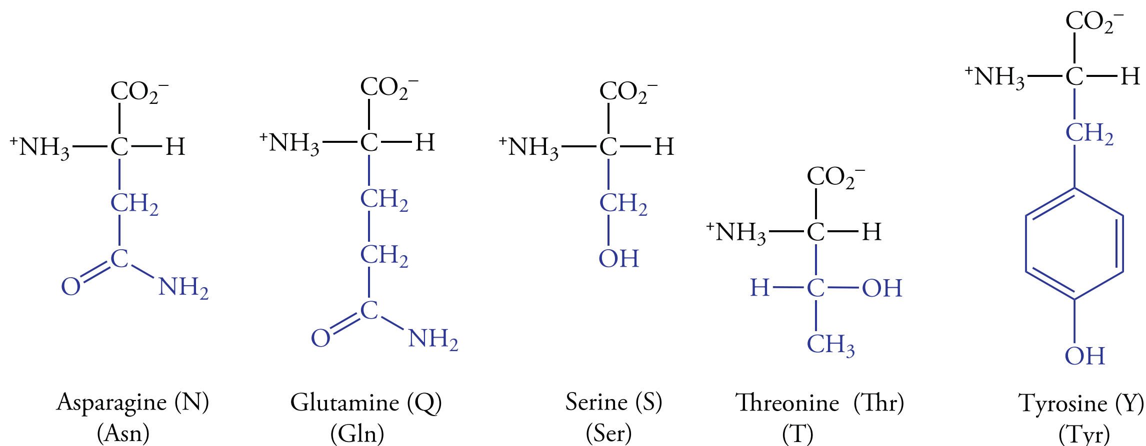
One amino acid has a structure that differs from the one shown in Figure 27.1, namely, proline. Its structure is shown in Figure 27.2.

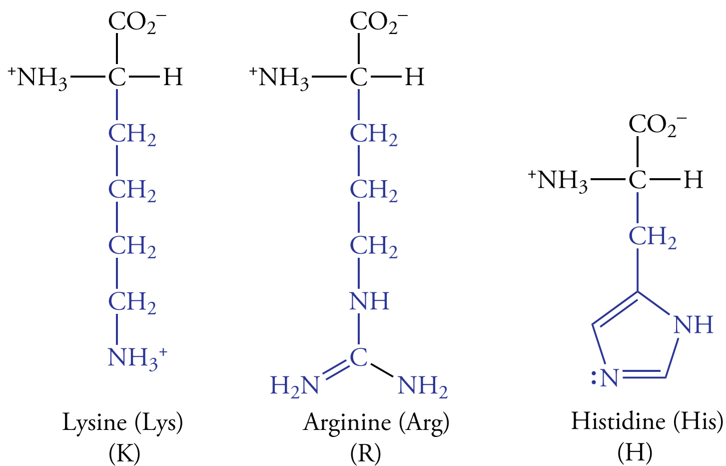
Organic Chemistry. http://dx.doi.org/10.1016/B978-0-12-800780-8.00027-9
Copyright © 2014 Elsevier Inc. All rights reserved.

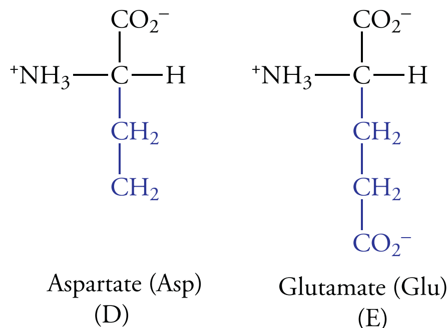
Figure 27.2 Structures of the α-Amino Acids at pH 7

At pH the α-amino and α-carboxyl groups are both ionized. The amino acids are classified by their side chain polarities. The amino acids have both one-letter and three-letter abbreviations.









Problem 27.1

What is the IUPAC name of the naturally occurring amino acid shown below?

Sample Solution

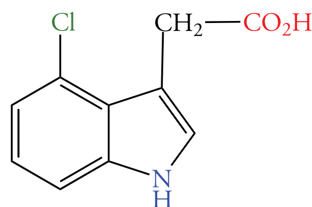
In IUPAC names, the carboxylic acid group takes precedence over amino groups. Therefore, C-1 is the carboxyl group, the parent compound is propanoic acid, and the name is therefore 3-aminopropanoic acid. Its common name is β-alanine, which is a component of pantothenic acid (vitamin B5) and some naturally occurring peptides.

Problem 27.2

Draw the structure of 1-aminocyclopropanecarboxylic acid. This compound undergoes an enzymatic decarboxylation to produce the plant hormone ethene, which is responsible for the initiation of fruit ripening.

Problem 27.3

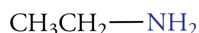
The following carboxylic acid and its methyl ester are found in green peas and many other plants. It is a plant hormone in the auxin family. Which amino acid is a likely (and in fact actual) precursor?



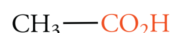
4-chloroindole-3-acetic acid

27.2 ACID–BASE EQUILIBRIA OF α-AMINO ACIDS

The α-amino acids have no *net* charge. However, their properties resemble those of salts rather than uncharged molecules. Amino acids have low solubilities in organic solvents but are moderately soluble in water, unlike most organic compounds of comparable molecular weight. The physical states of amino acids also differ from those of comparable carboxylic acids and amines. For example, ethyl amine is a gas, and acetic acid is a liquid at room temperature. In contrast, glycine is a solid.



ethyl amine
mp –84 ºC

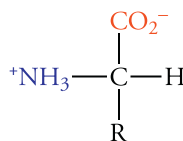


acetic acid
mp 16 ºC

glycine
mp 232 ºC

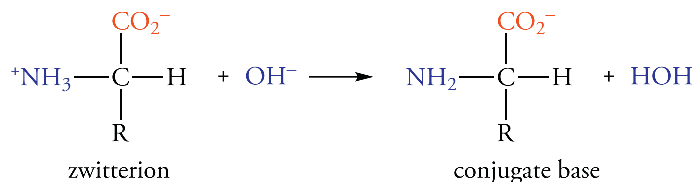
Ionic Form of Amino Acids

When an amino acid dissolves in an aqueous buffer at pH 7, its α-carboxyl group ionizes to give a carboxylate ion, and its α-amino group ionizes to give an ammonium ion. Thus, it exists as a dipolar ion, sometimes called a **zwitterion** (German, *zwitter*, hybrid). In fact, glycine is a dipolar ion in the solid state and that accounts for its high melting point. The dipolar ion acts both as an acid and a base. Thus, it is **amphoteric**.

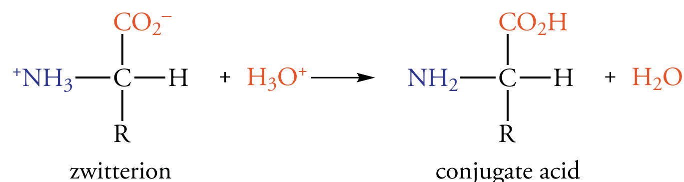


structure of a dipolar ion (zwitterion)

When an amino acid dissolves in basic solution, the carboxylate group exists as an anion and the ammonium ion exists as an unprotonated amino group. This species is the *conjugate base* of the original amino acid. It has a net charge of –1.



When an amino acid dissolves in acid solution, the carboxylate group exists as carboxylic acid group and the amino group exists as an ammonium ion. This species is the *conjugate acid* of the original amino acid. It has a net charge of +1.

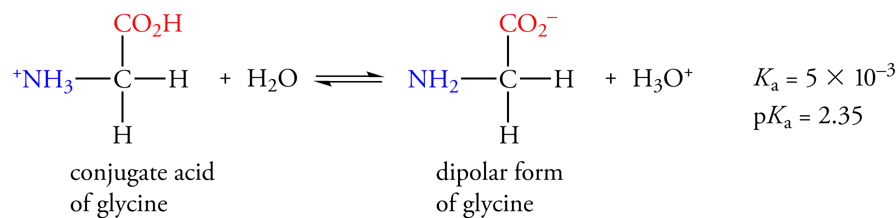


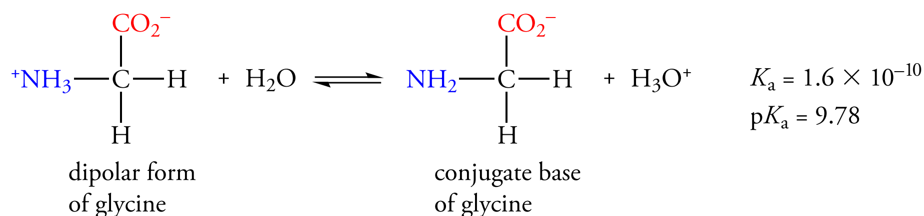
pK_a Values of α-Amino Acids

The pK_a values of the amino acids depend upon the structure of the amino acid. The pK_a values of the α-carboxyl groups range from 1.81 for histidine to 2.58 for phenylalanine. The pK_a values of the α-amino groups range from 8.8 for histidine to 10.78 for tyrosine (Table 27.1).

Table 27.1
pK_a Values of Acidic and Basic Groups in α-Amino Acids

Amino Acid	*α-CO_2H Group*	*α-NH_3^+ Group*	*Side Chain*
Glycine	2.35	9.78	
Alanine	2.35	9.87	
Valine	2.29	9.72	
Leucine	2.33	9.74	
Isoleucine	2.32	9.76	
Methionine	2.17	9.27	
Proline	1.95	10.64	
Phenylalanine	2.58	9.24	
Tryptophan	2.43	9.44	
Serine	2.19	9.44	
Threonine	2.09	9.10	
Cysteine	1.89	10.78	8.53
Tyrosine	2.20	9.11	10.11
Asparagine	2.02	8.80	
Glutamine	2.17	9.13	
Aspartate	1.99	10.00	3.96
Glutamate	2.13	9.95	4.32
Lysine	2.16	9.20	10.80
Arginine	1.82	8.99	12.48
Histidine	1.81	9.15	6.00





When an amino acid dissolves in solution, several species usually exist. When the pH of the solution equals the pK_a of the ionizing group, the concentrations of the conjugate acid and the dipolar form are equal. For example, the pK_a of the carboxyl group of glycine is 2.35, and at pH 2.35, the concentrations of the conjugate acid and the dipolar ion are equal. The pK_a of the ammonium ion of glycine is 9.78, and at pH 9.78, the concentrations of the conjugate base and the dipolar ion are equal. At pH values between 2.35 and 9.78, the dipolar ion is the major ionic form of glycine in solution.

Problem 27.4

What are the structures of the dipolar ion and conjugate base of alanine?

Problem 27.5

In what form does serine exist in 0.1 M HCl?

27.3 ISOIONIC POINT AND TITRATION OF α-AMINO ACIDS

Isoionic Points of Amino Acids

The isoionic point, pH_I, is the pH of the solution at which the concentration of the dipolar ion is a maximum. The relation of pH_I to the concentrations of the various ionic forms of an amino acid are as follows.

1. When pH_I equals pH, the amino acid has no *net* charge, and the dipolar ion is the predominant form of the amino acid.
2. When the pH is greater than pH_I, the conjugate base is the predominant form in solution and the amino acid has a net charge of –1.
3. When the pH is less than pH_I, the conjugate base is the predominant form in solution and the amino acid has a net charge of +1.

The isoionic point of an amino acid equals one-half the sum of the pK_a values of the carboxylate group and the amino group if it does not have an ionizing side chain. For example, the pK_a of the carboxyl group of alanine is 2.4, and the pK_a value of its amino group is 9.9. The isoionic point of alanine is 6.1. Table 27.2 lists the isoionic points of some amino acids.

The isoionic point of the acidic amino acids — aspartic acid and glutamic acid — equals one-half the sum of the pK_a values of the α-CO_2H group and the side chain carboxyl group. Similarly, the isoionic points of the basic amino acids— histidine, lysine, and arginine— equals one-half the sum of the pK_a values of the α-NH_3^+ group and the side chain group. Table 27.2 lists the isoionic points of some amino acids.

The pK_a values of ionizable side chains in proteins often differ from those of the free amino acids. Two factors alter pK_a values. First, α-NH_3^+ and α-CO_2H groups lose their charges when they are linked by peptide bonds in proteins, so they no longer exert strong inductive effects on their neighboring side chains. Second, the position of an ionizable side chain within the three-dimensional structure of a protein can affect its pK_a. For example, the enzyme ribonuclease A has four histidine residues. Each side chain has a slightly different pK_a value because each is in a slightly different environment.

Table 27.2 Isoionic Points

Amino Acid	pH_I
Glycine	5.97
Alanine	6.10
Valine	5.96
Leucine	5.98
Isoleucine	6.02
Methionine	5.74
Proline	6.30
Phenylalanine	5/48
Tryptophan	5.89
Serine	5.68
Threonine	5.60
Cysteine	5.07
Tyrosine	5.66
Asparagine	5.41
Glutamine	5.65
Aspartic acid	2.77
Glutamic acid	3.22
Lysine	9.74
Arginine	10.76
Histidine	7.59

Titration of Amino Acids

The pK_a values of the α-carboxyl and α-amino groups and the pK_I can be determined by titrating the conjugate acid with base. Figure 27.3 shows the titration curve for glycine, which is typical for amino acids without an ionizing side chain. As base is added, the pH increases, and some of the conjugate acid is converted to the dipolar ion. When 0.5 equivalent of base has been added, the concentrations of the α-CO_2H and α-CO_2^- groups are equal, and the pH equals pK_1. After one equivalent of base has been added, the dipolar ion is the major ionic form in solution. When 1.5 equivalents of base have been added, the concentrations of the α-NH_3^+ and α-NH_2 groups are equal, and the pH equals pK_2.

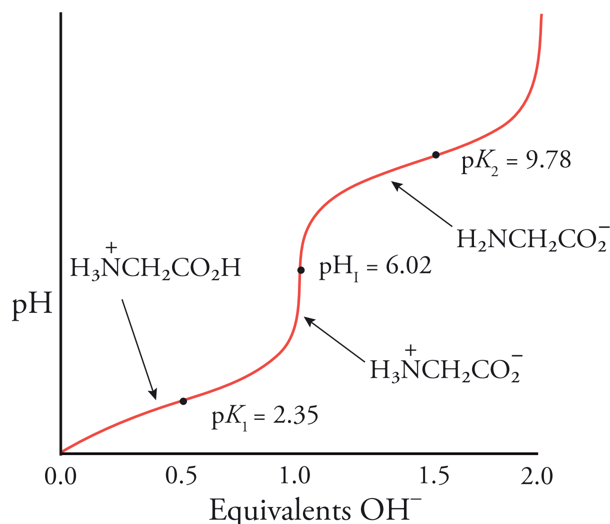


Figure 27.3 Titration Curve of Glycine

Isoionic Points of Proteins

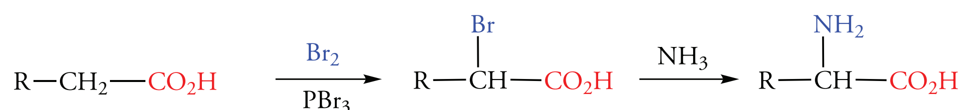
The isoionic point of a protein depends upon its amino acid composition. At its isoionic point, a protein has no net charge, and its solubility is at a minimum. As a consequence, a protein tends to precipitate form solution at its isoionic point. For example, casein, a protein in milk, has a negative charge at pH 6.3. Casein has many glutamic acid and aspartic acid residues. If acid is added to milk, these side chains are protonated, and casein precipitates. Casein is used in making cheese, and it is obtained by adding acid to milk or by adding bacteria that make lactic acid, which has the same effect.

27.4 SYNTHESIS OF α-AMINO ACIDS

Classical methods for the synthesis of amino acids illustrate many of the principles of organic synthesis that we considered in earlier chapters. Modern methods use organometallic catalysts to synthesize chiral amino acids to give high enantiomeric purity.

Amination of α-Halocarboxylic Acids

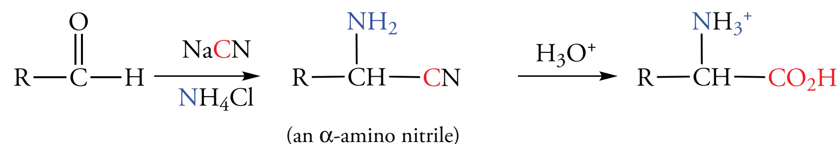
The oldest method of synthesizing α-amino acids is nucleophilic substitution of the halogen of an α-halocarboxylic acid by ammonia. The α-halocarboxylic acid is prepared by treating a carboxylic acid with Br_2 and PBr_3. This reaction, which we considered earlier, is named after its inventors, the Hell-Volhard-Zelinsky reaction. It produces a carboxylic acid with a bromine atom at the α position. The α-bromo compound reacts with ammonia by an S_N2 mechanism to give an α-amino acid.



We recall that direct substitution of alkyl halides by ammonia is often complicated by multiple alkylation of the nitrogen atom (Section 3.7). However, multiple alkylation does not occur in the synthesis of amino acids. The nitrogen atom in the amino acid is less nucleophilic than the nitrogen than in ammonia because the carboxyl group withdraws electrons from it by an inductive effect.

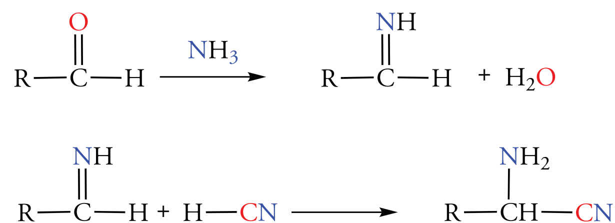
The Strecker Synthesis

A second early method of synthesizing α-amino acids starts with an aldehyde which have one less carbon atom than the desired amino acid. In the first step, an α-amino nitrile is prepared by treating a carboxylic acid with cyanide and ammonia or an ammonium salt. The nitrile is hydrolyzed to a carboxylic acid.



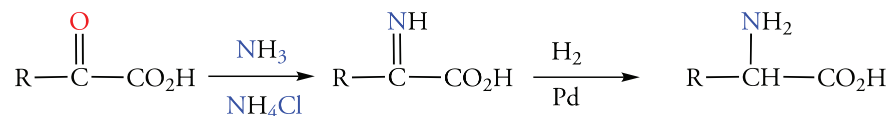
(an α-amino nitrile)

In the first step of this reaction sequence, an imine forms that is in equilibrium with an aldehyde (Section 19.10). The imine then reacts with HCN (the proton in HCN is provided by the ammonium salt). This reaction is similar to the reaction of HCN with a carbonyl group that we discussed in Section 19.2



Reductive Amination

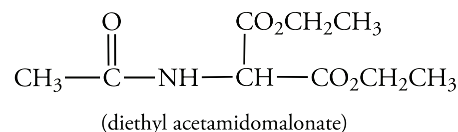
Reductive amination of aldehydes or ketones is an excellent method of producing amines, especially on an industrial scale. To form amino acids on a laboratory scale, the starting material is an α-keto acid. Ammonia reacts with the α-keto acid to give an imine. Reduction of the imine with H_2 in the presence of a palladium catalyst gives the amino acid.



The entire reaction is carried out in a single step with all reagents present. Although a carbonyl group can be reduced by hydrogen gas at high pressure, the imine is more easily reduced, and conditions are chosen to prevent reduction of the carbonyl group in the α-keto acid.

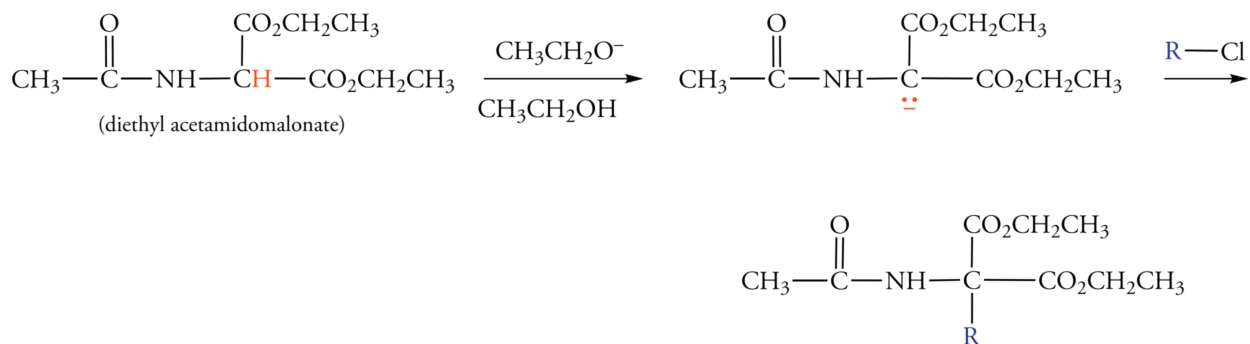
Acetamidomalonate Synthesis

One of the best methods for synthesizing amino acids is based on the chemistry of malonate esters (Section 22.17) and a modification of the Gabriel synthesis of amines (Section 23.7). Diethyl acetamidomalonate has a nitrogen atom bonded to the α-carbon of the malonate ester. This nitrogen eventually becomes the nitrogen of the final amino acid product.

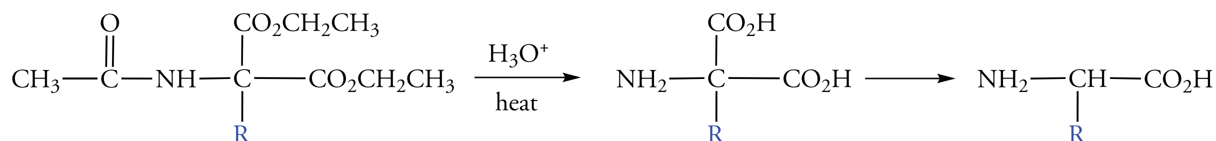


(diethyl acetamidomalonate)

We recall that the α-hydrogen of the malonate ester can be removed by an alkoxide base. In this case, the alkoxide base of choice is ethoxide since the malonate is a diethyl ester. The resulting product is an ester enolate that can be alkylated by an alkyl halide. The alkyl group of the halide is the same as the side chain of the desired amino acid.



Acid-catalyzed hydrolysis of the ester give a dicarboxylic acid. Under the reaction conditions, the amide also hydrolyzes. The resulting malonic acid spontaneously decarboxylates to give the amino acid.



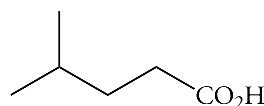
All of the methods for synthesizing amino acids we have described produce racemic mixtures. We recall, however, that racemic mixtures can be resolved by chiral chromatography (Section 8.8). We can also use organometallic reagents as chiral catalysts (Section 17.8) to produce chiral amino acids, as we will see in the next section.

Problem 27.6

What carboxylic acid is required to synthesize leucine using the amination of an α-halo carboxylic acid?

Sample Solution

In this synthesis, a bromine atom is substituted at the α position on the HVZ reaction. The halogen is then replaced by ammonia. Thus, the required acid is a carboxylic acid having the same carbon skeleton as leucine: 4-methylpentanoic acid.



4-methylpentanoic acid

Problem 27.7

What reagents are required for the Strecker synthesis of phenylalanine?

Problem 27.8

What keto acid is required to produce glutamic acid by reductive amination?

Problem 27.9

What reagents are required to synthesize methionine by the acetamidomalonate method?

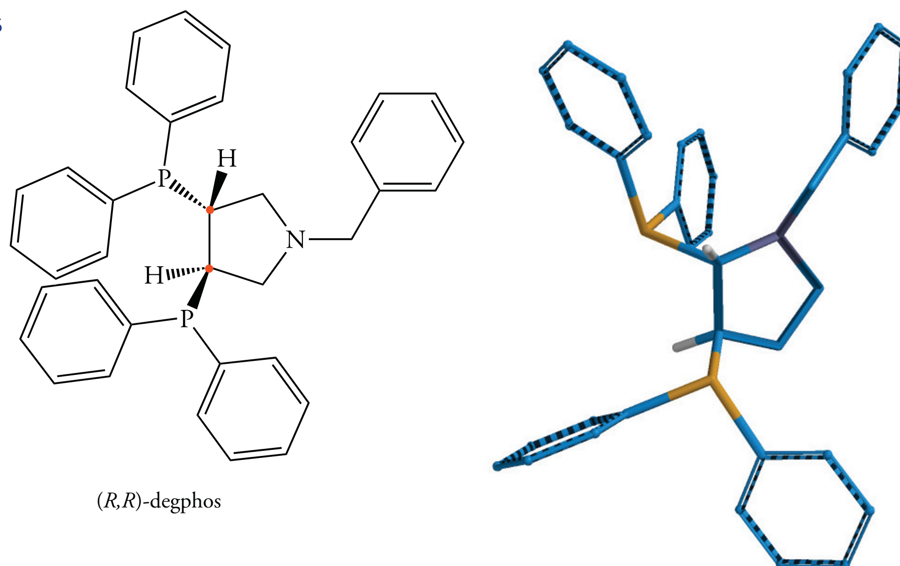
Sample Solution

The α carbon atom, the α amino group, and the carboxyl group are derived from diethyl acetamidomalonate. The side chain of the amino acid, the R group, is derived from an alkyl halide. In this case, the side chain contains a thiomethyl group.

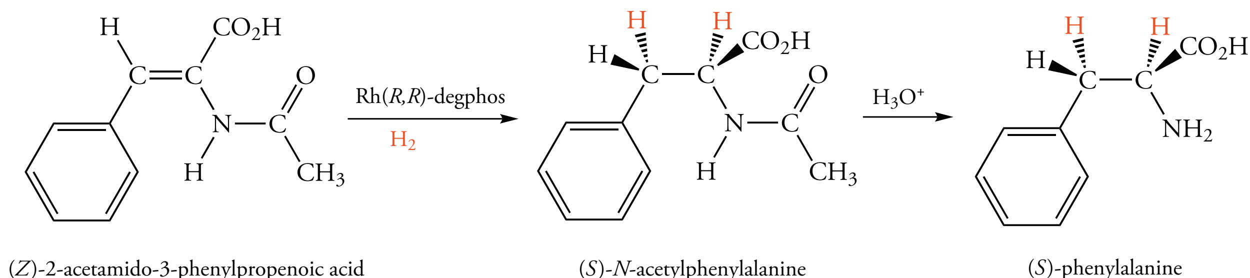
27.5 CHIRAL SYNTHESIS OF α-AMINO ACIDS

We recall that a chiral ruthenium catalyst with binaphthyl ligands can be employed to carry out chiral hydrogenation reactions (Section 17.8). Many chiral ligands have been developed, and one that is widely used for the chiral synthesis of amino acids is called **degphos**. It can be prepared with either an (*R*,*R*) or an (*S*,*S*) configuration. Since the ligand is chiral, the transition states leading to either an *R* or *S* amino acid are diastereomers. Therefore, there energies differ, and the rate of formation of one enantiomer is favored over the other.

Figure 27.4 Structure of (*R*,*R*)-degphos



Let's consider the catalytic reduction of the double bond of the enamide (*Z*)-2-acetamidophenylpropenoic acid by the ruthenium complex of (*R*,*R*)-degphos (Figure 27.4). This reaction leads exclusively to the *S* isomer when hydrogen adds to the *re,re* face of the alkenyl group. That is, in the reaction shown below, hydrogen adds from the bottom face to give (*S*)-*N*-acetylphenylalanine. Hydrolysis of the acetyl group gives *S*-phenylalanine. The net reaction has more than a 99% enantiomeric excess of the *S* isomer.

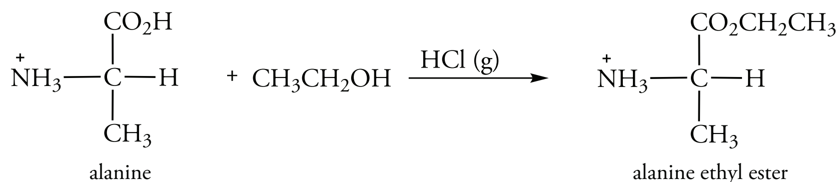


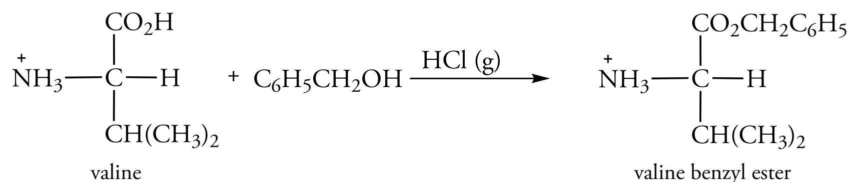
27.6 REACTIONS OF α-AMINO ACIDS

Each functional group of an amino acid undergoes characteristic reactions that we have discussed in previous chapters provided that conditions are chosen to prevent the simultaneous reaction of other functional groups. In this section, we will consider reactions of the α-carboxyl and α-amino groups that are used to synthesize peptides. These reactions are esterification of the carboxyl group and acylation of the amino group.

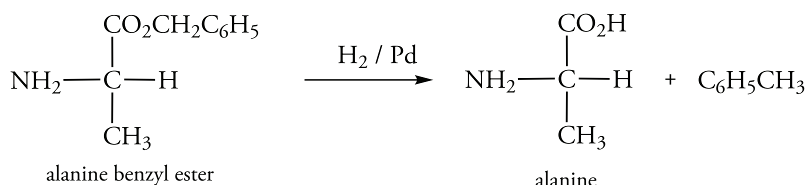
Esterification of the α-Carboxyl Group

The α-carboxyl group can be converted to an ester by reaction of an alcohol with gaseous HCl. Under these conditions, the α-amino group is protonated, so it is unreactive. Ethyl or benzyl esters are commonly prepared. They protect the carboxyl group from other reactions that can be carried out at the amino group.



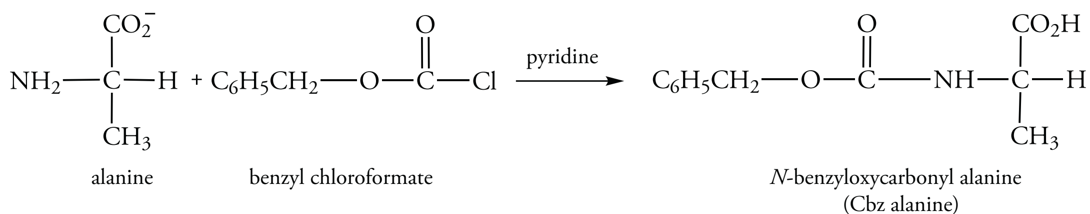


After the carboxyl group has been protected, the amino group can be covalently modified. Then the ethyl or benzyl group can be removed by acid hydrolysis. These groups can also be cleaved by catalytic hydrogenation, a process called **hydrogenolysis**. For example, hydrogenolysis of the benzyl ester of an amino acid released toluene. This reaction occurs under neutral conditions with no competing reactions.

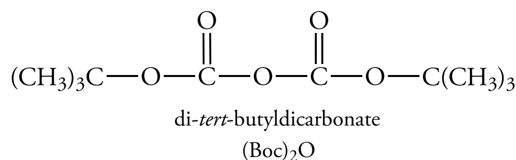


Acetylation of the α-Amino Group

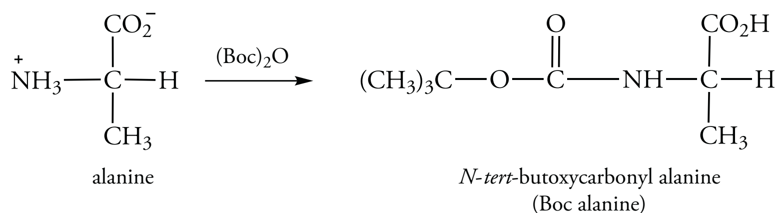
The α-amino group can be converted to an amide by acylation with an anhydride such as acetic anhydride or an acyl chloride. When the amino group is thus protected, it is possible to carry out reactions at the carboxyl group. However, we recall that it is difficult to hydrolyze amides, and the deprotection of an amino group as an *N*-acetyl derivative may well also affect other functional groups. And, since we are particularly interested in reactions that are important in peptide synthesis, these reactions are highly undesirable. We'll consider two reagents that yield easily removed protecting groups. One of these is **benzyl chloroformate**, which acylates an amino group to give a benzyloxycarbonyl (Cbz) derivative. Pyridine is required to convert the amino group to a neutral form that can act as a nucleophilic.



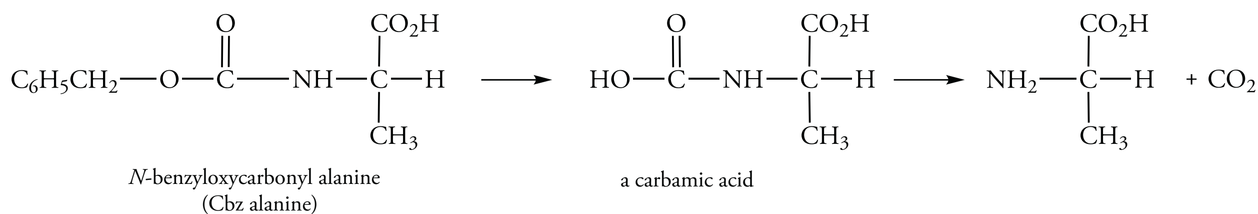
A second easily removed protecting group is ***tert*-butoxycarbonyl**, commonly abbreviated (Boc). The acid chloride of Boc, *tert*-butoxycarbonyl chloride, is highly unstable, so the Boc group is derived from its anhydride, di-*tert*-butyldicarbonate.



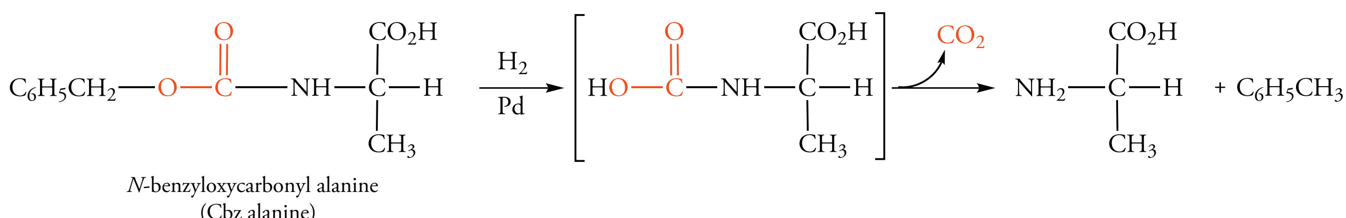
The Boc group is structurally similar to the Cbz group, but it contains a *tert*-butyl group instead of a benzyl group.



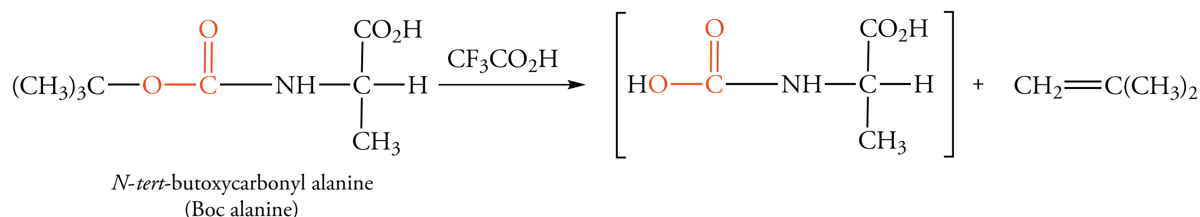
The Boc and the Cbz groups are carbamate esters. A carbamic acid is a highly unstable compound that easily decarboxylates to give an amine. Therefore, when a carbamate ester is converted to a carbamic acid, it decomposes.



Hydrogenolysis of the Cbz derivative of an amino acid releases the amino acid, CO_2 and toluene. The carbamic acid forms as an intermediate in this process.



The Boc group is very sensitive to acid. Treating the Boc derivative of an amino acid with trifluoroacetic acid leads to a carbamic acid intermediate that decarboxylates. The *tert*-butyl group is converted to 2-methylpropene (isobutylene) in the decarboxylation reaction.



27.7 PEPTIDES

Peptide Nomenclature

A peptide is a chain of amino acids in which the α-amino group of one amino acid is bonded to the α-carboxyl group of the next. Thus, each bond linking the amino acids is a secondary amide, called a **peptide bond**. If a peptide made from two amino acids is a **dipeptide**, one made from three is a **tripeptide**, and so forth. As we have seen many times, the prefixes, *di-*, *tri-*, *tetra-*, etc., indicate the number of amino acid units from which the chain is made. Peptides that contain only few amino acids are called **oligopeptides**; peptides with many amino acids are **polypeptides**, a term synonymous with protein.

A peptide has two ends: the end with a free amino group is called the **N-terminal amino acid residue**. The end with a free carboxyl group is called the **C-terminal amino acid residue**. Peptides are named from the N-terminal acid residue to the C-terminal amino acid. Two examples of isomeric dipeptides that contain glycine and alanine are shown below and in Figure 27.5.

glycine residue alanine residue alanine residue glycine residue

$$H{-}NH{-}CH(H){-}C({=}O){-}NH{-}CH(CH_3){-}C({=}O){-}OH$$

glycylalanine (Gly-Ala)

$$H{-}NH{-}CH(CH_3){-}C({=}O){-}NH{-}CH(H){-}C({=}O){-}OH$$

alanylglycine (Ala-Gly)

N-terminal residue C-terminal residue

$$^{+}NH_3{-}CH_2{-}C({=}O){-}NH{-}CH(CH_3){-}C({=}O){-}O^{-}$$

peptide bond

glycylalanine (Gly-Ala or GA)

(a)

N-terminal residue C-terminal residue

$$^{+}NH_3{-}CH(CH_3){-}C({=}O){-}NH{-}CH_2{-}C({=}O){-}O^{-}$$

alanylglycine (Ala-Gly or Ag)

(b)

$$^{+}NH_3{-}C_{\alpha_1}H(R_1){-}C({=}O){-}NH{-}C_{\alpha_2}H(R_2){-}C({=}O){-}NH{-}C_{\alpha_3}(R_3){-}$$

aminoacyl group

(c)

Figure 27.5 Peptide Nomenclature
(a) Structure of glycylalanine. The N-terminal α-amino group and the C-terminal α-carboxyl group and are ionized in aqueous solution at pH 7. Three-letter and one-letter abbreviations for the amino acids are commonly used. (b) Structure of alanylglycine. (c) An aminoacyl group consists of the —NH—CHR—CO— group of each amino acid in the peptide.

The number of isomers of peptides containing one each of *n* different amino acids is equal to *n!*, where

$$n! = 1 \times 2 \times 4 \ldots. (n{-}1) \times n$$

There are thus six possible isomers in a tripeptide that contains three different amino acids. The isomeric tripeptides containing glycine, alanine, and valine are Gly-Ala-Val, Gly-Val-Ala, Val-Ala-Gly, Ala-Gly-Val, and Ala-Val-Gly. For a peptide that contains one each of 20 different amino acids, there are 2,432,902,008,176,640,000 isomers! (That is, on the order of 2×10^{18} isomers.) Proteins often contain hundreds of amino acid residues, and they contain two or more amino acids of the same kind, but the number of possibilities is still astronomically large. Viewed from an evolutionary standpoint, we can say that nature has just begun to experiment.

Biological Functions of Peptides

Cells contain many peptides and polypeptides whose size range from 4 or 5 amino acids to 50 amino acids. (Chains longer than that are designated either as polypeptides or proteins.) They have a huge range of functions. Table 27.3 lists a few examples. These oligopeptides are released in small amounts. Many have very short half-lives. For example, the 14 residue peptide somatostatin, regulates the release of other hormones that play an important role in glucose metabolism, including insulin, glucagon, and secretin. Somatostatin has a half-life of only 4 min.

Hormones exert their effects by binding **receptor proteins**, most of which are located in cell membranes. The complex of the ligand, called an **agonist**, and the receptor protein then affect cell signaling pathways. A very large class of receptor proteins contain seven helical regions of the protein that are located in the membrane. These proteins, which act by similar mechanisms, are called **guanine protein coupled receptor** (GPCR). Subtle changes in the amino acid sequences of the membrane region of the receptors alter the specificity of the response (see Section 27.12).

This is an important area of research in many branches of biological chemistry. However, the details of hormone receptor interactions and cell signalling pathways are beyond the scope of an organic chemistry text.

Enkephalins are peptides that bind specific receptor proteins the brain cells to reduce pain. Enkephalin receptor proteins have a high affinity for opiates, including heroin, morphine, and structurally similar substances. These pain relievers are highly addictive, and the misuse of opiates causes thousands of deaths every year.

Peptides are produced by many tissues. For example, kidney cells secrete angiotensin II, which increases blood pressure by constricting blood vessels. Angotensin II is a potent vasoconstrictor, and the production of excess angiotensin II is responsible for some forms of hypertension.

Oxytocin, which is produced in the pituitary gland, causes the contraction of smooth muscle, including the uterus. It is used clinically to induce labor or to increase the strength of uterine contractions. Vasopressin, another pituitary hormone, regulates the secretion of water by the kidneys and affects blood pressure. The structures of vasopressin and oxytocin differ by only two amino acids. They are cyclic peptides that are linked by a disulfide bond between two cysteine residues.

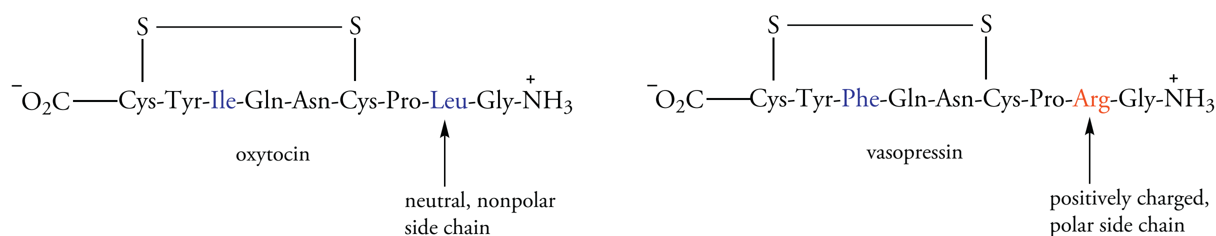


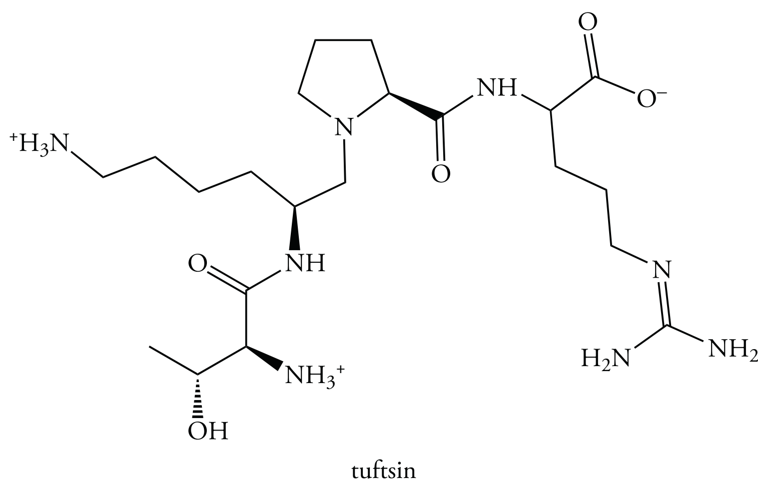
Table 27.3
Peptide Hormones

Hormone	*Amino Acid Residues*	*Receptor*	*Function*
Tuftsin	4	Immune system, cleaved from IgG	Stimulates phagocytosis
Met-enkephalin	5	δ-opioid receptor, (GCRP)	Analgesic activity
Angiotensin II	8	Angiotensin receptor AT_1, GCPR, G_q	Vasoconstriction, increased vasopressin secretion
Oxytocin	8	OXTR, GCRP, G_q	Affects uterine contractions
Vasopressin	8	V_1 receptor, GCRP	An antidiuretic
Bradykinin	9	Bradykinin receptor B_1, GCRP	Produced in response to tissue injury
Somatostatin	14	Somastostatin receptor 1 (human), GCRP	Inhibits release of other hormones
Gastrin	17	Gastrin releasing peptide receptor, GCPR	Leads to pepsin secretion
Secretin	27	Human secretin receptor (GCRP)	Stimulates pancreatic secretions
Glucagon	29	Glucagon receptor, GCRP G_s	Stimulates glucose production from glycogen
Calcitonin	32	Calcitonin receptor (CT), GCRP, G_s, G_q	Decreases calcium level in blood
Relaxin	48	RXPF1, GCRP	Relaxation of pubic joints
Insulin	51	Insulin receptor (IR), transmembrane helix (not GCRP)	Affects blood sugar level

The structures of the two peptides are similar, so it might seem surprising that their functions are so different. However, closer inspection shows that there is one small difference and one major difference in their amino acid composition. Both have a neutral, nonpolar side chain at residue 3, but residue 8 in oxytocin is the nonpolar amino acid leucine, whose side chain is a *sec*-butyl group, but residue 8 in vasopressin is arginine, whose side chain has a positive charge. As a result, the for oxytocin has a very low affinity for vasopressin and the receptor for vasopressin has a very low affinity for oxytocin. Since they bind different receptors, they have different functions.

Problem 27.10

(a) Identify each of the amino acids of tuftsin. (b) Write the name of tuftsin as three-letter abbreviations. (c) Write the name of tuftsin without abbreviations.



tuftsin

Problem 27.11

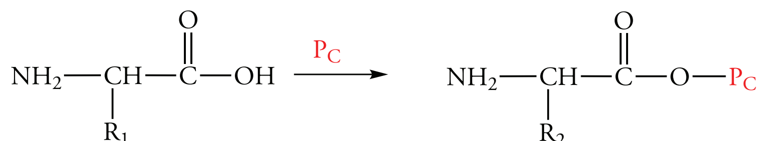
(a) How many isomeric peptides exist that contain one alanine and two glycine residues? (c) Write their names as three-letter abbreviations.

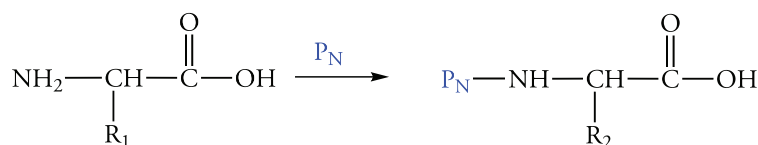
27.8 OVERVIEW OF PEPTIDE SYNTHESIS

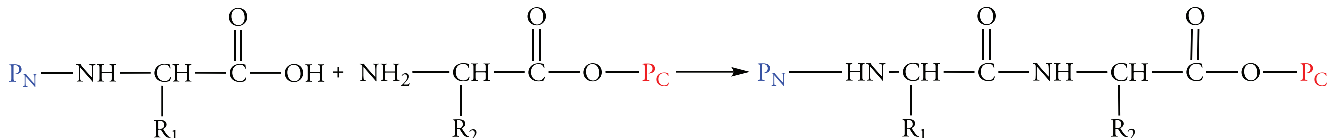
Peptides have such a wide range of physiological functions that their study is an important part of biological chemistry. Many companies provide peptides for researchers who lack the facilities of an organic chemistry laboratory and the training of synthetic organic chemists. Thus, peptide synthesis is a lucrative part of the biotechnology industry. In this section, we will consider the basic reactions required to synthesize a peptide.

We cannot simply react two amino acids under conditions that allow formation of a peptide bond if we wish to synthesize a specific dipeptide because an amino acid has two reactive positions, the α-amino group and the α-carboxyl group. For instance, reacting glycine with alanine would yield Gly-Gly, Gly-Ala, and Ala-Gly. Also, the amino acids and peptides in the reaction mixture can continue to react to give a host of other products.

The synthesis of a dipeptide having a specific sequence requires modifying both amino acids. One amino acid is protected at its carboxyl group— by a reagent we will call P_C — leaving the amino group free. The second amino acid is protected at its amino group— by a reagent we will call P_N — leaving the carboxyl group available for peptide bond formation. Only one condensation reaction is then possible.





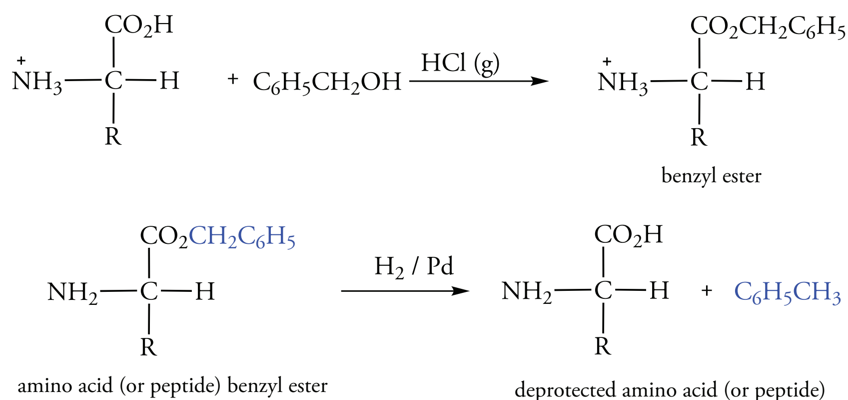


This method of peptide synthesis has several requirements.

1. The carboxyl group of one amino acid must be protected.
2. The amino group of one amino acid must be protected.
3. A reagent must be chosen to form the peptide bond.
4. Conditions must be chosen to free one protecting group selectively so that the sequence can be repeated.

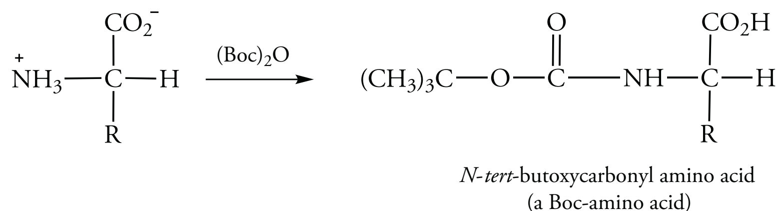
Protecting the Carboxyl Group

The carboxyl group can be protected by converting it to a benzyl ester. We also saw that the benzyl ester can be removed by hydrogenolysis without affecting other functional groups. Hence, the carboxyl group can be easily deprotected at the end of the synthesis.

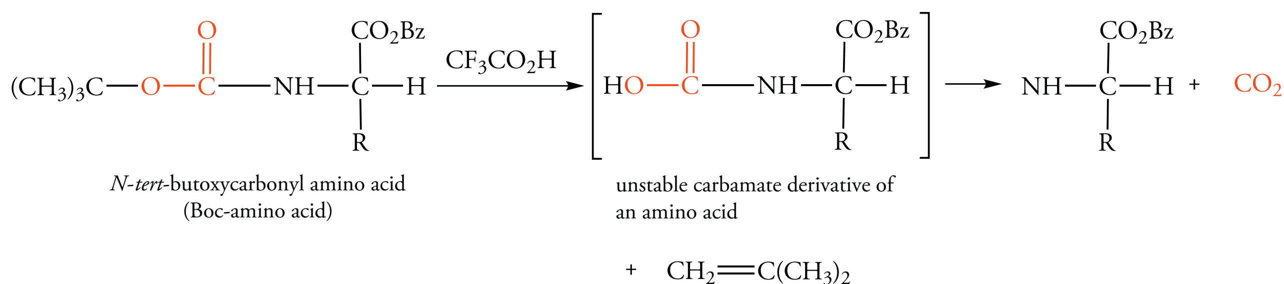


Protecting the Amino Group

We also recall that several protecting groups have been developed to protect the amino terminus of an amino acid; the *tert*-butoxycarbonyl (Boc) derivative is one example. Reaction of an amino acid with di-*tert*-butyl dicarbonate gives a Boc-amino acid.

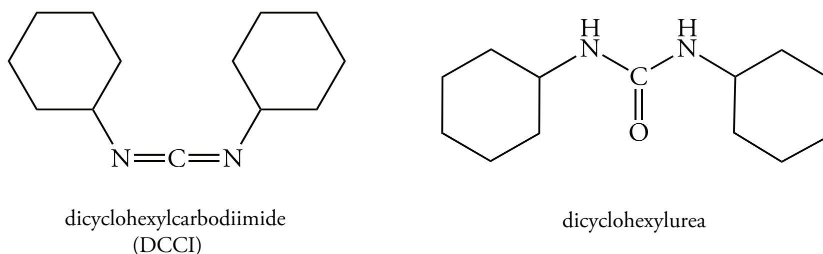


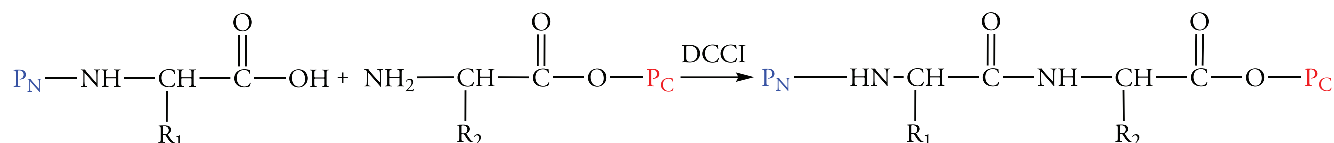
Note that the carbonyl group of the Boc group is bonded to both an oxygen atom and a nitrogen atom. This functional group is a carbamate, which is more easily hydrolyzed than amides or esters. The Boc group can be removed with trifluoroacetic acid. Both the amide bonds of a peptide and the protected carboxyl group are unaffected by this reaction. The by-products of the reaction are CO_2 and 2-methylpropene (isobutylene). Both are gases that escape from the reaction, pulling it to completion.



Peptide Bond Synthesis

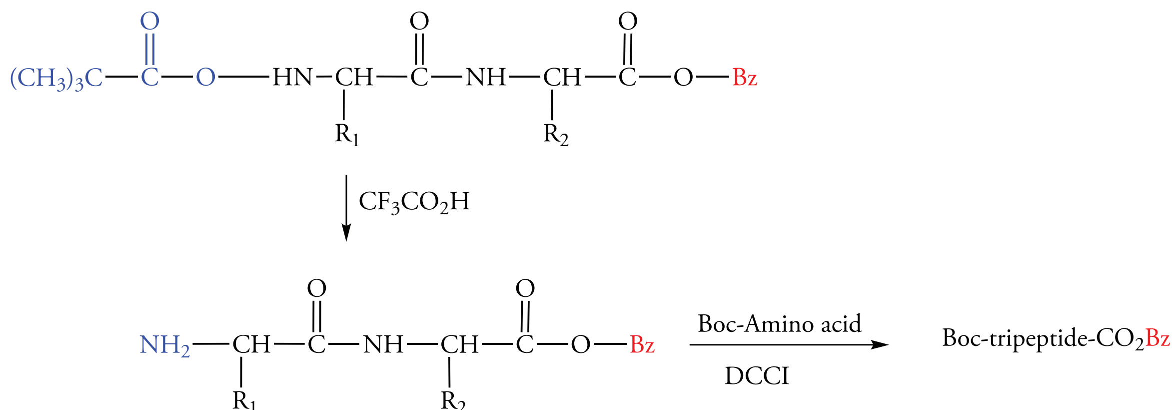
The protecting groups of both the amino and the carboxyl group are sensitive to acids and bases, so the condensation of two protected amino acids to form a peptide bond must be carried out under neutral conditions. It turns out that a reagent called dicyclohexylcarbodiimide (DCCI) causes condensation of two amino acids by removing the elements of water. The reaction has a very high yield, and no other functional groups are modified. The by-product of the reaction is dicyclohexylurea.





Polypeptide Synthesis

The dipeptide that is protected at both the carboxyl and amino terminus is deprotected by hydrolysis of the Boc group at the N-terminal amino acid. The dipeptide can only react at the free amino group. Reaction with another Boc-amino acid and DCCI yields a tripeptide. At the end of the synthesis, the final peptide is released by hydrolysis with base.



Experimental Limitations

One limitation of every synthetic method is mechanical losses that result from the isolation and purification of products. The product, in this case a peptide, must be separated from remnants of protecting groups, coupling agents, and by-products. Thus even reactions that yield a single regiospecific product may not produce a high isolated yield. The problem is compounded when many consecutive reactions are required in peptide synthesis. For example, a synthetic sequence required to prepare a peptide that contains 25 amino acid residues required a total of 100 steps. If the product of each step is isolated in 90% yield, the final yield would be extremely small because the amount of each product formed is controlled by the product that formed in the previous reaction. We obtain the fraction of product by multiplying the yields for each step. Thus, after 25 steps, the yield would be an infinitessimal 0.0026%.

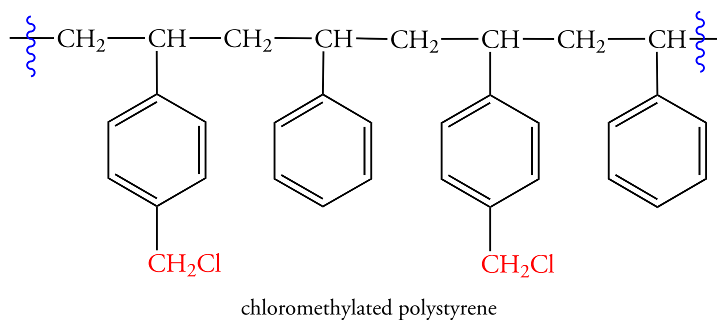
$$(0.90)^{100} = 0.000026$$

The problem becomes even worse as the number of amino acid residues in a polypeptide increases. For relatively large peptides that contain in excess of chemical synthesis by conventional means is clearly out off the question.

27.9 SOLID-PHASE PEPTIDE SYNTHESIS

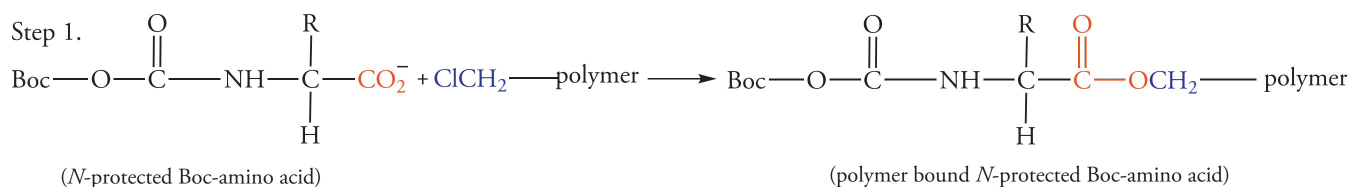
The solid-phase synthesis of polypeptides was developed by R. B. Merrifield at Rockefeller University beginning in the 1960s. This method has undergone continuous development and refinement in the ensuing decades, but the overall method has remained the same. The solid-state method uses a polymer with reactive sites that chemically bind to the developing peptide chain. This technique circumvents the problems associated with low yields due to separation and purification. Because the polymer is very insoluble, it can be filtered and washed without mechanical losses. The developing protein chain attached to the polymer is "dangling" off the polymer and is in contact with any reagents added in solution. As a result, a large number of steps can be carried out on the peptide, and the product remains linked to a solid that can be separated from impurities in solution.

One polymer used in solid-phase synthesis is an addition polymer of styrene in which some of the benzene rings have a —CH_2Cl group. As few as 1 out of 10 rings bear chloromethyl groups. The general structure of the chloromethylated polystyrene is shown below.

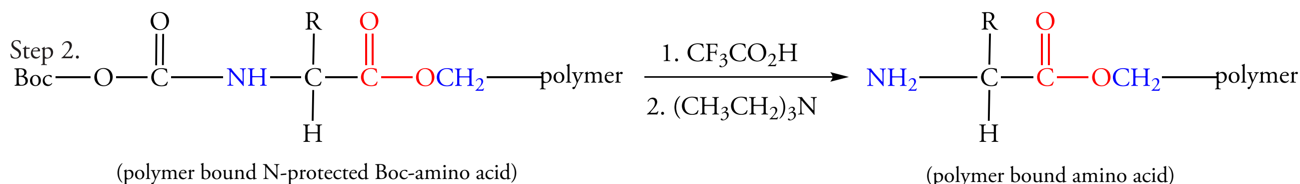


chloromethylated polystyrene

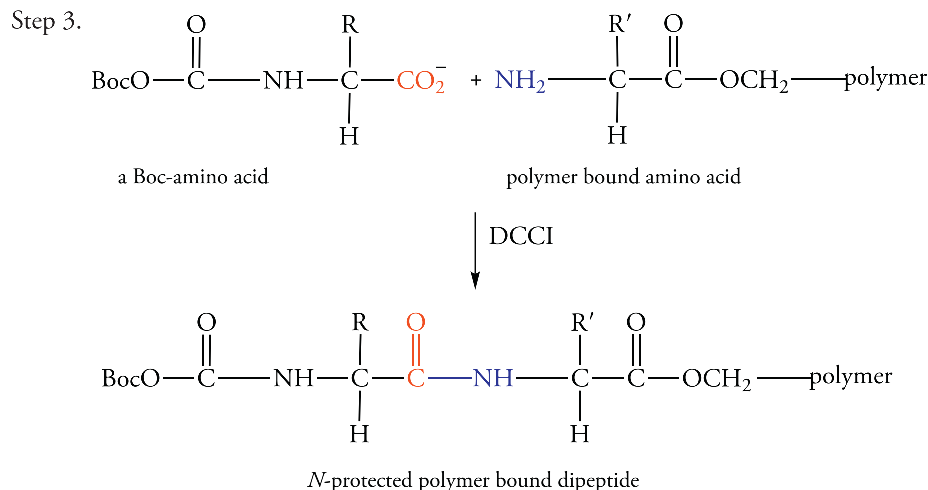
We recall that benzyl halides are reactive in substitution reactions. Even relatively weak nucleophiles such as carboxylate salts react with benzyl halides to yield benzyl esters. Thus, a solution of a carboxylate salt of an N-protected amino acid in an aprotic solvent such as DMF readily gives an ester. This first step, using a shorthand representation of the polymer, is shown below.



The polymer-bound N-protected amino acid is filtered and then washed with solvent. The product is then treated with CF_3CO_2H to deprotect the amino group by removing the Boc group in step 2. Subsequent treatment with an amine base neutralizes the ammonium group of the amino acid and yields a polymer-bound amino acid. No impurities remain in the solution.



The polymer-bound amino acid is then reacted with a solution of an N-protected amino acid and DCCI in step 3, yielding a polymer-bound N-protected dipeptide.



The polymer-bound amino acid is then reacted with a solution of an N-protected amino acid and DCCI in step 3, yielding a polymer-bound N-protected dipeptide. This cycle can be repeated many times. Peptide containing up to 80 amino acid residues can be synthesized in reasonable yield. R. B. Merrifield was awarded the Nobel Prize in Chemistry in 1984 for inventing solid state peptide synthesis.

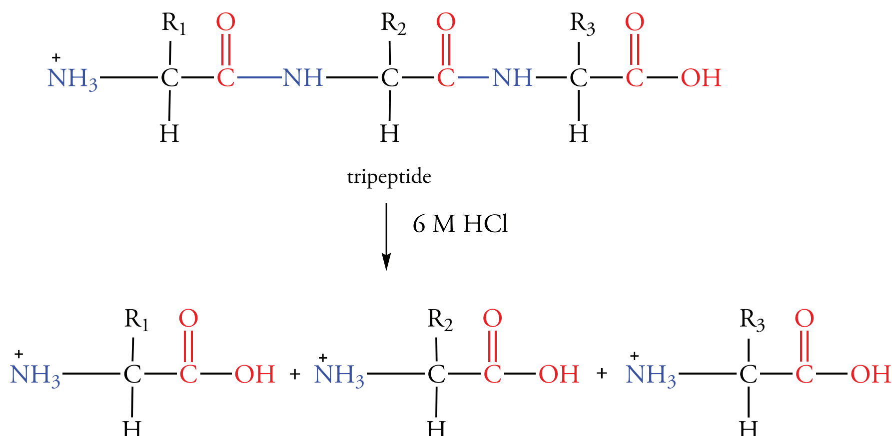
27.10 DETERMINATION OF THE AMINO ACID COMPOSITION OF PROTEINS

Determination of the Amino Acid Composition of Proteins by Chemical Methods

At one time, determining the amino acid composition was a difficult and time-consuming process. This analysis is not performed automatically in an instrument called an amino acid analyzer. This process has four steps. About 10 µg of protein are required. Modern instruments can detect about 5 nmol of a given amino acid.

1. Hydrolysis of the protein in HCl.
2. Synthesis of derivatives of the amino acids released in step 1.
3. Separation of the covalently modified amino acids by high-performance liquid chromatography (HPLC).
4. Analysis of the chromatographic data.

Step 1. Acid-catalyzed hydrolysis. We recall that amides, and therefore peptide bonds, are quite stable and the fairly stringent conditions are required for this reaction. The protein is hydrolyzed in 6 M HCl for about an hour at a temperature of 150 °C. This process is not as straightforward as it might seem since amino acid residues are not affected in the same way by acid hydrolysis. Thus, the Asn and Gln, which contain amide bonds in their side chains, are concerted to Asp and Glu, respectively. Tryptophan and cysteine are completely destroyed by acid hydrolysis. Therefore, a sample of the protein has to be hydrolyzed under a variety of conditions for an accurate analysis. Since we are primarily interested in basic principles, we will not consider the hydrolysis reactions in greater detail.



Step 2. Synthesis of amino acid derivatives. The amino acids released by hydrolysis of the protein are converted to phenylthiocarbamyl (PTC) amino acids by reaction with phenylisothiocyanate (PICT).

$$NH_2-CH(R_1)-C(=O)-OH + C_6H_5-N{=}C{=}S \longrightarrow C_6H_5-HN-C(=S)-NH-CH(R_1)-CO_2H$$

phenylisothiocyanate (PICT)

phenylisothiocarbamyl amino acid (PCT amino acid)

Step 3. Separation of the covalently modified amino acids by high performance liquid chromatography (HPLC). The PCT derivatives are separated by HPLC. The PCT derivatives are detected by UV spectroscopy; λ_{max} for the PCT derivatives is 254 nm. Figure 27.6 shows a sample result of the chromatographic separation.

Figure 27.6 HPLC Separation of PCT Amino Acids

Step 4. Analysis of the chromatographic data. Each chromatographic peak is identified by comparison to the retention times (in minutes) of known PCT derivatives. In commercial instruments, a system for data analysis is linked to the amino acid analyzer. The amount of each amino acid in the sample is calculated by dividing the peak area of each peak. Since the extinction coefficients for the amino acids differ, their absorbances also differ (Section 11.10), and the concentrations of are corrected to account for this difference.

Table 27.4 gives the amino acid composition of human lysozyme, also called α-lactalbumin. Lysozyme hydrolyzes the cell walls of gram-positive bacteria and provides a natural defense against bacterial infection. It is present, for instance, in the eye and helps to prevent eye infections. Lysozyme contains 120 amino acids.

Table 27.4
Amino Acid Composition of Human Lysozyme

Amino Acid	*Number of Amino Acids*	*Per Cent Composition*
Ala	5	4.1
Arg	1	0.8
Asn	4	3.3
Asp	12	9.8
Cys	8	6.5
Gln	7	4.9
Glu	8	6.5
Gly	6	9.8
His	2	11.4
Ile	12	9.8
Leu	14	11.4
Lys	12	9.8
Met	2	1.6
Phe	4	3.3
Pro	2	1.6
Ser	8	6.5
Thr	7	5.7
Trp	3	2.4
Tyr	4	3.3
Val	2	1.6

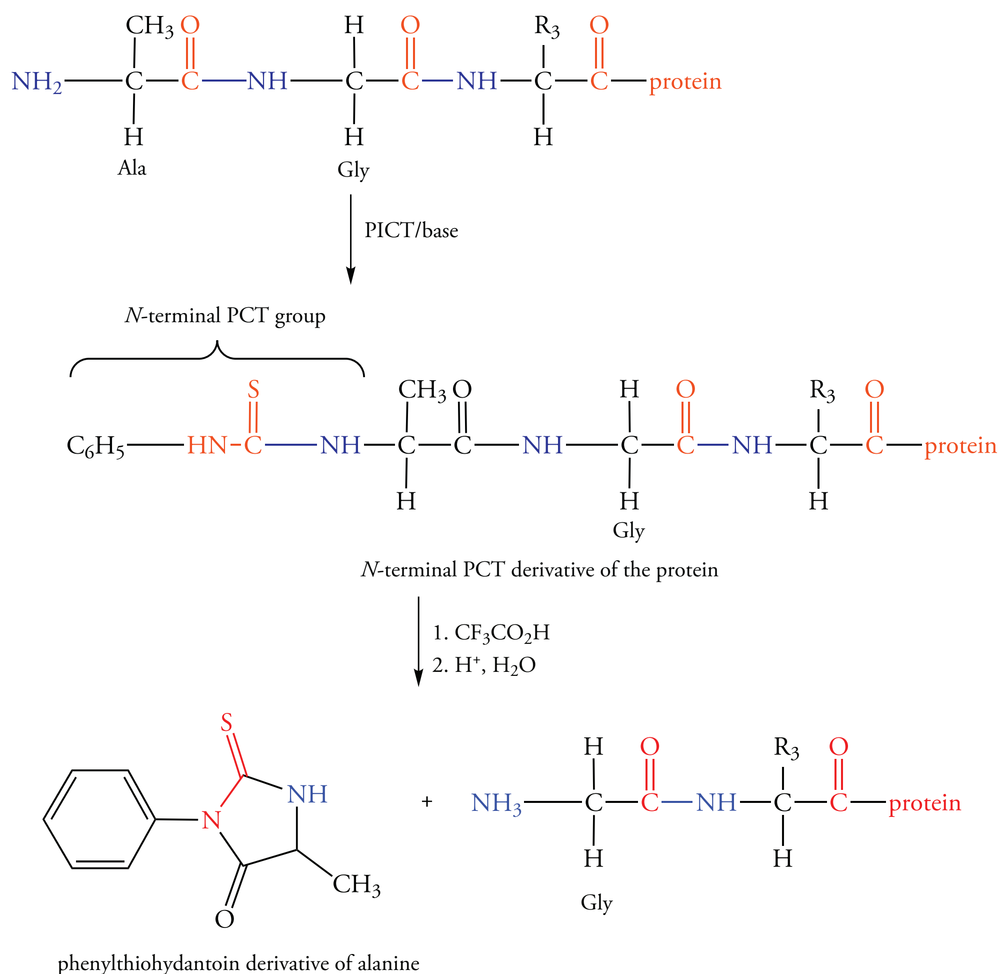
27.11 DETERMINATION OF THE AMINO ACID SEQUENCE OF PROTEINS

Chemical methods for determining the sequence of amino acids in a protein couples identification of the end N-terminal amino acid by an automated amino acid sequencer to an automated amino acid analyzer. Modern instruments can determine sequences for protein or peptide samples that contain 10–100 picomoles, pm (1 pm = 10^{-12} moles). The overall process requires several automated steps.

The Edman Degradation

The identity of the N-terminal amino acid of a polypeptide is determined by a method invented by Pehr Edman called the **Edman degradation**. In the Edman degradation, the polypeptide is treated with phenyl isothiocyanate—the Edman reagent— which reacts with the N-terminal amino acid to give an N-terminal PTC derivative or the protein. This derivative forms by addition of the terminal N—H bond across the C=N of the phenyl isothiocyanate. After the adduct has formed, anhydrous trifluoroacetic acid is added to the reaction mixture. This reagent cleaves the polypeptide at the N-terminal residue. Under these conditions, the peptide bonds in the protein do not break (Figure 27.7). Reaction of the N-terminal amino acid with phenylisothiocyanate (PITC) gives the N-terminal PCT derivative of the protein, which is exactly analogous to the reaction of amino acids with PICT. This derivative is treated with trifluoroacetic acid, and then water is added. These steps release the first amino acid as its phenylthiohydantoin (PTH) derivative. The other peptide bonds of the protein, which now contains one less amino acid, are not affected.

Figure 27.7 Edman Degradation
First, the peptide is converted to its N-terminal PCT derivative by treatment with phenylisothiocyanate. Next, the PCT protein is treated with trifluoroacetic acid, then with water to give the phenylthiohydantoin derivative. The N-terminal amino acid is released in this step. The other peptide bonds are not affected.

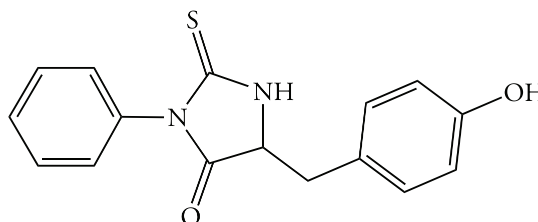


A complex cyclization reaction occurs to give a substituted phenylthiohydantoin. This ring contains the carbonyl carbon atom, the α-carbon atom, and the amino nitrogen atom. The R group of the amino acid is attached to the ring. The PTH derivative of the N-terminal amino acid is then automatically transferred to an amino acid analyzer. Comparison with the phenylthiohydantoin of known amino acids establishes the identity of the amino acid.

Because the Edman degradation does not cleave the peptide bonds in the protein, it can be repeated to sequentially identify the amino acids from the N-terminal amino acid of the molecule. The yield of the Edman degradation approaches 100%, and sequences of 30 residues of a polypeptide can be determined from 5-picomole (5×10^{-12} mole) samples. This means that the sequence of a peptide with 30 amino acid residues, with a molecular weight of about 3000, can be determined from a 15 nanogram sample!

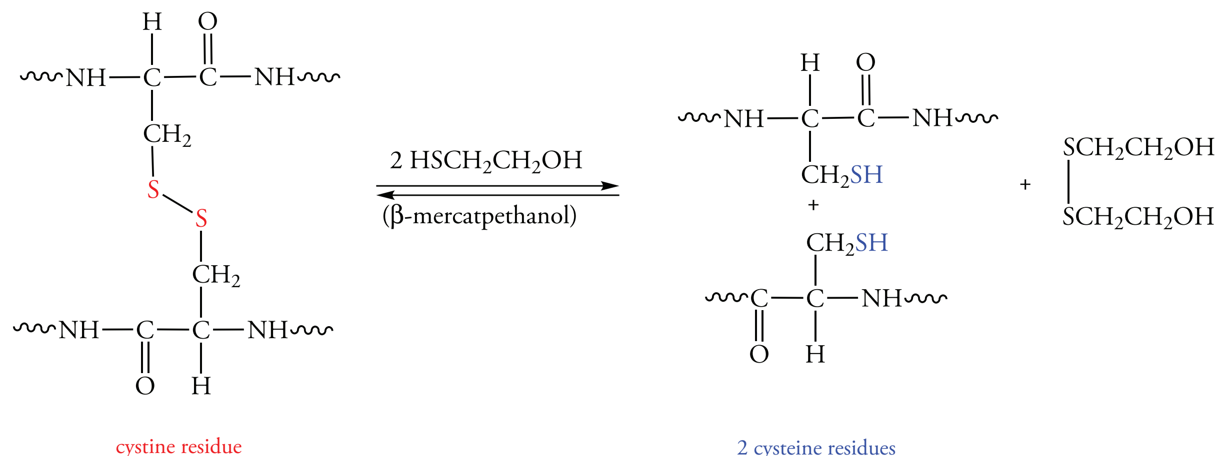
Problem 27.12

β-Endorphin, a peptide that contains 31 amino acid residues, has analgesic effects and promotes the release of growth hormone and prolactin. Treating β-endorphin with phenyl isothiocyanate followed by hydrolysis with anhydrous trifluoroacetic acid, and then with water, releases the following phenylthiohydantoin. What is the N-terminal amino acid of the peptide?

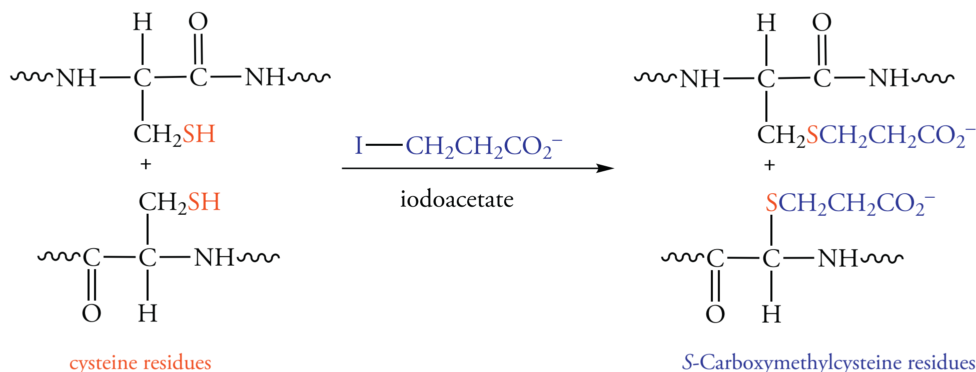


Blocking Cystine Residues

If a protein contains a chain internally linked by one or more cystine residues, the disulfide bonds of these residues must be cleaved. Treating the protein with excess β-mercaptoethanol converts cystine to two cysteine residues. This is a reversible disulfide bond exchange in which the protein is oxidized, and the β-mercaptoethanol is reduced.

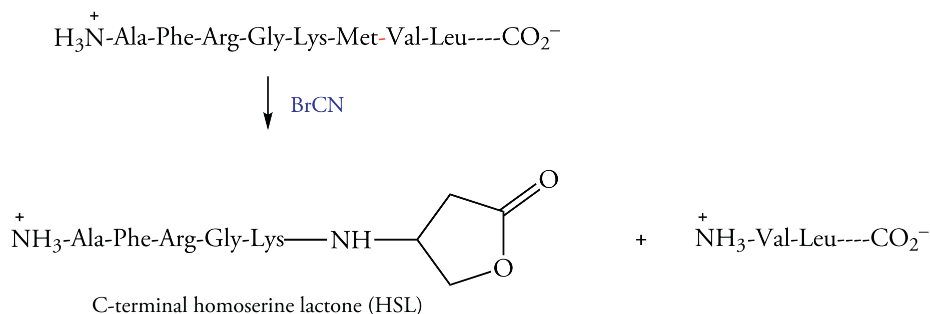


To prevent disulfide bonds from forming again, the protein is treated with iodoacetate, which converts the cysteine residues to *S*-carboxymethylcysteine residues. This reaction occurs by an S_N2 mechanism in which the nucleophilic sulfur atom displaces iodide.



Peptide Cleavage at Methionine Residues

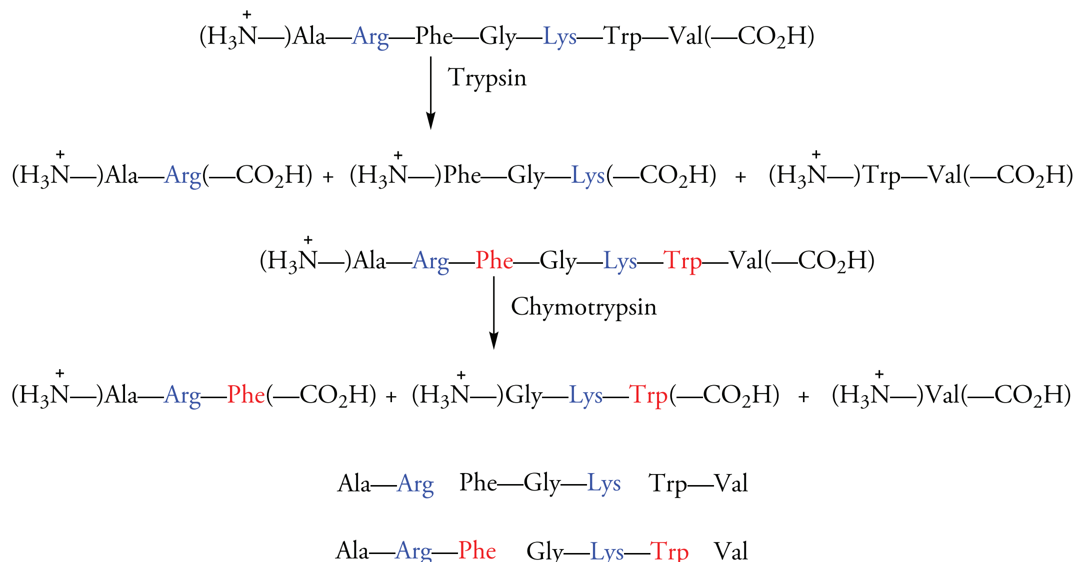
If a protein contains methionine residues, the polypeptide chain can be cleaved by cyanogen bromide (BrCN), which produces a C-terminal peptidyl homoserine lactone (HSL) residue. Most proteins contain only a few methionine residues, so only a few fragments result from this reaction.



Enzymatic Cleavage of Polypeptide Chains

Many proteins contain hundreds of amino acids. To determine their sequences other reactions are required to provide sequences short enough to be determined by Edman degradation. Enzymatic cleavage by two enzymes, trypsin and chymotrypsin, is used to produce smaller peptides. Trypsin cleaves polypeptide chains on the C-terminal side of basic residues such as arginine and lysine. Chymotrypsin cleaves the polypeptide on the C-terminal side of aromatic residues.

The sequences each oligopeptide fragment produced in these enzymatic reactions are determined by Edman degradation. Then, in the final step, the fragments are aligned to provide the entire sequence.



Primary Structures and Evolutionary Relationships

The primary structures of thousands of protein are known. Comparing the primary structures of proteins that are common to many species reveals evolutionary relationships. As organisms evolve, their genes change through mutation. Since the primary structure of a protein reflects the gene coding for it, differences among primary structures are a record of evolutionary change. Comparing the amino acid sequences of proteins found in different species thus opens a window to the past. In a sense, then, proteins can be regarded as living fossils.

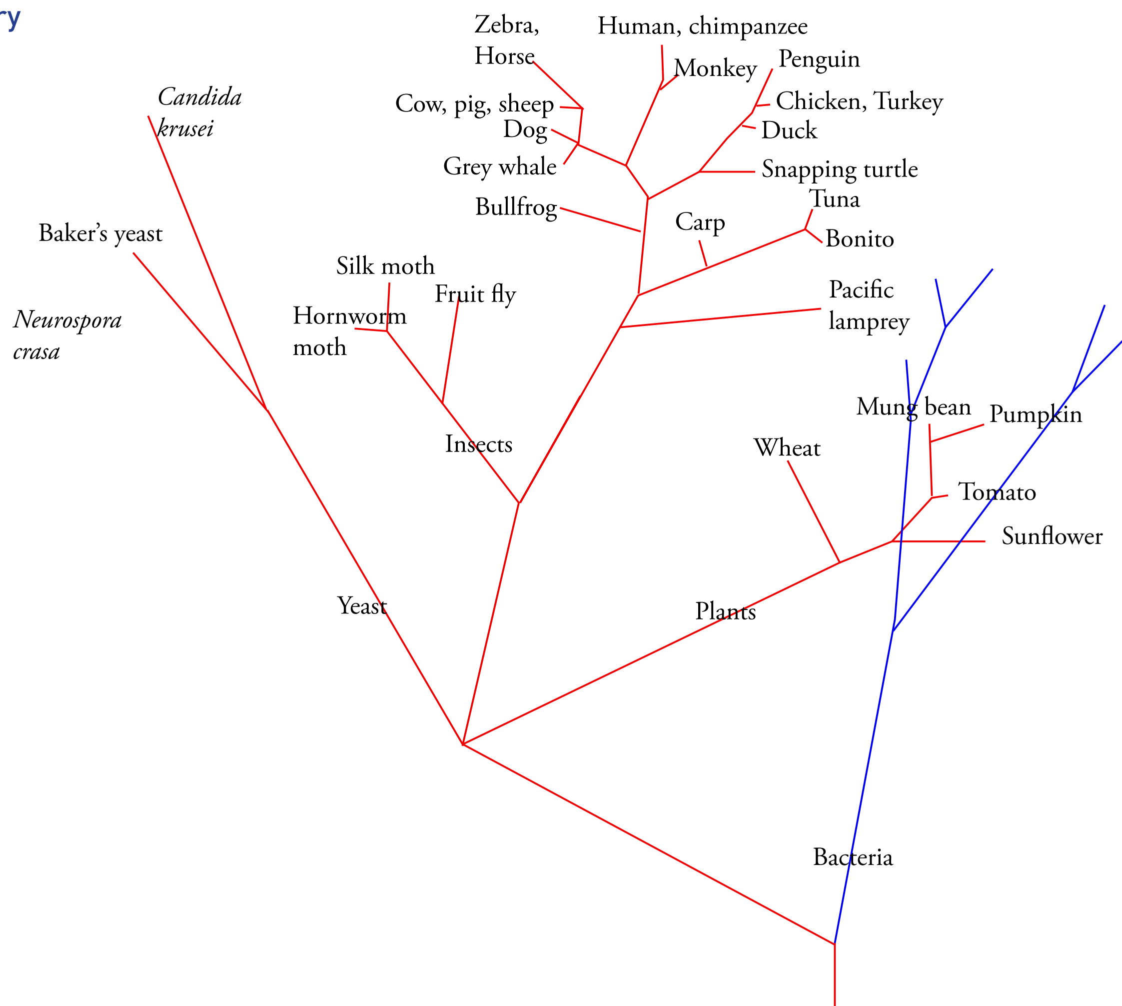
In closely related species, the primary structures of common proteins are similar. Counting the number of differences in amino acid sequences among these proteins gives some idea of how far various species have diverged in the course of evolution. For example, the protein cytochrome *c* is an excellent protein for evolutionary comparisons because it is found in the respiratory electron transport system, which is present in all aerobic organisms (Figure 27.8).

Figure 27.8 shows that as evolutionary lines diverge, the number of sequence variations increases so that closely related species have few differences and distantly related species have many difference in primary structure. Thus, human and chimpanzees have identical cyctochrome *c* sequences. The primary structures of cytochrome *c* molecules from California gray whales differ from that of pigs, cows, and sheeps by only two residues. We conclude that the whale has evolved from land animals related to modern hoofed animals. Gray whale cytochrome *c* differs from human cytochrome *c* 10 residues.

Peking duck and penguins also have cytochrome *c* sequences that differ by only three residues, but they differ by 11 residues from bullfrogs. Thus, these species are closely related to each other but distantly to bullfrogs.

The difference between human cytochrome *c* and baker's yeast cytochrome *c* is 45 residues, which we do not find particularly surprising since these are distantly related species. However, 59 of the 104 residues in cycotchrome *c* are identical. Identical residues are essential for the structure and function of the protein.

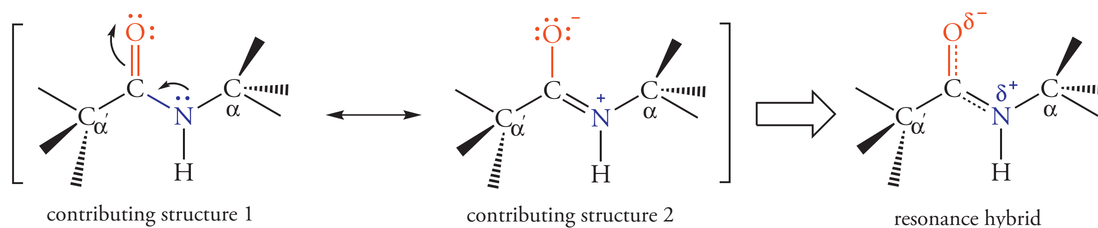
Figure 27.8 Evolutionary Family Tree for Cytcochrome *c*



27.12 BONDING IN PROTEINS

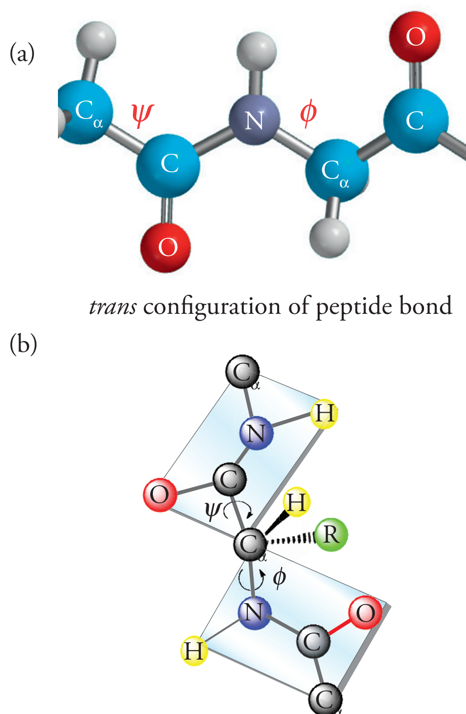
Structure of the Peptide Bond

The carbonyl group of one amino acid and the amino group of the next by secondary amides called peptide bonds. The lone pair on the nitrogen of the peptide bond is delocalized onto the carbonyl carbon, and the peptide bond is resonance stabilized.



The C—N bond length in a peptide bond has about 50% double bond character, and the formal charges on the nitrogen and oxygen atoms are about +½ and –½, respectively. We recall that rotation around double bonds does not occur. Similarly, rotation around the partial C—N double bond is restricted and does not occur at room temperature. As a result, the overwhelming majority of peptide bonds in proteins have *trans* configurations. However, free rotation does occur around the C—C_a and C—$C_{a'}$ single bonds. In peptide and protein nomenclature, the N—C_a bond is called phi (ϕ) and the C—C_a bond is called psi (ψ). Rotations around these bonds give rise to a vast number of conformations of the polypeptide chain (Figure 27.9).

Figure 27.9 Structure of the Peptide Bond
(a) Rotation around the C—N bond, which has 50% double bond character, does not occur at room temperature. However, rotation around the N—C_a bond (ϕ) and the C—C_a bond (ψ) is possible, and many conformations are possible in peptides and proteins.
(b) we can think of the α-carbon as a "hinge" between two planar peptide bonds. If one takes two note cards and links them with a swivel, it is easy to see that many arrangements are possible. However, some ϕ and ψ are not possible because of steric interference of the side chain R group. Glycine, for example, can assume many more conformations than amino acids like proline and tryptophan.



The α-Helix

One of the most common structural features in proteins is a repeating conformation called the **alpha helix**. In this conformation, the values of ϕ and ψ are –57° and 47°, respectively. Figure 27.10 shows the structure of an α-helix. The hydrogen bonds are approximately parallel to the long axis of the helix. In Figure 27.9a, amino acid side chains are shown. Figure 27.9b the polypeptide backbone and hydrogen bonding pattern in the helix. Some amino acids are more likely than others to be present in an α-helix. Proline, which lacks an N—H bond, and therefore cannot form a hydrogen bond, is never found in an α-helix. Furthermore, its five-membered pyrrolidine ring cannot assume the ϕ/ψ angles required for a helix. Nonpolar, hydrophobic amino acids such a leucine, valine, and phenylalanine are often found in α-helices; polar, charged side chains are less common. The carbonyl oxygen atom of the peptide bond is a hydrogen bond acceptor, and the peptide bond nitrogen atom is a hydrogen bond donor.

The α-helix is right-handed. Left-handed α-helices are not observed. Why not? We recall that a helix is a chiral structure so that right- and left-handed helices are mirror images, and we know that enantiomers have the same energy. However, the α-amino acids are chiral (except glycine), and a right-handed α-helix of L-amino acid residues and a left-handed α-helix of L-amino acid residues are *not* enantiomers, they are diastereomers. We know that diastereomers have different energies. A right-handed α-helix of L-amino acid residues is more stable than its diastereomer because in the left-handed configuration there is considerable steric hindrance among side chains, so it does not form.

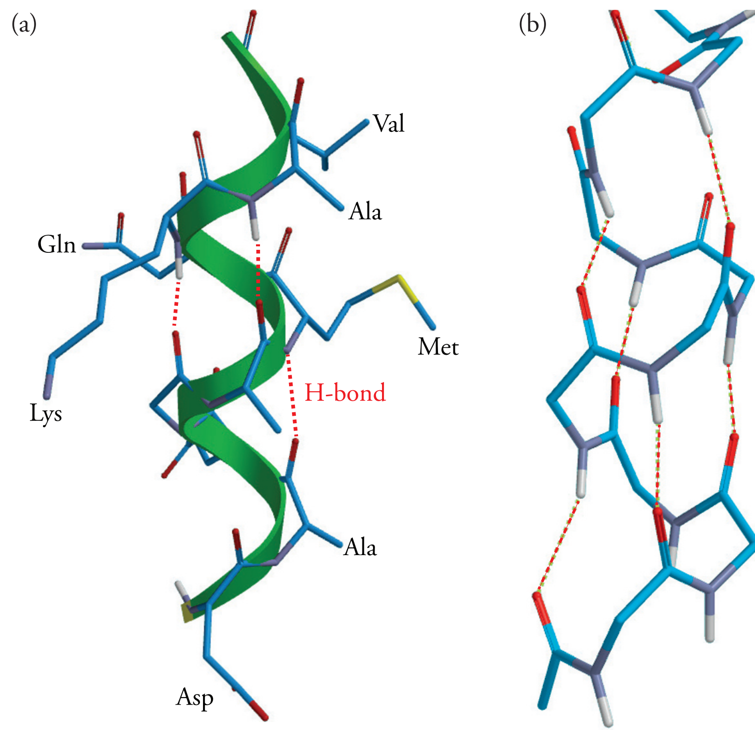


Figure 27.10 Dimensions of an α-Helix
The distance between amino acid residues in an α-helix is 0.15 nm. The distance required for one turn of the helix, its pitch, is 5.4 nm.

β-Pleated Sheets

Many proteins contain a type of secondary structure in which the polypeptide chain has a completely extended conformation. Two adjacent regions of fully extended polypeptide chains form hydrogen bonds at that link the chains at approximately right angles to the long axis of the chain. This type of secondary structure is called a **β-pleated sheet**. There are two kinds of β-pleated sheets, called **parallel** and **antiparallel**. In a parallel β-pleated sheet, the C- and N-termini of the sheet are together (Figure 27.11); in the antiparallel β-pleated sheet, they are opposed (Figure 27.12).

Figure 27.11 Hydrogen Bonding In Parallel β-Pleated Sheet

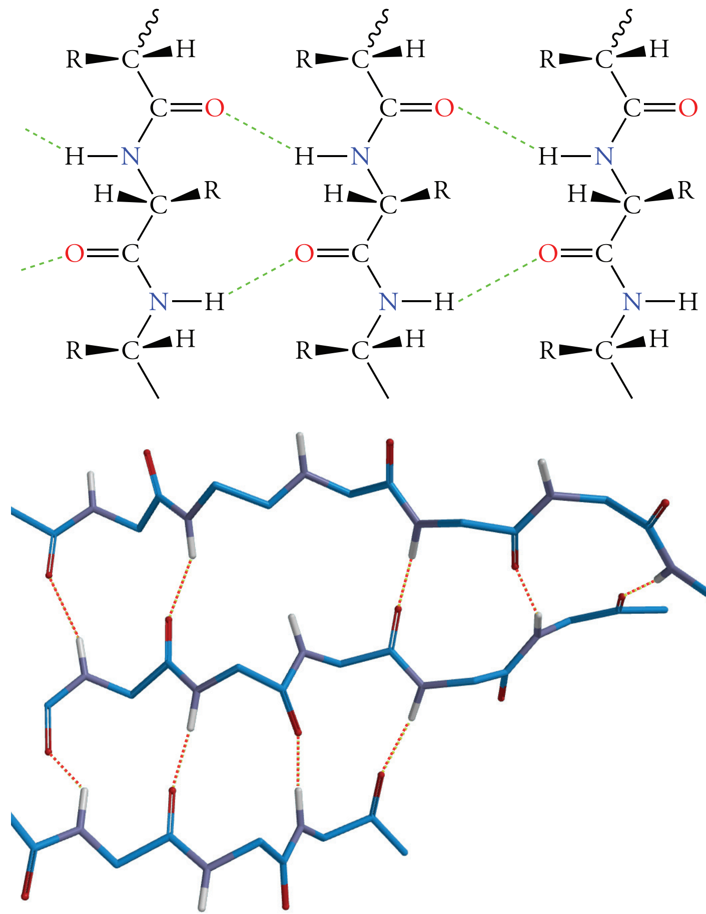
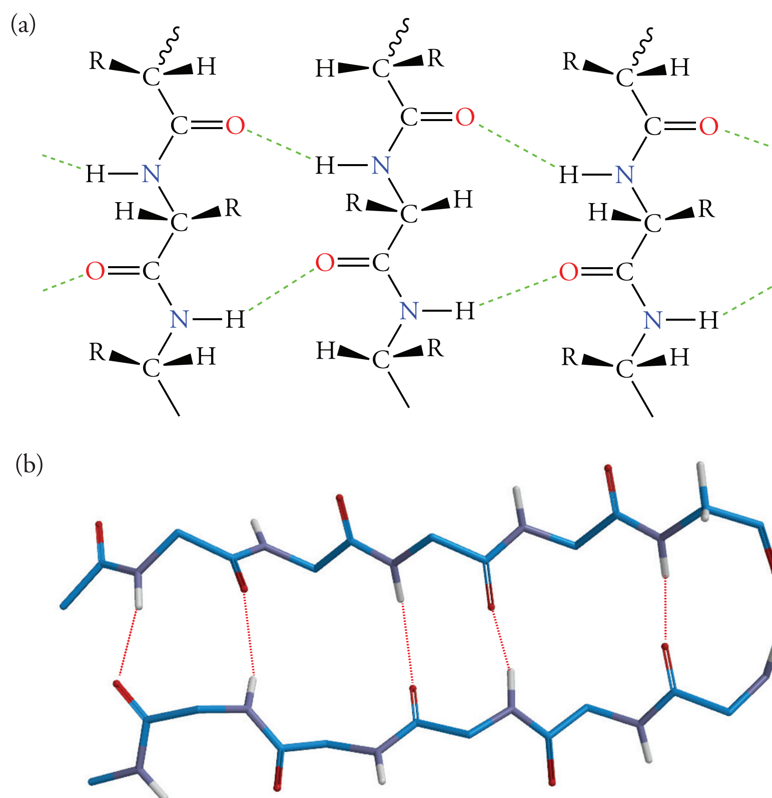
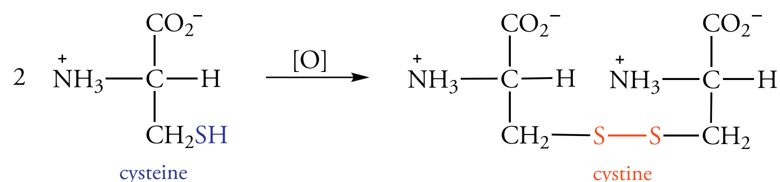


Figure 27.12 Hydrogen Bonding in an Antiparallel β-Pleated Sheet
(a) Bond-line structure of an antiparallel β pleated sheet.
(b) Molecular model of an antiparallel β pleated sheet showing only the polypeptide backbone and the hydrogen bonds.



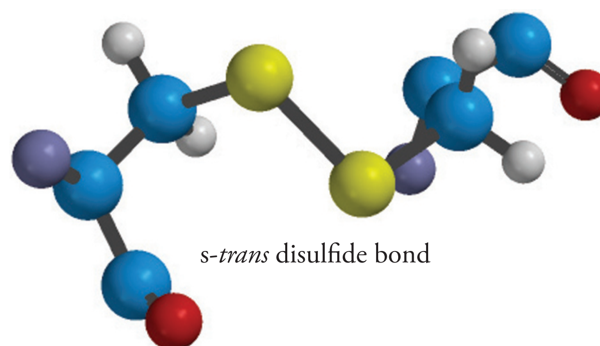
Disulfide Bonds

Many proteins—especially relatively small ones containing fewer than 100 amino acid residues—have a high cysteine content. Each of these cysteine residues has a sulfhydryl group (—SH) that can be oxidized to form a disulfide bond. The dimer of cysteine itself is called, somewhat confusingly, **cystine** (note the "missing e" in the name).



Disulfide bonds form after a protein has folded into its biologically active conformation. Once disulfide bonds have formed, the protein conformation is much less flexible. We recall that conformations around sigma bonds can be either s-*cis* or s-*trans* and that the s-*trans* conformation is usually more stable because it minimizes steric repulsion. When disulfide bonds exist in proteins, they nearly all have an s-*trans* conformation (Figure 27.13).

Figure 27.13 Conformation of an s-*trans* Disulfide Bond



Intrachain disulfide bonds occur in small peptides such as oxytocin and vasopressin, as we saw in Section 27.6. Disulfide bonds can also link a cysteine residue in one polypeptide chain with a cysteine residue in another polypeptide chain as in the polypeptide insulin.

Hydrophobic Interactions

Proteins contain many nonpolar side chains. These side chains are repelled by water and tend to associate with one another on the "inside" of a folded protein molecule, out of contact with water. The tendency of nonpolar side chains to collect out of contact with the solvent is called the **hydrophobic effect**. The hydrophobic interactions in proteins are similar to those in the micelle of a soap (Section 21.5) or the bilayer of lipids in membranes. Hydrophobic interactions among nonpolar side chains in proteins are weak, but abundant, and are primarily responsible for maintaining the folded conformation of a protein.

27.13 PROTEIN STRUCTURE

The highest operation in nature and in art is the attainment of significant form.
Goethe

Globular proteins are compact, more or less spherical molecules. The term globular sounds rather uninspiring, but globular proteins have a wonderful diversity of forms. The term globular does not in the least imply structural monotony or simplicity. Most globular proteins are soluble in the cytosol or in the lipid phase of biological membranes. Globular proteins are the primary agents of biological action in the cell. Most globular proteins are protein catalysts called enzymes. Some globular proteins transport oxygen and lipids in the blood. Some are hormones or membrane-bound receptors that mediate the action of hormones. The globular proteins called immunoglobulins, or antibodies, are the first lines of defense against pathogenic bacteria and viruses.

We can describe the structure of globular proteins in terms of four levels of structure. Each structural level has properties that cannot be deduced from a knowledge of the lower levels of structure.

1. The **primary structure** of a protein consists of its linear sequence of amino acid residues. Although it is the simplest level of structural organization, in many ways, it is the most important, since the primary structure determines the conformation and function of a protein.
2. The **secondary structure** of a protein consists of regularly repeating conformations of the polypeptide backbone such as α-helices and β-pleated sheets.
3. The **tertiary structure** of a protein consists of its three-dimensional conformation. Many proteins contain regions called **domains** that are relatively independent structural units.
4. Many proteins exist in cells in complexes that contain two or polypeptide chains. These complex assemblies are called the **quaternary structure** of the protein. Table 27.5 gives a few examples.

Table 27.5
Examples of Proteins Having Quatenary Structure

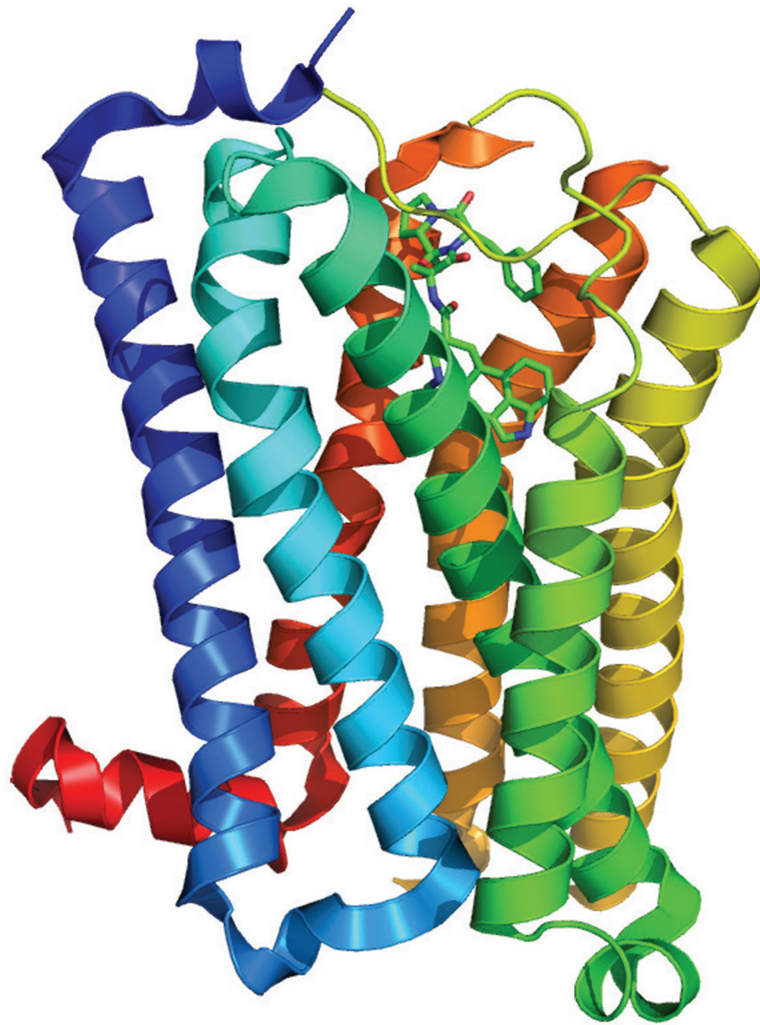
Protein	*Molecular Weight*	*Number of Subunits*	*Function*
Alcohol Dehydrogenase	80,000	4	Enzymatic reaction in fermentation
Aldolase	150,000	4	Enzymatic reaction in glycolysis
Fumarase	194,000	4	Enzymatic reaction in citric acid cycle
Hemoglobin	65,000	4	Oxygen transport in blood
Insulin	11,500	2	Hormone that regulates metabolism of glucose

In our discussion of peptide functions, we discussed a family of closely related proteins called guanine nucleotide, coupled receptor proteins. We noted that the hormone binding regions of these receptors were located in a region of seven helices within the membrane. Figure 27.14 shows the seven-helix bundle in the membrane region of the serotonin receptor.

We recall that the primary sequences of the guanine nucleotide protein receptors (GCPR) have similar sequences and that closely similar sequences correspond to close evolutionary ancestry. When we look at the sequences of the serotonin receptor, we find that membrane helix 5 plays an important part in ligand specificity. When we align the sequences of many GCPRs, we find that the secondary structures are also aligned, so sequence alignment can be extended to structure alignment (Figure 27.15).

Figure 27.14 Ribbon Diagram of the Membrane Region of the Serotonin Receptor

The seven helix region of the serotonin receptor is the site of serotonin binding. The serotonin receptor is a member of the G-coupled receptor protein family. These proteins have similar structures. Their different specificities depend upon differences in primary structure at the ligand bindings site.



A small protein called 1GB1, which contains the ligand (antigen) binding site at the N-terminus of an immunoglobulin, is shown in Figure 27.15. This region, which has 56 amino acid residues, contains antiparallel β-pleated with four strands. An α-helix sits on top of the β-pleated sheet.

Helices and sheets can combine in many other ways. For example, an enzyme that catalyzes a reaction in the degradation of glucose (glycolysis) called triose phosphate isomerase, which has 248 amino acid residues, contains many "strand-helix-strand" motifs β-α-β (Figure 27.16). The β strands form a parallel β-pleated sheet. Figure 27.17 shows the tertiary structure of triose phosphate isomerase.

Figure 27.15 Structure of 1GB1

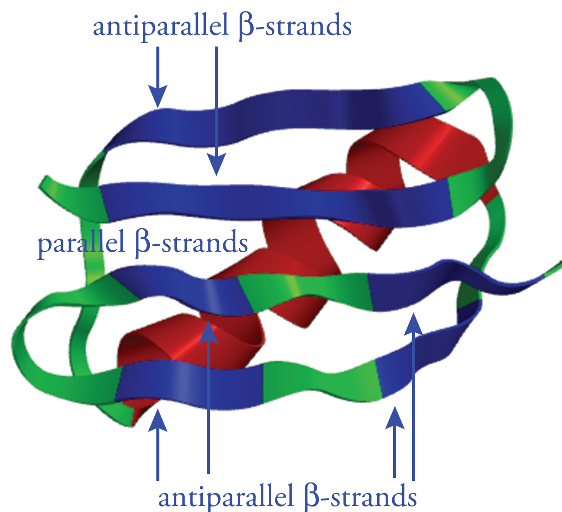


Figure 27.16 Parallel β Strands and an α-Helix in a β-α-β Arrangement

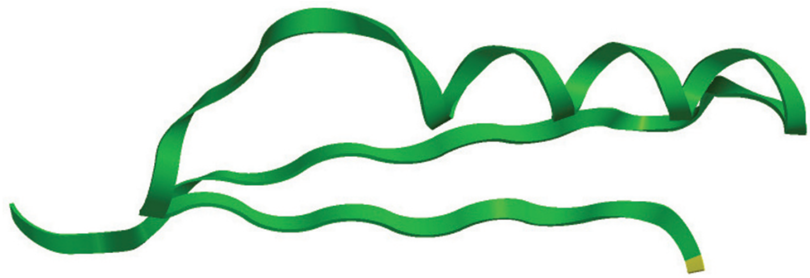
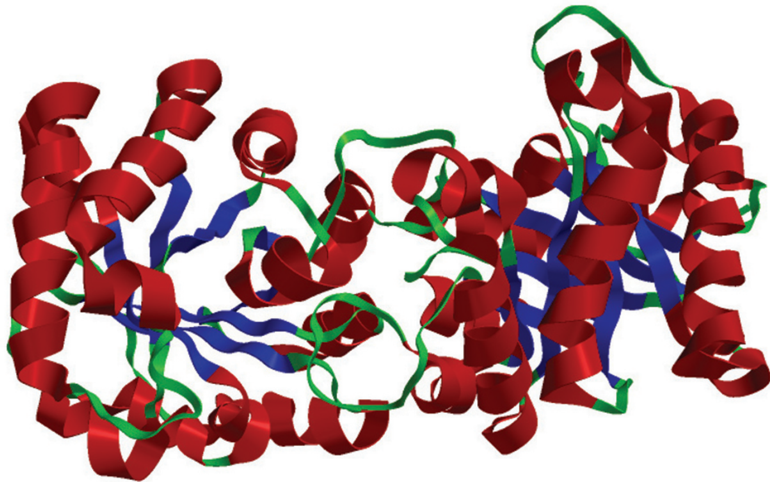


Figure 27.17 Tertiary Structure of Triose Phosphate Isomerase

The α-helices are shown in red, β-pleated sheets are blue, and less structured "loops" are shown in green.



27.14 OXYGEN STORAGE AND TRANSPORT: MYOGLOBIN AND HEMOGLOBIN

Humans and other vertebrates transport oxygen in red blood cells called **erythrocytes** (Greek, *erythro*, red; *kytos*, cell). A mature human erythrocyte is essentially a sack that carries hemoglobin. Hemoglobin transports oxygen throughout the body; myoglobin stores oxygen in cardiac and skeletal muscle until it is consumed during metabolism.

Myoglobin

We will begin our discussion with myoglobin. Myoglobin accounts for about 8% of total muscle protein in diving mammals such as seals and whales that store large amounts of oxygen for use during dives. These animals do not contain hemoglobin.

The structure of myoglobin was determined in 1960. Most of the amino acid residues in myoglobin are in α-helices (Figure 27.18). The oxygen binding site in myoglobin is not the protein itself, but a **heme** group. Heme contains Fe^{2+}, and O_2 is the ligand that binds it (Figure 27.19). Both carbon monoxide and cyanide have a higher affinity for heme than oxygen. In high concentrations, they inhibit oxygen binding; hence, they are fatal.

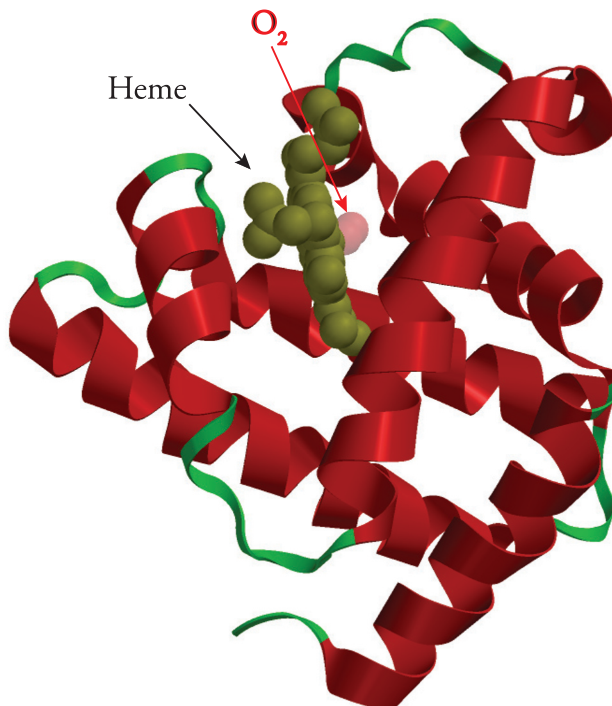


Figure 27.18 Structure of Oxymyoglobin.
The α-helices are shown in red, and less structured "loops" are shown in green.

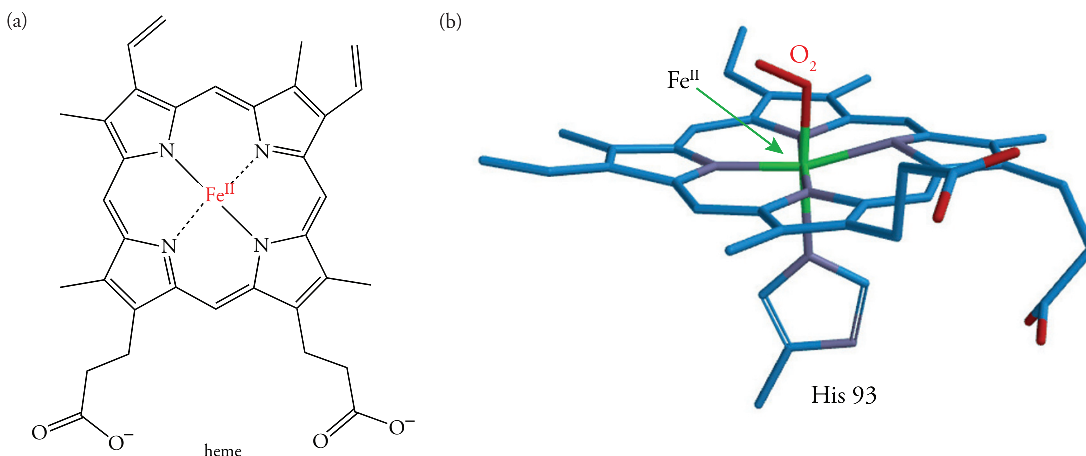


Figure 27.19 Heme
(a) Bond-line structure of heme. (b) Structure of the heme group bound to myoglobin via a bond from a nitrogen on histidine 93 and the Fe^{II} ion. Oxygen binds on the opposite side of the histidine.

Hemoglobin

Hemoglobin has a quaternary structure. It consists of two pairs of different proteins, designated the α and the β chains. There are 141 and 146 amino acids in the α and β chains of hemoglobin, respectively. As in myoglobin, each subunit is linked covalently to a molecule of heme. Thus, hemoglobin binds four O_2 molecules. The two identical α chains and the two identical β chains are arranged tetrahedrally (Figure 27.20). These units are held together by hydrophobic interactions, hydrogen bonding, and ion pairs (salt bridges) between oppositely charged amino acid side chains.

The subunits of hemoglobin do not act independently. When one subunit binds O_2, its conformation changes. When a change in conformation at one site of an oligomeric protein is caused by a change in a spatially separated site of the oligomer, the change is called an **allosteric** effect, and the protein is called an **allosteric protein**. Hemoglobin is an allosteric protein. When one heme group in hemoglobin binds oxygen, it is easier for successive oxygen molecules to bind at the remaining three sites. Thus, once oxygenation occurs at one heme, there is cooperation at all other sites in hemoglobin.

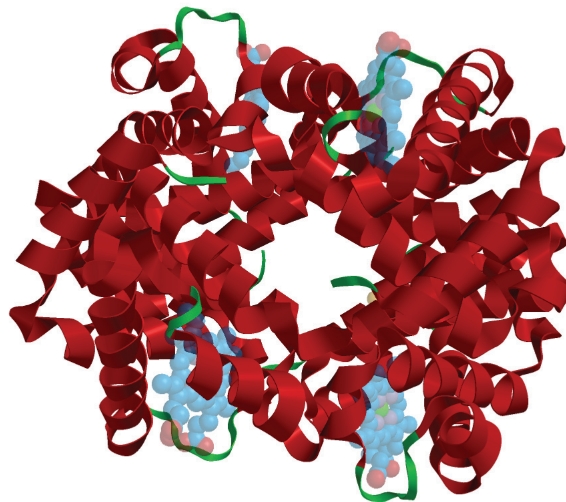


Figure 27.20 Structure of Deoxyhemoglobin.
The α and β subunits of hemoglobin interact cooperatively, and when one heme binds O_2, the each of the others rapidly binds O_2.

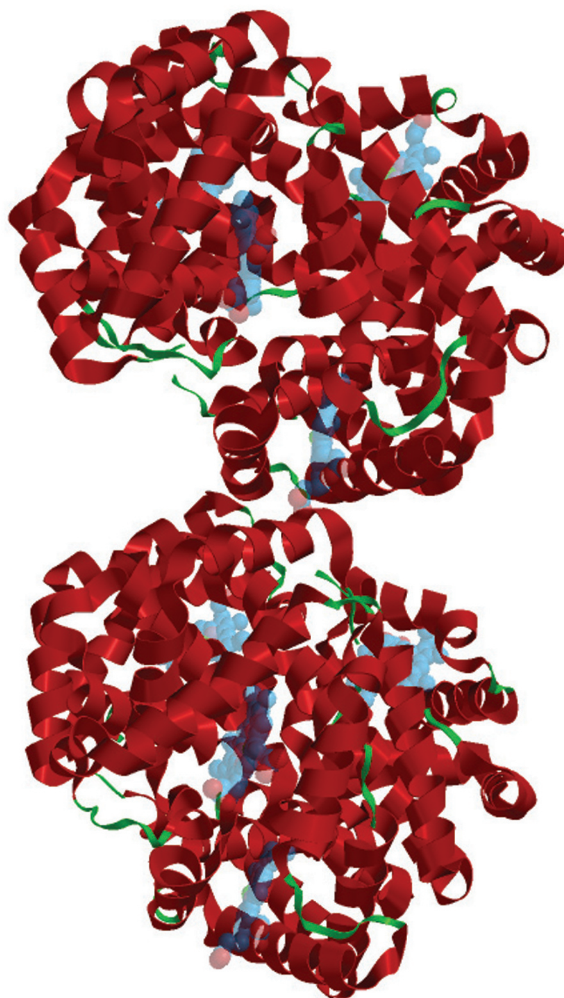
Sickle Cell Hemoglobin

In some people, glutamate 6 of the β chain has undergone a mutation to valine. This mutation changes the charge on the surface of hemoglobin. The mutant protein is called sickle cell hemoglobin, HbS. Valine residue 6 of the β chain of deoxy HbS lies on the surface of the protein. This hydrophobic residue, present in each β chain, forms a hydrophobic contact with a pocket in a neighboring β chain of another hemoglobin molecule. The mutation that replaces a glutamate residue by a valine residue decreases the solubility of deoxy-HbS.

	1	2	3	4	5	6	7	8
Hemoglobin A	Val-	His-	Leu-	Thr-	Pro-	Glu-	Glu-	Lys-
Hemoglobin S	Val-	His-	Leu-	Thr-	Pro-	Val-	Glu-	Lys-

The concentration of hemoglobin in red blood cells is high, and even in normal hemoglobin, it is near the limit of its solubility. The lower solubility of deoxy-HbS shifts the balance toward precipitation. The interaction of the valine residues leads to the formation of a polymeric fiber that forms when hemoglobin releases oxygen (Figure 27.21). The formation of the fibrous hemoglobin leads to abnormally shaped red blood cells. The cells tend to be sickle shaped, and as a result, their passage through the blood vessels is restricted. The associated circulatory problems are known as **sickle cell anemia**. That is why the mutant protein is called HbS.

Figure 27.21 Structure of Deoxyhemoglobin Dimer.
The β subunits of hemoglobin interact by van der Waals contact between the isopropyl side chains at residue 6 of sickle cell hemoglobin (HbS). Since each HbS has two β subunits on opposite sides of the tetramer, a fibrous polymer forms. HbS polymerizes when HbS releases O_2, which disorts the red blood cells into the shape of a sickle.



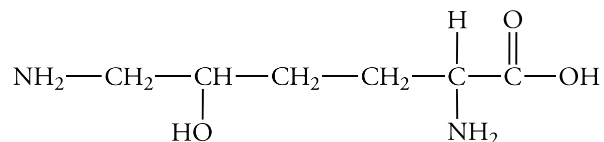
Exercises

Amino Acids

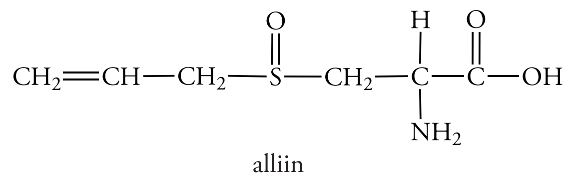
27.1 A D-Glutamic acid residue is found in some bacterial cell walls. Draw its Fischer projection formula.

27.2 Gramicidin S is a cyclic peptide antibiotic that contains a D-phenylalanine residue. Draw the projection formula of D-phenylalanine.

27.3 The following amino acid is present in collagen. From what amino acid is it derived?



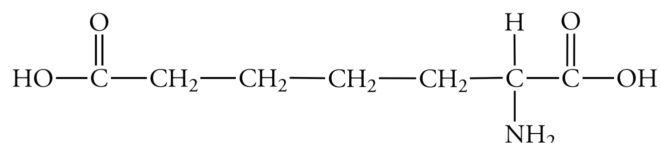
27.4 The following antibacterial agent, called allin, is present in garlic. From what amino acid might it be derived?



alliin

27.5 The following compound is an amino acid that acts as a neurotransmitter. Classify this amino acid and give its IUPAC and common name. (This neurotransmitter is universally known by its common name.)

27.6 The following compound is one of the amino acids formed in the biosynthesis of penicillin. Classify this amino acid and determine its common name.



Acid–Base Properties of Amino Acids

27.7 Draw the structures of alanine and glutamic acid at pH = 1 and pH = 12.

27.8 Draw the structures for the dipolar ions (zwitterions) of serine and valine.

27.9 How could you distinguish between aqueous solutions of asparagine and aspartic acid?

27.10 Would you expect an aqueous solution of lysine at pH 7 to be neutral, acidic, or basic? Explain.

27.11 One of the pK_a values of tyrosine is 9.11. What functional group is responsible for this acidic hydrogen atom?

27.12 One of the pK_a values of cysteine is 8.33. What functional group is responsible for this acidic hydrogen atom?

27.13 Explain why the pK_a for the —NH_3^+ group of tyrosine is slightly smaller than the corresponding pK_a of phenylalanine.

27.14 Explain why the pK_a for the side chain —CO_2H group of aspartic acid is smaller than the corresponding pK_a of glutamic acid.

27.15 Explain why the difference between the pK_a values of aspartic acid and asparagine is larger than the difference between the pK_a values of glutamic acid and glutamine.

27.16 Consider the amino and imino nitrogen atoms of the side chain of arginine. Which one would be protonated in acid solution? How can resonance stabilization account for the site of protonation?

Isoionic Points of Peptides and Proteins

27.17 Estimate the isoionic points of the following tripeptides.
(a) Ala-Val-Gly (b) Ser-Val-Asp (c) Lys-Ala-Val

27.18 Estimate the isoionic points of the following tripeptides.
(a) Glu-Val-Ala (b) Arg-Val-Gly (c) His-Ala-Val

27.19 Examine the structures of oxytocin and vasopressin in Section 27.7. Which one has the higher isoionic point?

27.20 Examine the structure of the enkephalin whose sequence is shown below and estimate its isoionic point.
Ala-Gly-Phe-Leu-Gly

27.21 The isoionic point of hen egg white lysozyme is 10.8. What does this value indicate about its amino acid composition?

27.22 The isoionic point of pepsin is 1.1. What does this value indicate about its amino acid composition?

Synthesis of Amino Acids

27.23 What haloalkane is required to synthesize isoleucine by the acetamidomalonate method? What side reaction might decrease the yield?

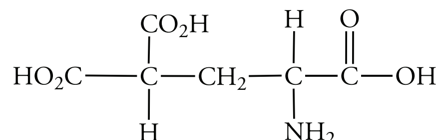
27.24 What reactants are required to synthesize phenylalanine by reductive amination?

27.25 3-Aminopropanoic acid, sometimes called β-alanine, is a nonsteroidal anti-inflammatory agent used in veterinary medicine. It is prepared by a conjugate addition reaction using ammonia and acrylonitrile (CH_2=CH—CN). The resulting nitrile is then hydrolyzed to give the product. Why is conjugated addition favored?

27.26 Methionine can be prepared from propenal in a multistep sequence. (a) Explain how is the thiomethyl group introduced? (b) Why is the carbon chain length increased by one carbon atom?

27.27 The structure of alliin is shown in Exercise 27.4. Propose a synthesis of alliin starting from an amino acid.

27.28 One of the amino acids in the blood-clotting protein prothrombin is shown below. It was difficult to detect because it decomposes under hydrolysis conditions. (a) What reaction occurs and (b) what is the product?



Peptides

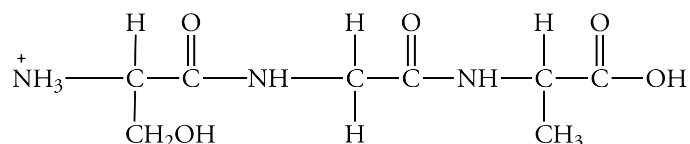
27.29 Write the bond line structure for alanylserine at pH 7.

27.30 How does glycylserine differ from serylglycine?

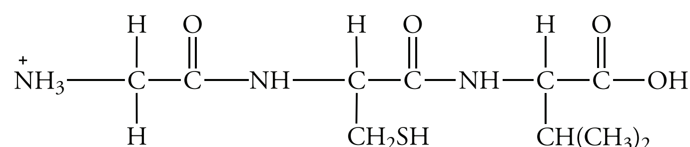
27.31 Which amino acids can form peptides with carboxylic acid groups or carboxylate groups at internal positions in the peptide chain?

27.32 Which amino acids can form peptides with amino groups or ammonium groups at internal positions in the peptide chain?

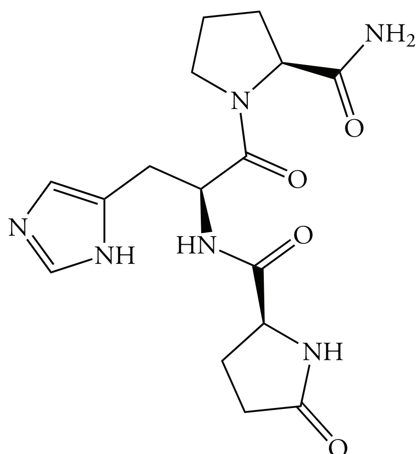
27.33 Identify the amino acids contained in the following tripeptide. Name the compound.



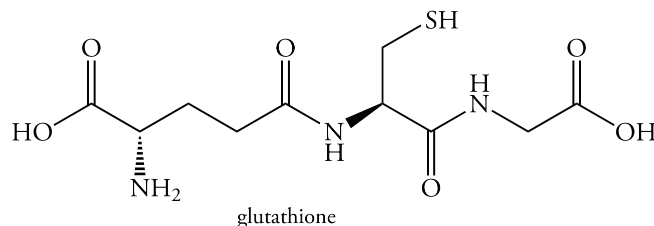
27.34 Identify the amino acids contained in the following tripeptide. Name the compound.



27.35 Thyrotropin-releasing hormone (TRH) causes the release of thyrotropin from the pituitary gland, which then stimulates the thyroid gland. Examine its structure and comment on one unusual structural feature.



27.36 The tripeptide glutathione, which is important in detoxifying metabolites, has an unusual structural feature. Identify it.



glutathione

27.37 How peptide many isomers with the composition Gly_2,Ala_2 are possible?

27.38 How peptide many isomers with the composition Gly_2,Ala,Leu are possible?

Peptide Hydrolysis and Primary Structure Determination

27.39 Assuming that only dipeptides are formed by partial hydrolysis, what is the minimum number that must be identified to establish the amino acid sequence of a pentapeptide?

27.40 Assuming that only tripeptides are formed by partial hydrolysis, what is the minimum number that must be identified to establish the amino acid sequence of an octapeptide?

27.41 The tetrapeptide tuftsin is hydrolyzed to produce Pro-Arg and Thr-Lys. Does this information establish the structure of tuftsin?

27.42 Partial hydrolysis of the octapeptide angiotensin II produces Pro-Phe, Val-Tyr-Ile, Asp-Arg-Val, and Ile-His-Pro. What is its amino acid sequence?

27.43 Treatment of somatostatin with the Edman reagent gives a derivative of alanine. Partial hydrolysis of the polypeptide gives the following oligopeptides. Write the structure of the polypeptide.
I: Phe-Trp II: Lys-Thr III: Thr-Ser-Cys IV: Thr-Phe-Thr-Ser-Cys
V: Asn-Phe-Phe-Trp-Lys VI: Ala-Gly-Cys-Lys-Asn-Phe

27.44 The amino acid composition of the peptide is (Arg_2, Gly, Phe_2, Pro_3, Ser). Treatment of bradykinin with the Edman reagent gives the PTC-derivative of arginine. Partial hydrolysis yields several fragments that include the following oligopeptides. What is the amino acid sequence of bradykinin?
I: Gly-Phe-Ser II: Arg-Pro-Pro-Gly III: Phe-Arg-Ser-Pro-Phe

Enzymatic Hydrolysis of Peptides

27.45 Which of the following tripeptides will be cleaved by trypsin? If cleavage occurs, name the products.
(a) Arg-Gly-Tyr (b) Glu-Asp-Gly (c) Phe-Trp-Ser (d) Ser-Phe-Asp

27.46 Which of the following tripeptides will be cleaved by trypsin? If cleavage occurs, name the products.
(a) Asp-Lys-Ser (b) Lys-Tyr-Cys (c) Asp-Gly-Lys (d) Arg-Glu-Ser

27.47 Indicate which of the tripeptides in Exercise 26.45 will be cleaved by chymotrypsin and name the products.

27.48 Indicate which of the tripeptides in Exercise 26.46 will be cleaved by chymotrypsin and name the products.

27.49 The tetrapeptide tuftsin is hydrolyzed by trypsin to produce Pro-Arg and Thr-Lys. Does this information establish the amino acid sequence of tuftsin?

27.50 The pentapeptide met-enkephalin is hydrolyzed by chymotrypsin to give Met, Tyr, and Gly-Gly-Phe. Does this information establish the amino acid sequence of met-enkephalin?

27.51 The nonapeptide known as the sleep peptide is hydrolyzed by chymotrypsin to produce Ala-Ser-Gly-Glu and Ala-Arg-Gly-Tyr and Trp. What two amino acid sequences are possible for the sleep peptide?

27.52 The sleep peptide is hydrolyzed by trypsin to produce Gly-Tyr-Ala-Ser-Gly-Glu and Trp-Ala-Arg. What is the amino acid sequence of the sleep peptide?

27.53 Feline gastrin, a hormone that stimulates secretion of gastric juice in cats, has the amino acid composition (Ala_2,Asp,Gly_2,$G1u_5$,Leu,Met,Phe,Pro,Trp2,Tyr).
End group analysis shows that the C-terminal and N-terminal amino acids are Phe and Glu, respectively. Hydrolysis with chymotrypsin yields the following four peptides. Write two possible amino acid sequences of feline gastrin.
I: Gly-Trp II: Met-Asp-Phe III: Glu-Gly-Pro-Trp IV: Leu-Glu-Glu-Glu-Glu-Ala-Ala-Tyr

27.54 Corticotropin, a pituitary hormone, stimulates the adrenal cortex. Hydrolysis by chymotrypsin yields six peptides:
I: Arg-Trp II: Ser-Tyr III: Ser-Met-Glu-His-Phe IV: Pro-Leu-Glu-Phe
V: Pro-Asp-Ala-Gly-Glu-Asp-Gln-Ser-Ala-Glu-Ala-Phe
VI: Gly-Lys-Pro-Val-Gly-Lys-Lys-Arg-Pro-Val-Lys-Val-Tyr
Hydrolysis by trypsin produces lysine, arginine, and five peptides:
I: Trp-Gly-Lys II: Pro-Val-Gly
III: Pro-Val-Gly-Lys IV: Ser-Tyr-Ser-Met-Glu-His-Phe-Arg
V: Val-Tyr-Pro-Asp-Ala-Gly-Glu-Asp-Gln-Ser-Ala-Glu-Ala-Phe-Pro-Leu-Glu-Phe
What is the amino acid sequence of corticotropin?

End Group Analysis

27.55 Edman degradation of tuftsin yields Thr as the N-terminal amino acid. Using the information in Exercise 27.49, what is the structure of tuftsin?

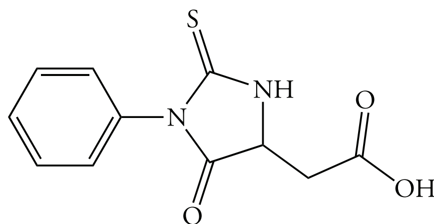
27.56 Hydrolysis of met-enkephalin with CNBr yields the homoserine lactone derivative of methionine and a tetrapeptide. Using the information in Exercise 27.50, what is the structure of met-enkephalin?

27.57 Explain why structure determination of insulin using the Edman method yields two phenylthiohydantoin products.

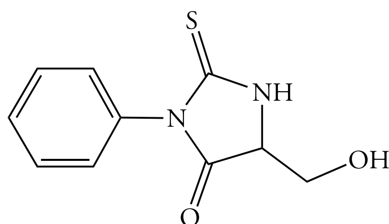
27.58 Cholecystokinin, a peptide that contains 33 amino acids, plays a role in reducing the desire for food, and its production is stimulated by food intake. Its N-terminal amino acid is lysine. Draw the structure of the phenylthiohydantoin product.

27.59 Reaction of angiotensin II with the Edman reagent yields the following product. What information has been established

27.60 Reaction of angiotensin II with the Edman reagent yields the following product. What is the N-terminal amino acid?



27.61 Reaction of corticotropin with the Edman reagent yields the following product. What information has been established?



Protein Structure

27.62 Which of the following amino acids are likely to exist in the interior of a protein dissolved in an aqueous solution?
(a) glycine (b) phenylalanine (c) glutamic acid (d) arginine

26.63 Which of the following amino acids are likely to exist in the interior of a protein dissolved in an aqueous solution?
(a) proline (b) cysteine (c) glutamine (d) aspartic acid

26.64 If a protein is embedded in a hydrophobic lipid bilayer of a biological membrane, which of the amino acids listed in Exercise 26.61 will be in contact with the interior of the bilayer?

26.65 If a protein is embedded in a lipid bilayer, which of the amino acids listed in Exercise 26.62 will be in contact with the interior of the bilayer?

26.66 Noting that proline is a secondary amine, explain how proline can disrupt the a helix of a protein.

26.67 Examine the structures of valine and glutamic acid and suggest a reason why human hemoglobin is affected by the substitution of valine for glutamic acid at position 6 in the β chain.