

# *Chemical Properties of Polypeptides*

---

---

# 1

Virtually every property that characterizes a living organism is affected by proteins. Nucleic acids, also essential for life, encode genetic information—mostly specifications for the structures of proteins—and the expression of that information depends almost entirely on proteins (though some RNA molecules with catalytic activity have been discovered recently).

Life forms make use of many chemical reactions to supply themselves continually with chemical energy and to use it efficiently, but by themselves these reactions could not occur fast enough under physiological conditions (aqueous solution, 37° C, pH 7, atmospheric pressure) to sustain life. The rates of these reactions are increased, by many orders of magnitude, in organisms by the presence of enzymes, which also are proteins. The subject of biochemistry is primarily a study of the roles of enzymes in living systems.

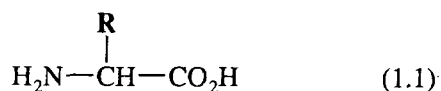
Proteins store and transport a variety of particles ranging from macromolecules to electrons. They guide the flow of electrons in the vital process of photosynthesis; as hormones, they transmit information between specific cells and organs in complex organisms; some proteins control the passage of molecules across the membranes that compartmentalize cells and organelles; proteins function in the immune systems of complex organisms to defend against intruders (the best known are the antibodies); and proteins control gene expression by binding to specific sequences of nucleic acids, thereby turning genes on and off. Proteins are the crucial components of muscles and other systems for converting chemical energy into mechanical energy. They also are necessary for sight, hearing, and the other senses. And many proteins are simply structural, providing the filamentous architecture within cells and the materials that are used in hair, nails, tendons, and bones of animals.

In spite of these diverse biological functions, proteins are a relatively homogeneous class of molecules. All are the same type of linear polymer, built of various combinations of the same 20 amino acids. They differ only in the sequence in which the amino acids are assembled into polymeric chains. The secret to their functional diversity lies partly in the chemical diversity of the amino acids but primarily in the diversity of the three-dimensional structures that these building blocks can form, simply by being linked in different sequences. The awesome functional properties of proteins can be understood only in terms of their relationship to the three-dimensional structures of proteins. That relationship is the topic of this volume.

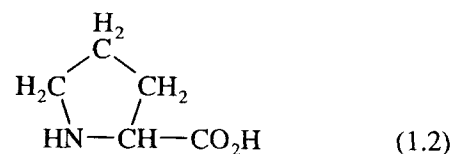
## 1.1 The Polymeric Nature of Proteins

Proteins are more complex than most linear polymers in that they can incorporate 20 different monomers in their construction instead of only one or two. In another sense, proteins are structurally less complex. Most chemical polymers are synthesized by chemically polymerizing a mixture of monomers, thereby producing a distribution of chain lengths and, if more than one type of monomer is present, an approximately random sequence of monomers. Natural proteins, however, are linear and unbranched and have precise lengths and exact sequences of amino acids. Indeed, it is only the differences in length and sequence that distinguish one protein from any other and that make possible a diversity of structures and functions. Most importantly, the linear polymeric chain of almost every natural protein has the crucial property of being able to assume a specific three-dimensional folded conformation, as will be described in later chapters.

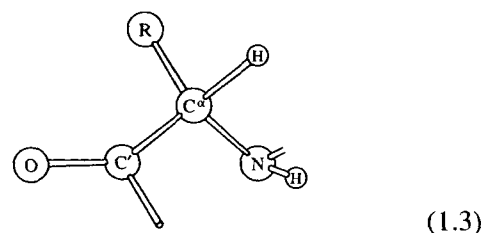
Of the 20 amino acids usually found in proteins, 19 have the general structure



and differ only in the chemical structures of the side chain **R**. The 20th natural amino acid, proline, is similar, but its side chain is bonded to the nitrogen atom to give the imino acid:



Except in glycine, where the side chain is only a hydrogen atom, the central carbon atom is asymmetric and is always the L isomer:



Unless indicated otherwise, all amino acids in this volume are L isomers.

The structures of the side chains of the 20 amino acids are illustrated in Figure 1.1. The central carbon atom depicted in Equation (1.1) is designated as  $\alpha$ , and the atoms of the side chains are commonly designated  $\beta$ ,  $\gamma$ ,  $\delta$ ,  $\epsilon$ , and  $\zeta$ , in order away from the  $\alpha$  carbon atom. Chemical groups, however, are usually designated by the carbon atom to which they are bonded; hence, the  $\zeta$ N atom of a Lys residue is part of the  $\epsilon$ -amino group. Table 1.1 and Figure 1.1 also give the three- and one-letter abbreviations commonly used for designating the amino acids when they are incorporated into proteins. Three-letter abbreviations are used throughout this volume because their designations are obvious. The relationships of the one-letter codes to the amino acids are somewhat less obvious, but they are commonly used in compilations of long sequences because they save space and are less likely to be confused (e.g., Gln, Glu, and Gly can easily be confused but not Q, E, and G).

**FIGURE 1.1**

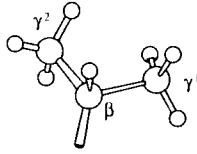
Side chains of the 20 amino acids that occur naturally in proteins. Small unlabeled spheres are hydrogen atoms, and large unlabeled spheres are carbon atoms; other atoms are labeled. Double bonds are black, and partial double bonds are shaded. In the case of Pro, the bonds of the polypeptide backbone are included and are black. Below the name of the amino acid are the three-letter and the one-letter abbreviations commonly used. Note that isoleucine and threonine have asymmetric centers in their side chains, and only the isomer illustrated is used biologically.



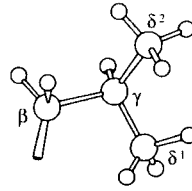
Glycine  
Gly  
G



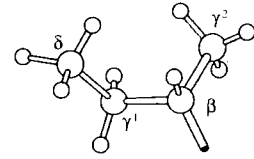
Alanine  
Ala  
A



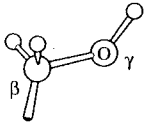
Valine  
Val  
V



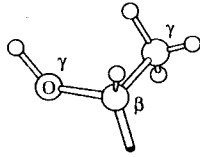
Leucine  
Leu  
L



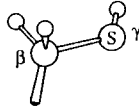
Isoleucine  
Ile  
I



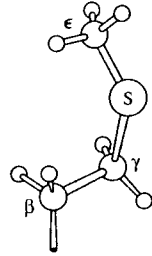
Serine  
Ser  
S



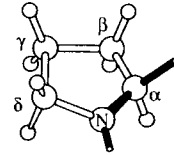
Threonine  
Thr  
T



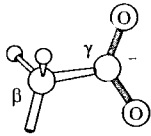
Cysteine  
Cys  
C



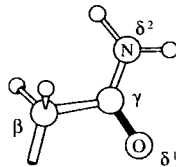
Methionine  
Met  
M



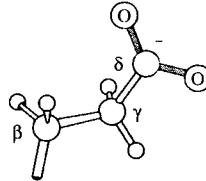
Proline  
Pro  
P



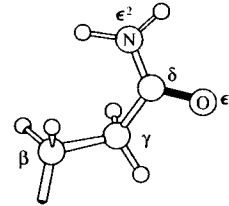
Aspartic acid  
Asp  
D



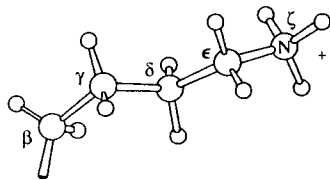
Asparagine  
Asn  
N



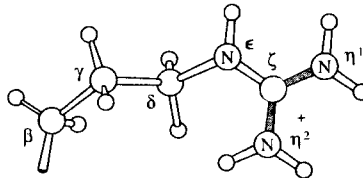
Glutamic acid  
Glu  
E



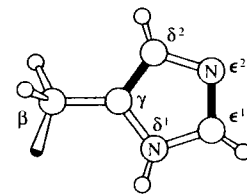
Glutamine  
Gln  
Q



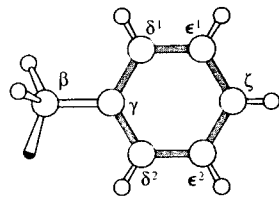
Lysine  
Lys  
K



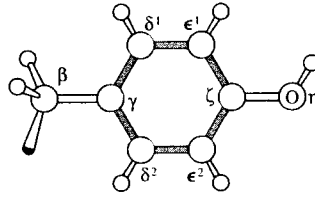
Arginine  
Arg  
R



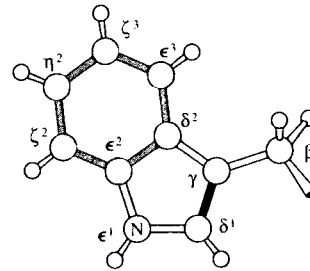
Histidine  
His  
H



Phenylalanine  
Phe  
F



Tyrosine  
Tyr  
Y



Tryptophan  
Trp  
W

**Table 1.1** Properties of Individual Amino Acid Residues

Residue	Residue mass <sup>a</sup> (daltons)	Van der Waals volume <sup>b</sup> (Å <sup>3</sup> )	Frequency in proteins <sup>c</sup> (%)
Ala (A)	71.09	67	8.3
Arg (R)	156.19	148	5.7
Asn (N)	114.11	96	4.4
Asp (D)	115.09	91	5.3
Cys (C)	103.15	86	1.7
Gln (Q)	128.14	114	4.0
Glu (E)	129.12	109	6.2
Gly (G)	57.05	48	7.2
His (H)	137.14	118	2.2
Ile (I)	113.16	124	5.2
Leu (L)	113.16	124	9.0
Lys (K)	128.17	135	5.7
Met (M)	131.19	124	2.4
Phe (F)	147.18	135	3.9
Pro (P)	97.12	90	5.1
Ser (S)	87.08	73	6.9
Thr (T)	101.11	93	5.8
Trp (W)	186.21	163	1.3
Tyr (Y)	163.18	141	3.2
Val (V)	99.14	105	6.6
Weighted average <sup>d</sup>	119.40	161	

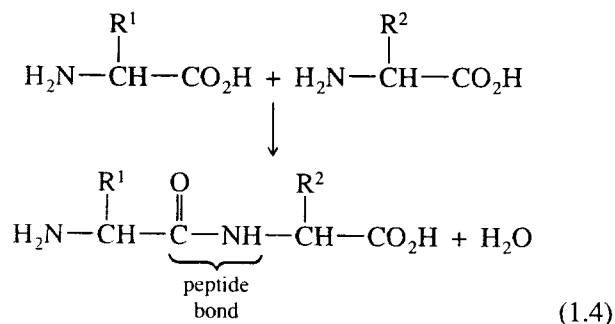
<sup>a</sup> Molecular weight of nonionized amino acid minus that of water.

<sup>b</sup> Volume enclosed by van der Waals radii, calculated with volumes of atoms given by F. M. Richards, *J. Mol. Biol.* 82:1–14 (1974).

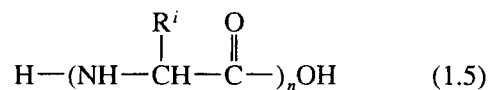
<sup>c</sup> Frequency of occurrence of amino acid residue in primary structures of 1021 unrelated proteins of known sequence (P. McCaldon and P. Argos, *Proteins* 4:99–122, 1988).

<sup>d</sup> Weighted by frequency of occurrence in proteins, to give the value for average residue in globular proteins.

The 20 amino acids are linked into proteins by the **peptide bond**, as illustrated here by the condensation of two amino acids:



Generally, between 50 and 3000 such amino acids are linked in this way to form a typical linear **polypeptide chain**. The polypeptide backbone is a repetition of the basic unit common to all amino acids. When the side chain is included, this unit is described as an amino acid **residue**:



Some of the most important properties of the amino acid residues are listed in Table 1.1.

All proteins and polypeptides have this basic structure. Proteins differ only in the number of amino acids linked together ( $n$  in Eq. 1.5) and the sequence in which the various amino acids occur. The sequence of amino acids in a polypeptide chain generally identifies a protein unambiguously.

Sequences of amino acids in proteins are usually written with either the three-letter or the one-letter abbreviations, starting with the N-terminal residue, which is at the left and is considered the first residue of the polypeptide chain. Hyphens are inserted between residues to indicate known peptide bonds, and other punctuation is used if the sequence is not known entirely: parentheses around segments of uncertain sequence, dots separating residues whose positions are almost certain, and commas between amino acid residues of unknown sequence. If it is not known whether a residue is Glu or Gln, the appropriate abbreviation is Glx or Z; for Asp or Asn, Asx or B is used.

Amino acid residues in proteins are properly referred to by adding *-yl* to the ends of their amino acid names (e.g., glycyl or alanyl residues). This complication is avoided by using the one- or three-letter abbreviations.

The following terms are used to describe the various types of polymerized amino acids:

- **Peptide** A short chain of residues with a defined sequence. There is no maximum number of residues in a peptide, but the term is appropriate to a chain if its physical properties are those expected from the sum of its amino acid residues and if there is no fixed three-dimensional conformation.
- **Polypeptide** A longer chain, usually of defined sequence and length.
- **Polyamino acids** Random sequences of varying lengths generally resulting from nonspecific polymerization of one or a few amino acids.
- **Protein** Usually reserved for those polypeptides that occur naturally and have a definite three-dimen-



**Table 1.2** Intrinsic  $pK_a$  Values of Ionizable Groups Found in Proteins

Group	Observed $pK_a^a$
$\alpha$ -Amino	6.8–8.0
$\alpha$ -Carboxyl	3.5–4.3
$\beta$ -Carboxyl (Asp)	3.9–4.0
$\gamma$ -Carboxyl (Glu)	4.3–4.5
$\delta$ -Guanido (Arg)	12.0
$\epsilon$ -Amino (Lys)	10.4–11.1
Imidazole (His)	6.0–7.0
Thiol (Cys)	9.0–9.5
Phenolic hydroxyl (Tyr)	10.0–10.3

<sup>a</sup> The ranges of values are given by different model compounds used to represent an isolated amino acid residue. The values for the terminal  $\alpha$ -amino and  $\alpha$ -carboxyl groups especially depend on the identity of the terminal residue. Values from C. Tanford, *Adv. Protein Chem.* 17:69–165 (1962); A. Bundi and K. Wüthrich, *Biopolymers* 18:285–297 (1979); J. B. Matthew et al., *CRC Crit. Rev. Biochem.* 18:91–197 (1985).

Resonance of the peptide bond (Eq. 1.7) tends to redistribute its electrons, and the polypeptide backbone is correspondingly polar. The H and N atoms appear to have, respectively, positive and negative equivalent charges of 0.20 electron, whereas C and O, respectively, have positive and negative equivalent charges of 0.42 electron. This gives the peptide bond a substantial permanent dipole moment of about 3.5 Debye units (corresponding to one unit charge separated by 0.73 Å; see Sec. 4.1.2.b). Note that the polypeptide backbone of each residue contains one potent hydrogen bond donor, —NH—, and a hydrogen bond acceptor, carbonyl —CO—. Chapters 5 and 6 will show this to be crucial for the three-dimensional architectures of proteins.

The other two types of bonds of the peptide backbone behave as normal C—C and C—N single bonds.

The peptide backbone is not very reactive chemically. The only groups usually ionized are the terminal  $\alpha$ -amino and carboxyl groups, which normally have  $pK_a$  values of about 7.4 and 3.9, respectively (Table 1.2), depending on the nature of the terminal amino acid residue. A proton is added or lost to internal peptide bonds only at extremes of pH. The apparent  $pK_a$  value of the amide NH for deprotonation is between 15 and 18 and is in the region of —8 to —12 for protonation. The oxygen atom of the carbonyl group is protonated more readily, with an apparent  $pK_a$  of about —1. These properties facilitate the exchange of hydrogen isotopes between the backbone and aqueous solvents, which is important to the study of protein conformational fluctuations in solution (Chap. 7).

At pH values for which protonation or deprotonation of the peptide bond becomes significant, the polypeptide chain is usually hydrolyzed to the substituent amino acids. For example, the standard method for determining the amino acid composition of proteins or peptides is to maintain them at about 105°C for about 24 h in 6 M HCl. The peptide bond is hydrolyzed very slowly in more physiological conditions, its half-life having been measured to be approximately 7 years at neutral pH and room temperature; but this decreases to about 1 min at 250°C.

Other chemical alterations of the peptide chain require either drastic conditions or the very close proximity of certain reactive groups attached to the amino acid side chains. Such reactions are useful in the laboratory for selective cleavage of the chain at certain residues (see Sec. 1.6.2.e).

### References

- Hydrogen exchange. S. W. Englander et al. *Ann. Rev. Biochem.* 41:903–924 (1972).
- The  $\alpha$ -helix as an electric macro-dipole. A. Wada. *Adv. Biophys.* 9:1–63 (1976).
- Coordinating properties of the amide bond. Stability and structure of metal ion complexes of peptides and related ligands. H. Sigel and R. B. Martin. *Chem. Rev.* 82:385–426 (1982).
- Hydrolytic stability of biomolecules at high temperatures and its implication for life at 250°C. R. H. White. *Nature* 310:430–432 (1984).
- The role of the  $\alpha$ -helix dipole in protein function and structure. W. G. J. Hol. *Prog. Biophys. Mol. Biol.* 45:149–195 (1985).
- Hydrolysis of a peptide bond in neutral water. D. Kahne and W. C. Still. *J. Amer. Chem. Soc.* 110:7529–7534 (1988).

## 1.3 Amino Acid Residues

The 20 different amino acid side chains possess a variety of chemical properties. This variety is greatly enhanced when the various groups are combined in various sequences in a single molecule, which gives a protein properties far beyond those of simpler molecules. The chemical properties of a protein molecule are far more complex than the sum of the properties of its constituent amino acids, but the properties of the 20 side chains are the starting point for understanding proteins. For example, if the thiol group of a Cys residue is important in the function of a protein, that function will be abolished by a reagent that reacts chemically with thiol groups (see Sec. 1.3.10). For such a test to be meaningful, of course,

one must be certain that the reagent does not also react with other groups on the protein.

Each of the side chains will now be discussed briefly and their chemical properties described. Remember, however, that residues in biologically active proteins may have chemical or physical properties very different from those described here. The chemical reactions undergone by the various side chains that are described here are those that are at least moderately selective and sufficiently specific to be useful in determining the roles of the residues in proteins.

Some pertinent chemical and physical properties of the amino acid residues are summarized in Tables 1.1 and 1.2.

### References

- Chemistry of the Amino Acids.* J. P. Greenstein and M. Winitz. John Wiley & Sons, New York, 1961.
- Chemical Modification of Proteins.* G. E. Means and R. E. Feeney. Holden-Day, San Francisco, 1971.
- The chemical modification of proteins by group-specific and site-specific reagents. A. N. Glazer. In *The Proteins*, 3rd ed., H. Neurath and R. L. Hill, eds., vol. 2, pp. 1–103. Academic Press, New York, 1976.
- Chemistry and Biochemistry of Amino Acids.* G. C. Barrett, ed. Chapman and Hall, New York, 1985.
- Amino acid side chain parameters for correlation studies in biology and pharmacology. J. L. Fauchere et al. *Intl. J. Peptide Protein Res.* 32:278–279 (1988).
- Chemical modification. T. Imoto and H. Yamada. In *Protein Function: A Practical Approach*, T. E. Creighton, ed., pp. 247–277. IRL Press, Oxford, 1989.

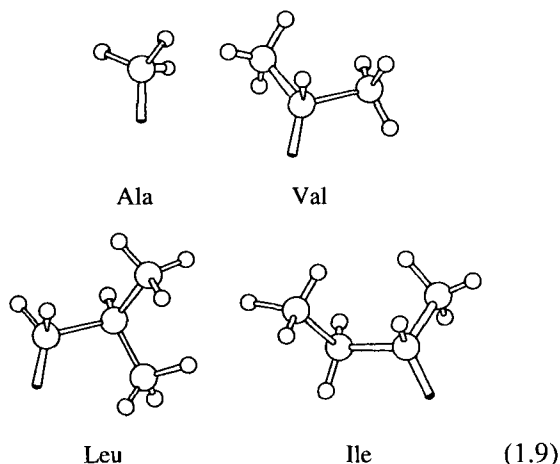
### 1.3.1 Gly

Glycine is the simplest amino acid, with only a hydrogen atom for a side chain. The absence of a larger side chain gives the polypeptide backbone at Gly residues much greater conformation flexibility than at other residues (discussed further in Chap. 5).

The alpha carbon atom of Gly is not asymmetric because it has two H atoms, so this amino acid does not occur as D or L isomers.

### 1.3.2 The Aliphatic Residues: *Ala, Val, Leu, Ile*

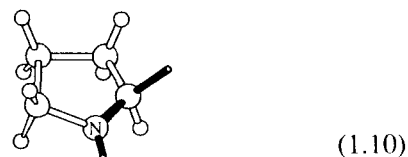
These amino acid residues have no reactive groups on their side chains, only inert methylene ( $-\text{CH}_2-$ ) and methyl ( $-\text{CH}_3$ ) groups. They have the important property, however, of *not* interacting favorably with water.



Instead, they interact much more favorably with each other and with other nonpolar atoms than with water, which is one of the main factors in stabilizing the folded conformations of proteins (Sec. 4.3). Consequently, the aliphatic residues can be considered the “bricks” around which the functional parts of a protein molecule are assembled. The variety of their shapes is important for this structural role.

Note that the Ile side chain has an extra center of asymmetry, at  $\text{C}^\beta$ , and that only the one isomer occurs naturally and is incorporated into proteins. It is curious that there are amino acids in proteins with one, three, or four carbon atoms in their side chains but not two. The corresponding amino acid,  $\alpha$ -amino butyric acid, is not used in protein biosynthesis.

### 1.3.3 The Cyclic Imino Acid: *Pro*

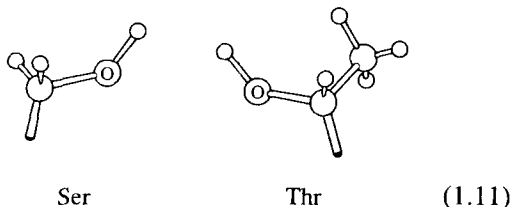


The side chain of proline is aliphatic like those of the preceding amino acids, with no functional groups. It is unique, however, in that it is bonded covalently to the nitrogen atom of the peptide backbone, which is indicated by the solid bonds in Equation (1.10). Therefore, the peptide backbone at Pro residues has no amide hydrogen for use as a donor in hydrogen bonding or in resonance stabilization of the peptide bond of which it is part. The cyclic five-membered ring also imposes rigid constraints on rotation about the  $\text{N}-\text{C}^\alpha$  bond of the backbone, and the peptide bond preceding a Pro residue is more likely to adopt the *cis* configuration (see Secs. 1.2, 5.2.1). Pro residues, consequently, have significant

effects on the conformation of the polypeptide backbone (discussed further in Chaps. 5 and 6).

The five-membered pyrrolidine ring of the Pro residue is invariably puckered, with the  $C^\alpha$ ,  $C^\beta$ ,  $C^\delta$ , and N atoms approximately coplanar but with the  $C^\gamma$  atom displaced from the plane by about 0.5 Å.

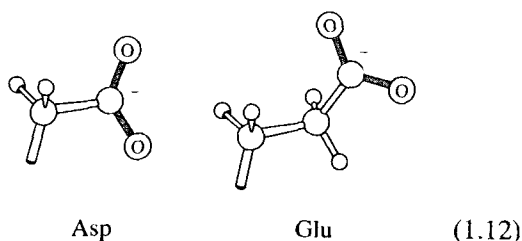
### 1.3.4 The Hydroxyl Residues: Ser and Thr



The side chains of Ser and Thr are small and aliphatic, except for the presence of a polar hydroxyl group on each. These hydroxyl groups are normally no more reactive chemically than ethanol, so there are few chemical reactions in which they participate that are useful with proteins. The only reaction that occurs readily is acetylation with acetyl chloride in aqueous trifluoroacetic acid.

Note that the side chain of Thr, like that of Ile, has a center of asymmetry at  $C^\beta$  and that only the one isomer occurs naturally.

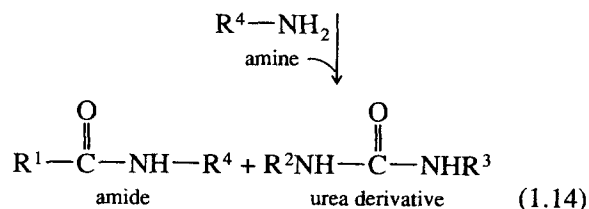
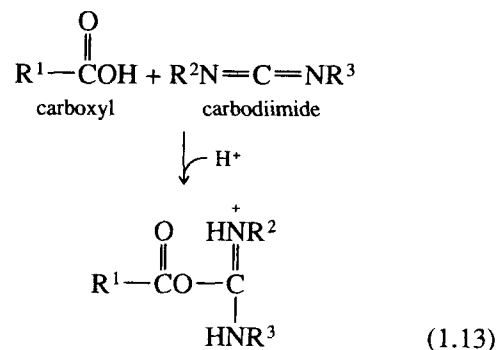
### 1.3.5 The Acidic Residues: Asp and Glu



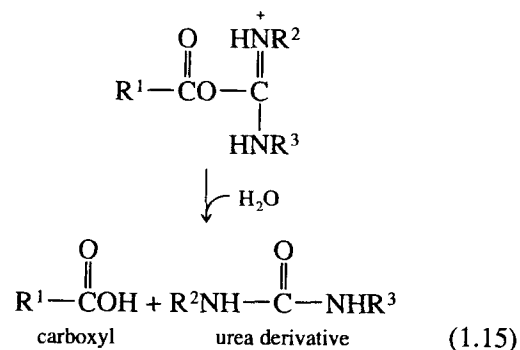
The side chains of Asp and Glu differ only in having one and two methylene groups, respectively. It might be thought, therefore, that they would be very similar chemically and functionally in proteins, but this is not the case. The slight difference in length of the side chains causes them to have different tendencies in their chemical interactions with the peptide backbone. Consequently, Asp and Glu residues have markedly different effects on the conformation and chemical reactivity of the peptide backbone.

The carboxyl groups of Asp and Glu are normally no more reactive than are those of corresponding organic molecules such as acetic acid. They can be chemi-

cally esterified, coupled with amino or other nucleophiles, and reduced to alcohols, but these reactions are somewhat harsh chemically. The most useful chemical modification of the carboxyl groups is their coupling to amines, which is catalyzed by carbodiimides:



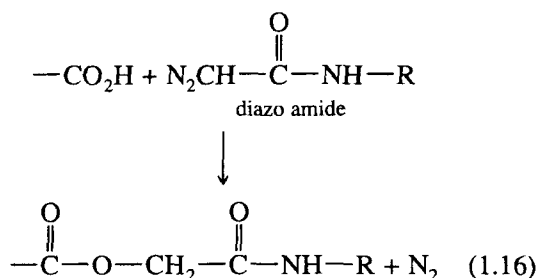
The carboxyl groups are first activated by reaction with the carbodiimide (Eq. 1.13); then they react with an added amine to generate the amide (Eq. 1.14). A competing reaction is hydrolysis of the activated carbodiimide to regenerate the carboxyl and to produce the urea derivative of the carbodiimide:



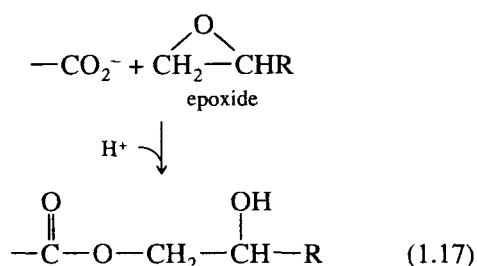
The carboxyl groups that are regenerated can react with another molecule of the carbodiimide. The carbodiimide reagents most useful with proteins have  $\text{R}_2$  and  $\text{R}_3$  groups that make the reagent soluble in water.

The carboxyl groups of Asp and Glu side chains ionize with intrinsic  $\text{pK}_a$  values of 3.9 and 4.3, respectively (Table 1.2), so these residues are ionized and very polar under physiological conditions. They can also be effective chelators of certain metal ions when held in appropriate proximity to the ions (Sec. 8.3.4.a).

The ionization state of a carboxyl can sometimes be determined by its susceptibility to modification by different reagents. Diazo compounds, such as diazoacetate esters and amides, react with the nonionized form:



In contrast, epoxides react with the ionized form:

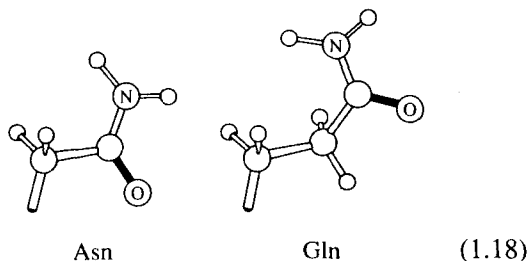


The carboxyl groups of Glu and Asp residues differ only marginally in their physical and chemical properties from the terminal  $\alpha$ -carboxyl group of the polypeptide chain.

### References

- Esterification. P. E. Wilcox. *Methods Enzymol.* 25:596-615 (1972).  
Carbodiimide modification of proteins. K. L. Carraway and D. E. Koshland. *Methods Enzymol.* 25:616-623 (1972).

### 1.3.6 The Amide Residues: Asn and Gln



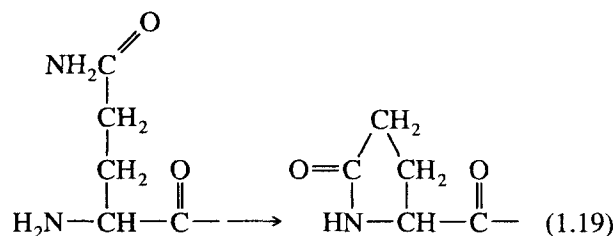
These residues are the amide forms of Asp and Glu, but the corresponding amino acids occur naturally and are incorporated directly into proteins; Asn and Gln residues do not occur as a result of amidation of Asp and Glu residues in proteins.

The Asn and Gln amide side chains do not ionize and are not very reactive chemically, but they are polar,

being both hydrogen-bond donors and acceptors. The amide groups are labile at extremes of pH and at high temperatures, and these residues can deaminate to Asp and Glu residues. At alkaline pH, the Asn residue is much more labile than the Gln residue because its side chain is sterically suited to interact with the  $-\text{NH}-$  group of the following residue to form transiently a cyclic succinimide derivative (Fig. 1.3). This derivative can undergo racemization and hydrolysis to cleave the peptide chain or to produce a mixture of D and L isomers of Asp and isoAsp residues. In an isoAsp residue, the peptide bond of the backbone is through the side-chain carboxyl group rather than the usual  $\alpha$ -carboxyl. The deamidation reaction of Asn occurs 30-50 times more rapidly if the following residue is Gly because the absence of a side chain favors succinimide formation. The rate of this reaction also depends markedly on the conformation of the polypeptide backbone because only some conformations permit succinimide formation. This reaction may be a limiting factor in the longevities of some proteins (Chap. 10).

At acidic pH values, deamidation occurs by other mechanisms, and Gln is more reactive than Asn.

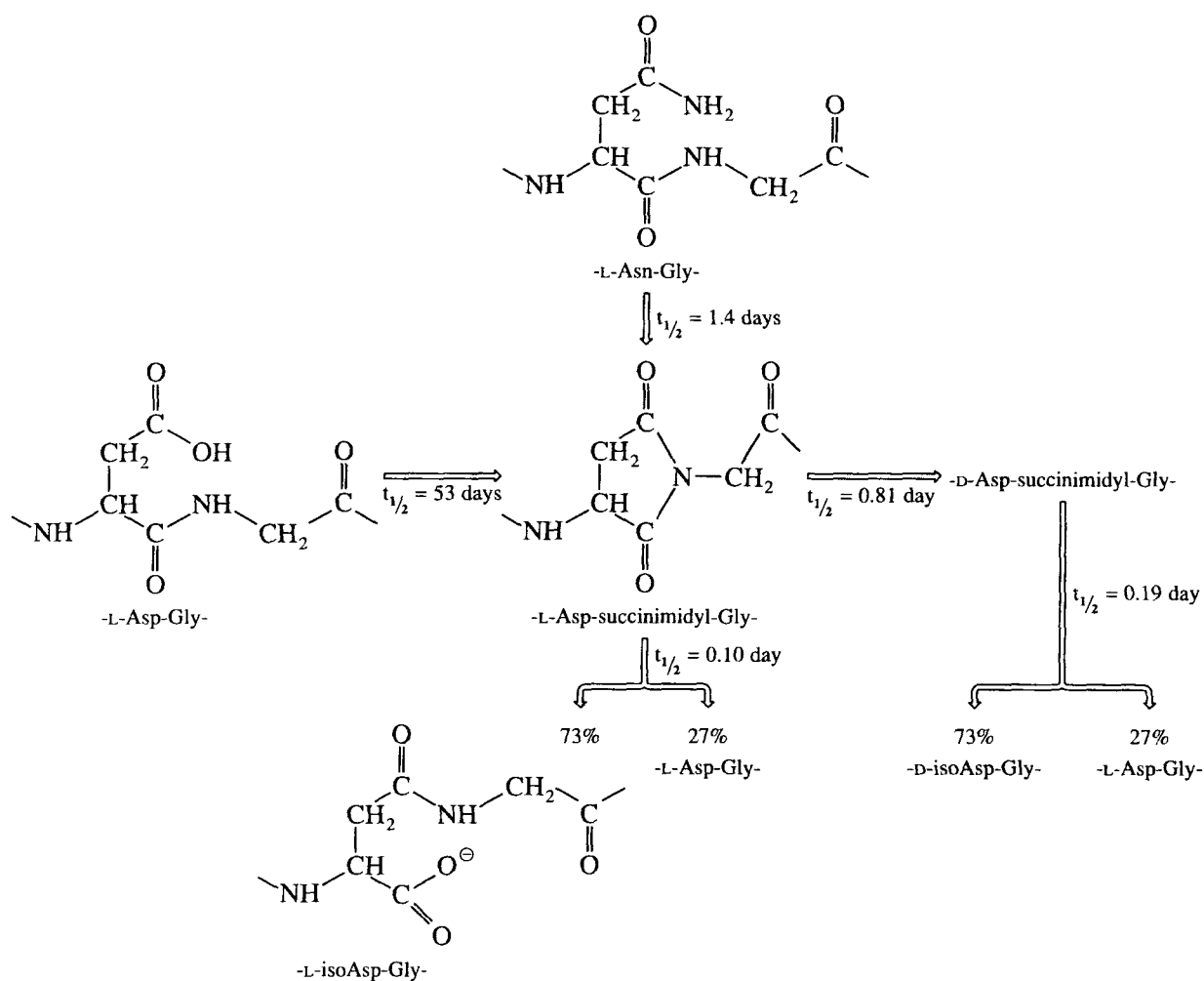
When Gln residues are at the N-terminus of the peptide chain, they spontaneously cyclize:



The resulting residue of pyrrolidone carboxylic acid makes the amino terminus unreactive in most procedures for determining amino acid sequences (Sec. 1.6.2); but it can be removed by the enzyme pyroglutaminyl amino peptidase.

### References

- Rates of nonenzymatic deamidation of glutamyl and asparaginyl residues in pentapeptides. A. B. Robinson et al. *J. Amer. Chem. Soc.* 95:8156-8159 (1973).  
Deamidation, isomerization, and racemization at asparaginyl and aspartyl residues in peptides. Succinimide-linked reactions that contribute to protein degradation. T. Geiger and S. Clarke. *J. Biol. Chem.* 262:785-794 (1987).  
Effect of protein conformation on rate of deamidation: ribonuclease A. S. J. Wearne and T. E. Creighton. *Proteins: Struct. Funct. Genet.* 5:8-12 (1989).  
Succinimide formation from aspartyl and asparaginyl peptides as a model for the spontaneous degradation of pro-

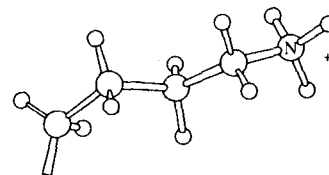


**FIGURE 1.3**

Spontaneous formation of peptide succinimides and the products of hydrolysis from Asp and Asn residues. The half-times shown for the various reactions are for the model peptides Val-Tyr-Pro-Asn-Gly-Ala and Val-Tyr-Pro-Asp-Gly-Ala at 37°C and pH 7.4. When the Gly residue is replaced by Leu, the rate of succinimide formation is 50 times slower. (From S. Clarke, *Int. J. Peptide Protein Res.* 30:808–821, 1987.)

teins. R. C. Stephenson and S. Clarke. *J. Biol. Chem.* 264:6164–6170 (1989).

Identification of an isoaspartyl linkage formed upon deamidation of bovine calbindin  $D_{9k}$  and structural characterization by 2D  $^1\text{H}$  NMR. W. J. Chazin et al. *Biochemistry* 28:8646–8653 (1989).



Lys (1.20)

### 1.3.7 The Basic Residues: Lys and Arg

#### a. Lys

The side chain of Lys is a hydrophobic chain of four methylene groups capped by an amino group that ionizes with an intrinsic  $\text{pK}_a$  value of 11.1 in the absence of

perturbing factors, so it is ionized under most physiological conditions. There is always a finite fraction of nonionized amino groups, however, which are potent nucleophiles. Consequently, the amino groups of Lys residues readily undergo a variety of acylation, alkylation, arylation, and amidation reactions. The rates of such reactions usually increase with increasing pH due



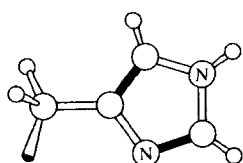


This reaction, however, is often accompanied by cleavage of the polypeptide backbone (Sec. 1.6.2.a).

### References

- Reversible blocking at arginine by cyclohexanedione. E. L. Smith. *Methods Enzymol.* 47:156–161 (1977).  
 Chemical modification of peptides by hydrazine. A. Honegger et al. *Biochem. J.* 199:53–59 (1981).  
 Modification of available arginine residues in proteins by *p*-hydroxyphenylglyoxal. R. B. Yamasaki et al. *Anal. Biochem.* 109:32–40 (1980).

### 1.3.8 His



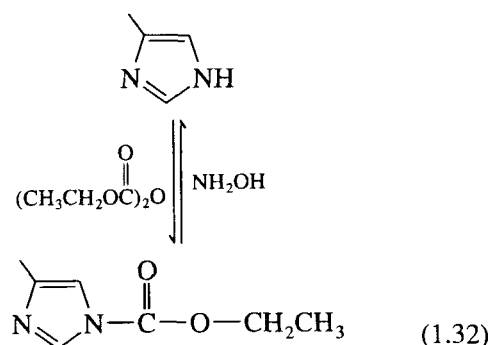
His (1.31)

The imidazole side chain of His residues possesses several special properties that make it extremely effective as a nucleophilic catalyst. It is an amine, which is much more reactive than hydroxide ion in terms of its basicity. Furthermore, it is a tertiary amine, which is intrinsically more nucleophilic than primary or secondary amines. The enhanced reactivity of tertiary amines is usually canceled by their greater steric hindrance, but in imidazole the atoms bonded to the nitrogens are held back in a five-membered ring and cause relatively little steric hindrance. Imidazole has a  $pK_a$  value near 7, so it is one of the strongest bases that can exist at neutral pH. A weaker base would have a lower nucleophilic reactivity, whereas a stronger base would be protonated to a greater extent at neutral pH and would be correspondingly less reactive.

In the nonionized form of the imidazole ring, the nitrogen with the hydrogen atom is an electrophile and donor for hydrogen bonding, and the other nitrogen atom is a nucleophile and acceptor for hydrogen bonding. Consequently, this one side chain is extremely versatile—almost the chemical equivalent of being ambidextrous.

The imidazole group is, in principle, capable of undergoing numerous reactions, but most of these reactions occur more readily with amino and thiol groups; very few are suitable for modifying His residues specifically. The classical approach to modifying His residues was oxidation by light in the presence of dye sensitizers such as methylene blue or rose bengal. The reaction products are probably an aspartic acid residue and urea,

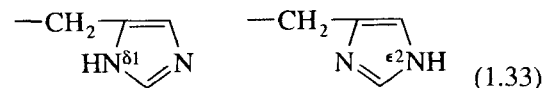
but the reaction is not well characterized. The most specific reaction is that with diethylpyrocarbonate, which can be reversed by hydroxylamine:



Reactions with amino groups can be minimized by carrying out the reaction at acidic pH.

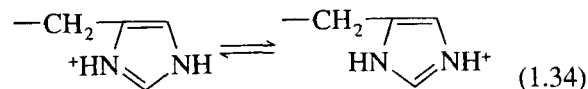
The two nitrogen atoms of the His side chain are designated here as  $\delta 1$  and  $\epsilon 2$ , but they are also known, respectively, as  $\pi$  and  $\tau$  or as N-1 and N-3. The last designation is often ambiguous because biochemists usually assign the number 1 to the nitrogen atom adjacent to the side chain whereas organic chemists tend to designate this atom as 3.

The nonionized imidazole ring can exist as two tautomers, with the hydrogen atom on either the  $\delta 1$  or the  $\epsilon 2$  nitrogen atom:

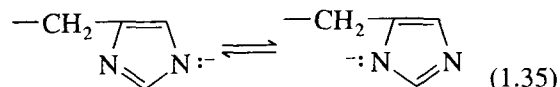


$^{13}\text{C}$ -NMR studies have shown that in model peptides the hydrogen atom is usually predominantly on the  $\epsilon 2$  nitrogen atom, which has a  $pK_a$  value about 0.6 pH unit higher than that of the  $\delta 1$  atom. The position of the hydrogen atom depends on the relative affinities of the two nitrogen atoms for protons, however, and can vary with conditions in the local environment; both forms are found in proteins.

The His side chain is readily protonated, with a  $pK_a$  value near 7 at the second N atom, which destroys its nucleophilicity. The positive charge is shared by the two N atoms by resonance:



Both of the imidazole nitrogen atoms can also be deprotonated simultaneously, with an apparent  $pK_a$  value of about 14.4, to give the aromatic anion:



**Table 1.3** Spectroscopic Properties of the Aromatic Amino Acids at Neutral pH

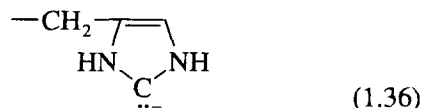
	Absorbance <sup>a</sup>		Fluorescence Emission <sup>b</sup>	
	$\lambda_{\max}$ (nm)	Molar absorbance ( $M^{-1}cm^{-1}$ )	$\lambda_{\max}$ (nm)	Quantum yield
Phenylalanine	257.4	197	282	0.04
Tyrosine	274.6	1420	303	0.21
Tryptophan	279.8	5600	348	0.20

<sup>a</sup> From J. E. Bailey, Ph.D. thesis, London University, 1966.

<sup>b</sup> From F. W. J. Teale and G. Weber, *Biochem. J.* 65:476–482 (1957).

This anion would be expected to be a potent nucleophile but is rarely present in substantial quantities.

The C<sup>ε1</sup> atom (often designated as C-2) is observed to exchange its hydrogen atom slowly with the solvent, indicating that it has a very small probability of being deprotonated:



This exchange reaction provides a useful probe of the environments of His residues in proteins (Chap. 7).

His residues are especially useful in <sup>1</sup>H-NMR studies of proteins because the hydrogen atom on the C<sup>ε1</sup> atom is usually well resolved from the multitude of resonances of the other hydrogen atoms in proteins in one-dimensional spectra. Its resonance is also usually shifted by about 1 ppm to a lower field strength upon protonation of the side chain, often making it relatively easy to determine the pK<sub>a</sub> values of individual His residues even in large proteins.

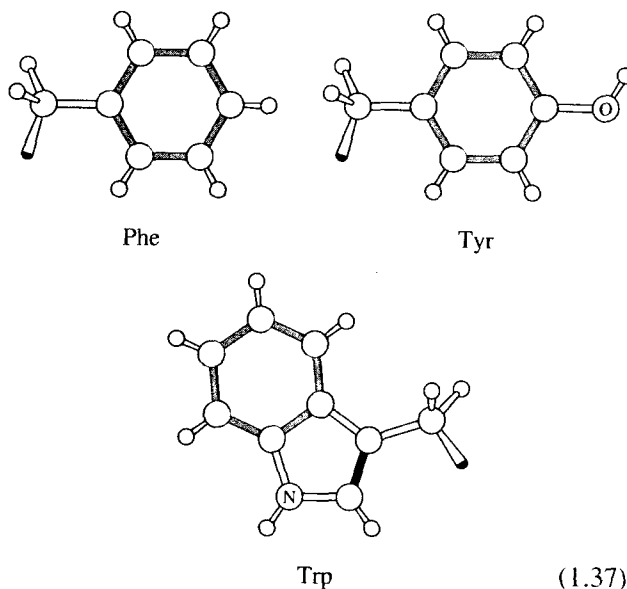
### References

- The roles of imidazole in biological systems. E. A. Barnard and W. D. Stein. *Adv. Enzymol.* 20:51–110 (1959).
- Titration behaviour and tautomeric states of individual histidine residues of myoglobin. D. J. Wilbur and A. Allerhand. *J. Biol. Chem.* 252:4968–4975 (1977).
- Modification of histidyl residues in proteins by diethylpyrocarbonate. E. W. Miles. *Methods Enzymol.* 47:431–442 (1977).
- Tautomeric states of the histidine residues of bovine pancreatic ribonuclease A. D. E. Walters and A. Allerhand. *J. Biol. Chem.* 255:6200–6204 (1980).
- <sup>1</sup>H-NMR study on the tautomerism of the imidazole ring of

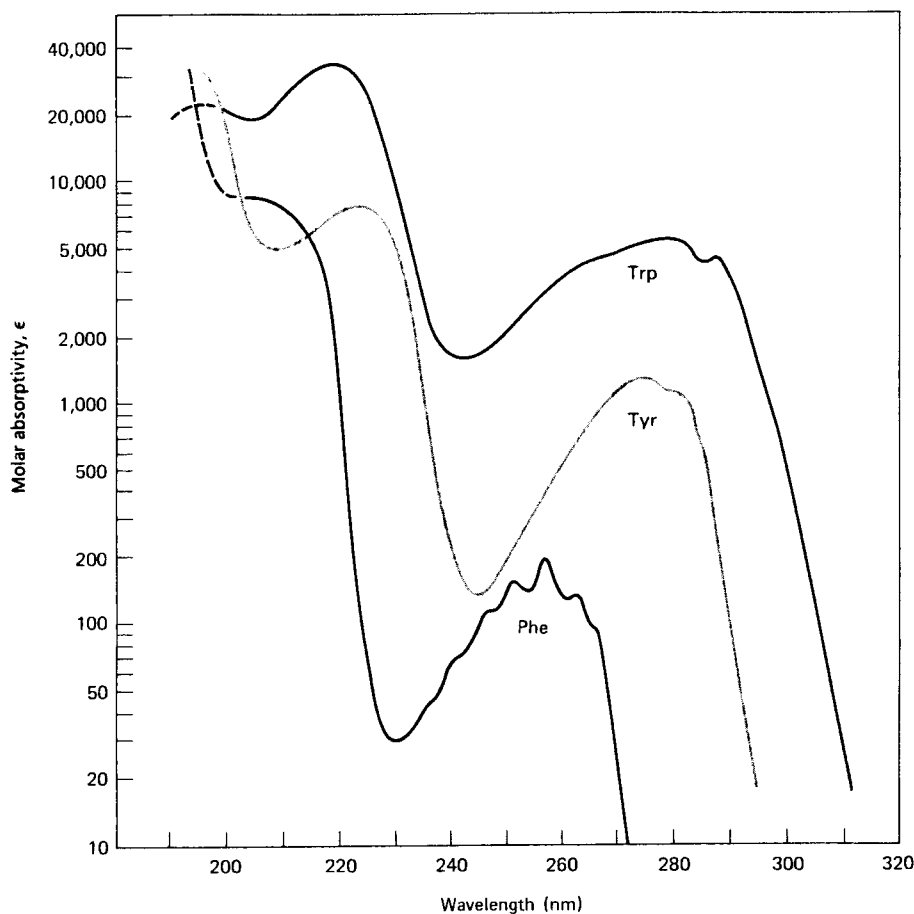
histidine residues. I. Microscopic pK values and molar ratios of tautomers in histidine-containing peptides. M. Tanokura. *Biochim. Biophys. Acta* 742:576–585 (1983).

Chemical modification of pig kidney 3,4-dihydroxyphenylalanine decarboxylase with diethyl pyrocarbonate. P. Dominici et al. *J. Biol. Chem.* 260:10583–10589 (1985).

### 1.3.9 The Aromatic Residues: Phe, Tyr, and Trp



These aromatic side chains are responsible for most of the ultraviolet absorbance and fluorescence properties of proteins. The spectral properties of the side chains (Table 1.3 and Fig. 1.4) are very sensitive to the immedi-



**FIGURE 1.4**

Ultraviolet absorbance spectra of the aromatic amino acids at pH 6. Incorporation of these amino acids into peptides has little direct effect on the absorbance properties of their side chains, unless they are placed into different environments. (From D. B. Wetlaufer, *Adv. Protein Chem.* 17:303–390, 1962.)

ate environment of the side chains, which makes them useful probes of protein structure (Chaps. 5 and 7).

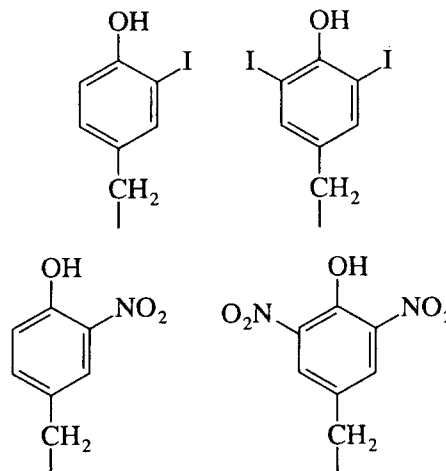
**a. Phe**

The aromatic ring of Phe residues is chemically comparable to that of benzene or toluene. Consequently, it is nonpolar and is chemically reactive only under extreme conditions that are not applicable to proteins.

**b. Tyr**

The hydroxyl group of the phenolic ring of Tyr residues makes this aromatic ring relatively reactive in electrophilic substitution reactions. These usually occur at the positions designated here as  $\epsilon 1$  and  $\epsilon 2$ , which elsewhere are often numbered as 3 and 5, respectively. Conse-

quently, Tyr side chains can be readily nitrated and iodinated to form the following derivatives:



(1.38)

Table 1.4 Spectroscopic Properties of Tyrosine and Several Derivatives

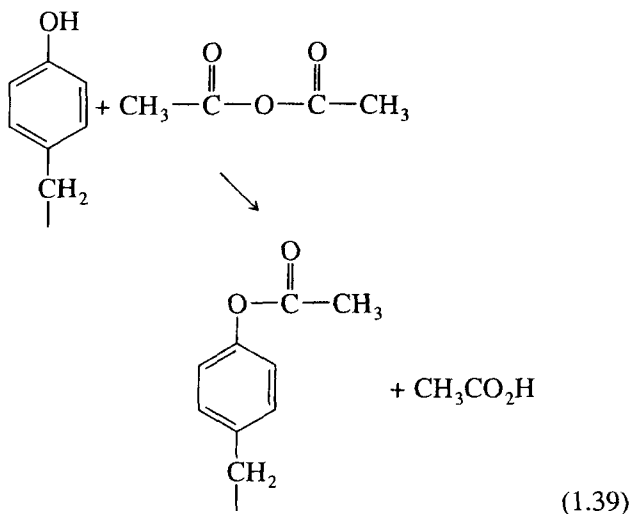
	$pK_{app}$ of —OH	Nonionized OH		Ionized Hydroxyl	
		$\lambda_{max}$ (nm)	Molar absorbance ( $M^{-1}cm^{-1}$ )	$\lambda_{max}$ (nm)	Molar absorbance ( $M^{-1}cm^{-1}$ )
Tyrosine	10.1	274.5	1400	293	2400
$\epsilon$ -Iodotyrosine	8.2	283	2750	305	4100
$\epsilon 1, \epsilon 2$ -Diiodotyrosine	6.5	287	2750	311	6250
$\epsilon$ -Nitrotyrosine	7.2	360	2790	428	4200
$\epsilon$ -Aminotyrosine	10.0 <sup>a</sup>	275	1600	320	4200
O-Acetyltirosine	—	262	262	—	—

<sup>a</sup> The  $pK_{app}$  of the aromatic amino group is approximately 4.8.

From A. N. Glazer, in H. Neurath and R. L. Hill (eds.), *The Proteins*, 3rd ed. vol. 2, pp. 1–103. New York, Academic Press, 1976.

The spectrophotometric properties of these groups are altered substantially, making spectrophotometric analysis particularly useful (Table 1.4).

The hydroxyl group of the Tyr side chain ionizes at alkaline pH values, with an intrinsic  $pK_a$  of 11.1 and with a change in its spectral properties (Table 1.4). The hydroxyl group can participate in hydrogen bonding and also can be acetylated by acetic anhydride:

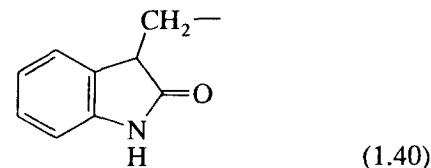


### c. Trp

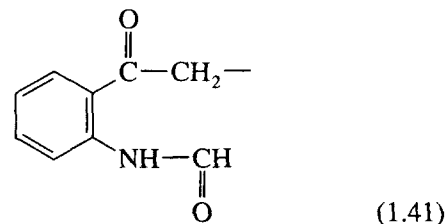
The indole side chain of Trp residues is the largest and the most fluorescent of the side chains of proteins. This amino acid also occurs least frequently, so proteins often have only one or a few Trp residues. Their spectral properties are often useful probes of protein structure (Chap. 7). Their fluorescence properties are especially

sensitive to the environment of the side chain, but in largely unpredictable ways.

The indole ring is susceptible to irreversible oxidation. In particular, iodine and *N*-bromosuccinimide oxidize the Trp indole ring to that of oxindolealanine:

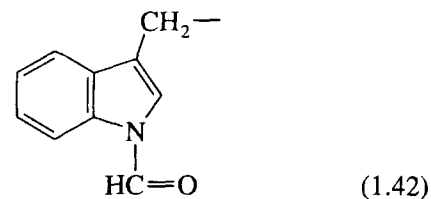


Ozone opens the indole ring to that of *N*-formylkynurenine:

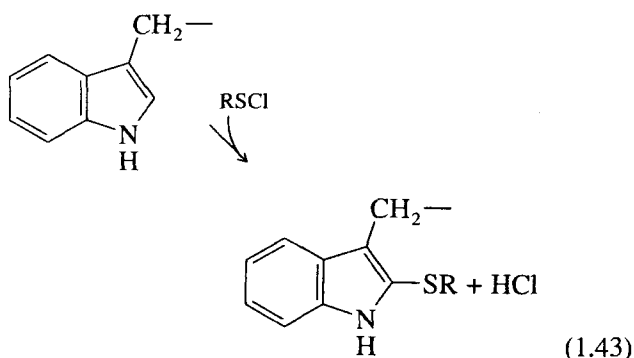


These reagents, however, also rapidly oxidize Cys thiol groups and react more slowly with Tyr and His residues.

The nitrogen atom of the indole ring can be reversibly formylated by anhydrous formic acid containing HCl to yield



Certain reactive benzyl and sulfenyl halides have been designed that alkylate the C<sup>β1</sup> atom of the indole ring:



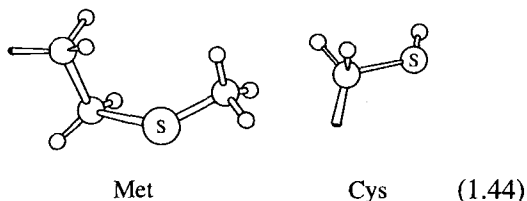
A variety of groups with other spectral properties can be introduced into Trp residues in this way.

The pyrrole nitrogen atom of the indole ring may be a hydrogen bond donor. This is also the only side chain capable of participating as a donor in charge-transfer complexes with pyridinium compounds and other electrophiles.

### References

- Ultraviolet fluorescence of the aromatic amino acids. F. J. W. Teale and G. Weber. *Biochem. J.* 65:476–482 (1957).  
 Ultraviolet spectra of proteins and amino acids. D. B. Wetlaufer. *Adv. Protein Chem.* 17:303–390 (1962).  
 Nitration with tetranitromethane. J. F. Riordan and B. L. Vallee. *Methods Enzymol.* 25:515–521 (1972).  
 Fluorescence of aminotyrosyl residues in peptides and helical proteins. R. L. Seagle and R. W. Cowgill. *Biochim. Biophys. Acta* 439:461–469 (1976).

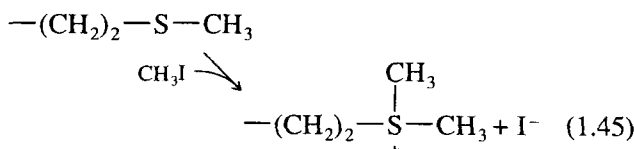
### 1.3.10 The Sulfur-Containing Residues: Met and Cys



#### a. Met

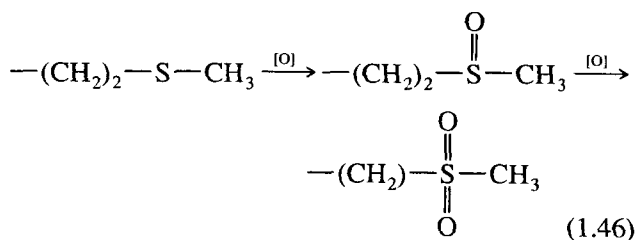
The long side chain of Met residues is nonpolar and relatively unreactive and is the only unbranched nonpolar side chain of all the natural amino acids. The sulfur atom is somewhat nucleophilic, but unlike other nucleophiles in proteins it cannot be protonated. Conse-

quently, it is the most potent nucleophile in proteins at acidic pH. It can be selectively modified under such conditions, readily forming sulfonium salts with alkylating agents such as methyl iodide:



This reaction can be reversed by thiols. The methyl group removed is equally likely to be the original one or that introduced by the methyl iodide, so this reaction offers the possibility of introducing an isotopic label in 50% of the residues by using labeled methyl iodide. The sulfur atom also has a tendency to interact noncovalently with the platinum derivative  $\text{PtCl}_4^{2-}$ , which is a useful heavy-atom derivative in protein crystallography (Chap. 6).

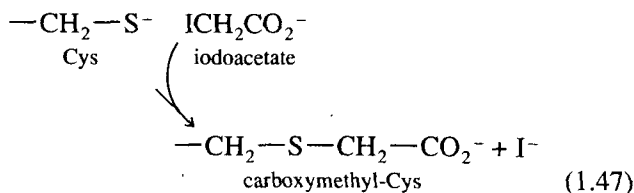
The sulfur atom of Met residues is also susceptible to oxidation by air and more potent oxidants such as peroxides. The sulfoxide is formed first, followed by the sulfone:



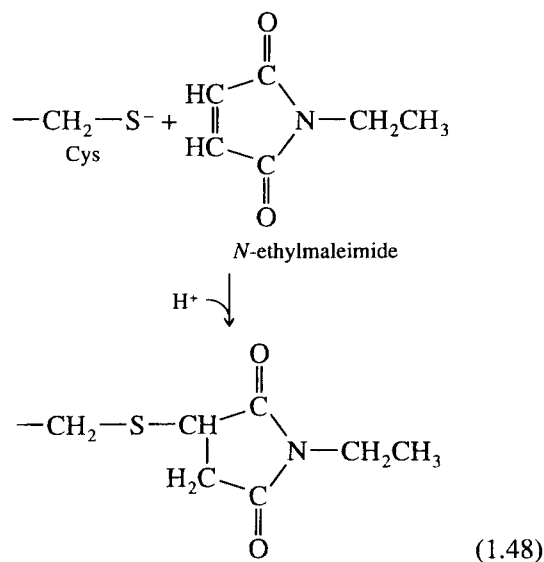
The first step, but not the second, can be reversed by sulfides and thiols. Either oxidation state makes the Met residue unreactive with alkylating reagents (Eq. 1.45).

#### b. Cys

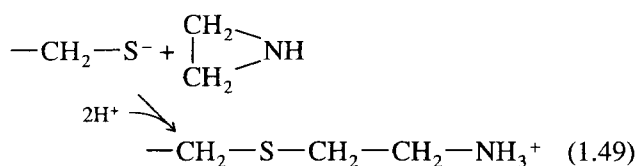
The thiol group of Cys residues is the most reactive of any amino acid side chain, so only a few of the most important and useful examples can be given here. The Cys thiol usually ionizes at slightly alkaline pH values with an intrinsic  $\text{pK}_a$  in the region of 9.0 to 9.5. The thiolate anion formed is the reactive species in most instances. It reacts rapidly with alkyl halides, such as iodoacetate, iodoacetamide, and methyl iodide, to give the corresponding stable alkyl derivatives:



The thiol group can also add across double bonds, such as those of *N*-ethylmaleimide or maleic anhydride:

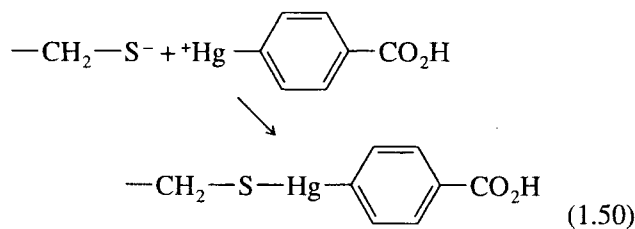


It can also open the ring of ethyleneimine:



The resulting positively charged side chain provides a new site for cleavage of polypeptides by trypsin, in addition to those of Lys and Arg, which is useful for amino acid sequence determination (Sec. 1.6.2).

The thiols of Cys residues form complexes of varying stability with a variety of metal ions. The most stable are those with divalent mercury,  $\text{Hg}^{2+}$ , but complexes with a variety of stoichiometries are formed. Consequently, univalent organic mercurials of the type  $\text{R---Hg}^+$  tend to be used instead because they more reproducibly form one-to-one complexes with thiols. The best known is *p*-mercuribenzoic acid:

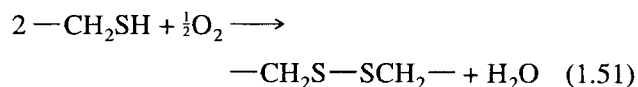


It has been used to titrate thiol groups by measuring the spectral change that takes place when binding occurs. Such reactions with mercurial compounds are also the most obvious and useful ways to make heavy-atom derivatives for protein crystallography (Chap. 6).

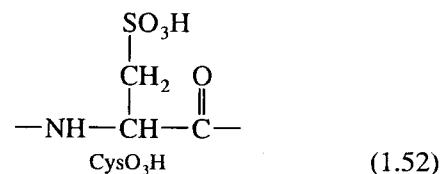
Thiol complexes with silver are less stable than those with mercury, but univalent  $\text{Ag}^+$  reacts stoichiometrically and can be used to titrate thiols. Copper, iron, zinc, cobalt, molybdenum, manganese, and cadmium ions all form various complexes with the thiol groups of Cys residues.

Thiols are readily oxidized by oxygen, especially in the presence of trace amounts of metal ions such as  $\text{Cu}^{2+}$ ,  $\text{Fe}^{2+}$ ,  $\text{Co}^{2+}$ , and  $\text{Mn}^{2+}$ ; it is likely that the metal complexes are the actual reactants with oxygen.

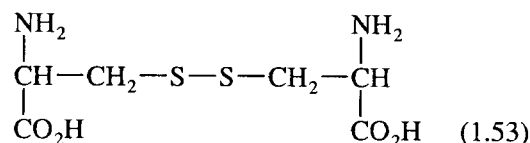
The sulfur atom of Cys residues can exist in a variety of oxidation states, but some of them are unstable. Besides the thiol form, only two oxidation states are generally encountered, the disulfide and the sulfonic acid. The disulfide is usually the end product of air oxidation:



The sulfonic acids are produced by more potent oxidizing agents. For example, performic acid oxidizes both thiol and disulfide forms of Cys residues to cysteic acid residues (abbreviated  $\text{CysO}_3\text{H}$  or  $\text{Cya}$ ):



Disulfide bonds between Cys residues occur in some proteins; two such residues linked by a disulfide bond are often designated as a **cystine** residue, after the amino acid cystine:

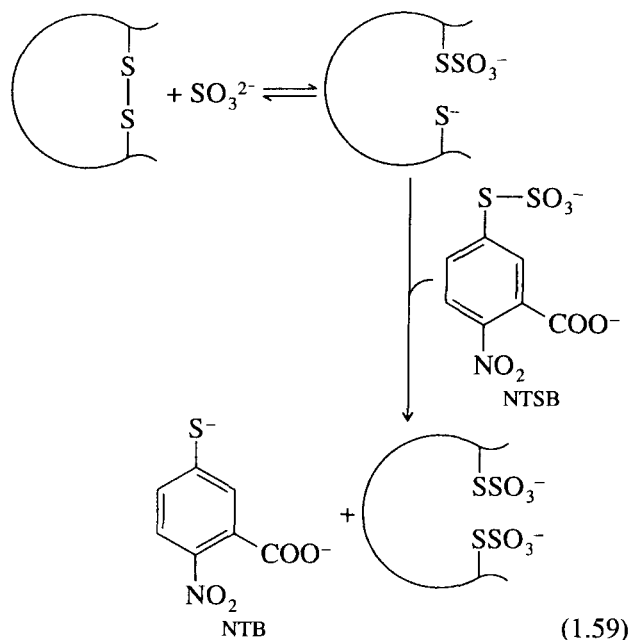


and individual Cys residues are often designated as  $\frac{1}{2}$ -cystines. It is now clear, however, that cystine is not incorporated into proteins as such (Chap. 2). Instead, the thiol form, cysteine, is used in protein biosynthesis, and disulfide bonds between Cys residues can be added later. The designation of cystine as an amino acid of proteins is therefore somewhat misleading, but references to it in the literature are still common.

Disulfide bonds are covalent and can be kept intact under appropriate conditions. The angle of rotation (see Sec. 5.1) about the disulfide bond in its preferred (most stable) geometry has a value close to either  $+90^\circ$  or  $-90^\circ$ . Other angles are unstable by as much as 10 kcal/



The equilibria for these reactions are such that it is difficult to drive the reactions to completion by simply adding reagent. Complete reaction can be accomplished, however, by adding a second reagent that reacts with the thiol generated. When this is nitrothio-sulfobenzoate (NTSB),



the reaction can go to completion, and 1 mol of the colored NTB is generated for each mole of protein disulfide originally present. This is the most convenient assay for protein disulfides.

### References

- Dithiothreitol, a new protective reagent for SH groups. W. W. Cleland. *Biochemistry* 3:480-482 (1964).
- Biochemistry of the SH Group*. P. C. Jocelyn. Academic Press, New York, 1972.
- Disulfide bond formation in proteins. T. E. Creighton. *Methods Enzymol.* 107:305-329 (1984).
- Selective oxidation and reduction of methionine residues in peptides and proteins by oxygen exchange between sulfoxide and sulfide. Y. Schechter. *J. Biol. Chem.* 261:66-70 (1986).
- Detection and quantitation of biological sulfhydryls. A. Russo and E. A. Bump. *Methods Biochem. Anal.* 33:165-241 (1988).

## 1.4 Detection of Amino Acids, Peptides, and Proteins

The experimental study of proteins requires first that they be detectable. Assays are required to determine the

quantity of protein present in a solution and to detect all proteins present after separation by techniques such as chromatography and electrophoresis.

Purified proteins are most easily detected and quantified by their ultraviolet (UV) absorbance in the aromatic region around 280 nm (Table 1.3), so long as no interfering substances that significantly absorb UV at the same wavelength are present. The absorbance of each protein depends on the number and positions of its aromatic amino acid residues Phe, Tyr, and Trp and on any disulfide bonds between Cys residues; so the molar absorbance coefficient of each protein must be known for this UV absorbance procedure to measure the amount of protein present quantitatively. Fortunately, a fully unfolded protein has absorbance properties in the aromatic region that are close to those of its constituent amino acids, making it possible to calculate the molar absorbance coefficient of a protein under unfolding conditions if the content of aromatic amino acids and disulfides is known. Conversely, such a spectrum can be used to determine the amino acid composition of a protein.

In the ideal detection procedure, all proteins would react to the same extent, and thus the assay would give an accurate measure of the amount of every protein present. Such an ideal assay would make use of some property common to all protein molecules, such as the peptide bonds of the backbone. Few of the procedures commonly used are ideal, however; that is, the response of various proteins depends to some extent on their amino acid compositions. Moreover, the assays usually make use of known quantities of one particular protein, often bovine serum albumin (BSA), to produce a standard curve. The accuracy of the results obtained consequently depends very much on how closely the standard protein corresponds to the protein being measured.

One procedure that does measure primarily peptide bonds, and is therefore specific to proteins, is the **biuret reaction**. A dilute solution of cupric sulfate in strongly alkaline tartrate is added to the protein solution. A purplish-violet compound with maximum absorbance at 540 nm forms. The nature of this colored compound is uncertain, but its color is probably due to the formation of a cupric ion complex coordinated with adjacent peptide groups:

