Genetic Defects in the Urea Cycle Can Be Life-Threatening

People with genetic defects in any enzyme involved in urea formation cannot tolerate protein-rich diets. Amino acids ingested in excess of the minimum daily requirements for protein synthesis are deaminated in the liver, producing free ammonia that cannot be converted to urea and exported into the bloodstream, and, as we have seen, ammonia is highly toxic. The absence of a urea cycle enzyme can result in hyperammonemia or in the buildup of one or more urea cycle intermediates, depending on the enzyme that is missing. Given that most urea cycle steps are irreversible, the absent enzyme activity can often be identified by determining which cycle intermediate is present in elevated concentration in the blood and/or urine. Although the breakdown of amino acids can have serious health consequences in individuals with urea cycle deficiencies, a protein-free diet is not a treatment option. Humans are incapable of synthesizing half of the 20 common amino acids, and these **essential amino acids** (Table 18-1) must be provided in the diet.

A variety of treatments are available for individuals with urea cycle defects. Careful administration of the aromatic acids benzoate or phenylbutyrate in the diet can help lower the level of ammonia in the blood. Benzoate is converted to benzoyl-CoA, which combines with glycine to form hippurate (**Fig. 18-14**, left). The glycine used up in this reaction must be regenerated, and ammonia is thus taken up in the glycine synthase reaction. Phenylbutyrate is converted to phenylacetate by β oxidation. The phenylacetate is then converted to phenylacetyl-CoA, which combines with glutamine to form phenylacetylglutamine (Fig. 18-14, right). The resulting removal of glutamine triggers its further synthesis by glutamine synthetase (see Eqn 22-1) in a reaction that takes up ammonia. Both hippurate and phenylacetylglutamine are nontoxic compounds that are excreted in the urine. The pathways shown in Figure 18-14 make only minor contributions to normal metabolism, but they become prominent when aromatic acids are ingested.

TABLE 18-1 Nonessential and Essential Amino Acids for Humans and the Albino Rat

Nonessential	Conditionally essential[a]	Essential
Alanine	Arginine	Histidine
Asparagine	Cysteine	Isoleucine
Aspartate	Glutamine	Leucine
Glutamate	Glycine	Lysine
Serine	Proline	Methionine
	Tyrosine	Phenylalanine
		Threonine
		Tryptophan
		Valine

[a] Required to some degree in young, growing animals and/or sometimes during illness.

It is hardly surprising that much of the early biochemical research was concerned with the study of proteins. Proteins form the class of biological macromolecules that have the most well-defined physicochemical properties, and consequently they were generally easier to isolate and characterize than nucleic acids, polysaccharides, or lipids. Furthermore, proteins, particularly in the form of enzymes, have obvious biochemical functions. The central role that proteins play in biological processes has therefore been recognized since the earliest days of biochemistry. In contrast, the task of nucleic acids in the transmission and expression of genetic information was not realized until the late 1940s and their catalytic function only began to come to light in the 1980s, the role of lipids in biological membranes was not appreciated until the 1960s, and the biological functions of polysaccharides are still somewhat mysterious.

In this chapter we study the structures and properties of the monomeric units of proteins, the **amino acids.** It is from these substances that proteins are synthesized through processes that we discuss in Chapter 32. Amino acids are

A. General Properties

The pK values of the 20 "standard" α-amino acids of proteins are tabulated in Table 4-1. Here pK_1 and pK_2, respectively, refer to the α-carboxylic acid and α-amino groups, and pK_R refers to the side groups with acid–base properties. Table 4-1 indicates that the pK values of the α-carboxylic acid groups lie in a small range around 2.2 so that above pH 3.5 these groups are almost entirely in their carboxylate forms. The α-amino groups all have pK values near 9.4 and are therefore almost entirely in their ammonium ion forms below pH 8.0. This leads to an important structural point: *In the physiological pH range, both the carboxylic acid and the amino groups of α-amino acids are completely ionized* (Fig. 4-2). An amino acid can therefore act as either an acid or a base. Substances with this property are said to be **amphoteric** and are referred to as **ampholytes** (*ampho*teric electr*olytes*). In Section 4-1D, we shall delve a bit deeper into the acid–base properties of the amino acids.

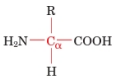


Figure 4-1 **General structural formula for α-amino acids.** There are 20 different R groups in the commonly occurring amino acids (Table 4-1).

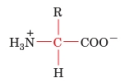
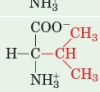
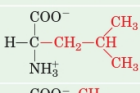
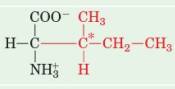
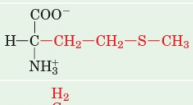
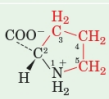
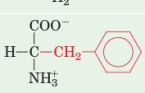
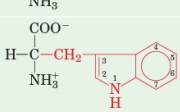


Figure 4-2 **Zwitterionic form of the α-amino acids that occurs at physiological pH values.**

Table 4-1 **Covalent Structures and Abbreviations of the "Standard" Amino Acids of Proteins, Their Occurrence, and the pK Values of Their Ionizable Groups**

Name, Three-Letter Symbol, and One-Letter Symbol	Structural Formula[a]	Residue Mass (D)[b]	Average Occurrence in Proteins (%)[c]	pK_1 α-COOH[d]	pK_2 α-NH_3^+[d]	pK_R Side Chain[d]
Amino acids with nonpolar side chains						
Glycine Gly G		57.0	7.1	2.35	9.78	
Alanine Ala A		71.1	8.3	2.35	9.87	
Valine Val V		99.1	6.9	2.29	9.74	
Leucine Leu L		113.2	9.7	2.33	9.74	
Isoleucine Ile I		113.2	6.0	2.32	9.76	
Methionine Met M		131.2	2.4	2.13	9.28	
Proline Pro P		97.1	4.7	1.95	10.64	
Phenylalanine Phe F		147.2	3.9	2.20	9.31	
Tryptophan Trp W		186.2	1.1	2.46	9.41	

(*continued*)

[a]The ionic forms shown are those predominating at pH 7.0 (except for that of histidine[e]), although residue mass is given for the neutral compound. The C_α atoms, as well as those atoms marked with an asterisk, are chiral centers with configurations as indicated according to Fischer projection formulas. The standard organic numbering system is provided for heterocycles.

[b]The residue masses are given for the neutral residues. For molecular masses of the parent amino acids, add 18.0 D, the molecular mass of H_2O, to the residue masses. For side chain masses, subtract 56.0 D, the formula mass of a peptide group, from the residue masses.

[c]The average amino acid composition in the complete SWISS-PROT database (http://www.expasy.ch/sprot/relnotes/relstat.html), Release 55.11.

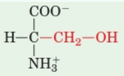
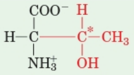
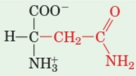
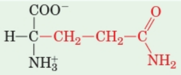
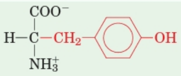
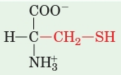
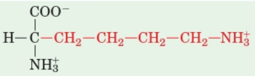
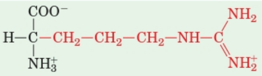
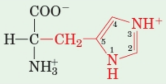
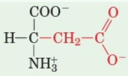
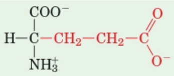
[d]From Dawson, R.M.C., Elliott, D.C., Elliott, W.H., and Jones, K.M., *Data for Biochemical Research* (3rd ed.), *pp.* 1–31, Oxford Science Publications (1986).

[e]Both the neutral and protonated forms of histidine are present at pH 7.0 because its pK_R is close to 7.0. The imidazole ring of histidine is numbered here according to the biochemistry convention. In the IUPAC convention, N3 of the biochemistry convention is designated N1 and the numbering increases clockwise around the ring.

[f]The three- and one-letter symbols for asparagine *or* aspartic acid are Asx and B, whereas for glutamine *or* glutamic acid they are Glx and Z. The one-letter symbol for an undetermined *or* "nonstandard" amino acid is X.

[c]The average amino acid composition in the complete SWISS-PROT database (http://www.expasy.ch/sprot/relnotes/relstat.html), Release 55.11.
[d]From Dawson, R.M.C., Elliott, D.C., Elliott, W.H., and Jones, K.M., *Data for Biochemical Research* (3rd ed.), *pp.* 1–31, Oxford Science Publications (1986).
[e]Both the neutral and protonated forms of histidine are present at pH 7.0 because its pK_R is close to 7.0. The imidazole ring of histidine is numbered here according to the biochemistry convention. In the IUPAC convention, N3 of the biochemistry convention is designated N1 and the numbering increases clockwise around the ring.
[f]The three- and one-letter symbols for asparagine *or* aspartic acid are Asx and B, whereas for glutamine *or* glutamic acid they are Glx and Z. The one-letter symbol for an undetermined *or* "nonstandard" amino acid is X.

Table 4-1 ***(Continued)***

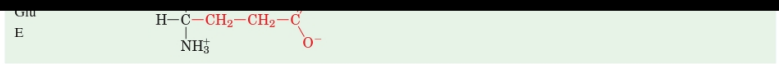
Name Three-Letter Symbol, and One-Letter Symbol	Structural Formula[a]	Residue Mass (D)[b]	Average Occurrence in Proteins (%)[c]	pK_1 α-COOH[d]	pK_2 α-NH_3^+[d]	pK_R Side Chain[d]
Amino acids with uncharged polar side chains						
Serine Ser S		87.1	6.5	2.19	9.21	
Threonine Thr T		101.1	5.3	2.09	9.10	
Asparagine[f] Asn N		114.1	4.0	2.14	8.72	
Glutamine[f] Gln Q		128.1	3.9	2.17	9.13	
Tyrosine Tyr Y		163.2	2.9	2.20	9.21	10.46 (phenol)
Cysteine Cys C		103.1	1.4	1.92	10.70	8.37 (sulfhydryl)
Amino acids with charged polar side chains						
Lysine Lys K		128.2	5.9	2.16	9.06	10.54 (ε-NH_3^+)
Arginine Arg R		156.2	5.5	1.82	8.99	12.48 (guanidino)
Histidine[e] His H		137.1	2.3	1.80	9.33	6.04 (imidazole)
Aspartic acid[f] Asp D		115.1	5.4	1.99	9.90	3.90 (β-COOH)
Glutamic acid[f] Glu E		129.1	6.8	2.10	9.47	4.07 (γ-COOH)

Molecules that bear charged groups of opposite polarity are known as **zwitterions** (German: *zwitter*, hybrid) or **dipolar ions.** The zwitterionic character of the α-amino acids has been established by several methods including spectroscopic measurements and X-ray crystal structure determinations (in the solid state the α-amino acids are zwitterionic because the basic amine group abstracts a proton from the nearby acidic carboxylic acid group). Because amino acids are zwitterions, their physical properties are characteristic of ionic compounds. For instance, most α-amino acids have melting points near 300°C, whereas their nonionic derivatives usually melt around 100°C. Furthermore, amino acids, like other ionic compounds, are more soluble in polar solvents than in nonpolar solvents. Indeed, most α-amino acids are very soluble in water but are largely insoluble in most organic solvents.

B. Peptide Bonds

of polypeptides are also linear polymers. This permits the direct correspondence between the monomer (nucleotide) sequence of a nucleic acid and the monomer (amino acid) sequence of the corresponding polypeptide without the added complication of specifying the positions and sequences of any branching chains.

With 20 different choices available for each amino acid residue in a polypeptide chain, it is easy to see that a huge number of different protein molecules can exist. For example, for dipeptides, each of the 20 different choices for the first amino acid residue can have 20 different choices for the second amino acid residue, for a total of $20^2 = 400$ distinct dipeptides. Similarly, for tripeptides, there are 20 possibilities for each of the 400 choices of dipeptides to yield a total of $20^3 = 8000$ different tripeptides. A relatively small protein molecule consists of a single polypeptide chain of 100 residues. There are $20^{100} = 1.27 \times 10^{130}$ possible unique polypeptide chains of this length, a quantity vastly greater



Molecules that bear charged groups of opposite polarity are known as **zwitterions** (German: *zwitter*, hybrid) or **dipolar ions.** The zwitterionic character of the α-amino acids has been established by several methods including spectroscopic measurements and X-ray crystal structure determinations (in the solid state the α-amino acids are zwitterionic because the basic amine group abstracts a proton from the nearby acidic carboxylic acid group). Because amino acids are zwitterions, their physical properties are characteristic of ionic compounds. For instance, most α-amino acids have melting points near 300°C, whereas their nonionic derivatives usually melt around 100°C. Furthermore, amino acids, like other ionic compounds, are more soluble in polar solvents than in nonpolar solvents. Indeed, most α-amino acids are very soluble in water but are largely insoluble in most organic solvents.

B. *Peptide Bonds*

The α-amino acids polymerize, at least conceptually, through the elimination of a water molecule as is indicated in Fig. 4-3. The resulting CO—NH linkage, which was independently characterized in 1902 by Emil Fischer and Franz Hofmeister, is known as a **peptide bond.** Polymers composed of two, three, a few (3–10), and many **amino acid residues** (alternatively called **peptide units**) are known, respectively, as **dipeptides, tripeptides, oligopeptides,** and **polypeptides.** These substances, however, are often referred to simply as "peptides." *Proteins are molecules that consist of one or more polypeptide chains*. These polypeptides range in length from ~40 to ~34,000 amino acid residues (although few have more than 1500 residues) and, since the average mass of an amino acid residue is ~110 D, have molecular masses that range from ~40 to over ~3700 kD.

Polypeptides are ***linear polymers;*** that is, each amino acid residue is linked to its neighbors in a head-to-tail fashion rather than forming branched chains. This observation reflects the underlying elegant simplicity of the way living systems construct these macromolecules for, as we shall see, the nucleic acids that encode the amino acid sequences of polypeptides are also linear polymers. This permits the direct correspondence between the monomer (nucleotide) sequence of a nucleic acid and the monomer (amino acid) sequence of the corresponding polypeptide without the added complication of specifying the positions and sequences of any branching chains.

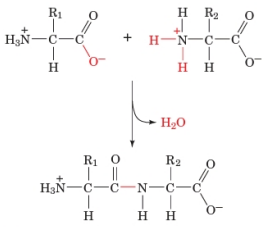


Figure 4-3 **Condensation of two α-amino acids to form a dipeptide.** The peptide bond is shown in red.

With 20 different choices available for each amino acid residue in a polypeptide chain, it is easy to see that a huge number of different protein molecules can exist. For example, for dipeptides, each of the 20 different choices for the first amino acid residue can have 20 different choices for the second amino acid residue, for a total of $20^2 = 400$ distinct dipeptides. Similarly, for tripeptides, there are 20 possibilities for each of the 400 choices of dipeptides to yield a total of $20^3 = 8000$ different tripeptides. A relatively small protein molecule consists of a single polypeptide chain of 100 residues. There are $20^{100} = 1.27 \times 10^{130}$ possible unique polypeptide chains of this length, a quantity vastly greater than the estimated number of atoms in the universe (9×10^{78}). Clearly, nature can have made only a tiny fraction of the possible different protein molecules. Nevertheless, *the various organisms on Earth collectively synthesize an enormous number of different protein molecules whose great range of physicochemical characteristics stem largely from the varied properties of the 20 "standard" amino acids.*

C. *Classification and Characteristics*

The most common and perhaps the most useful way of classifying the 20 "standard" amino acids is according to the polarities of their side chains **(R groups).** This is because proteins fold to their native conformations largely in response to the tendency to remove their hydrophobic side chains from contact with water and to solvate their hydrophilic side chains (Chapters 8 and 9). According to this classification scheme, there are three major types of amino acids: (1) those with nonpolar R groups, (2) those with uncharged polar R groups, and (3) those with charged polar R groups.

a. The Nonpolar Amino Acid Side Chains Have a Variety of Shapes and Sizes

Nine amino acids are classified as having nonpolar side chains. **Glycine** (which, when it was found to be a component of gelatin in 1820, was the first amino acid to be identified in protein hydrolyzates) has the smallest possible side chain, an H atom. **Alanine** (Fig. 4-4), **valine, leucine,** and **isoleucine** have aliphatic hydrocarbon side chains ranging in size from a methyl group for alanine to isomeric butyl groups for leucine and isoleucine. **Methionine** has a thiol ether side chain that resembles an *n*-butyl group in many of its physical properties (C and S have nearly equal electronegativities and S is about the size of a methylene group). **Proline,** a cyclic secondary amino acid, has conformational constraints imposed by the cyclic nature of its pyrrolidine side chain, which is unique among the "standard" 20 amino acids. **Phenylalanine,** with its phenyl moiety (Fig. 4-4), and **tryptophan,** with its indole group, contain

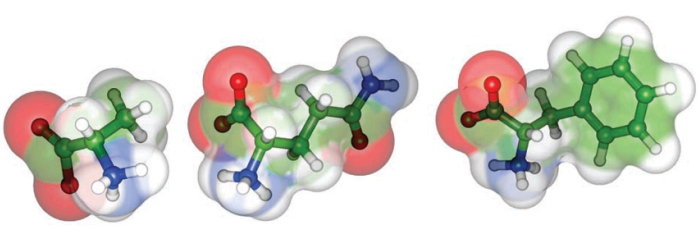


Figure 4-4 **Structures of the α-amino acids alanine, glutamine, and phenylalanine.** The amino acids are shown as ball-and-stick models embedded in their transparent space-filling models. The atoms are colored according to type with C green, H white, N blue, and O red.

aromatic side chains, which are characterized by bulk as

acid **cystine** because they were originally thought to form a

One- and Three-Letter Symbols for the Amino Acids[a]

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

[a]The one-letter symbol for an undetermined or nonstandard amino acid is X.

Thermodynamic Constants and Conversion Factors

***Joule* (J)**
$1\ J = 1\ kg \cdot m^2 \cdot s^{-2}$ $1\ J = 1\ C \cdot V$ (coulomb volt)
$1\ J = 1\ N \cdot m$ (newton · meter)

***Calorie* (cal)**
1 cal heats 1 g of H_2O from 14.5 to 15.5°C
$1\ cal = 4.184\ J$

***Large calorie* (Cal)**
$1\ Cal = 1\ kcal$ $1\ Cal = 4184\ J$

***Avogadro's number* (*N*)**
$N = 6.0221 \times 10^{23}$ molecules · mol^{-1}

***Coulomb* (C)**
$1\ C = 6.241 \times 10^{18}$ electron charges

***Faraday* ($\mathcal{F}$)**
$1\ \mathcal{F} = N$ electron charges
$1\ \mathcal{F} = 96{,}485\ C \cdot mol^{-1} = 96{,}485\ J \cdot V^{-1} \cdot mol^{-1}$

***Kelvin temperature scale* (K)**
0 K = absolute zero 273.15 K = 0°C

***Boltzmann constant* (k_B)**
$k_B = 1.3807 \times 10^{-23}\ J \cdot K^{-1}$

***Gas constant* (*R*)**
$R = Nk_B$ $R = 1.9872\ cal \cdot K^{-1} \cdot mol^{-1}$
$R = 8.3145\ J \cdot K^{-1} \cdot mol^{-1}$ $R = 0.08206\ L \cdot atm \cdot K^{-1} \cdot mol^{-1}$

The Standard Genetic Code

First Position (5′ end)	*Second Position*				*Third Position (3′ end)*
	U	C	A	G	
U	UUU Phe	UCU Ser	UAU Tyr	UGU Cys	U
	UUC Phe	UCC Ser	UAC Tyr	UGC Cys	C
	UUA Leu	UCA Ser	UAA Stop	UGA Stop	A
	UUG Leu	UCG Ser	UAG Stop	UGG Trp	G
C	CUU Leu	CCU Pro	CAU His	CGU Arg	U
	CUC Leu	CCC Pro	CAC His	CGC Arg	C
	CUA Leu	CCA Pro	CAA Gln	CGA Arg	A
	CUG Leu	CCG Pro	CAG Gln	CGG Arg	G
A	AUU Ile	ACU Thr	AAU Asn	AGU Ser	U
	AUC Ile	ACC Thr	AAC Asn	AGC Ser	C
	AUA Ile	ACA Thr	AAA Lys	AGA Arg	A
	AUG Met[a]	ACG Thr	AAG Lys	AGG Arg	G
G	GUU Val	GCU Ala	GAU Asp	GGU Gly	U
	GUC Val	GCC Ala	GAC Asp	GGC Gly	C
	GUA Val	GCA Ala	GAA Glu	GGA Gly	A
	GUG Val	GCG Ala	GAG Glu	GGG Gly	G

[a]AUG forms part of the initiation signal as well as coding for internal Met residues.

Some Common Biochemical Abbreviations[a]

A	adenine
aa	amino acid
aaRS	aminoacyl–tRNA synthetase
ACAT	acyl-CoA:cholesterol acyltransferase
ACh	acetylcholine
AChE	acetylcholinesterase
ACP	acyl-carrier protein
ADA	adenosine deaminase
ADH	alcohol dehydrogenase
AdoCbl	5′-deoxyadenosylcobalamin
AdoMet	adenosylmethionine
ADP	adenosine diphosphate
ADPNP	adenosine-5′-(β,γ-imido)triphosphate
AIDS	acquired immunodeficiency syndrome
AKAP	A-kinase anchoring protein
ALA	δ-aminolevulinic acid
AMP	adenosine monophosphate
AMPK	AMP-dependent protein kinase
E4P	erythrose-4-phosphate
EPR	electron paramagnetic resonance
ER	endoplasmic reticulum
ESI	electrospray ionization
EST	expressed sequence tag
ETF	electron-transfer flavoprotein
FAD	flavin adenine dinucleotide, oxidized form
FADH·	flavin adenine dinucleotide, radical form
$FADH_2$	flavin adenine dinucleotide, reduced form
FAS	fatty acid synthase
FBP	fructose-1,6-bisphosphate
FBPase	fructose-1,6-bisphosphatase
Fd	ferredoxin
FGF	fibroblast growth factor
FH	familial hypercholesterolemia
fMet	*N*-formylmethionine
FMN	flavin mononucleotide
FNR	ferredoxin-NADP$^+$ reductase