12.2: STRUCTURES OF AMINO ACIDS

OBJECTIVES

After completing this section, you should be able to

1. identify the structural features present in the 20 amino acids commonly found in proteins.

   Note: You are not expected to remember the detailed structures of all these amino acids, but you should be prepared to draw the structures of the two simplest members, glycine and alanine.

2. draw the Fischer projection formula of a specified enantiomer of a given amino acid.

   Note: To do so, you must remember that in the *S* enantiomer, the carboxyl group appears at the top of the projection formula and the amino group is on the left.

3. classify an amino acid as being acidic, basic or neutral, given its Kekulé, condensed or shorthand structure.
4. draw the zwitterion form of a given amino acid.
5. account for some of the typical properties of amino acids (e.g., high melting points, solubility in water) in terms of zwitterion formation.
6. write appropriate equations to illustrate the amphoteric nature of amino acids.

KEY TERMS

Make certain that you can define, and use in context, the key terms below.

- α-amino acids
- amphoteric
- essential amino acids
- zwitterion

STUDY NOTES

This is a good point at which to review some of the principles of stereochemistry presented in Chapter 5. Be sure to make full use of molecular models when any stereochemical issues arise.

You should recognize that a three-letter shorthand code is often used to represent individual amino acids. You need not memorize this code.

The distinction between essential and nonessential amino acids is not as clear-cut as one might suppose. For example, arginine is often regarded as being nonessential.

INTRODUCTION TO AMINO ACIDS

Amino acids form polymers through a nucleophilic attack by the amino group of an amino acid at the electrophilic carbonyl carbon of the carboxyl group of another amino acid. The carboxyl group of the amino acid must first be activated to provide a better leaving group than OH^-. (We will discuss this activation by ATP later in the course.) The resulting link between the amino acids is an amide link which biochemists call a peptide bond. In this reaction, water is released. In a reverse reaction, the peptide bond can be cleaved by water (hydrolysis).

- Structure and Property of the Naturally-Occurring Amino Acids (Too large to include in text: print separately)

When two amino acids link together to form an amide link, the resulting structure is called a dipeptide. Likewise, we can have tripeptides, tetrapeptides, and other polypeptides. At some point, when the structure is long enough, it is called a protein. There are many different ways to represent the structure of a polypeptide or protein, each showing differing amounts of information.

Figure: Different Representations of a Polypeptide (Heptapeptide)

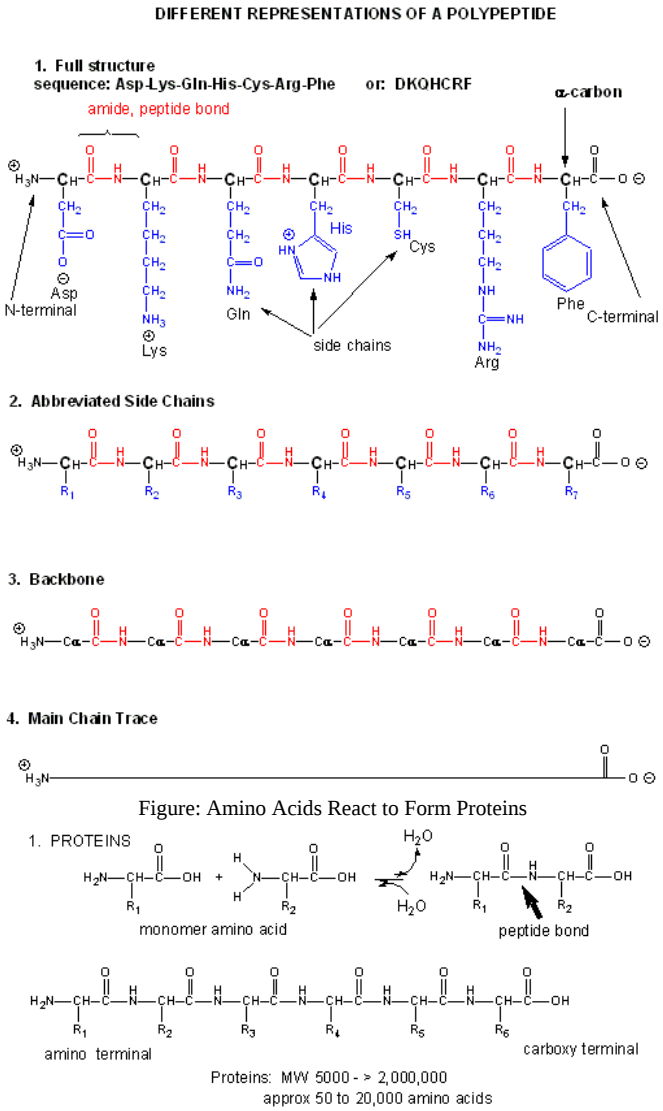


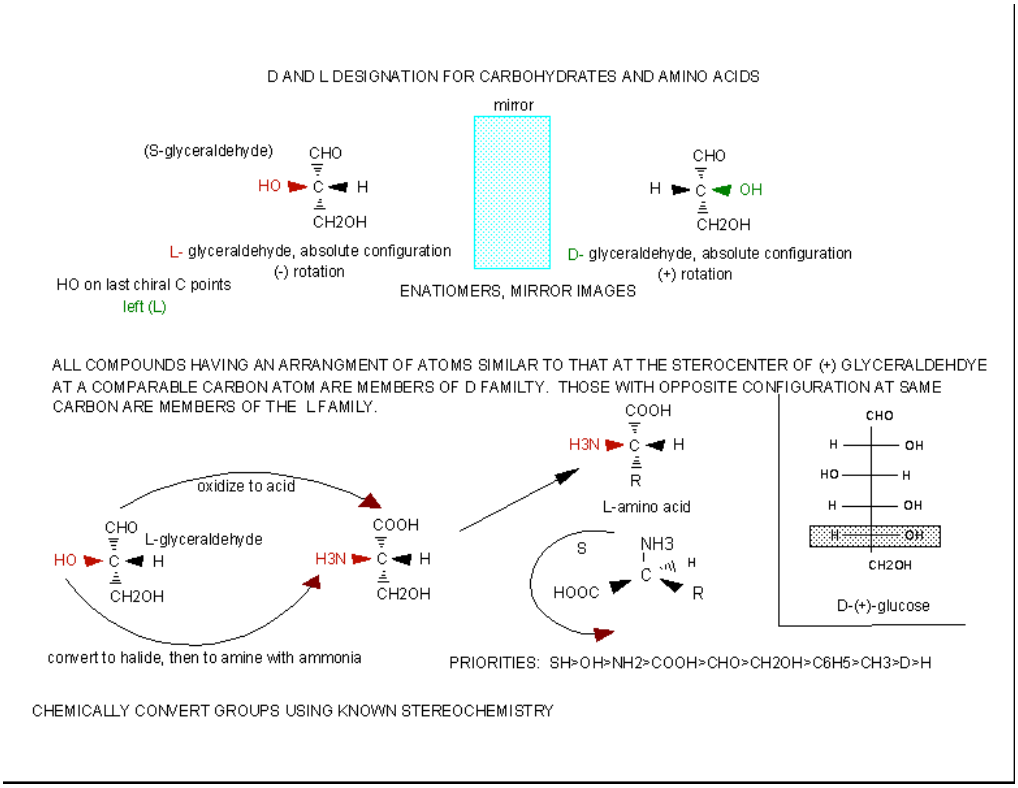
Figure: Amino Acids React to Form Proteins

(Note: above picture represents the amino acid in an unlikely protonation state with the weak acid protonated and the weak base deprotonated for simplicity in showing removal of water on peptide bond formation and the hydrolysis reaction.) Proteins are polymers of twenty naturally occurring amino acids. In contrast, nucleic acids are polymers of just 4 different monomeric nucleotides. Both the sequence of a protein and its total length differentiate one protein from another. Just for an octapeptide, there are over 25 billion different possible arrangements of amino acids. Compare this to just 65536 different oligonucleotides of 8 monomeric units (8mer). Hence the diversity of possible proteins is enormous.

STEREOCHEMISTRY

The amino acids are all chiral, with the exception of glycine, whose side chain is H. As with lipids, biochemists use the L and D nomenclature. All naturally occuring proteins from all living organisms consist of L amino acids. The absolute stereochemistry is related to L-glyceraldehyde, as was the case for triacylglycerides and phospholipids. Most naturally occurring chiral amino acids are S, with the exception of cysteine. As the diagram below shows, the absolute configuration of the amino acids can be shown with the H pointed to the rear, the COOH groups pointing out to the left, the R group to the right, and the NH_3 group upwards. You can remember this with the anagram CORN.

Figure: *Stereochemistry of Amino Acids.*



Why do biochemists still use D and L for sugars and amino acids? This explanation (taken from the link below) seems reasonable.

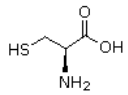
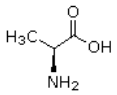
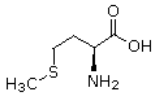
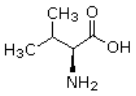
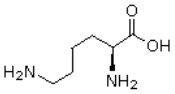
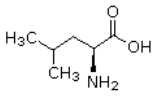
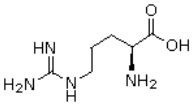
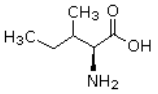
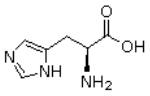
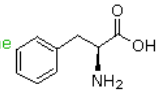
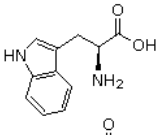
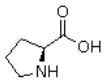
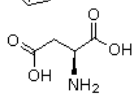
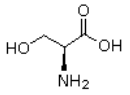
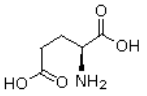
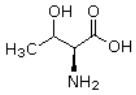
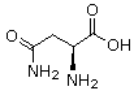
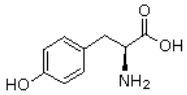
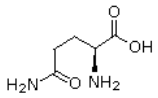
"In addition, however, chemists often need to define a configuration unambiguously in the absence of any reference compound, and for this purpose the alternative (R,S) system is ideal, as it uses priority rules to specify configurations. These rules sometimes lead to absurd results when they are applied to biochemical molecules. For example, as we have seen, all of the common amino acids are L, because they all have exactly the same structure, including the position of the R group if we just write the R group as R. However, they do not all have the same configuration in the (R,S) system: L-cysteine is also (R)-cysteine, but all the other L-amino acids are (S), but this just reflects the human decision to give a sulphur atom higher priority than a carbon atom, and does not reflect a real difference in configuration. Worse problems can sometimes arise in substitution reactions: sometimes inversion of configuration can result in no change in the (R) or (S) prefix; and sometimes retention of configuration can result in a change of prefix.

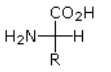
It follows that it is not just conservatism or failure to understand the (R,S) system that causes biochemists to continue with D and L: it is just that the DL system fulfils their needs much better. As mentioned, chemists also use D and L when they are appropriate to their needs. The explanation given above of why the (R,S) system is little used in biochemistry is thus almost the exact opposite of reality. This system is actually the only practical way of unambiguously representing the stereochemistry of complicated molecules with several asymmetric centres, but it is inconvenient with regular series of molecules like amino acids and simple sugars. "

NATURAL A-AMINO ACIDS

Hydrolysis of proteins by boiling aqueous acid or base yields an assortment of small molecules identified as α-aminocarboxylic acids. More than twenty such components have been isolated, and the most common of these are listed in the following table. Those amino acids having green colored names are **essential** diet components, since they are not synthesized by human metabolic processes. The best food source of these nutrients is protein, but it is important to recognize that not all proteins have equal nutritional value. For example, peanuts have a higher weight content of protein than fish or eggs, but the proportion of essential amino acids in peanut protein is only a third of that from the two other sources. For reasons that will become evident when discussing the structures of proteins and peptides, each amino acid is assigned a one or three letter abbreviation.

NATURAL A-AMINO ACIDS

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


L-Amino Acid

Some common features of these amino acids should be noted. With the exception of proline, they are all 1°-amines; and with the exception of glycine, they are all chiral. The configurations of the chiral amino acids are the same when written as a Fischer projection formula, as in the drawing on the right, and this was defined as the **L-configuration** by Fischer. The R-substituent in this structure is the remaining structural component that varies from one amino acid to another, and in proline R is a three-carbon chain that joins the nitrogen to the alpha-carbon in a five-membered ring. Applying the Cahn-Ingold-Prelog notation, all these natural chiral amino acids, with the exception of cysteine, have an **S**-configuration. For the first seven compounds in the left column the R-substituent is a hydrocarbon. The last three entries in the left column have hydroxyl functional groups, and the first two amino acids in the right column incorporate thiol and sulfide groups respectively. Lysine and arginine have basic amine functions in their side-chains; histidine and tryptophan have less basic nitrogen heterocyclic rings as substituents. Finally, carboxylic acid side-chains are substituents on aspartic and glutamic acid, and the last two compounds in the right column are their corresponding amides.

The formulas for the amino acids written above are simple covalent bond representations based upon previous understanding of mono-functional analogs. **The formulas are in fact incorrect**. This is evident from a comparison of the physical properties listed in the following table. All four compounds in the table are roughly the same size, and all have moderate to excellent water solubility. The first two are simple carboxylic acids, and the third is an amino alcohol. All three compounds are soluble in organic solvents (e.g. ether) and have relatively low melting points. The carboxylic acids have pK_a's near 4.5, and the conjugate acid of the amine has a pK_a of 10. The simple amino acid alanine is the last entry. By contrast, it is very high melting (with decomposition), insoluble in organic solvents, and a million times weaker as an acid than ordinary carboxylic acids.

https://chem.libretexts.org/@go/page/500499

Physical Properties of Selected Acids and Amines

Compound	Formula	Mol.Wt.	Solubility in Water	Solubility in Ether	Melting Point	pK_a
isobutyric acid	$(CH_3)_2CHCO_2H$	88	20g/100mL	complete	-47 ºC	5.0
lactic acid	$CH_3CH(OH)CO_2H$	90	complete	complete	53 ºC	3.9
3-amino-2-butanol	$CH_3CH(NH_2)CH(OH)CH_3$	89	complete	complete	9 ºC	10.0
alanine	$CH_3CH(NH_2)CO_2H$	89	18g/100mL	insoluble	ca. 300 ºC	9.8

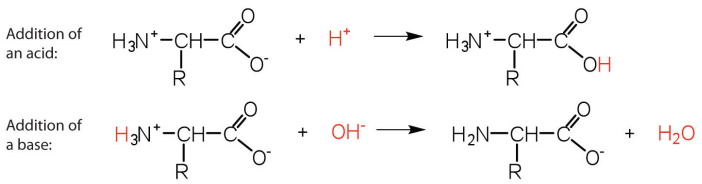
ZWITTERION

These differences above all point to internal salt formation by a proton transfer from the acidic carboxyl function to the basic amino group. The resulting ammonium carboxylate structure, commonly referred to as a **zwitterion**, is also supported by the spectroscopic characteristics of alanine.

$$CH_3CH(NH_2)CO_2H \rightleftharpoons CH_3CH(NH_3)^{(+)}CO_2^{(-)}$$

As expected from its ionic character, the alanine zwitterion is high melting, insoluble in nonpolar solvents and has the acid strength of a 1º-ammonium ion. Examples of a few specific amino acids may also be viewed in their favored neutral zwitterionic form. Note that in lysine the amine function farthest from the carboxyl group is more basic than the alpha-amine. Consequently, the positively charged ammonium moiety formed at the chain terminus is attracted to the negative carboxylate, resulting in a coiled conformation.

The structure of an amino acid allows it to act as both an acid and a base. An amino acid has this ability because at a certain pH value (different for each amino acid) nearly all the amino acid molecules exist as zwitterions. If acid is added to a solution containing the zwitterion, the carboxylate group captures a hydrogen (H^+) ion, and the amino acid becomes positively charged. If base is added, ion removal of the H^+ ion from the amino group of the zwitterion produces a negatively charged amino acid. In both circumstances, the amino acid acts to maintain the pH of the system—that is, to remove the added acid (H^+) or base (OH^-) from solution.

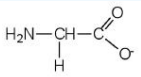


EXAMPLE 26.1

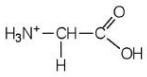
a. Draw the structure for the anion formed when glycine (at neutral pH) reacts with a base.
b. Draw the structure for the cation formed when glycine (at neutral pH) reacts with an acid.

Solution

a. The base removes H^+ from the protonated amine group.

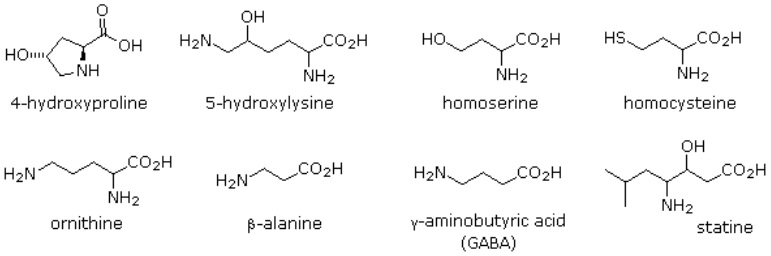


The acid adds H^+ to the carboxylate group.



OTHER NATURAL AMINO ACIDS

The twenty alpha-amino acids listed above are the primary components of proteins, their incorporation being governed by the genetic code. Many other naturally occurring amino acids exist, and the structures of a few of these are displayed below. Some, such as hydroxylysine and hydroxyproline, are simply functionalized derivatives of a previously described compound. These two amino acids are found only in collagen, a common structural protein. Homoserine and homocysteine are higher homologs of their namesakes. The amino group in beta-alanine has moved to the end of the three-carbon chain. It is a component of pantothenic acid, $HOCH_2C(CH_3)_2CH(OH)CONHCH_2CH_2CO_2H$, a member of the vitamin B complex and an essential nutrient. Acetyl coenzyme A is a pyrophosphorylated derivative of a pantothenic acid amide. The gamma-amino homolog GABA is a neurotransmitter inhibitor and antihypertensive agent.



Many unusual amino acids, including D-enantiomers of some common acids, are produced by microorganisms. These include ornithine, which is a component of the antibiotic bacitracin A, and statin, found as part of a pentapeptide that inhibits the action of the digestive enzyme **pepsin**.

EXERCISES

QUESTIONS

Q26.1.1

Why is cysteine the only L amino acid with an R configuration at the alpha carbon?

Q26.1.2

Isoleucine has two stereogenic centers.

(a) Draw a Fischer projection of isoleucine.

(b) Draw a Fischer projection of an isoleucine diastereomer, and label each stereocenter as R or S.

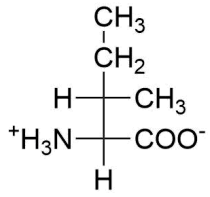
SOLUTIONS

S26.1.1

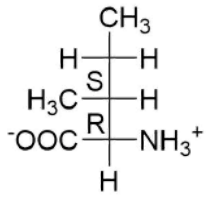
The sulfur atom in the side chain causes the side chain to have higher priority than the other substituents.

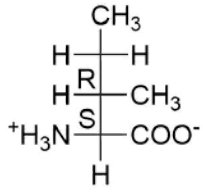
S26.1.2

(a)



(b)





This page titled 12.2: Structures of Amino Acids is shared under a CC BY-SA 4.0 license and was authored, remixed, and/or curated by Steven Farmer, Dietmar Kennepohl, William Reusch, Henry Jakubowski, & Henry Jakubowski (Cañada College) .

- **26.2: Structures of Amino Acids** by Dietmar Kennepohl, Henry Jakubowski, Steven Farmer, William Reusch is licensed CC BY-SA 4.0.