22: Peptides, Proteins, and α-Amino Acids

- *Peptides*
- *Protein Structure and Organization*
- *Properties of α-Amino Acids*
- *Enzymes and Enzyme Catalysis*

Preview

Proteins are a major class of bioorganic molecules present in all organisms. They contain one or more **polypeptide** chains with the repeating general structure -(NH-CHR-C(=O))-. These repeating units come from 20 different chiral **α-amino acids** with the general structure H_2N-CHR-CO_2H. The R groups play a major role in determining conformations of the peptide chains and the shapes of proteins. Free α-Amino acids are polyprotic acids because they have at least two functional groups (CO_2H and NH_2) with acid and conjugate base forms. Some of these 20 α-amino acids also have an acid-base functional group in R. *Enzymes* are proteins that catalyze biochemical reactions. These enzyme-catalyzed reactions take place in specific regions of the enzymes called **active sites.**

22.1 Peptides

Peptides are polymers of **α-amino acids** joined by *amide* or *peptide bonds.* [graphic 22.1]

Peptide Structure (22.1A)
Peptides with two α-*amino acid* components are **dipeptides**, those with three are **tripeptides**, and so on. Peptides with 3-10 α-amino acid components are called **oligopeptides**, while those with many α-amino acid components are **polypeptides**. These terms are similar to the terms *disaccharide*, *oligosaccharide*, and *polysaccharide* that we learned when we studied carbohydrates (Chapter 20).

α-Amino Acids in Peptides. Hydrolysis of peptides cleaves amide bonds (Chapters 15 and 16) releasing the individual α-amino acids. [graphic 22.2] α-Amino acids in naturally occurring peptides generally have one R group and one

H on C_α (H_2N-C_αHR-CO_2H). We will see later in this chapter that the NH_2 and CO_2H groups of free α-amino acids exist as NH_3^+ and CO_2^- so the general structure of these amino acids is actually ^+H_3N-CHR-CO_2^-. [graphic 22.3]

α-Amino Acids Can be D or L. When R is a group other than H, these α-amino acids are chiral compounds with two enantiomeric forms because C_α is chiral (Chapter 4). We can identify the configuration of C_α as *R* or *S* if we know the structure of the C_α-R group, but we usually describe the two enantiomers as D or L. [graphic 22.4] Most α-Amino acids in naturally occurring peptides have L-configurations.

> **The Definition of D and L for α-Amino Acids**. Definitions of D and L for enantiomers of α-amino acids are based on those of D and L-glyceraldehyde (Chapter 20). [graphic 22.5] We define C(=O)O^- of the amino acid as equivalent to C(=O)H of glyceraldehyde (both have C=O) and place it at the top of the α-amino acid Fischer projection. We define R of the amino acid as equivalent to CH_2OH of glyceraldehyde and place it at the bottom of the Fischer projection. The result is a pair of enantiomers with NH_3^+ to the right (defined as D) or to the left (defined as L) on C_α. Although we will see that one α-amino acid (*proline*, see Figure (graphic 22.6)) has an additional bond between the aminium group and R ($^+H_2$N-R), its D and L assignments similarly depend on the "right" or "left" orientation of its C_α-N bond.

The R Groups. While α-amino acids have many different R groups, we are interested in the 20 R groups of the "**standard**" α-amino acids (Figure (graphic 22.6)) in naturally occurring peptides. [graphic 22.6] There are a few other R groups in naturally occurring peptides, but these "**nonstandard**" groups arise from biosynthetic modification of a *"standard"* R group already present in a peptide. [graphic 22.7] You can see that "standard" R groups are diverse. They include alkyl and aryl groups, heterocyclic rings, sulfur containing groups, alcohols, aminium ions, carboxylates, and amides. The R group of *proline* differs from the others because it chemically binds to its α-amino group. *Isoleucine* and *threonine* each have an additional chiral C* (C3) in their R groups (Figure (graphic 22.6)). The naturally occurring stereoisomer of *L-isoleucine* is *S* at C3, while C3 of *L-threonine* is *R*. [graphic 22.9][~~graphic 22.8~~]

Categories of "Standard" Amino Acids. In order to learn their names and R group structures, it is helpful for us to group the "standard" amino acids in the three categories shown in Figure (graphic 22.6). R groups of the 5 "**charged polar**" amino acids are electrically charged (- or +) at physiological pH values (about pH 7), while those of the remaining 15 amino acids are electrically neutral under the same conditions. R groups of the 5 "**uncharged polar**" amino acids form hydrogen bonds to each other and to water, while R groups of the "**nonpolar**" amino acids do not form hydrogen bonds.

Abbreviated Names. Each "standard" amino acid name has a *three-letter abbreviation* as well as a *one-letter designation* (Figure (graphic 22.6)). The three-letter abbreviations are usually the first three letters of the full name, while one-letter designations correspond to the first letter of the name where possible. You will often see three-letter abbreviations used interchangeably with full names.

> **Glx and Asx**. The abbreviation *Glx* refers to both *Glu* and *Gln*, and *Asx* refers to both *Asp* and *Asn*. Amide groups of *Gln* and *Asn* sometimes hydrolyze to carboxylate groups of *Glu* and *Asp* during amino acid analysis (described later in the chapter). [graphic 22.10] As a result, the relative amounts of carboxylate and amide side chains in a naturally occurring peptide may be uncertain so they are grouped together as *Glx* or *Asx*.

Peptide Synthesis (22.1B)

Laboratory syntheses of peptides make use of familiar reactions illustrated below. *Biosynthesis* of peptides involves *nucleic acids* so we defer this topic until Chapter 23.

General Considerations. If we wish to make a dipeptide from two amino acids (AA_x and AA_y), we must recognize that it can have two possible structures (AA_x-AA_y and AA_y-AA_x). [graphic 22.11] Chemists write peptide formulas so that the first amino acid in the sequence is **N-terminal** (it has the unreacted NH_2 group), while the last one is **C-terminal** (it has the terminal CO_2H group). As a result, AA_x is the *N-terminal* amino acid in AA_x-AA_y while AA_y is the *N-terminal* amino acid in AA_y-AA_x. Each amino acid has an amino group and a carboxylic acid group, so we might expect that direct reactions of AA_x and AA_y (or of appropriate derivatives) may give not just AA_x-AA_y and AA_y-AA_x, but AA_x-AA_x and AA_y-AA_y as well [graphic 22.12] Since the number of possible

combinations of amino acids increases rapidly as the size of the desired peptide increases, chemists do not use direct reactions of amino acids to make peptides.

Automated Peptide Synthesis. We can avoid the problem of multiple products by using **automated peptide synthesis**. We illustrate its general features here for the synthesis of the tripeptide AA_z-AA_y-AA_x: [graphic 22.13]

(1) Chemically bind AA_x to a **solid insoluble resin**.
(2) Couple AA_y to AA_x-Resin.
(3) Couple AA_z to AA_y-AA_x-Resin.
(4) Remove AA_z-AA_y-AA_x from the resin.

To accomplish these general steps we **protect** and **deprotect** NH_2 groups and **activate** CO_2H groups. One way chemists *protect* NH_2 groups is with ***tert*-butyloxycarbonyl** (***t*-Boc**) groups from ***tert*-butyloxycarbonyl chloride** (**t-BocCl**). [graphic 22.14] Trifluoroacetic acid in methylene chloride *deprotects* NH_2 groups by removing the *t*-Boc groups. We *activate* CO_2H groups on N-protected amino acids (*t*-Boc-AA) by treating them with **dicyclohexylcarbodiimide** (**DCCD**). [graphic 22.15] The resulting carboxylic acid derivatives (C(=O)-O-DCCD) react readily with NH_2 groups to form amide bonds (C(=O)-NH).

The detailed steps of *automated peptide synthesis* are:

(1) Attach AA_x to the Solid Support
Couple CO_2H groups of N-protected AA_x (*t*-Boc-AA_x) with benzyl chloride groups (Ar-CH_2-Cl) on the *resin solid support* to form esters. [graphic 22.16] Deprotect *t*-Boc-AA_x-Resin to give reactive NH_2 group on each AA_x.

(2) Couple AA_y to AA_x
Activate N-protected AA_y (*t*-Boc-AA_y) using DCCD and couple it with AA_x-Resin. [graphic 22.17] An amide bond forms between AA_y and AA_x leading to *t*-Boc-AA_y-AA_x-Resin. Subsequent deprotection gives AA_y-AA_x-Resin with a reactive NH_2 group on each AA_y.

(3) Couple AA_z to AA_y
Activate N-protected AA_z (*t*-Boc-AA_z) using DCCD and couple it with AA_y-AA_x-Resin. [graphic 22.18] An amide bond forms between AA_z and AA_y leading to *t*-Boc-AA_z-AA_y-AA_x-Resin.

(4) Isolate AA_z-AA_y-AA_x
Treat *t*-Boc-AA_z-AA_y-AA_x-Resin with liquid HF cleaving *t*-Boc from AA_z, as well as the ester bond between AA_x and the resin, giving free AA_z-AA_y-AA_x. [graphic 22.19]

Protection of Functional Groups on R. We must also protect functional groups such as NH_2, OH, SH, and CO_2H on amino acid side chains (R) during peptide synthesis by converting them into derivatives such as benzyl groups. [graphic 22.20] These benzyl groups cleave along with the *t*-Boc group and the ester linkage to the resin during treatment with liquid HF in Step 4.

22.2 Protein Structure and Organization

Proteins are biological molecules made up of one or more polypeptides. The amino acids of the polypeptides, and their relative order in the polypeptide, are the protein's **primary** (**1**°) structure. The 3-dimensional structure of localized segments of the protein's polypeptides is **secondary** (**2**°) structure, while the overall shape of a protein is its **tertiary** (**3**°) structure. **Quaternary** (**4**°) structure describes the way that the individual polypeptides interact with each other in a protein with multiple polypeptides.

Primary (1°) Structure (22.2A)
Amino acid content and **amino acid sequence** are our first concern in elucidation of the structure of a protein.

Content. We identify and quantify the amino acids in a polypeptide using an **automated amino acid analyzer** that performs the following steps illustrated in Figure (graphic 22.21). [graphic 22.21]

(1) Hydrolysis of the polypeptide to its component amino acids by heating in strong acid (6 M HCl) or strong base (2 to 4 M NaOH).

(2) Separation of the resulting mixture of amino acids into its individual amino acids by column chromatography.

(3) Derivatization of the amino acids so they are detected by spectroscopic methods and displayed in a spectrum. [graphic 22.22] (Sometimes derivatization is done before chromatographic separation (step 2)).

(4) Identification of each spectral signal in the spectrum as that of a specific amino acid and determination of the amount of amino acid present from the signal intensity.

Sequence. We determine the *sequence* of amino acids in a polypeptide by identifying its component amino acids one at a time starting at its N-terminus or C-terminus. An important **end-group analysis** procedure for N-terminal amino acids is **Edman Degradation** that first derivatizes and then cleaves the N-terminal amino acid of a polypeptide leaving the rest of the chain intact. [graphic 22.23] *Automated sequenators* carry out N-terminal amino acid derivatization, cleavage, and analysis in a repetitive stepwise fashion allowing accurate determination of as many as 100 amino acid units in a peptide chain.

We can analyze C-terminal amino acids using **carboxypeptidase enzymes**. Such enzymes specifically hydrolyze (cleave) the peptide bond between the C-terminal amino acid and the rest of the chain. [graphic 22.24] They are useful for analyzing only the first few amino acids at the C-terminal end of a peptide because they cleave different C-terminal amino acids at different rates. As a result, C-terminal amino acids of shortened chains formed during the analysis procedure may cleave more rapidly than those of the original polypeptide. This can quickly complicate the reaction mixture with a variety of individual amino acids and shortened peptide chains. **Aminopeptidase enzymes** similarly operate on the N-terminal end of a peptide and have the same limitations as *carboxypeptidases*.

When peptide chains are too long for accurate determination of their full sequence by a method such as Edman degradation, we can cleave them into smaller peptides using **specific peptide cleavage reactions**. We determine

the sequences of these smaller peptides and use them to reconstruct the original sequence of the complete peptide. One of the most specific peptide cleavage reactions uses the enzyme **trypsin** to cleave the peptide bond in $CH(R_x)C(=O)$—$NHCH(R_y)$ where R_x is from *Lys* or *Arg*, while R_y is from any amino acid except *Pro*. As a result, the C-terminus of each cleaved segment is *Lys* or *Arg*, with the exception of the segment with the original C-terminus. [graphic 22.25]

> **More about Enzymatic Peptide Cleavage**. The *carboxypeptidases* and *aminopeptidases* that we have just described are **exopeptidases** because they cleave the *terminal* amide bonds connecting N-terminal or C-terminal amino acids to the rest of the peptide chain. In contrast, *trypsin* is one of several enzymes that cleave amide bonds "inside" the peptide chain (**endopeptidases**). Others are **α-chymotrypsin**, **elastase**, and **pepsin**. *Enzymes* are *proteins* that catalyze specific types of reactions. We will discuss *enzyme catalysis* in the last section of this chapter.

Separation of Individual Peptide Chains. Polypeptide chains of proteins with two or more peptide chains often covalently bind to each other with disulfide bonds (S-S) formed from neighboring SH groups of the amino acid *Cys*. [graphic 22.26] In order to determine content and sequence of individual polypeptide chains of such proteins, we must break these disulfide bonds by *reduction* or *oxidation*. We then separate the chains using **denaturing agents** and isolate them using chromatography. We describe *denaturing agents* later in this chapter.

Secondary (2°) Structure (22.2B)

Secondary (2°) protein structure includes such localized structural features of peptide chains as the **α-helix**, the **β-pleated sheet**, and the *planarity* of amide groups.

Planarity of Amide Groups. We learned in Chapter 15 that amide groups have significant electron delocalization. [graphic 22.27] This causes them to be planar since rotational barriers about their C-N bonds are 60 to 75 kJ/mol (rotational barriers about C-N single bonds are 12 to 20 kJ/mol). While substituents on amide bonds are *cis* or *trans*, the most favorable conformation in peptides is that where C_α's of the chain are *trans*. As a result, we can represent peptide chains as a series of "planes", containing the amide groups, that connect

at the C_α's. [graphic 22.28] Since these planes have a restricted range of relative orientations, amide groups impart significant rigidity to a polypeptide.

Helical Structures. The α-*helix* structure that Pauling and Corey proposed in 1951 is a result of repeating planar *trans* amide groups in peptides. The C=O of each amide group in the α-helical region hydrogen bonds to an N-H of another amide group. Two intervening amide groups separate the hydrogen bonded amide groups leading to a favorable distance of 2.8 Å between O and N in the resulting C=O·····H-N hydrogen bonds. [graphic 22.29] Since C_α's have L-configurations, the α-helix adopts a **right-handed** twist so that the C_α-R groups point away from the chain. α-Helical regions contain on the order of 12 amino acid residues.

β-*Pleated Sheets*. Segments of peptide chains separated by many intervening amino acids also hydrogen bond to each other to give β-*pleated sheets*. [graphic 22.30] These β-*pleated sheets*, that Pauling and Corey also proposed in 1951, are **parallel** or **antiparallel** and the hydrogen bonded segments can be from the same or different peptide chains. A "sheet" includes 2 to 15 peptide chain segments that average about 6 amino acid residues in length. You can see from the figure that the C=O···· H-N bonds in *parallel* β-pleated sheets are distorted and they are less stable than those in *antiparallel* sheets where there is no distortion.

Other Structures. α-*Helical* regions and β-*pleated sheets* connect using other peptide structures called **coils** or **loop conformations**. α-*Helices* and β-*pleated sheets* make up about 50% of peptide chains.

Tertiary (3°) Structure (22.2C)

We generally describe proteins at the *3° structural level* as **globular** or **fibrous**. Proteins of both types contain all of the 2° structural elements we have just described.

Fibrous Proteins. The *fibrous* protein of skin, bone, and muscle has an elongated "fiber-like" shape. An example is **α-keratin** of hair, horns, and nails. [graphic 22.31] Two α-helical polypeptide chains wind together and group with other polypeptide dimers to form **protofilaments**. Two **protofilaments**

comprise a **protofibril** and four **protofibrils** wind together in a **microfibril**. Microfibrils cluster to form **macrofibrils** which come together in groups to form **cells**. A strand of hair includes many cells grouped together in a sheath.

Globular Proteins. The shapes of *globular* proteins also reflect their name. Most *enzymes* are *globular* proteins, and there are many others such as the familiar oxygen transport proteins **myoglobin** and **hemoglobin**. [graphic 22.32] *Myoglobin* has one peptide chain with 153 amino acid residues surrounding a heme group containing an Fe(II) atom. In muscle tissue, it accepts O_2 from oxygenated *hemoglobin* in blood. *Hemoglobin* has four polypeptides with heme groups that have globular tertiary structures like *myoglobin*. *Hemoglobin* binds oxygen in the lungs, transports it to myoglobin, and returns to the lungs with CO_2 formed as a metabolic product in muscle tissue.

Factors that Determine Protein Shape (22.2D)
The 3° structure of a protein depends on the interactions of the amino acid side chains (the R groups) with each other and with the medium surrounding the protein. These interactions involve *electrostatic forces*, *hydrogen bonding*, **hydrophobic bonding**, and *disulfide bonds*. [graphic 22.33] We will see below that the same forces also bind together the multiple peptide chains of a protein in its *quaternary (4°)* structure.

Hydrophobic Bonding. A major influence on the shape and stability of a protein is the desire of *nonpolar* amino acid side chains to minimize their exposure to H_2O. This leads them to the interior of a protein where they interact with each other by *hydrophobic bonding*. For example, nonpolar alkyl groups induce complementary dipoles in each other. While they are individually weak, these attractive induced dipolar interactions taken together have a major effect on the shape and stability of a protein (Figure (graphic 22.33)). Side chains with permanent dipolar groups also interact attractively with each other and induce dipoles in nonpolar groups as well.

> **Thermodynamics of Hydrophobic Bonding**. You may be surprised to learn that the enthalpy (H) of a hydrocarbon in water is more favorable (lower) than the enthalpy of the same hydrocarbon in a hydrocarbon medium. In spite of this, hydrocarbons prefer to dissolve in hydrocarbons and not water because the entropy (S) of a hydrocarbon in

water is less favorable (much lower) than it is in a hydrocarbon medium. Water molecules form highly structured cages around hydrocarbon molecules leading to a decrease in volume of the aqueous solution and a decrease in the entropy of the system. The net result of these effects on H and S is that the free energy (G) for the hydrocarbon in water is greater (less favorable) than in a hydrocarbon medium. The unfavorable $T\Delta S_{(water \rightarrow hydrocarbon)}$ term overwhelms the favorable $\Delta H_{(water \rightarrow hydrocarbon)}$ term in the equation $\Delta G = \Delta H - T\Delta S$ so that $\Delta G_{(water \rightarrow hydrocarbon)}$ is unfavorable.

Electrostatic Interactions and Hydrogen Bonding. In contrast to individually weak *hydrophobic* interactions, *ionic* interactions between *charged polar* side chains within a protein are individually strong (Figure (graphic 22.33)). However, since charged polar groups readily hydrogen bond to H_2O on the exterior of a protein, these groups predominate on the exterior surface of a protein. Uncharged polar groups such as OH and $C(=O)NH_2$ form hydrogen bonds with each other in the interior of a protein, but they also hydrogen bond to H_2O molecules on the surface of proteins and are present in both locations.

Disulfide Bonds. The SH groups of *Cys* stabilize and determine the shape of proteins by forming disulfide bonds (S-S). Their effect on the shapes of proteins is the basis of the "**permanent wave**" process that straightens or curls hair. [graphic 22.34] Application of a reducing agent to the hair breaks S-S bonds to give individual SH side chains. With disulfide bonds broken, curlers or straighteners mechanically modify the shape of the hair. Treatment with an oxidizing agent reforms S-S bonds that hold the hair in a new shape. Disulfide bonds periodically cleave and reform over time allowing the strands of hair to slowly resume their natural configuration.

Quaternary (4°) Structure. Most proteins have more than one polypeptide chain and the ways that these multiple chains interact with each other determine the *quaternary (4°) structure* of the protein. The forces that hold these individual polypeptide chains together are the same as those that determine protein 3° structure. *Hemoglobin* (Figure (graphic 22.32)) has four peptide subunits that come together to form its *4° structure.*

Denaturation. We often refer to a protein with the same shape and activity that it has in an organism as the **native state** or **native form** of the protein.

Denaturation of a protein involves a change in the 4°, 3°, and/or 2° structure of the *native state*. [graphic 22.35] While some proteins denature and then return to their native state (**reversible denaturation**), 2°, 3°, and/or 4° structures of many proteins are sufficiently fragile that denaturation is irreversible. Biochemists studying protein structure and function try to avoid protein denaturation, except to separate individual peptide chains of a protein for content and sequence determination.

Native proteins are often just slightly more thermodynamically stable than *denatured* proteins, so denaturation frequently occurs with relative ease. Denaturating agents, or conditions, include temperature increases, pH changes, detergents, water soluble organic compounds such as alcohols and urea, and ions such as ClO_4^-, SCN^-, Ca^{+2}, Ba^{+2}, and guanidinium ion. [graphic 22.36] A temperature increase provides energy to overcome attractive forces between R groups, while a change in pH alters the charge on acidic or basic side chains (next section) and the attractive interactions between R groups. The chemical agents provide *intermolecular* competition with the *intramolecular* attractive forces between R groups in the native form.

22.3 Properties of α-Amino Acids

We have learned the structures of the 20 "standard" amino acids that organisms use to make peptides and proteins, and have just seen how these R groups influence the structures of peptides and proteins. In this section we examine the properties of α-amino acids as acids and bases as well as their synthetic and biosynthetic origins.

α-Amino Acids Are Polyprotic Acids (22.3A)

The "standard" α-amino acids are diprotic ("H_2A") or triprotic acids ("H_3A").

Diprotic α-Amino Acids. Most "standard" α-amino acids are diprotic acids that we can represent by the general formula ("H_2A"). Since these "H_2A" forms are positively charged, they have the specific formula H_2A^{+1}.

$$H_2A^{+1} \underset{+H^+}{\overset{-H^+}{\rightleftarrows}} HA^0 \underset{+H^+}{\overset{-H^+}{\rightleftarrows}} A^{-1}$$

H_2A^{+1} is in equilibrium with the uncharged monoprotonated form HA^0 and it is in equilibrium with the unprotonated form A^{-1}. An example is *glycine* where H_2A^{+1} is $^+H_3N\text{-}CH_2\text{-}CO_2H$, and A^{-1} is $H_2N\text{-}CH_2\text{-}CO_2^-$ (NH_3^+ and CO_2H have each lost a proton). [graphic 22.37] While these are the only possible structures for H_2A^{+1} and A^{-1}, HA^0 could be either the *zwitterion* form $^+H_3N\text{-}CH_2\text{-}CO_2^-$, or the *uncharged* form $H_2N\text{-}CH_2\text{-}CO_2H$. We stated at the beginning of the chapter that the structure of HA^0 is the *zwitterion* form. This seems reasonable since we would expect a *basic* NH_2 group and *acidic* CO_2H group in the *uncharged* structure to react with each other to form the NH_3^+ and CO_2^- groups of the *zwitterion* form. [graphic 22.38]

Quantitative support for this notion comes from $pK_{a1} = 2.3$ for the equilibrium between H_2A^{+1} and HA^0, and $pK_{a2} = 9.6$ for the equilibrium between HA^0 and A^{-1}. [graphic 22.39] If HA^0 is the *zwitterion* , $pK_{a1} = 2.3$ is that for ionization of CO_2H in H_2A^{+1}. However, if HA^0 is the *uncharged* form, $pK_{a1} = 2.3$ is that for ionization of the NH_3^+ group in H_2A^{+1}. Simple carboxylic acids (RCH_2CO_2H) have pK_a's ≈ 5, while simple aminium ions ($RCH_2NH_3^+$) have pK_a's between 10 and 11. We can see that $pK_{a1} = 2.3$ is much closer to that for CO_2H ionization ($pK_a \approx 5$) than for NH_3^+ ionization ($pK_a \approx 10$ to 11). This means that $pK_{a2} = 9.6$ corresponds to ionization of NH_3^+ in the zwitterion form of HA^0 consistent with the pK_a 10 to 11 for ionization of NH_3^+ in a simple aminium ion ($RCH_2NH_3^+$).

The lower pK_a for ionization of CO_2H in $^+H_3N\text{-}CH_2\text{-}CO_2H$, compared to ionization of CO_2H in RCH_2CO_2H, results from the *electron withdrawing inductive effect* of the positively charged NH_3^+ group (Chapter 14) that stabilizes $^+H_3N\text{-}CH_2\text{-}CO_2^-$. This effect causes the energy difference between $^+H_3N\text{-}CH_2\text{-}CO_2H$ and $^+H_3N\text{-}CH_2\text{-}CO_2^-$ to be less than that between RCH_2CO_2H and $RCH_2CO_2^-$ resulting in a lower pK_a for the amino acid. [graphic 22.40] Besides *Gly*, there are 12 other diprotic "standard" amino acids (Table 22.1). [Table 22.1, next page] Their values of pK_{a1} ($C_\alpha\text{-}CO_2H$) and of pK_{a2} ($C_\alpha\text{-}NH_3^+$) are remarkably similar to those of *Gly* (an exception is pK_{a2} for *Pro* that is structurally different from the others). The average value of pK_{a1} is about 2.2 while that of pK_{a2} (excluding *proline*) is about 9.3. We will discuss the pI values later in this section.

Diprotic Amino Acid Forms at Different pH Values. The pK_a values for the H_2A^{+1}, HA^0, and A^{-1} forms of the 13 diprotic amino acids allow us to predict the pH dependence of their concentrations. [graphic 22.41]

(1) H_2A^{+1} (^+H_3N-CHR-CO_2H) is the major form in solution at pH values smaller than pK_{a1} (below approximately 2.2).

(2) The concentrations of H_2A^{+1} [^+H_3N-CHR-CO_2H] and HA^0 [^+H_3N-CHR-CO_2^-] are equal when the solution pH is equal to pK_{a1} (at approximately 2.2).

Table 22.1. Acid Dissociation Constants for Diprotic Amino Acids

Name	**pK_{a1}** (C_α-CO_2H)	**pK_{a2}** (C_α-NH_3^+)	**pI**
Nonpolar R			
alanine	2.3	9.7	6.0
glycine	2.3	9.6	6.0
isoleucine	2.4	9.6	6.0
leucine	2.4	9.6	6.0
methionine	2.3	9.2	5.7
phenylalanine	1.8	9.1	5.5
proline	2.0	10.6	6.3
tryptophan	2.8	9.4	5.9(?)
valine	2.3	9.6	6.0
Uncharged Polar R			
asparagine	2.0	8.8	5.4
glutamine	2.2	9.1	5.7
serine	2.2	9.2	5.7
threonine	2.1	9.1	5.6
Average	(2.2)	(9.3)*	(5.8)*

* Excluding *proline*

(3) HA^0 (^+H_3N-CHR-CO_2^-) is the major form when the solution pH is between pK_{a1} and pK_{a2} (between approximately 2.2 and 9.3)

(4) The concentrations of HA^0 [^+H_3N-CHR-CO_2^-] and A^{-1} [H_2N-CHR-CO_2^-] are equal when the solution pH is equal to pK_{a2} (approximately 9.3).

(5) A^{-1} (H_2N-CHR-CO_2^-) is the major form when the solution pH is greater than pK_{a2} (above approximately 9.3).

Triprotic α-Amino Acids. The remaining 7 "standard" amino acids (Table 22.2) are *triprotic* acids ("H_3A") because their R's have acidic or basic functional groups in addition to the C_α-CO_2H and C_α-NH_3^+ groups. [Table 22.2] The charges on these forms depend on the specific R group.

$$\underset{\text{"}H_3A\text{"}}{} \overset{-H^+}{\underset{+H^+}{\rightleftarrows}} \underset{\text{"}H_2A\text{"}}{} \overset{-H^+}{\underset{+H^+}{\rightleftarrows}} \underset{\text{"HA"}}{} \overset{-H^+}{\underset{+H^+}{\rightleftarrows}} \underset{\text{"A"}}{}$$

Table 22.2. Acid Dissociation Constants for Triprotic Amino Acids

Name	**pK_{a1}** (C_α-CO_2H)	**pK_{a2}** (C_α-NH_3^+)	**pK_R**	**pI**
Charged Polar R				
aspartic acid	1.9	9.6	3.7	2.8
glutamic acid	2.2	9.7	4.3	3.2
arginine	2.2	9.0	12.5	10.8
histidine	1.8	9.2	6.0	7.6
lysine	2.2	9.0	10.5	9.7
Uncharged Polar R				
cysteine	2.0	10.3	8.2	5.1
tyrosine	2.2	9.1	10.1	5.7
Overall Average	(2.1)	(9.4)		

We continue to use pK_{a1} and pK_{a2} for ionization of C_α-CO_2H and C_α-NH_3^+, and introduce $\mathbf{pK_{aR}}$ as the acid dissociation constant of the acid-base group in R. In spite of major differences in R, the average values of pK_{a1} (2.1), and pK_{a2} (9.4) for these 7 triprotic amino acids are very similar to those of the diprotic amino acids (Tables 22.1 and 22.2). pK_{a1} (C_α-CO_2H) is always the lowest acid dissociation constant for all the standard amino acids (Tables 22.1 and 22.2), but pK_{aR} (C_α-R) can be smaller or larger than pK_{a2} (C_α-NH_3^+) (Table 22.2). We discuss pI values later.

Once we specify R for a triprotic amino acid, we can write unique structures and electrical charges for "H_3A" and "A". However we also need to know relative

values of pK_{aR} and pK_{a2} in order to write structures and electrical charges for "H_2A" and "HA". Because their structures and charges depend on R, we will examine the triprotic amino acids in three separate categories based on their R groups.

Aspartic and Glutamic Acid. *Glu* and *Asp* each have an acidic CO_2H group in R. Since they differ only in the number of CH_2 groups connecting that CO_2H to C_α, their acid dissociation equilibria and associated pK_a values are very similar. [graphic 22.42] Their pK_{aR} values (≈ 4) are closer to pK_a's of simple carboxylic acids (≈ 5) than pK_{a1} values (≈ 2) because one or two CH_2 groups separate the CO_2H groups in R from NH_3^+. The relative concentrations of the four protonated forms of *Glu* or *Asp* depend on pH as illustrated in Figure (graphic 22.43) and their specific electrical charges are H_3A^{+1}, H_2A^0, HA^{-1}, and A^{-2}. [graphic 22.43] At physiological pH (approximately pH 7), their major forms are HA^{-1} where the R grtoups ($CH_2CO_2^-$ or $CH_2CH_2CO_2^-$) have an electrical charge of -1.

Lysine, Arginine, and Histidine. In contrast to *Asp* and *Glu*, the three amino acids *Lys*, *Arg*, and *His* all have *basic* functional groups in R. That group in *Lys* is NH_2, *Arg* has a **guanidino** group ($NHC(=NH)NH_2$), while *His* has an **imidazole** ring. [graphic 22.44] These groups are all positively charged when they are protonated (Figure (graphic 22.44)) so their four forms are H_3A^{+2}, H_2A^{+1}, HA^0, and A^{-1}. Although each R is basic, the pK_{aR} values are significantly different (Table 22.2). The pK_a order for *Lys* and *Arg* is $pK_{a1} < pK_{a2} < \underline{pK_{aR}}$, while that for *His* is $pK_{a1} < \underline{pK_{aR}} < pK_{a2}$. We illustrate the relative concentrations of their four forms as a function of pH in Figure (graphic 22.45). [graphic 22.45] At approximate physiological pH 7, the electrical charge on the R groups of *Arg* and *Lys* is +1. In contrast, the R of *His* is predominantly neutral at pH 7, but a significant amount of the protonated +1 form is also present. (Remember that R in *Glu* and *Asp* is -1 at pH 7.)

Cysteine and Tyrosine. The remaining triprotic amino acids are *Cys* and *Tyr*. While their R groups are acidic, they are much weaker acids than the CO_2H groups of *Glu* and *Asp*. [graphic 22.46] The SH group of *Cys* has an approximate pK_{aR} of 8, while the phenol group of *Tyr* has an approximate pK_{aR} of 10. As a result, the pK_a order for *Cys* is $pK_{a1} < \underline{pK_{aR}} < pK_{a2}$ while that for *Tyr* is $pK_{a1} <$

$pK_{a2} < \underline{pK_{aR}}$. In spite of the differences in pK_a order for these two amino acids, the net electrical charges on the four forms of each of them are H_3A^{+1}, H_2A^0, HA^{-1}, and A^{-2}. We show their relative concentrations as a function of pH in Figure (graphic 22.47). You can see that at pH 7 their R groups predominantly exist in their uncharged protonated forms (CH_2-SH or CH_2-Ph-OH). [graphic 22.47]

Isoelectric Points (22.3B)

The **isoelectric point** (**pI**) for any amino acid is the pH value where the electrically neutral form has its highest concentration. pI values allow us to predict the behavior of amino acids in electrical fields. When we place a solution of amino acids between positive and negative electrodes, amino acids with a net *(+) charge* migrate toward the negative electrode, those with a net *(-) charge* migrate toward the positive electrode, while those that are *uncharged* do not migrate. We can adjust the direction of migration of amino acids in an electric field by raising or lowering pH with respect to their pI values in order to facilitate their separation in solution.

pI Values of Diprotic Amino Acids. We have seen that the electrically neutral form of all diprotic amino acids is HA^0. It is the major form at pH values between pK_{a1} and pK_{a2} (Figure (graphic 22.41)) and its maximum concentration is at a pH (its pI value) midway between these two pK_a values. For example, the pI for *glycine* (R = H) is 6.0 ($pI_{Gly} = (pK_{a1} + pK_{a2})_{Gly}/2 = (2.3 + 9.6)/2 = 6.0$). pI's of all the other diprotic amino acids are also about 6 since their pK_{a1} and their pK_{a2} values are very similar (Table 22.1).

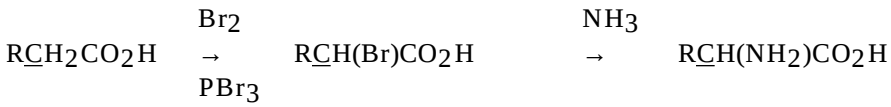
pI Values of Triprotic Amino Acids. In contrast to diprotic amino acids, the electrically neutral forms and pI values of *triprotic* amino acids depend on the structure of R (Table 22.2). To a good approximation, the highest concentration of the electrically neutral form (H_nA^0) of any triprotic acid occurs at a pH midway between the two pK_a's associated with that neutral form ($pK_{a(n+1)}$ and $pK_{a(n)}$). [graphic 22.48] As a result, *Asp* and *Glu* have relatively low pI values of about 3 (see Figure (graphic 22.43)), *Arg*, *Lys*, and *His* have relatively high pI values ranging from 8 to 11 (see Figure (graphic 22.45)), and *Cys* and *Tyr* have pI values between 5 and 6 (see Figure (graphic 22.47)).

pI Values of Proteins. A protein also has a pI value equal to the pH where the number of its positively charged (+) R groups (from *Arg, Lys,* and *His*) is exactly equal to the number of its negatively charged (-) side chains (from *Asp*, *Glu*, *Cys* and *Tyr*). In contrast to amino acids, protein pI values cover the broad range from <1 to >12 because pKa's of their R groups depend on whether R is folded inside or outside the protein and how it interacts with other R's. Proteins often have their lowest solubility at pH = pI and biochemists use this behavior to facilitate their isolation and purification.

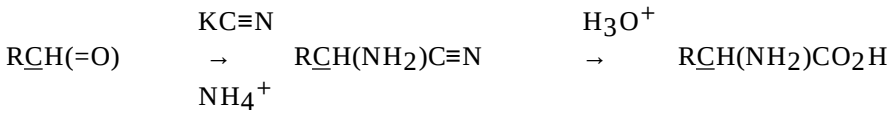
Laboratory Synthesis of Amino Acids (22.3C)

Chemists use the following reactions to synthesize α-amino acids in the laboratory. They give racemic mixtures of the D and L enantiomers, so resolution is required to obtain individual enantiomers.

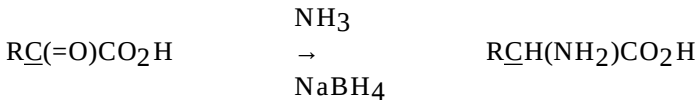
Amination of α-Bromo acids (Chapter xx).



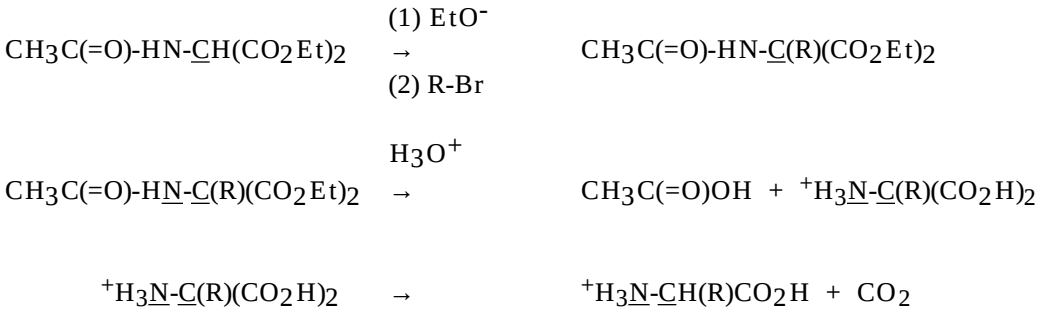
Strecker Synthesis (Chapter xx).



Reductive Amination (Chapter xx).



Diethylacetamidomalonate Synthesis.



Biosynthesis of α-Amino Acids (22.3D)

Humans and other mammals biosynthesize just 10 of the "standard" amino acids in amounts necessary for biosynthesis of proteins. They are called "**non-essential**" amino acids to contrast with the 10 "**essential**" amino acids that we cannot biosynthesize and must obtain directly or indirectly from plants or other sources (Table 22.3).

Table 22.x. Essential and Non-Essential Amino Acids

Non-Essential		**Essential**	
Ala	Gln	Arg*	Met
Asn	Gly	His	Phe
Asp	Pro	Ile	Thr
Cys	Ser	Leu	Trp
Glu	Tyr**	Lys	Val

* We biosynthesize Arg in small quantities.
** Tyr requires Phe for biosynthesis.

Non-Essential Amino Acids. Biosynthesis of *Ala, Asp, Glu,* and *Ser* occurs by NH_2 transfer to α-ketocarboxylates ($RC(=O)CO_2^-$) from α-amino acids ($R'CH(NH_3^+)CO_2^-$) already present in an organism. [graphic 22.49]

$$R\text{-}\underline{C}(=O)CO_2^- + R'\text{-}CH(NH_3^+)CO_2^- \rightarrow R\text{-}\underline{C}H(NH_3^+)CO_2^- + R'\text{-}C(=O)CO_2^-$$

existing ketocarboxylate	existing amino acid	new amino acid	new ketocarboxylate

These "amino transfers" are redox reactions that reduce the $\underline{C}$ of the $\underline{C}$=O group to $\underline{C}$ of the new $\underline{C}H(NH_3^+)$ group. *Asp, Glu,* and *Ser* then serve as biosynthetic precursors to the other non-essential amino acids except *Tyr*. It forms from the *essential* amino acid *Phe*. [graphic 22.50]

Essential Amino Acids. The flow charts in Figures (graphic 22.51) through (graphic 22.54) outline the biosynthetic origins of the essential amino acids in plants and microorganisms. The atoms in the starting materials and products have marks so that you can trace their participation in the overall transformations. Each overall transformation includes many intermediate steps that that biochemistry texts describe in detail. [graphics 22.51-22.54]

22.4 Enzymes and Enzyme Catalysis

Enzymes are proteins that catalyze biochemical reactions in organisms. We will first examine general aspects of enzymes and enzyme catalysis and then the specific mechanistic details of catalysis by α-*chymotrypsin*.

General Features (22.4A)

All enzymes and enzyme catalyzed reactions share a number of general features.

Active Sites. In enzyme catalyzed reactions, the reactant biomolecule (**substrate**) binds to a region of the enzyme called its **active site**. The *active site* is an indentation or cleft in the enzyme where R groups on amino acid residues interact with the *substrate* by noncovalent attractive forces. [graphic 22.55] These attractive forces include *hydrophobic bonding*, *hydrogen bonding*, and *electrostatic interactions*. They are the same as those we described earlier for interactions between R groups of polypeptides.

Enzyme Catalysis Mechanism. Once in the *active site*, a series of reactions transforms the *substrate* into product. These reactions may use amino acid R groups in the active site as reagents, as well as other reactants that diffuse into the active site. The general scheme involves reversible formation of an enzyme-substrate complex (ES) from enzyme (E) and substrate (S), followed by its conversion into product (P) and regeneration of the enzyme.

$$E + S \underset{k_{-1}}{\overset{k_1}{\rightleftarrows}} ES \xrightarrow{k_2} E + P$$

The k_2 step is generally not a single reaction, but includes a number of sequential molecular transformations. We examine such a mechanism for α-chymotrypsin catalyzed hydrolysis of peptides at the end of this section.

Substrate Specificity. Enzyme catalyzed reactions have *stereochemical* and *geometric* specificity. Enzyme active sites have specific stereochemical configurations because their peptide chains contain only L-amino acids. As a result, active sites only interact with specific stereoisomers of chiral substrates, or they only catalyze stereospecific reactions on achiral substrates. As an

example, the enzyme **yeast alcohol dehydrogenase** exclusively removes H_a from the CH_2 group of ethanol giving acetaldehyde containing only H_b. [graphic 22.56] Besides their *stereospecificity*, enzymes often catalyze reactions on one or only a few specific members of a general class of compounds. This *geometric specificity* varies from enzyme to enzyme. Although *yeast alcohol dehydrogenase* slowly dehydrogenates (oxidizes) a number of simple primary alcohols to aldehydes, it overwhelmingly favors *ethanol* as its substrate. In contrast, α-*chymotrypsin* effectively hydrolyzes amide bonds of peptides, amide bonds of simple amides, and ester bonds.

Types of Enzymes. Enzymes often have common names with the ending *ase* added to the name of a substrate, or the name of the reaction, that they catalyze. In addition, systematic names classify them by the general type of process they catalyze. **Oxidoreductases** oxidize or reduce substrates, **transferases** catalyze functional group transfers, **hydrolases** hydrolyze functional groups, **lyases** form double bonds, **isomerases** cause isomerization reactions, and **ligases** make chemical bonds.

α-*Chymotrypsin* (22.4B)

α-*Chymotrypsin* is one of several *hydrolase* enzymes (commonly called **proteases**) that catalyze hydrolysis of amide bonds of peptides to give smaller peptide fragments. [graphic 22.57] It is a globular enzyme composed of 241 amino acid residues.

α-Chymotrypsin Active Site. The active site of α-*chymotrypsin* contains the R groups of its amino acids *His 57*, *Asp 102*, and *Ser 195*. Amino acid residues in polypeptides have sequential numbers, so many other amino acids separate *His 57*, *Asp 102*, and *Ser 195*. In spite of this, their R groups are close neighbors in the active site. They form hydrogen bonds with each other because of the folded 3° structure of the protein. [graphic 22.58]

General Hydrolysis Mechanism. The OH group of *Ser 195* adds to the C=O group of a peptide bond initiating the series of reactions that leads to hydrolysis of the peptide (amide) bond. [graphic 22.59] We represent the enzyme schematically as "E-OH" where the OH is that of Ser 195 and R-C(=O)-NHR' represents the peptide. Hydrolysis cleaves the original peptide into a new N-

terminal peptide fragment H_2NR' and new C-terminal peptide fragment $R\text{-}CO_2H$.

Detailed Hydrolysis Mechanism. The detailed mechanism in Figure (graphic 22.60) shows how the other two R groups in the active site participate. [graphic 22.60]

(1) The peptide (S) forms a complex (ES) with α-*chymotrypsin* (E) in which an amide bond is close to the OH of Ser 195.

(2) The OH of *Ser 195* attacks C=O of the amide bond giving a tetrahedral intermediate.

(3) The C-N bond of the tetrahedral intermediate breaks giving the N-terminal peptide fragment ($R'NH_2$) and an acylated enzyme. *His 57* activated by *Asp 102* provides acid catalysis.

(4) $R'NH_2$ diffuses from the active site and is replaced by H_2O.

(5) H_2O activated by hydrogen bonding to His 57, nucleophilically attacks C=O of the acyl-enzyme intermediate giving a new tetrahedral intermediate.

(6) The C-O-E bond of the tetrahedral intermediate breaks giving the C-terminal peptide fragment RCO_2H that diffuses out of the active site of the enzyme.

Chapter Review

Peptides

(1) Peptides contain α-amino acids ($^+H_3N\text{-}C_\alpha HR\text{-}C(=O)O^-$) joined by amide bonds. (2) When R ≠ H, C_α's are chiral and have L-configurations. (3) There are 20 "standard" α-amino acids biosynthetically incorporated into naturally occurring polypeptides. (4) We can classify R groups of the 20 "standard" α-amino acids as "nonpolar", "uncharged polar", and "charged polar". (5) Names of amino acids have three-letter abbreviations and one-letter designations. (6) Peptides are synthesized in the laboratory by automated peptide synthesis that starts with the *C*-

terminal amino acid bound to a solid resin support and sequentially adds amino acids, using N-protection and carbonyl activation.

Protein Structure and Organization

(1) *Primary (1°)* protein structure includes amino acid *content* and *sequence.* (2) Content is determined using automated amino acid analyzers that hydrolyze proteins, chromatographically separate the individual amino acids, and provide a spectral display of their derivatives. (3) Sequence is determined using Edman degradation (N-terminal), and carboxypeptidases (C-terminal), in conjunction with peptide cleavage reactions catalyzed by endopeptidases. (4) Before determining content or sequence, disulfide bonds are cleaved by reduction or oxidation. (5) *Secondary (2°)* protein structure includes planar electron delocalized amide groups, as well as α-helices and β-pleated sheets resulting from hydrogen bonding between separated amide groups. (6) The *fibrous* and *globular tertiary (3°)* structures of proteins result from interactions of the amino acid R groups with each other and with water. (7) R group interactions include hydrophobic bonding, electrostatic interactions, hydrogen bonding, and disulfide bond formation. (8) Protein quaternary (4°) structure is the result of interactions between individual polypeptides in a protein with two or more peptide chains. (9) Native states of proteins are denatured by heat, pH changes, certain organic compounds, and ions, because they disrupt favorable interactions between R groups.

Properties of α-Amino Acids

(1) Depending on their R group, α-amino acids are diprotic or triprotic acids. (2) For the 13 *diprotic* α-amino acids, pK_{a1} for $C_\alpha\text{-}NH_3^+ \approx 2.2$, while pK_{a2} for $C_\alpha\text{-}CO_2H \approx 9.3$. (3) For diprotic acids, H_2A^+ predominates below pH ≈ 2.2, HA predominates between pH ≈ 2.2 to 9.3, while A^- predominates above pH ≈ 9.3. (4) Values of pK_{a1} ($C_\alpha\text{-}NH_3^+$) and pK_{a2} ($C_\alpha\text{-}CO_2H$) for triprotic amino acids are about the same as those of diprotic acids, but pK_{aR} values depend on R. (5) Fully protonated forms of the triprotic acids *Asp* and *Glu* have the formula H_3A^+ and their R groups are (-) at physiological pH, those of *Lys*, *Arg*, and *His* are H_3A^{+2} and their R groups are (+) at physiological pH, while those of *Cys* and *Tyr* are H_3A^+ and their R groups are uncharged at physiological pH. (6) pI values for diprotic amino acids are approximately 6, those of the triprotic acids *Asp* and *Glu* are about 3, those for *Lys*, *Arg*, and *His* are between 8 and 11, while those of *Cys* and *Tyr* are between 5 and 6. (7) Racemic mixtures of α-amino acids arise from (a) amination of α-bromocarboxylic acids, (b) the Strecker synthesis, (c) reductive amination of α-ketocarboxylic acids, and (d) the diethylacetamidomalonate synthesis. (8) Humans and other animals biosynthesize "non-essential" "standard" α-amino acids *Ala*, *Asp*, *Glu*, and *Ser*, by amino transfer reactions to α-ketocarboxylates, and these in turn serve as biosynthetic

precursors for *Asp*, *Glu*, and *Ser*. (9) Plants and microorganisms, but not humans or other animals, biosynthesize "essential" "standard" α-amino acids.

Enzymes and Enzyme Catalysis

(1) Enzymes are proteins that catalyze biochemical reactions. (2) Substrates bind to the active sites of enzymes and the resulting enzyme-substrate complexes undergo reactions leading to the final product and regenerate active enzyme. (3) Enzyme-catalyzed reactions are stereochemically and geometrically specific. (4) Enzymes are classified as *oxidoreductases*, *transferases*, *hydrolases*, *lyases*, *isomerases*, and *ligases*. (5) *α-chymotrypsin* is a hydrolase for peptide and amide bonds. (6) In the active site of α-chymotrypsin, *Ser 195* nucleophilically attacks the C=O of an amide bond leading to formation of a tetrahedral intermediate, C-N cleavage is catalyzed by *His 57* giving an N-terminal peptide fragment, H_2O attacks the acyl-enzyme, and the resulting tetrahedral intermediate decomposes to give a C-terminal peptide fragment and active enzyme.