

Chap. 3. "Amino Acids and the Primary Structures of Proteins"

Reading Assignment: pp. 52-72 & 78-80.

Problem Assignment: 1, 4, 6, 8-9, 14, 16.

I. Introduction to protein functions.

Proteins carry out most of the work and provide much of the supporting framework of cells. Examples of functions performed are listed below.

- A. Enzymatic catalysis
 1. ATP synthase (make ATP in mitochondria)
 2. trypsin (food digestion)
 3. P450 (drug metabolism)
- B. Transport and storage
 1. hemoglobin (transports O₂ to tissues)
 2. HDL (lipids from liver to tissues)
- C. Motion
 1. actin & myosin (muscular contraction)
 2. flagellin (bacterial flagella)
- D. Mechanical support
 1. collagen (tendon connective tissue)
 2. elastin (ligaments)
 3. fibroins (silk)
- E. Protection
 1. antibodies (neutralize invading bacteria)
 2. toxins (many species)
- F. Nerve signal transmission
 1. acetylcholine receptor (synapses of cholinergic nerves)
 2. opiate receptor (pain diminution)
 3. multiple forms of each
- G. Growth and differentiation
 1. insulin (cell growth, carbohydrate and fat utilization)
 2. thyroid stimulating hormone (body metabolic rate)

II. General structure of amino acids.

Proteins are polymers of the 20 common (or "standard") amino acids. The chemical properties of proteins are dictated by the chemical properties of the amino acids they contain. Amino acids contain both amino (-NH₂) and carboxyl (-COOH) functional groups (**Fig. 3.1**). The amino group is located in the α position relative to the -COOH group. A hydrogen and a side-chain or R-group also are always attached to the α -carbon. The acidic -COOH group is ionized to -COO⁻, and the basic -NH₂ group is protonated to -NH₃⁺ at physiological pH. Because both are charged at the same time, amino acids are sometimes referred to as zwitterions (dipolar ions). A

form sometimes shown in structure representations in which the amino group and carboxyl group both are uncharged occurs very rarely in solution. The side-chain is the structurally variable part of an amino acid. Many side-chain functional groups can be protonated or deprotonated depending on the pH. An α , β , γ , etc. numbering system is used to label carbons. Carbons within the side-chain are numbered outward starting with position β .

In all amino acids except glycine, four different substituent groups are attached to the α -carbon, which is the chiral carbon. Although two configurations of substituents are theoretically possible (two possible enantiomers, or non-superimposable mirror images (**Fig. 3.2**)), only the *s* absolute configuration is observed for the α -carbons of the 20 standard amino acids (**Box 3.1**). The 20 standard amino acids used in ribosomal protein synthesis are assigned the L configuration based on the reference compound, L-glyceraldehyde, whose chiral carbon also has the *s* configuration. The configurations of L-serine and D-serine are compared in **Fig. 3.2**. When drawn in Fisher form, L-amino acids have the amino group on the left, and D-amino acids have the amino group on the right.

III. Structures of the 20 common amino acids.

Amino acids are classified according to the type of R-group they contain (see structures pp. 57-60).

1. right up front have to memorize the names, both 3 and 1 letter abbreviations and the R group structure, just like the alphabet for writing

2. the R groups are the key to how proteins fold, their shapes, their functions and their evolution

3. if it helps you memorize then use the groups in the book but functionally these are the ones that work

a. small neutral: G, A, and P (FIGURES)

b. polar: S, T, Q, N, and C

c. hydrophobic: V, L, I, F, Y, W, and M

d. charged: H, K, R, D, and E

4. Small neutral

a. Gly is smallest and has the greatest ability to rotate in the protein, so is used for close contact

b. Ala is slightly larger so can be use in close contact, esp. with hydrophobic amino acids or can be reasonably put in water

c. Pro is cyclic and is really an IMINO acid, it causes the protein polymer to kink at about 180°

5. Polar: all are capable of H-bonding and dipole interactions

a. Ser has an OH group that is generally uncharged but the H can be removed to give a very strong nucleophile frequently used in enzymes

b. Thr also has an OH group but is used to interact with water, other polar or charged amino acids or polar compounds

c. Gln is the amide form of Glu

d. Asn is the amide form of Asp

e. Cys is the sulphur analog of Ser, it participates in some enzyme reactions like Ser and also can form a bond with another Cys FIG 3.4

9. All of the hydrophobic amino acids are rarely found exposed to water and are usually buried in the interior of the protein forming hydrophobic and Van der Waals interactions TABLE 3.1

a. V, L, I and F have different shapes but similar uses, note F is aromatic and that can provide additional interactions

b. Y and W have some polarity but are still basically hydrophobic and rarely found exposed to water, the H on Y OH can be removed but only with substantial effort

c. M has sulfur instead of carbon but basically is analogous to L except that the terminal CH₃ can be removed and put back under some conditions

10. Positively charged

a. K and R are positively charged under all physiological conditions, K due to its primary amine can be modified by a variety of different enzymes

b. H is unique in that its pK_a is close to physiological pH thus it plays a unique role in many enzymes by donating or accepting H⁺

11. Negatively charged

a. Asp is always ionized and provides water interactions

b. Glu is also always ionized and is used frequently for interaction with positively charged compounds the enzymes use

A scale is presented in **Table 3.1** that compares the relative hydrophobicity of the side chains of amino acids. The scale is based on the free energy change for transfer of an amino acid from the interior of a membrane to water. The scale is quite useful for predicting whether an unknown protein may be associated with a cell membrane, etc.

IV. Other amino acids and amino acid derivatives.

Amino acids can be converted to many other important compounds such as neurotransmitters γ -aminobutyrate and hormones (epinephrine and thyroxine) (**Fig. 3.5**). Two rare amino acids that are incorporated just like the usual 20 are shown in the Fig. on pg 62. The D-forms of amino acids can be synthesized by bacteria, and often are incorporated into the cell wall. The linkages between D-amino acids are more resistant to attack by common proteases.

V. Ionization of amino acids.

The charge properties of amino acids depend on the pH of the solution and the values of pK_as for their side-chains. Simple amino acids, such as glycine and alanine, contain only the α -carboxyl and α -amino functional groups. However, about half of the amino acids also contain R groups with a third ionizable functional group (**Table 3.2**). The pK_a of the α -COOH group (pK₁) is typically about 2.4. The pK_a of the α -NH₃⁺ group (pK₂) is typically about 9.9. pK_Rs depend on the chemistry of the functional group and may be either acidic (Glu and Asp) or basic (Lys and Arg) (**Table 3.2**). As explained in **Fig. 3.8**, glutamate and aspartate side-chains are acidic because the formation of the dissociated (carboxylate) form of their side-chains is favored by delocalization of the negative charge across the 3 atoms of the carboxylate group. On the other hand, the protonated positively charged form of the guanidinium ion of the arginine side chain can be readily formed due to delocalization of the positive charge among the atoms of this group.

Two buffering plateaus occur on the titration curves of simple amino acids (e.g., alanine **Fig. 3.6**). Amino acids with ionizable R-groups display three buffering plateaus (e.g., histidine, **Fig. 3.7**). The only generalization that can be made about these curves is that the α -carboxyl group always is the most acidic group and is the first to titrate on addition of strong base. The order in which the remaining functional groups titrate depends on their specific pK_a values. When the pH is such that the net charge on the amino acid is zero, the titration curves are at the so-called isoelectric points (pI)-- the pH at which the net charge of the amino acid is zero. At this pH, an amino acid will not migrate in an electric field. The pI occurs precisely halfway between the pK_{as} of the ionizable groups that when titrated produce, or remove the zero-charge form from solution (**Figs. 3.6 & 3.7**).

Solutions of amino acids are good buffers when the solution pH equals either pK_1 , pK_2 , or pK_R . Solutions are very poor buffers at the pI. The relative proportions of charged forms can be determined for any point on an amino acid titration curve using the Henderson-Hasselbalch equation. An example of these calculations will be given in class.

VI. Peptide bonds.

Peptides and proteins are formed by the polymerization of amino acids into chains. The linear sequence of amino acids in a protein is referred to as its primary structure. Amino acids are joined together in chains via amide linkages known as peptide bonds (**Fig. 3.9**). The α -amino group and α -carboxyl group lose their charge and ability to ionize at physiological pH when combined together in a peptide bond. Depending on the number of amino acids, the polymer is called a peptide (typically < 10 amino acids), an oligopeptide (10 - 20 amino acids), or a polypeptide (protein, > 20 amino acids). Note that the cut-offs for the number of residues is highly arbitrary. Short peptides of 2, 3, and 4 amino acids are referred to as di-, tri-, and tetrapeptides. The artificial sweetener aspartame is a dipeptide composed of aspartate and phenylalanine methyl ester (**Fig. 3.10**).

The alternating $-NH-C_\alpha-CO-$ atoms of the polypeptide chain are called its backbone. The R-groups of amino acids in peptides and polypeptides still maintain approximately the same ionization properties as they have in individual amino acids. Chains have a polarity; the sequence H_2N -ala-gly-trp-COOH is not the same as H_2N -trp-gly-ala-COOH. The end of the chain with the free α -amino group is called the N-terminus. The end having the free α -carboxyl group is called the C-terminus. These ends maintain their respective positive and negative charges in the polymer. Peptides and polypeptides are numbered from the N- to the C-terminus. It is standard convention to write the sequence of a polypeptide using three-letter or single-letter abbreviations, left-to-right from the N- to C-terminus. Always assume this is the case when reading amino acid sequences, unless otherwise stated.

VII. Protein purification techniques.

Purification methods are very important for determination of the structure and function of proteins and enzymes. Examples of the kinds of information that can be learned about the protein are its size, charge, and ligand binding affinity. In all purifications, one starts with a tissue/cell that may be enriched in the protein of interest, disrupts the sample, and through a series of steps obtains an increasingly purer preparation in which unwanted contaminating proteins are

removed. Purification procedures require a method for monitoring protein purity, such as a specific enzyme assay, or an antibody detection technique. In this section, some of the procedures commonly used in protein purification are discussed. These methods exploit the often subtle differences in the physical properties of the proteins in the original mixture.

A. Crude separation procedures.

Proteins differ with respect to their solubilities at high salt concentrations. Crude separations can be achieved by incrementally increasing the salt concentration until the protein of interest precipitates. Most proteins can be recovered by re-dissolving the precipitate in a buffer lacking salt. Ammonium sulfate commonly is used as precipitant, because it rarely harms the precipitated protein. Typically, only a two- to three-fold purification is achieved by ammonium sulfate precipitation.

Residual salt in the re-dissolved protein sample often is removed by a procedure such as dialysis. In dialysis, the protein sample is placed inside a bag made from a semi-permeable membrane which allows molecules below a certain molecular weight cut-off to pass through. The bag is placed inside a container of buffer, and the small solutes diffuse into the buffer over time. After several hours, the protein becomes free of unwanted solutes.

B. Column chromatography procedures.

Higher levels of purity usually are achieved by separating and purifying proteins on the basis of their size, charge and ligand binding properties. In each case, the procedure of column chromatography typically is used (**Fig. 3.11**). In this procedure, a solution containing the protein is applied to the top of a column containing a matrix of insoluble "resin" which accomplishes the purification as the protein passes through. The purified protein comes out at the bottom in the solution called the eluate. The eluate is collected in fractions, and the fraction containing the protein is identified by an assay or by the absorbance at 280 nm, if the protein contains tryptophan.

1. Ion-exchange chromatography (see Fig. on p. 29). Some resins separate proteins based on their net charge. Protein charge depends on amino acid composition and the pH. A mixture of proteins in a buffer at a given pH will have different charges because they have different contents of acidic and basic amino acid residues. Thus, they will differ in their ability to bind to a positively or negatively charged resin. As an example, an acidic, negatively charged protein will pass rapidly through a negatively charged resin, whereas a basic, positively charged protein will be bound. Thus, the two proteins can be separated. Proteins bound to the resin can be eluted by changing their charge by manipulating the buffer pH. A protein also can be eluted by adding a salt that competes with the protein for binding to the charged groups of the resin. A commonly used positively charged resin is diethylaminoethyl (DEAE) cellulose, which is called an "anion exchange" resin because it binds anionic (negatively charged) proteins at neutral pH. Carboxymethyl (CM) cellulose is a commonly used negatively charged resin, which is used as a "cation exchange" resin, because it binds cationic (positively charged) proteins at neutral pH.

2. Gel-filtration chromatography (see Fig. on p. 30). Some resins separate proteins by size, and in this case the procedure is called gel-filtration chromatography. Proteins are passed through a column containing porous polymer beads suspended in aqueous medium. Small proteins can partition in and out of the pores on the surfaces of the beads, whereas proteins above a certain size cut-off are restricted to the buffer outside the beads. Larger proteins therefore pass down the column faster than smaller proteins, resulting in separation.

3. Affinity chromatography (see Fig. on p. 31). In this procedure, the protein of interest is purified on a resin containing a "ligand" to which the protein binds with high affinity. Other proteins applied to the resin pass through the column and are removed. The bound protein subsequently is eluted by adding a solution of free ligand which competes with resin-associated ligand for binding to the protein. Affinity chromatography can sometimes purify a protein by 1,000-fold or more in a single step.

C. Analytical procedures.

The state of purity of a protein preparation can be monitored by a technique called sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS PAGE) (**Fig. 3.12**). Electrophoresis refers to the migration of charged molecules in an electric field. Molecules move toward the electrode having the opposite charge. The negatively charged electrode is called the cathode because it attracts cations. The positively charged electrode is called the anode because it attracts anions. The rate of migration is proportional to the charge of the molecule and the strength of the electric field.

Although SDS PAGE is an electrophoresis technique, it separates proteins by size. SDS is a negatively charged detergent that binds to proteins at ~1 molecule of SDS per 2 amino acid residues. Binding imparts a uniformly negative charge to the protein, and the protein adopts a rod-like shape in which SDS coats the surface. Treated proteins have roughly the same charge per unit mass. Disulfide bonds typically are broken prior to analysis to allow proteins to become fully unfolded (**Fig. 3.17**). Polyacrylamide is a porous gel matrix formed by free radical polymerization of acrylamide in the presence of the cross-linker, methylenebisacrylamide. Pore size is determined by the percentage of methylenebisacrylamide in the polymerization reaction. Large proteins move relatively slowly through an acrylamide gel matrix compared to small proteins. Separated proteins are detected by staining using various dyes.

VIII. Determination of amino acid composition.

Proteins differ in amino acid composition. Thus, the amino acid composition of a protein can be useful in establishing its identity. To determine composition, proteins first are broken down (hydrolyzed) to their individual amino acid residues by cleaving peptide bonds using strong acid (6 M HCl) (**Fig. 3.13**). Then the amino acids are reacted with a chromogenic reagent such as phenylisothiocyanate (PITC) (**Fig. 3.14**) which tags them so that they can be detected by virtue of their absorbance at 254 nm. The PTC-tagged amino acids are separated by chromatography, and the percentages of each amino acid determined by comparing the relative areas of the peaks appearing in the chromatogram (**Fig. 3.15**).

The method has a few limitations in that a fraction of the serine, threonine, and tyrosine residues are lost, and tryptophan is completely destroyed and is not detected. In addition, the amino groups of the amide side chains of asparagine and glutamine are cleaved off resulting in the lumping together of these amines with their acid counterparts, aspartate and glutamate. Nonetheless, the technique can be used to confirm the ID of an unknown as illustrated by comparing the data shown in **Table 3.3**.

IX. Determination of amino acid sequence.

A. Protein sequence determination by DNA sequencing.

Due to the technical difficulties associated with protein sequencing, protein sequences typically are determined today by first cloning the gene for the protein of interest and sequencing the DNA. The protein sequence then can be determined by decoding the DNA sequence using the genetic code. Nonetheless, peptide sequencing remains a mainstay of modern biochemistry research. It typically is used to obtain part of the sequence of a protein so that a DNA probe can be constructed for gene cloning procedures. The use of protein sequencing in this application will be covered in the final part of the course when we discuss recombinant DNA technology.

X. Phylogenetic and evolutionary relationships revealed by protein sequencing.

Comparison of protein sequences is important for predicting the function of a newly identified protein and establishing evolutionary relationships between organisms. Indeed, phylogenetic trees can be constructed by comparing the sequences of highly conserved proteins such as cytochrome *c* (**Fig. 3.21 & 3.22**). Sequence analysis also can be used to identify signature sequences that are characteristic of a particular metabolic function or cellular location.