

Amino Acids, Peptides, and Proteins

Objectives:

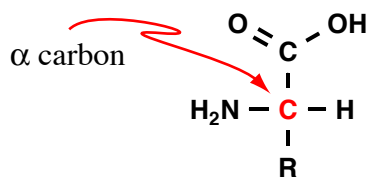
- I. Know the general structure of an α amino acid.
- II. Define the ionic / pH properties of amino acids.
 - A. Conditions where the amino acid is a zwitterion.
 - B. Conditions where the amino acid is at its isoelectric point.
 1. How to calculate the isoelectric point of an amino acid given the pKa's of the ionizable groups.
 - C. Conditions where the amino acid is in its cationic form.
 - D. Conditions where the amino acid is in its anionic form.
- III. Classify a given amino acid as hydrophobic, polar and uncharged, polar with a negative charge (acidic) or polar with a positive charge (basic).
 - A. Subclassifications
 1. Aromatic amino acids.
 - a) Properties conferred to proteins by the presence of these amino acids.
 2. Sulfur containing amino acids.
- IV. Amino acids with a chiral carbon (tetrahedral stereocenter) in their structure.
 - A. D / L enantiomeric forms versus R / S enantiomeric forms.
- V. Describe the structure of the peptide bond.
- VI. Describe the primary structure of proteins.
 - A. Conditions where peptides and/or proteins are at its isoelectric point.
 - B. Conditions where peptides and/or proteins are in a cationic form.
 - C. Conditions where peptides and/or proteins are in an anionic form.
- VII. Classification according to size.
 - A. Peptides
 - B. Oligopeptides
 - C. Polypeptides
 - D. Proteins
 1. Monomeric Proteins
 2. Multimeric Proteins
 - a) Homomultimeric Proteins
 - b) Heteromultimeric Proteins
 - c) Subunits
- VIII. Be familiar with the convention for drawing the structure and/or writing the amino acid sequence of peptides, oligopeptides, and proteins.
- IX. Classification of proteins based upon biological function.



The α Amino Acids

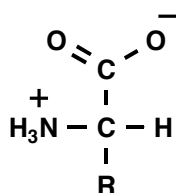
As the name implies, amino acids contain a carboxylic acid group and an amino group. If this was the only criteria for a biologically active amino acid, there would be thousands of possible amino acids. The 20 amino acids found in nature are called alpha (α) amino acids based upon an archaic system of carboxylic acid nomenclature. In this system the carboxyl carbon is carbon 1 and the carbon adjacent to it (carbon 2) is

designated the α carbon. The general structure of the amino acids found in proteins could be depicted as:



Typical monoprotic organic acids have ionization constants $\leq 10^{-3}$, pK_a 's greater than 3. The carboxylic acid groups of the α amino acids are stronger acids than typical monoprotic organic acids. These carboxyl groups have pK_a 's in the range of 1.8 to 2.4. Amino groups are weak bases. Typical monoamines have ionization constants $\leq 10^{-10}$, pK_a 's ≥ 10 . The α amino group is a weaker base (stronger acid) than most simple amines, having a pK_a in the range of 8.8 to 10.8.

Since the carboxyl group is an acid and the α amino group is a base, in solution amino acids react with and neutralize each other. The general structure given above for the amino acids is incorrect. A more correct structure for an amino acid in solution, in the cell is:

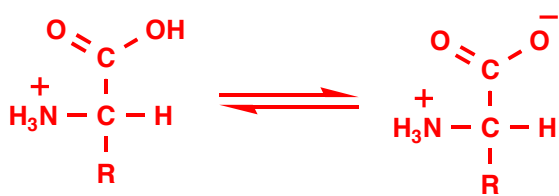
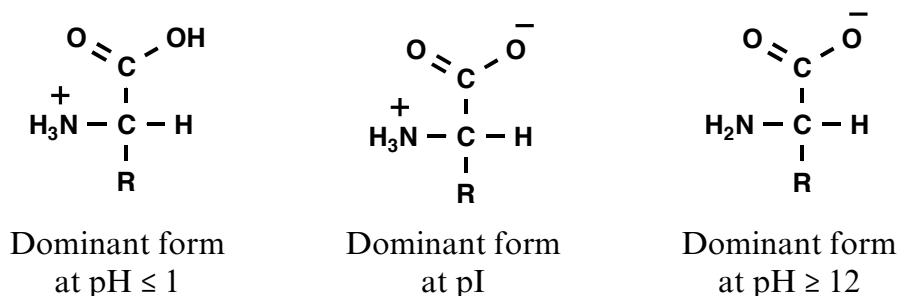
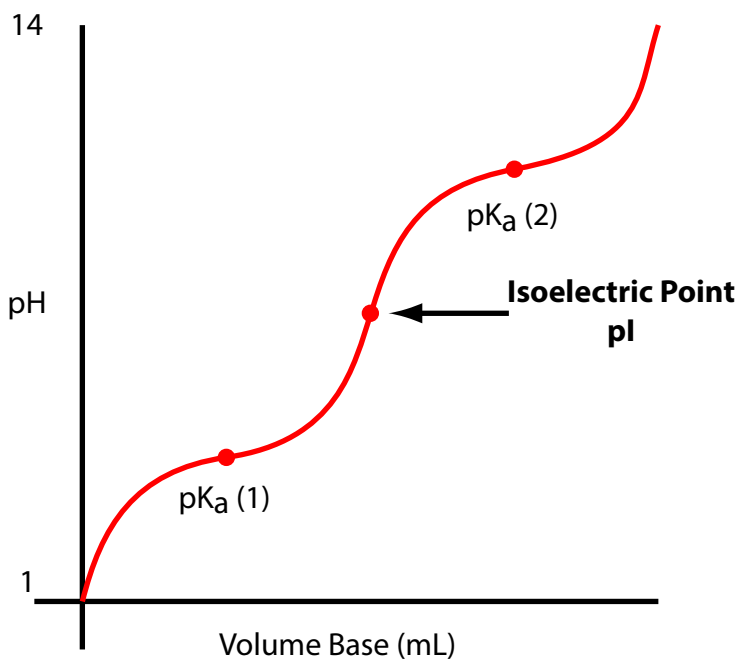


In this structure the carboxyl group has released a hydrogen ion and the amino group has accepted a proton forming a DIPOLAR ION. This dipolar ionic form exists when amino acids are in solution and when they are in the solid state. The dipolar ionic form explains many of the properties of the amino acids. For example their melting points, boiling points, and solubility more closely resemble those of ionic salts rather than organic molecules. The dipolar ionic form of an amino acid is also called a ZWITTERION. Any molecular compound that contains a mixture of positive and negative charges is a DIPOLAR ION, a ZWITTERION.

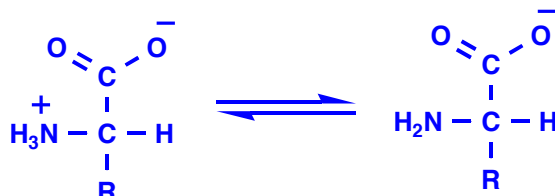
pH Properties of the Amino Acids

The ionic state of the amino acids depends upon the pH of the solution. At low pH ($\text{pH} \leq 1.0$) the amino acid is in the cationic form because the amino and carboxylic acid groups are protonated. This form will migrate toward the negative pole (cathode) if placed in an electric field. As the pH is increased the carboxylic acid group begins to ionize, lose its proton. When the pH is equal to the pK_a of the carboxyl group the amino acid exists as a 50:50 mixture of the cationic and zwitter ionic forms. Adding more base results in continued ionization of the carboxylic acid group until the zwitter ionic form is the predominant form of the amino acid in solution. The pK_a of the amino group is reached as more base is added and at this point the amino acid exists as a 50:50 mixture of the zwitter ionic form and the anionic form. As the pH is increased further the amino group continues to lose its proton and ultimately, at high pH ($\text{pH} \geq 12.0$), the anionic form is the predominant form in solution. The anionic form will migrate toward the positive pole (anode) if placed in an electric field. This titration curve demonstrates that an amino acids never exist in an uncharged form.

As the amino acid is titrated, there is a solution pH where the amount of positive charge on the molecule is exactly balanced by the amount of negative charge. At this pH the dissolved amino acid behaves as if it has no net charge because the amount of positive charge is exactly balanced by an equal amount of negative charge. The solution pH at which the opposite charges exactly balance one another is called the ISOELECTRIC pH or ISOELECTRIC POINT (pI) of the amino acid.



Forms present in 50:50 mix
at pK_a(1)



Forms present in 50:50 mix
at pK_a(2)

At the isoelectric point an amino acid will not migrate in an electric field since it has no net charge. At solution pH's below the isoelectric point, the amino acid has a net positive charge and it will migrate toward the negative pole of an electric field, toward the cathode. At solution pH's above the isoelectric point the

molecule has a net negative charge and it will migrate toward the positive pole of an electric field, toward the anode.

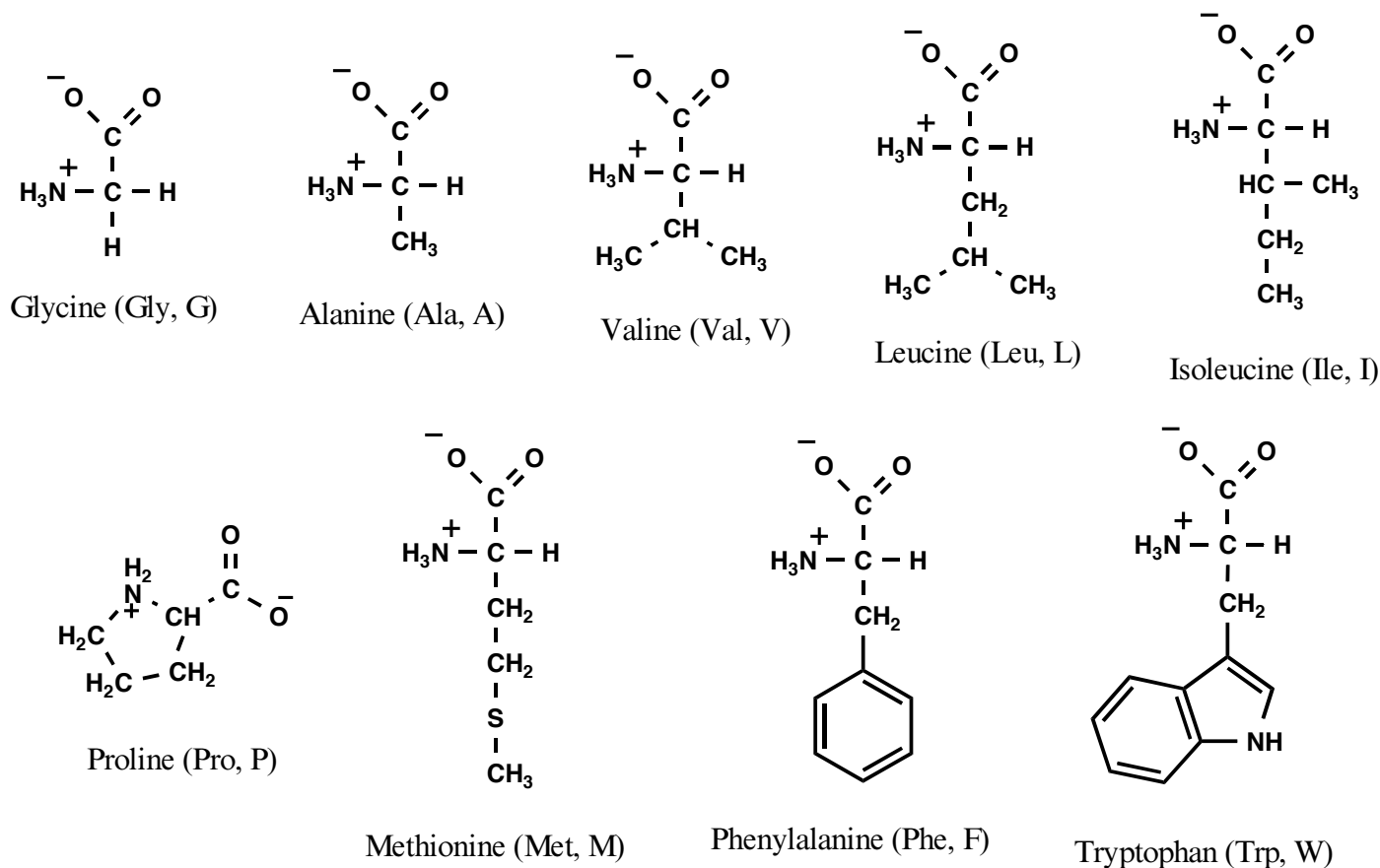
The R Groups - The Amino Acid Side Chains

The different R groups, the different side chains of the amino acids, give each of the 20 amino acids different chemical and physical properties. Different combinations of the amino acids within a protein give the protein its unique chemical and physical properties. The unique sequence of amino acids defines the structure and function of the protein.

Its time to examine the structure of the side chains of the 20 amino acids naturally occurring in proteins.

The amino acids can be divided into two major groups. Nonpolar (Hydrophobic) Amino Acids have side chains comprised primarily of carbon and hydrogen. The other major group, the Polar (Hydrophilic) Amino Acids, have electronegative elements (O, N, S) as part of their side chains. This second group can be divided into three subgroups: Polar But Not Charged Amino Acids, Amino Acids with Negatively Charged Side Chains at physiological pH (Acidic Amino Acids), and Amino Acids with Positively Charged Side Chains at physiological pH (Basic Amino Acids).

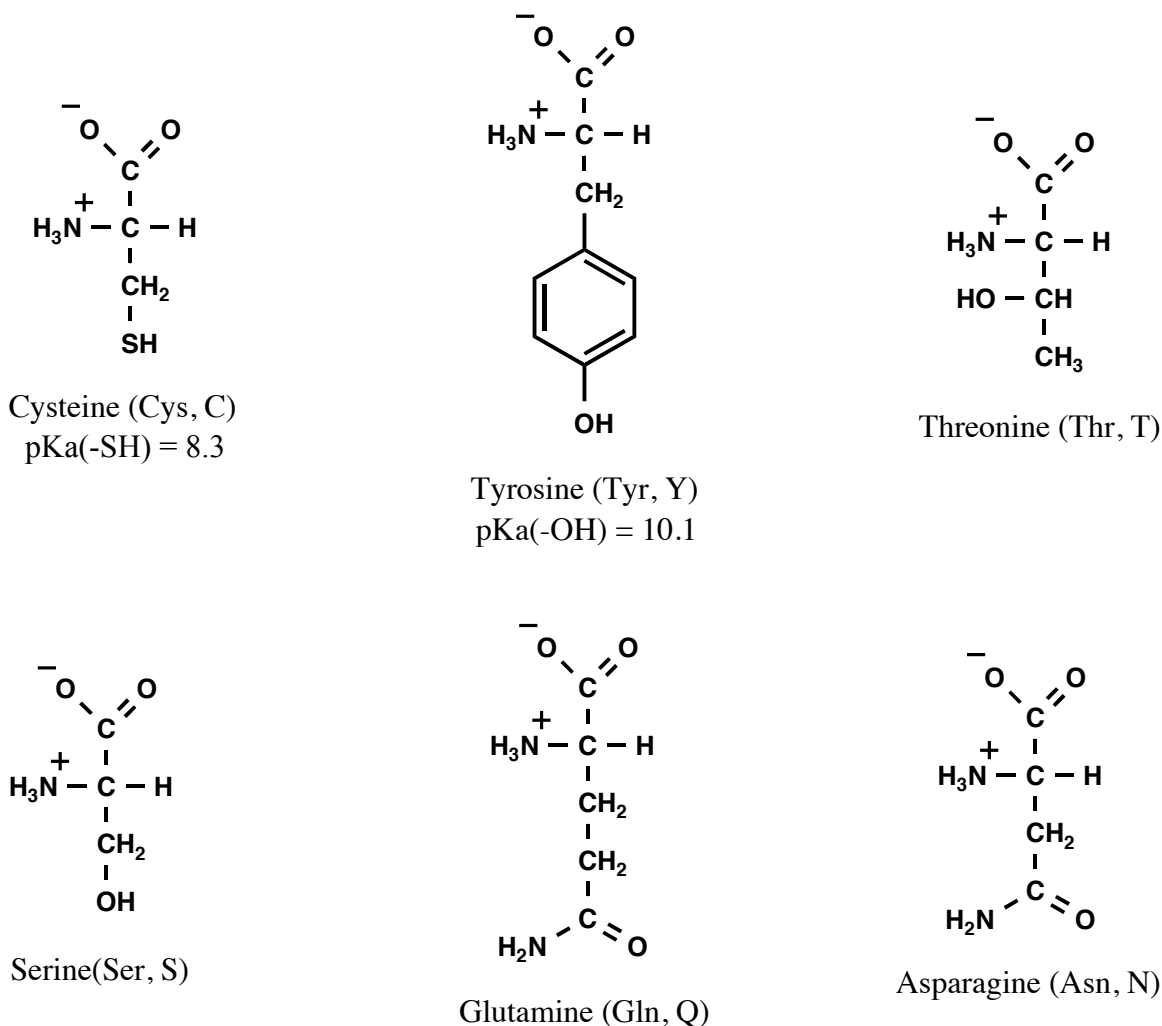
The Hydrophobic (Nonpolar) Amino Acids



Note 1: The amino acids Tryptophan (Trp) and Glycine (Gly) are sometimes segregated from the other

amino acid groups. Tryptophan is sometimes separated because the nitrogen that is part of the indole ring system makes the side chain very slightly polar. However, the hydrocarbon part of the complex aromatic ring system shields much of the polar nature of the nitrogen. Depending upon the external environment and/or the environment in the protein surrounding the Tryptophan, the Trp can have properties somewhere between a nonpolar and polar but not charged amino acid. Glycine is sometimes segregated because its side chain is a small hydrogen atom. Glycine can likewise have physical and chemical properties that fall somewhere between those of the nonpolar amino acids and the polar but not charged group of amino acids.

The Polar But Not Charged Amino Acids



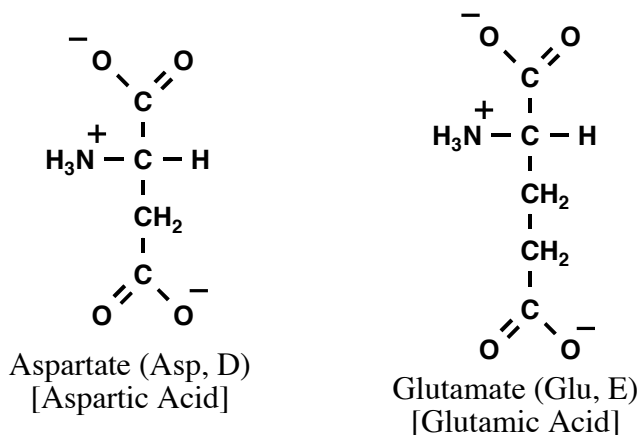
Note 2: The side chains of cysteine and tyrosine have weak acid / base properties *in vitro*. When cysteine is titrated on the bench top, the thiol (-SH) group loses its proton with a pK_a of about 8.3. Similarly, titration of tyrosine results in the hydroxyl (-OH) group of the side chain losing its proton at a pK_a of about 10.1. These pK_a's can be / are measured *in vitro*. However, physiological pH, pH 7.4, is far too low for these side chains ionize *in vivo*. In the cell these side chains are always in the protonated, uncharged form.

Note 3: Tryptophan, Tyrosine, and Phenylalanine are the aromatic amino acids. The side chains of these amino acids all contain an aromatic ring as part of their structure. The aromatic groups of Trp, Tyr, and to a

lesser extent Phe absorb light in the ultraviolet region of the spectrum, at 280 nm. This property is employed to detect and quantify proteins in solutions and/or in biological fluids. Light absorbance 280 nm is the only nondestructive method for detecting and quantifying proteins.

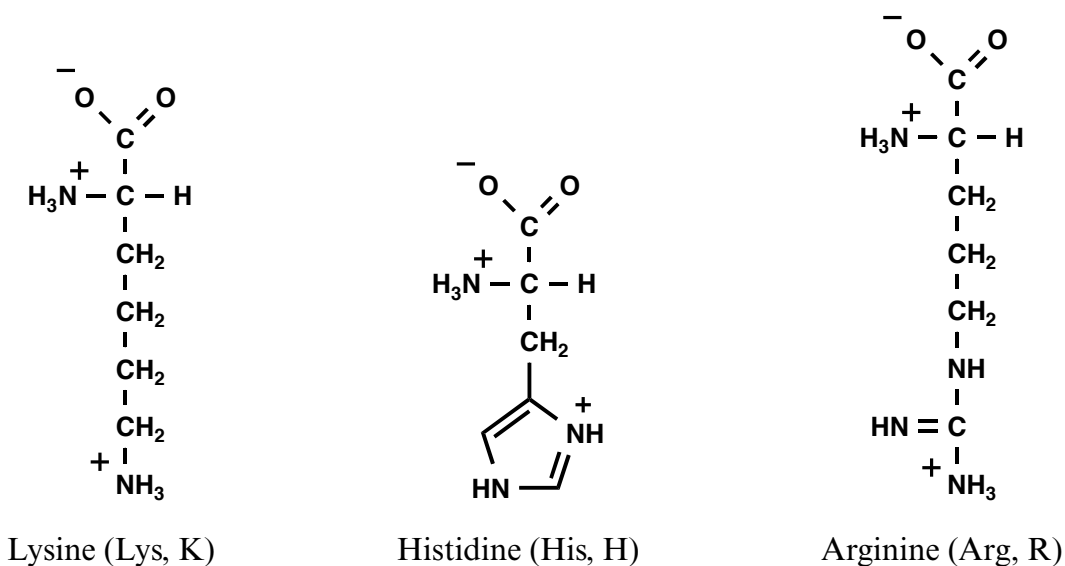
Note 4: Methionine (Met) and Cysteine (Cys) can be grouped as the Sulfur containing amino acids,

The Amino Acids with Negatively Charged Side Chains (Acidic Amino Acids)



The side chain carboxyl group of aspartate (the β -COOH group) has a pK_a of 3.9 and the side chain carboxyl group of glutamate (the γ -COOH group) has a pK_a of 4.3. These values are characteristic of pK_a 's of organic acids. At physiological pH the carboxyl side chain of these two amino acids is always ionized so they are named as the carboxylate ion ("ate" ending) rather than the acid ("ic acid" ending).

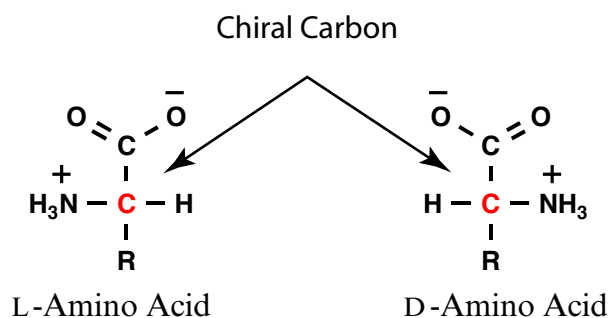
The Amino Acids with Positively Charged Side Chains (Basic Amino Acids)



The imidazole ring of histidine has a pK_a of 6.0, the side chain amino group of lysine (the ϵ -amino group) has a pK_a of 10.5, and the guanidinium group of arginine has a pK_a of 12.5. The imidazole side chain of histidine is the only side chain that has buffering properties at physiological pH.

Chiral Carbon

When the R group, the side chain, of an amino acid is a functional group other than a hydrogen (H) atom, the α carbon of the amino acid is an asymmetric carbon; a chiral carbon, a tetrahedral stereocenter. Having a chiral carbon in the molecule means that the molecule can exist as a pair of stereoisomers. These stereoisomers have configurations that are non-superimposable mirror images; they exist as a pair of enantiomers. For a general α amino acid the pair of enantiomers has the following configurations:



In nature, when two enantiomers exist, one is synthesized and used by cells in preference to the other. In the case of the amino acids, the L configuration is the isomer that primarily exists in nature and is used by living cells. Some bacteria and fungi synthesize and use D configuration amino acids. The molecules synthesized with these D-form amino acids are often toxic to other living organisms.

Calculation of Amino Acid Isoelectric Point

For the nonpolar and polar amino acids with two pK_a 's, the isoelectric point is calculated by taking the numerical average of the carboxyl group pK_a and the α -amino group pK_a .

The titration curves of seven amino acids (Cys, Try, Glu, Asp, His, Lys, & Arg) demonstrate three inflection points, three pK_a 's. One for the α -amino group, one for the carboxyl group of carbon 1, and one for the ionizable side chain. The isoelectric point for these amino acids is calculated by taking the numerical average of the pK_a 's of the two groups that when ionized have like charge. For example to calculate the isoelectric point of Glu, the pK_a 's of the two carboxyl groups are averaged. To calculate the isoelectric point of Arg, the pK_a 's of the α -amino group and the guanidinium group are averaged. Remember, at the isoelectric point the amino acid has no net charge. For the amino acids with three ionizable groups, the total charge on the groups with like charge must equal one (1) so that it can be balanced by the one (1) opposite charge present on the molecule.

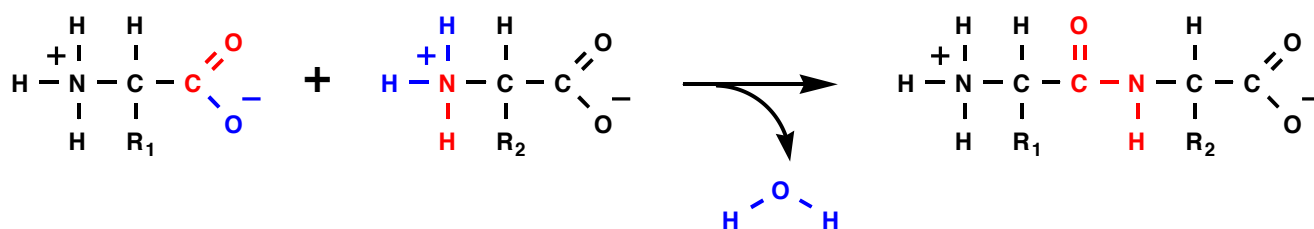
Proteins: Their Primary Structure and Biological Functions

The word protein comes from the Greek *PROTEIOS* which means first or primary. Proteins are of primary importance to cells and organisms. Proteins can be compared to words of the English language. The words that are read, written, and spoken are composed of the 26 letters of the alphabet. The unique sequence of letters is the primary structure of the word. For proteins the alphabet is the 20 amino acids. The unique sequence of amino acids is the primary structure of the molecule. The average length of a human protein is

375 amino acids. Using this sequence length and doing some simple math, there are 20^{375} unique amino acid sequences of this length. From this calculation it should be obvious that proteins are the most diverse group of biomolecules with respect to structure and biological function. Depending upon the cell type, it can contain from a few hundred to greater than 2,000 different proteins. Each of these proteins contains a unique sequence of amino acids, has a unique three dimensional structure, and a unique biological function. Nowhere in biochemistry is structure and function more closely linked. If the structure of a protein is changed its function will be changed and if a new function is required a new structure must be developed, must evolve.

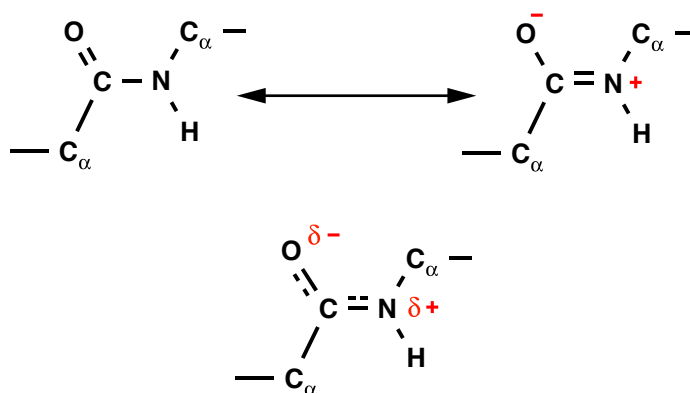
Peptide Bonds

Amino acids are linked in proteins in a head to tail manner by a condensation reaction between the carboxyl group of one amino acid and the α amino group of the second. The “backbone” of a protein consists of the repeating sequence $-C-N-C_{\alpha}-$. The amino acid side chains of the amino acids project perpendicularly from the back bone of the molecule.



The bond between amino acids in a protein is an amide bond. Since this amide bond holds peptides and proteins together in biochemistry it is called a PEPTIDE BOND. This amide linkage, the peptide bond, has no acid or base properties. It will neither donate nor accept a proton. However, the peptide bond is very polar.

Physical chemical studies have shown that the peptide bond exists in two resonance forms:



The real nature of the peptide bond lies somewhere between these two extremes, it has partial double bond character. The partial double bond character of the peptide bond restricts rotation about the carbon-nitrogen bond limiting the possible number of conformations that the peptide or protein can assume. It also places the six atoms of the peptide bond in the same plane with the carbonyl oxygen *trans* to the amido hydrogen

Peptide / Protein Terminology

The unique sequence of amino acids in a peptide or protein is termed the PRIMARY (1°) STRUCTURE of the Protein. Genes contain the information necessary for the directed synthesis of a protein, for the assembly of the primary structure of the protein. The primary structure contains the information necessary for the protein to fold into its final three dimensional conformation and once correctly folded to assume its cellular function. If the 1° structure of a protein is changed very often the final folded shape changes resulting in a less functional or nonfunctional polypeptide.

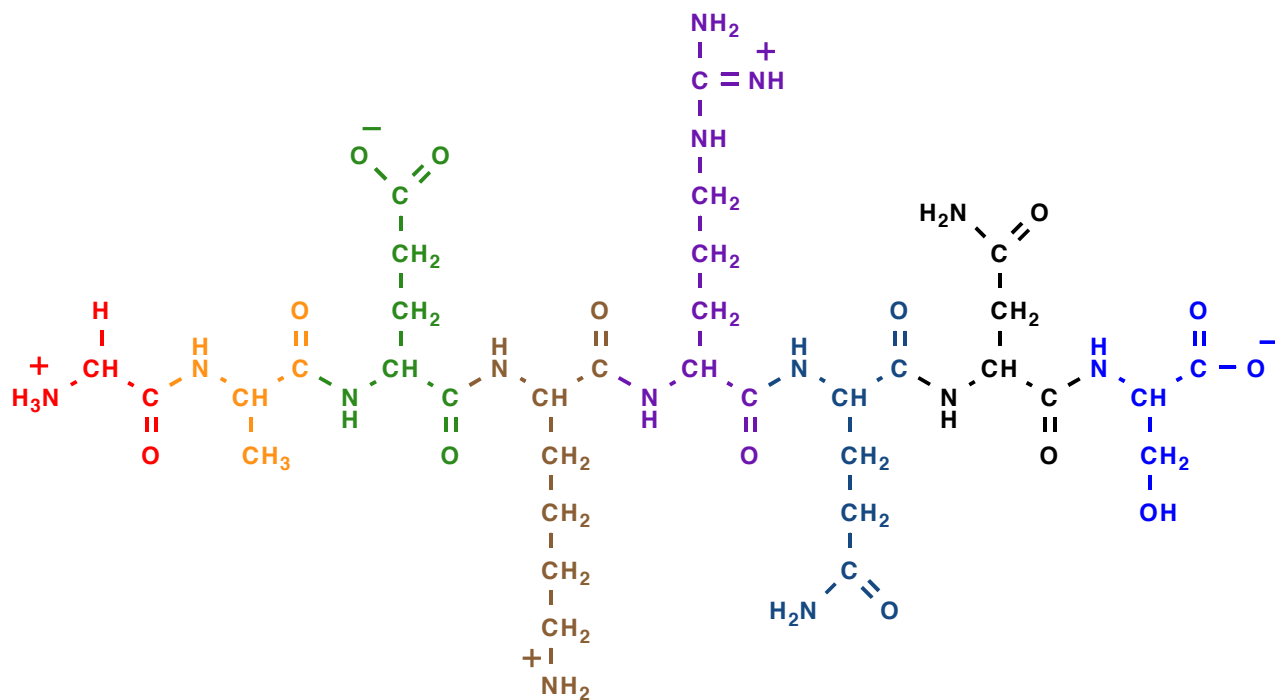
A molecule containing two amino acids joined by a single peptide bond is a DIPEPTIDE; one with three amino acids held together by two peptide bonds is a TRIPEPTIDE; a molecule composed of four amino acids is a TETRAPEPTIDE; etc. In general, PEPTIDES contain 12 or fewer amino acid residues. An OLIGOPEPTIDE contains between 12 and 20 amino acids and a POLYPEPTIDE contains greater than 20 amino acids. Twenty appears to be a magic number with respect to peptide/protein structure. Oligopeptides with 20 or fewer amino acids do not fold into a “single” low energy conformation, rather they exist in numerous random shapes. Molecules with greater than 20 amino acids most often fold into a single stable low energy conformation. The terms polypeptide and protein are often used interchangeably. However, the term PROTEIN is reserved for molecules that when correctly folded perform some cellular function.

MONOMERIC PROTEINS contain a single correctly folded polypeptide chain. Some proteins are supramolecular complexes composed of two or more polypeptide chains. These proteins are called MULTIMERIC PROTEINS. HOMOMULTIMERIC PROTEINS are composed of two or more polypeptides with the same sequence of amino acids, with the same primary structure. HETEROMULTIMERIC PROTEINS are composed of two or more different polypeptides. Each of the individual polypeptides in a heteromultimeric protein can have a different primary structure or there can be some set ratio of different polypeptides necessary to form the heteromultimeric protein. The individual polypeptide chains that comprise a multimeric protein are termed SUBUNITS.

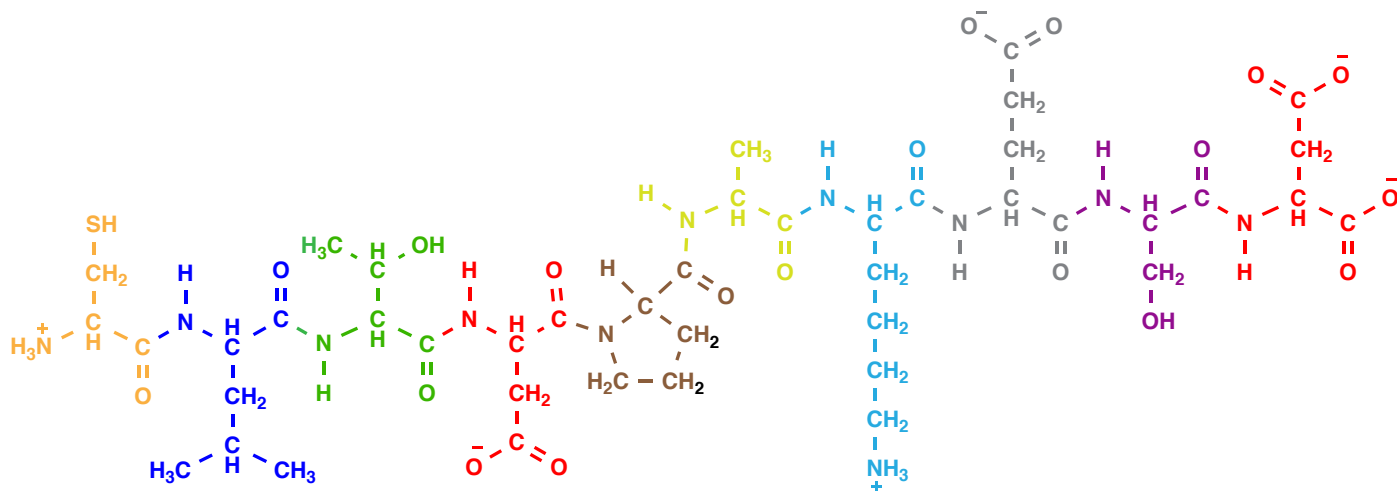
Sequence Convention

A protein is a long linear sequence of amino acids joined by peptide bonds. One end of the molecule has a free amino group, an amino group not involved in the formation of a peptide bond. This is the AMINO TERMINUS or N-TERMINUS of the peptide/oligopeptide/polypeptide/protein. The opposite end of the molecule has a free carboxyl group, a carboxyl group not involved in a peptide bond formation. This end is the CARBOXY TERMINUS, CARBOXYL TERMINUS or C-TERMINUS of the molecule.

When the sequence of a protein is written either the three letter abbreviations or the one letter abbreviations for the amino acids is used. The three letter abbreviations are usually separated by hyphens. The amino terminus is always the Left most amino acid of the sequence and the carboxyl terminus is always the Right most amino acid.



gly-ala-glu-lys-arg-gln-asn-ser



cys-leu-thr-asp-pro-ala-lys-glu-ser-asp

pH Properties and the Isoelectric Point (pI) of Peptides & Proteins

Like the amino acids, peptides, oligopeptides, polypeptides, and proteins have acid base properties. At physiological pH, pH 7.4, the amino terminus along with the side chains of the basic amino acids carry a positive charge. The carboxyl terminus and the side chains of the acidic amino acids carry a negative charge. Just like amino acids, peptides, oligopeptides, polypeptides, and proteins when titrated will reach a solution pH where the amount of positive charge on the molecule is exactly balanced by the amount of negative charge. Peptides and proteins have an isoelectric point (pI). Since these molecules are composed

of a unique sequence of amino acids, the pI is unique for each molecule. The isoelectric point of peptides and proteins is difficult to calculate from the pK_a 's of the amino acids. Usually it is determined experimentally. At the isoelectric point, the peptide/protein will not migrate in an electric field and this property can be used to isolate and/or identify proteins. When the solution pH is less than the pI the protein will have a net positive charge and when placed in an electric field will migrate toward the negative pole (cathode). When the solution pH is greater than the pI the molecule will have a net negative charge and migrate toward the positive pole (anode).

Biological Functions of Proteins

1. ENZYMES are the biological catalysts of the chemical reactions that occur within the cell. Enzymes will be examined in detail shortly.
2. REGULATORY PROTEINS regulate the activities of the cell and the ability of other proteins to carry out their cellular function. The *peptide and protein hormones* are regulatory proteins. They play a role in regulating overall metabolism, growth, development, and maintenance of the organism. *Allosteric enzymes* can be considered regulatory proteins since they control key cellular reactions. *Gene inducers* and *gene repressors* are also considered regulatory proteins. Gene inducers stimulate gene expression, they turn genes “on”. Gene repressors inhibit gene expression, they turn genes “off”. Initiation factors for Replication and Transcription are regulatory proteins.
3. TRANSPORT PROTEINS carry specific substances from one place to another. *Membrane transporters* carry polar molecules across cell membranes. *Hemoglobin* transports oxygen from the lungs to the tissues; *serum albumin* carries a wide variety of drugs and metabolites throughout the body.
4. STORAGE PROTEINS provide a reservoir of an essential nutrient. *Myoglobin* stores oxygen in skeletal muscle tissue; *ferritin* stores iron in the liver and bone marrow; the protein *ovalbumin* serves as a storage depot of amino acids for developing birds; and *casein*, the major protein in mammalian milk stores amino acids for growing infants.
5. STRUCTURAL PROTEINS provide strength, support, and form to cells, tissues, and organisms. *Collagen* is the major protein of bone, tendons, and cartilage. α -*Keratin* is the protein of skin, hair, horns, hooves, and fingernails. β -*Keratin* is a structural protein of the skin, scales, beaks, claws, and feathers. There are cytoskeletal protein fibers adjacent to the cell membrane that give the cell shape and support. *Tubulin*, *actin*, and *spectrin* are some of the proteins that make up the cytoskeleton.
6. CONTRACTILE and MOTILE PROTEINS provide the cell / organism with motion. Examples include *actin* and *myosin* of muscle cells and *tubulin* of cilia and flagella. The proteins *dynein* and *kinesin* drive the movement of vesicles and organelles along cytoskeletal tracks within the cell. A single protein can serve several functions within a cell. For example actin and tubulin can be classified as structural proteins or motile proteins depending upon their location within the cell.
7. SCAFFOLD PROTEINS act as bridges by binding to and localizing specific proteins to specific sites

within the cell. They act as a site upon which supramolecular complexes are formed.

8. PROTECTIVE PROTEINS play an active role in cell defense or in defense of the organism. *Antibodies* of the immune system protect the organism from foreign invaders. The *hemostasis cascade* of proteins protects the organism from excessive blood loss. *Toxins*, such as snake venoms, could also be placed into this class of proteins. The *Keratins* of the skin could be considered a protective protein.
9. STRESS PROTEINS endow the organism with the capacity to survive abiotic stresses. *Cytochrome P₄₅₀* is an important factor in the biological detoxification of xenobiotics. *Metallothionein* is a cysteine rich protein that functions by binding and sequestering toxic heavy metals; e.g., Cd, Hg, Ag. *Heat Shock Proteins (Hsp)* aid in the (re)folding of proteins damaged by heat (a fever) or signal for their destruction if too severely damaged.
10. EXOTIC PROTEINS display functions that do not fit the other classifications. One example is the *glue protein* secreted by mussels. This protein allows the mussels to anchor to hard surfaces.