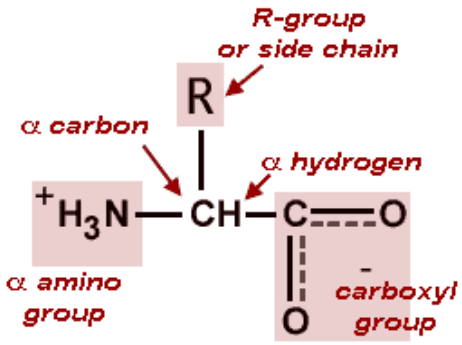
Amino acids

Proteins are the most abundant and functionally diverse molecules in living systems. Virtually every life process depends on this class of molecules. For example, enzymes and polypeptide hormones direct and regulate metabolism in the body, whereas contractile proteins in muscle permit movement. In bone, the protein collagen forms a framework for the deposition of calcium phosphate crystals, acting like the steel cables in reinforced concrete. In the bloodstream, proteins, such as hemoglobin and plasma albumin, shuttle molecules essential to life, whereas immunoglobulins fight infectious bacteria and viruses. In short, proteins display an incredible diversity of functions, yet all share the common structural feature of being linear polymers of amino acids. This lecture describes the properties of amino acids and explores how these simple building blocks are joined to form proteins that have unique three-dimensional structures, making them capable of performing specific biologic functions.

Amino acids are organic compounds that combine to form proteins. They play central roles both as building blocks of proteins and as intermediates in metabolism. Although more than 300 different amino acids have been described in nature, only 20 are commonly found as constituents of mammalian proteins.

Basic Structure of an Amino Acid:

Each amino acid (except for proline, which has a secondary amino group) has a carboxyl group, a primary amino group, and a distinctive side chain (R-group) bonded to the α-carbon atom. At physiologic pH (approximately pH 7.4), the carboxyl group is dissociated, forming the negatively charged carboxylate ion (COO^-), and the amino group is protonated (NH_3^+). All amino acids found in proteins have the following basic structure, which differing only in the structure of the R-group or the side chain. In fact, the simplest, and smallest, amino acid found in proteins is glycine for which the R-group is hydrogen (H).



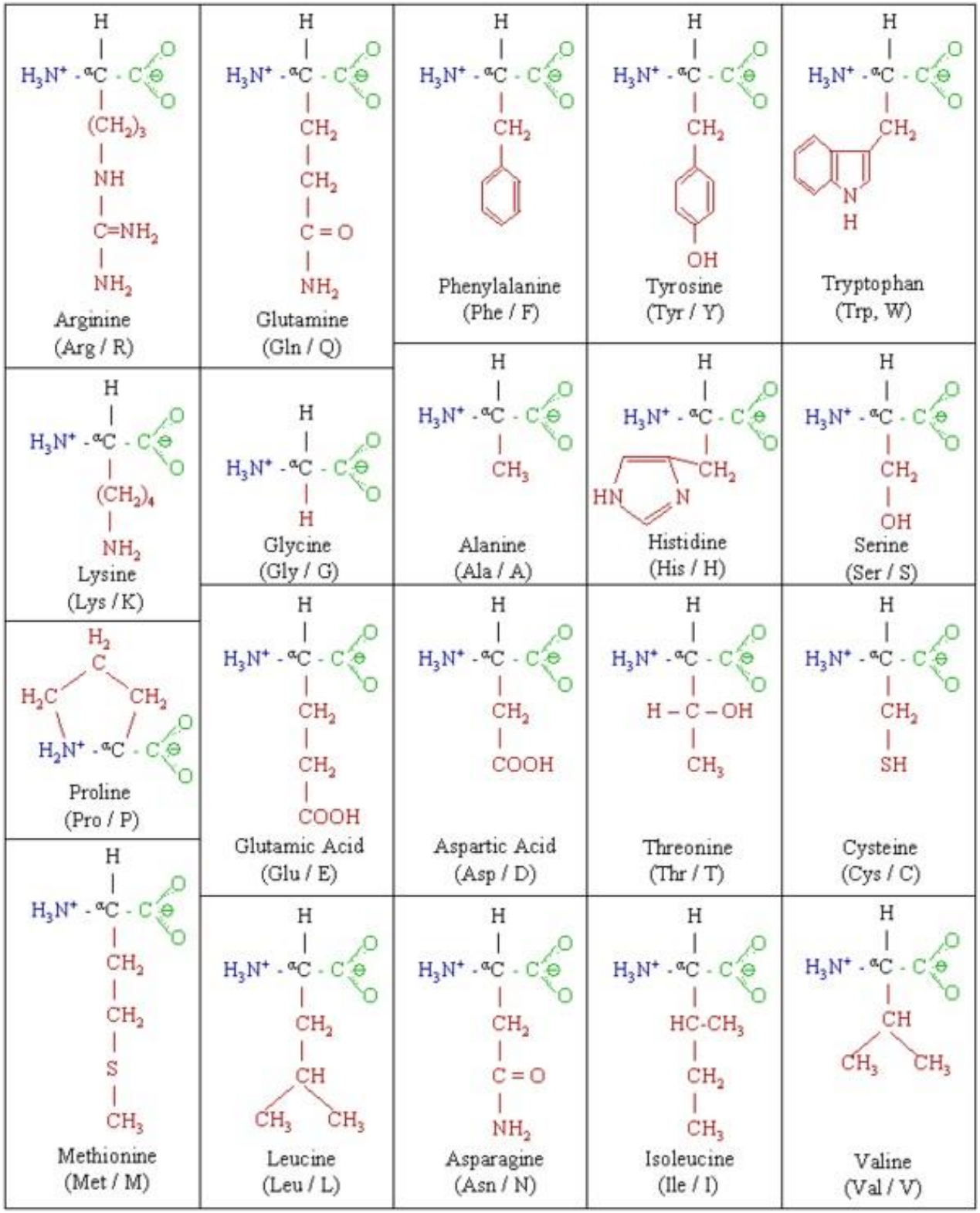
Classification of amino acids:

The most common and perhaps the most useful way of classifying the 20 "standard" amino acids is according to the chemical nature of their side chains (R groups).

1- Aliphatic amino acids: Glycine (Gly), Alanine (Ala), Valine (Val), Leucine (Leu), Isoleucine (Ile).
2- Hydroxyl or Sulfur containing: Serine (Ser), Cysteine (Cys), Threonine (Thr), Methionine (Met).
3- Aromatic amino acids: Phenylalanine (Phe), Tyrosine (Tyr), Tryptophan (Trp).
4- Acidic amino acids and their amide: Aspartic acid (Asp), Glutamic acid (Glu), Asparagine (Asn), Glutamine (Gln).
5- Basic amino acids: Histidine (His), Lysine (Lys), Arginine (Arg).
6- Cyclic amino acid: Proline (Pro).

Also, amino acids can be classified in the category of either essential or non-essential.

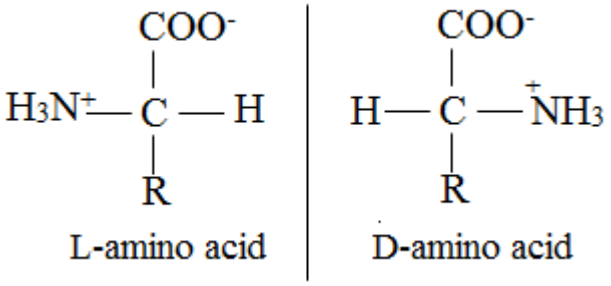
1. Essential amino acids are amino acids that cannot be synthesized by the human body at the level needed for normal growth, so they must be obtained from food. The 8 amino acids that humans cannot synthesize are phenylalanine, valine, threonine, tryptophan, methionine, leucine, isoleucine and lysine.
2. Non-Essential amino acids: The rest of 20 amino acids that the human body can synthesize them from other compounds at the sufficient levels.

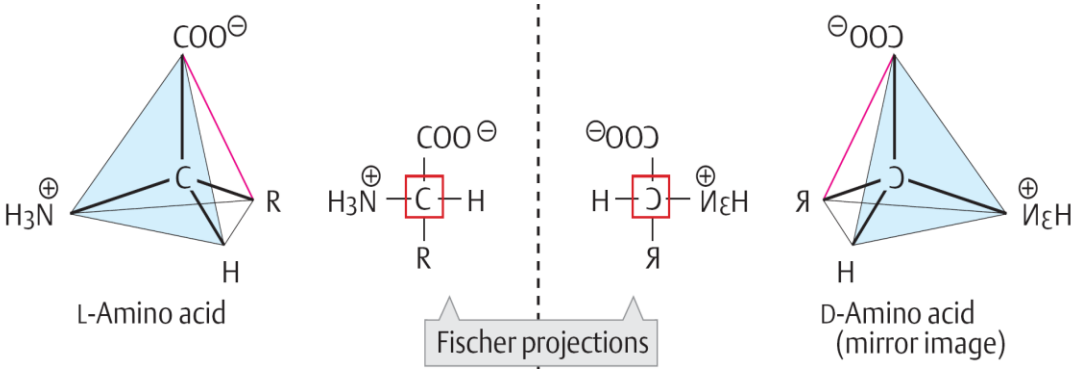


List of 20 amino acids with their structures, and three symbol and one-letter codes

Optical properties of amino acids:

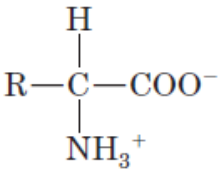
The α-carbon of each amino acid is attached to four different chemical groups and is, therefore, a chiral or optically active carbon atom. Glycine is the exception because its α-carbon has two hydrogen substituents and, therefore, is optically inactive. Amino acids that have an asymmetric center at the α-carbon can exist in two forms, designated D and L that are mirror images of each other. The two forms in each pair are termed stereoisomers, optical isomers, or enantiomers. All amino acids found in proteins are of the L-configuration. However, D-amino acids are found in some antibiotics and in plant and bacterial cell walls.





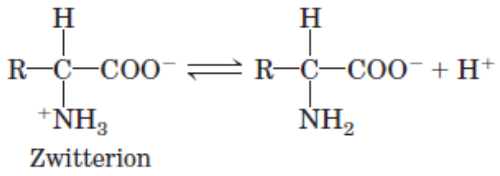
What are zwitterions?

An amino acid has COOH and NH_2 groups in the same molecule. Therefore, in water solution, the COOH donates a proton to the NH_2 so that an amino acid actually has the structure.

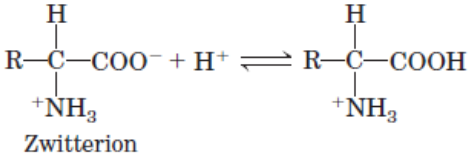


Compounds that have a positive charge on one atom and a negative charge on another are called zwitterions, from the German word zwitter, meaning "hybrid". Amino acids are zwitterions, not only in water solution but also in the solid state. They are therefore ionic compounds that is, internal salts. Un-ionized $RCH(NH_2)COOH$ molecules do not actually exist, in any form.

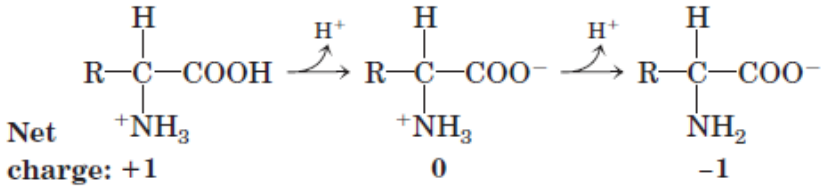
When an amino acid is dissolved in water, it exists in solution as the dipolar ion, or zwitterion. A zwitterion can act as either an acid (proton donor):



Or a base (proton acceptor):



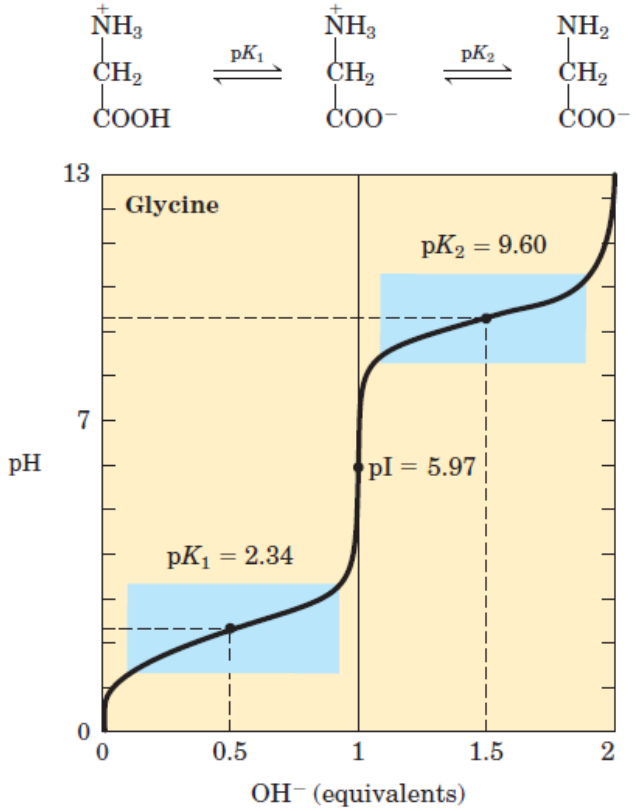
Substances having this dual nature are amphoteric and are often called ampholytes (from amphoteric electrolytes). A simple mono-amino mono-carboxylic amino acid, such as alanine, is a diprotic acid when fully protonated it has two groups, the COOH group and the NH_3 group, that can yield protons:



Acidic and Basic Properties of Amino Acids:

Amino acids and proteins have conspicuous acid–base properties. The amino acids have two or, for those with ionizable side groups, three acid–base groups. The following figure shows

the titration curve of the diprotic form of glycine. The plot has two distinct stages, corresponding to deprotonation of two different groups on glycine.



At very low pH, the predominant ionic species of glycine is the fully protonated form, NH_3^+-CH_2-COOH. At the midpoint in the first stage of the titration, in which the COOH group of glycine loses its proton, equimolar concentrations of the proton-donor (NH_3^+-CH_2-COOH) and proton-acceptor (NH_3^+-CH_2-COO^-) species are present. At the midpoint of any titration, a point of inflection is reached where the pH is equal to the p*Ka* of the protonated group being titrated. For glycine, the pH at the midpoint is 2.34, thus its COOH group has a p*Ka* (labeled pK_1 in the Fig.) of 2.34. As the titration proceeds, another important point is reached at pH 5.97. Here there is another point of inflection, at which removal of the first proton is essentially complete and removal of the second has just begun. At this pH glycine is present largely as the dipolar ion NH_3^+-CH_2-COO^-. We shall return to the significance of this inflection point in the titration curve (labeled pI in the Fig.) shortly. The second stage of the titration corresponds to the removal of a proton from the NH_3 group of glycine. The pH at the midpoint of this stage is 9.60, equal to the

p*Ka* (labeled pK_2 in the Fig.) for the NH_3 group. The titration is essentially complete at a pH of about 12, at which point the predominant form of glycine is $H_2NCH_2COO^-$

From the titration curve of glycine we can derive several important pieces of information. First, it gives a quantitative measure of the p*Ka* of each of the two ionizing groups: 2.34 for the COOH group and 9.60 for the NH_3 group. The second piece of information provided by the titration curve of glycine is that this amino acid has two regions of buffering power. One of these is the relatively flat portion of the curve, extending for approximately 1 pH unit on either side of the first p*Ka* of 2.34, indicating that glycine is a good buffer near this pH. The other buffering zone is centered around pH 9.60. Within the buffering ranges of glycine, the following Henderson-Hasselbalch equation can be used to calculate the proportions of proton-donor and proton-acceptor species of glycine required to make a buffer at a given pH.

$$\mathrm{pH} = \mathrm{p}K + \log\left(\frac{[\mathrm{A}^-]}{[\mathrm{HA}]}\right)$$

Another important piece of information derived from the titration curve of an amino acid is the relationship between its net electric charge and the pH of the solution. At pH 5.97, the point of inflection between the two stages in its titration curve, glycine is present predominantly as its dipolar form, fully ionized but with no net electric charge. The characteristic pH at which the net electric charge is zero is called the isoelectric point or isoelectric pH, designated pI. For glycine, which has no ionizable group in its side chain, the isoelectric point is simply the arithmetic mean of the two p*Ka* values:

$$\mathrm{pI} = \frac{1}{2}(\mathrm{p}K_1 + \mathrm{p}K_2) = \frac{1}{2}(2.34 + 9.60) = 5.97$$

Electrophoresis:

Electrophoresis is a separation technique based on the movement of charged ions under the influence of an electrical field. This technique is primarily used for the separation of amino acids and peptides on the basis of their charge. All amino acids contain ionizable groups that cause the amino acids, in solution, to act as charged polyelectrolytes that can migrate in an

electric field. The amino acids with a net positive charge will migrate toward the negative electrode. Those with a negative net charge will move toward the positive electrode.

As is evident in the Figure above, glycine has a net negative charge at any pH above its pI and will thus move toward the positive electrode (the anode) when placed in an electric field. At any pH below its pI, glycine has a net positive charge and will move toward the negative electrode (the cathode). The farther the pH of a glycine solution is from its isoelectric point, the greater the net electric charge of the population of glycine molecules. At pH 1.0, for example, glycine exists almost entirely as the form H_3NCH_2COOH, with a net positive charge of 1.0. At pH 2.34, where there is an equal mixture of H_3NCH_2COOH and H_3NCH_2COO, the average or net positive charge is 0.5. The sign and the magnitude of the net charge of any amino acid at any pH can be predicted in the same way.

Example: Predict the direction of migration in an electrophoresis for the amino acids (Ser, His, Val and Asp) at pH 2.0, 4.0, 6.0 and 12?

Peptides and Proteins

Proteins, from the Greek *proteios*, meaning first, are a class of organic compounds which are present in and vital to every living cell. They are linear polymers built of monomer units called amino acids, which are linked end to end by peptide bonds.

Proteins are the most abundant class of biological macromolecules as they represent over 50% of the dry weight of cells. They perform a variety of functions, including the following:

1. **Catalysis:** Catalytic proteins called the enzymes accelerate thousands of biochemical reactions in such processes as digestion, energy capture, and biosynthesis. These molecules have remarkable properties.
2. **Structure:** Structural proteins often have very specialized properties. For example, collagen (the major components of connective tissues) and fibroin (silkworm protein) have significant mechanical strength.
3. **Movement:** Proteins are involved in all cell movements. Actin, tubulin, and other proteins comprise the cytoskeleton. Cytoskeletal proteins are active in cell division, endocytosis, exocytosis, and the ameboid movement of white blood cells.
4. **Defense:** A wide variety of proteins are protective. In vertebrates, keratin, a protein found in skin cells, aids in protecting the organism against mechanical and chemical injury. The blood-clotting proteins fibrinogen and thrombin prevent blood loss when blood vessels are damaged. The immunoglobulins (or antibodies) are produced by lymphocytes when foreign organisms such as bacteria invade an organism.
5. **Regulation:** Binding a hormone molecule or a growth factor to cognate receptors on its target cell changes cellular function. For example, insulin and glucagon are peptide hormones that regulate blood glucose levels.
6. **Transport:** Many proteins function as carriers of molecules or ions across membranes or between cells. Transport proteins include hemoglobin, which carries O_2 to the tissues from the lungs, and the lipoproteins LDL and HDL, which transport water insoluble lipids in the blood from the liver. Transferrin and ceruloplasmin are serum proteins that transport iron and copper, respectively.

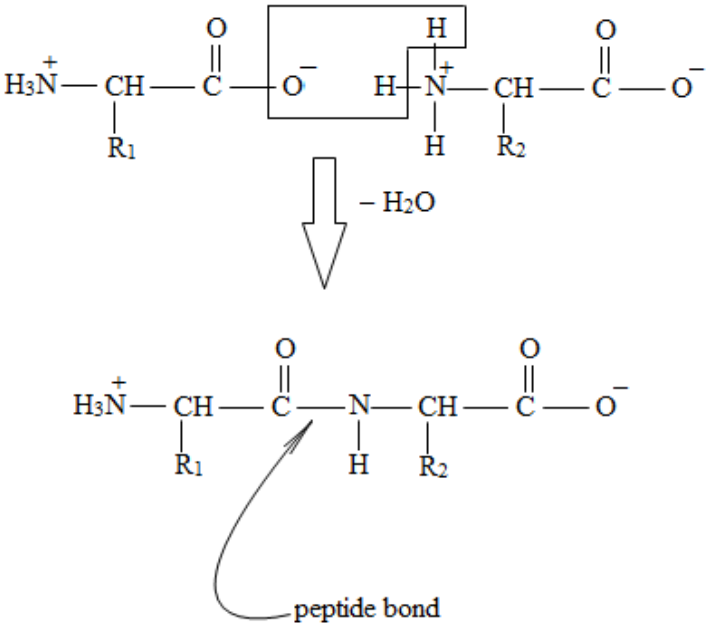
7. **Storage:** Certain proteins serve as a reservoir of essential nutrients. For example, ovalbumin in bird eggs and casein in mammalian milk are rich sources of organic nitrogen during development.

Protein structure:

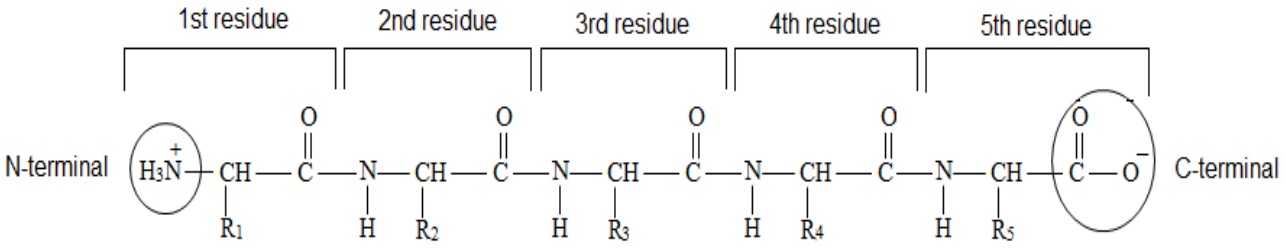
Amino acid molecules can be covalently joined through substituted amide linkages, termed peptide bonds, to yield a protein. Amide linkage (peptide bond) is formed by linking the carboxyl group of one amino acid to the amino group of another amino acid with removal of the elements of water (dehydration). Two amino acid molecules can be covalently joined through a substituted amide linkage, termed a peptide bond, to yield a dipeptide. Three amino acids can be joined by two peptide bonds to form a tripeptide; similarly, amino acids can be linked to form tetrapeptides, pentapeptides, and so forth.

In fact, when a few amino acids are joined in this fashion, the structure is called an oligopeptide. When many amino acids are joined, the product is called a polypeptide. Proteins may have thousands of amino acid residues. Although the terms "protein" and "polypeptide" are sometimes used interchangeably, molecules referred to as polypeptides generally have molecular weights below 10,000, and those called proteins have higher molecular weights.

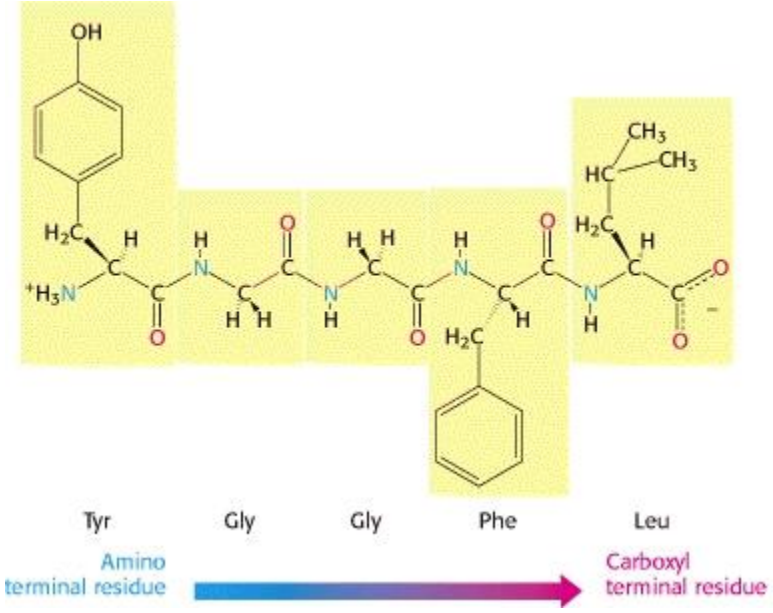
The following figure shows a dipeptide formation by peptide bond. The linking of two amino acids is accompanied by the loss of a molecule of water.



And the following figure shows a peptide chain that consists of five amino acid residues.



A series of amino acids joined by peptide bonds form a polypeptide chain, and each amino acid unit in a polypeptide is called a residue (the part left over after losing a hydrogen atom from its amino group and the hydroxyl moiety from its carboxyl group). A polypeptide chain has polarity because its ends are different, with an α-amino group at one end and an α-carboxyl group at the other end of the chain. By convention, the amino end is taken to be the beginning of a polypeptide chain, and so the sequence of amino acids in a polypeptide chain is written starting with the amino-terminal residue. It is the universal custom to write peptide and protein chains with the N-terminal residue on the left. Thus, in the pentapeptide Tyr-Gly-Gly-Phe-Leu, phenylalanine is the amino-terminal (N-terminal) residue and leucine is the carboxyl-terminal (C-terminal) residue.



To name peptides, the names of acyl groups ending in yl are used. Thus if the amino acids glycine and alanine condense together, the dipeptide formed is named glycylalanine. Higher peptides are named similarly, e.g. alanylleucyltryptophan. Thus the name of the peptide begins with the name of the acyl group representing the *N*-terminal residue, and this is followed in order by the names of the acyl groups representing the internal residues. Only the *C*-terminal residue is represented by the name of the amino acid, and this ends the name of the peptide. Formulas should normally be written in the same order, with the *N*-terminal residue on the left, and the *C*-terminal on the right.

Classification of Proteins:

Proteins may be classified according to their chemical composition, their shape, or their function. Protein composition and function are treated in detail in a biochemistry course. For now, we briefly survey the types of proteins and their general classifications. Proteins are grouped into simple and conjugated proteins according to their chemical composition.

1. **Simple proteins:** are those that hydrolyze to give only amino acids. All the protein structures we have considered so far are simple proteins. Examples are insulin, ribonuclease, oxytocin, and bradykinin.
2. **Conjugated proteins:** are bonded to a non-protein prosthetic group such as a sugar, a nucleic acid, a lipid, or some other group.
    - **Glycoproteins:** are proteins that contain oligosaccharide chains (glycans) covalently attached to polypeptide side-chains, such as mucins.
    - **Phosphoproteins:** are proteins which are chemically bonded to a substance containing phosphoric acid, such as casein.
    - **Metalloprotein:** A protein that contains a metal ion as cofactor, for example: ferritin and hemoglobin.
    - **Lipoproteins:** A lipoprotein is a biochemical assembly that contains both proteins and lipids water-bound to the proteins, such as chylomicrons.
    - **Nucleoproteins:** A proteins that is structurally associated with nucleic acids such as ribosomes, viruses.

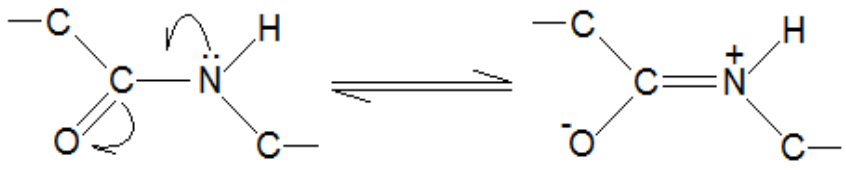
Also, proteins can be classified according to their shape into fibrous or globular proteins.

1. **Fibrous proteins** are stringy, tough, and usually insoluble in water. They function primarily as structural parts of the organism. Examples of fibrous proteins are α-keratin in hooves and fingernails, and collagen in tendons.
2. **Globular proteins** are somewhat water-soluble (forming colloids in water), unlike the fibrous proteins. They are folded into roughly spherical shapes. They usually function as enzymes, hormones, or transport proteins.

Types of Chemical Bonds in Proteins:

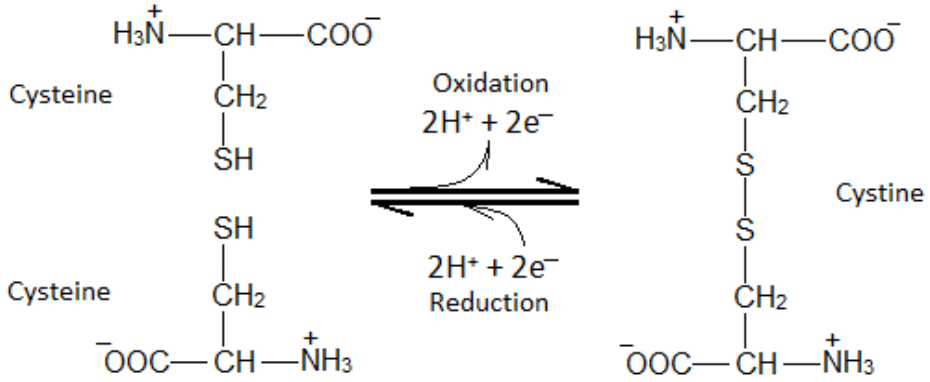
First: Two types of covalent bonds that include.

1. **Peptide bonds:** A peptide bond is a chemical bond formed between two molecules when the carboxyl group of one amino acid reacts with the amino group of another amino acid, releasing a molecule of water (H_2O). Examination of the geometry of the protein backbone reveals that the peptide bond is essentially rigid and planar. Thus for a pair of amino acids linked by a peptide bond, six atoms lie in the same plane: the α-carbon and CO group of the first amino acid and the NH group and α-carbon atom of the second amino acid. The peptide bond has considerable double-bond character due to a resonance structure (as shown in the figure below), which prevents rotation about this bond and thus constrains the conformation of the peptide backbone. The double bond character is also expressed in the length of the bond between CO and the NH groups. The C-N distance in a peptide bond is typically 1.32 Å, which is between the values expected for a C-N single bond (1.49 Å) and a C=N double bond (1.27 Å).



2. **Disulfide bonds:** Second kind of covalent bonds is possible between any cysteine residues present. Cysteine residues can form disulfide bridges (also called disulfide linkages) which can join two chains or link a single chain into a ring. Mild oxidation joins two molecules of a thiol into a disulfide, forming a disulfide linkage between the

two thiol molecules. This reaction is reversible, and a mild reduction cleaves the disulfide. Similarly, two cysteine sulfhydryl groups are oxidized to give a disulfide linked pair of amino acids. This disulfide-linked dimer of cysteine is called cystine.

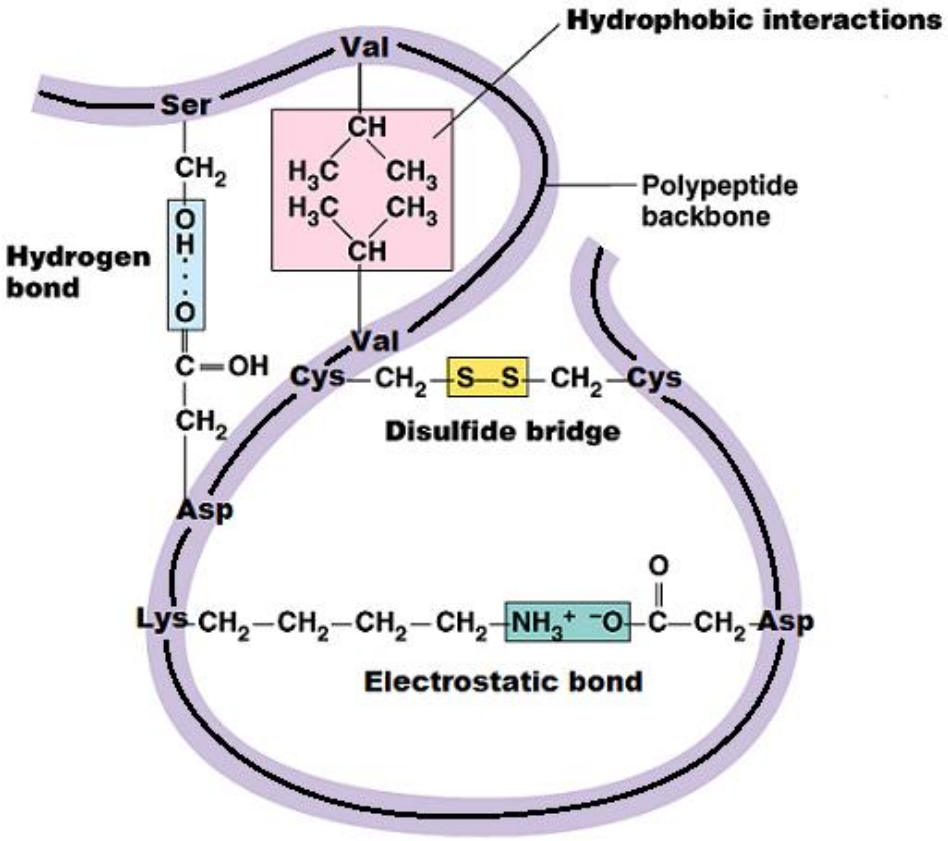


Second: Non-covalent bonds, which usually include.

1. **Hydrogen bonds:** Linus Pauling first suggested that H bonds (between water and the protein and within the protein itself) would play a dominant role in protein folding and stability. Hydrogen bonds form between the oxygen of the C=O of each peptide bond in the strand and the hydrogen of the N-H group of the peptide bond four amino acids below it in the helix. The hydrogen bonding in a *ß-sheet* is between strands (inter-strand) rather than within strands (intra-strand). The carbonyl oxygens in one strand hydrogen bond with the amino hydrogens of the adjacent strand. The hydrogen bonds make the secondary structure of protein especially stable.
2. **Electrostatic Forces:** Electrostatic forces are mainly of charge-charge and charge-dipole interactions. Typical charge-charge interactions that favor protein folding are those between oppositely charged R-groups such as Lysine or Arginine and Glutamic acid or Aspartic acid. A substantial component of the energy involved in protein folding is charge-dipole interactions. This refers to the interaction of ionized R-groups of amino acids with the dipole of the water molecule.
3. **Hydrophobic interactions:** The hydrophobic interactions of non-polar side chains are believed to contribute significantly to the stabilizing of the structures in proteins. This interaction is really just an application of the solubility rule that "likes dissolve likes".

The non-polar groups mutually repel water and other polar groups and results in a net attraction of the non-polar groups for each other. Hydrocarbon alkyl groups on Ala, Val, Leu, and Ile interact in this way. In addition, aromatic ring on Phe can "stack" together. In many cases this results in the non-polar side chains of amino acids being on the inside of a globular protein, while the outside of the proteins contains mainly polar groups.

The following figure shows the types of chemical bonds that play important roles in stabilizing protein structure.



Levels of structure in proteins:

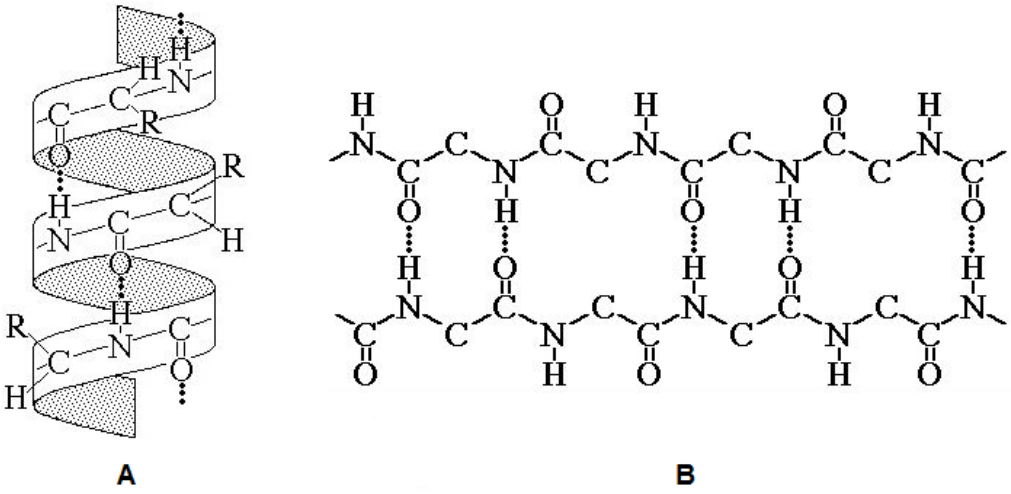
Structural features of proteins are usually described at four levels of complexity:

Primary structure: The primary protein structure refers to the unique sequence of amino acids and the location of disulfide bonds. The amino acids when linked by peptide bonds are referred to as residues. Short chains of amino acid residues are often called (oligo-) peptides. Simply, the

primary structure of a protein is what is encoded in the DNA. Thus, all the properties of the protein are determined, directly or indirectly, by the primary structure.

Secondary structure: Protein structures are also classified by their secondary structure. Secondary structure refers to regular, local structure of the protein backbone, stabilized by intramolecular and sometimes intermolecular hydrogen bonding of amide groups.

There are two common types of secondary structure. The most prevalent is the alpha helix. The alpha helix (α has a right-handed spiral conformation, in which every backbone N-H group donates a hydrogen bond to the backbone C=O group of the amino acid four residues before it in the sequence. The other common type of secondary structure is the beta strand. A Beta strand (β strand) is a stretch of polypeptide chain, typically 3 to 10 amino acids long, with its backbone in an almost fully extended conformation. Two or more parallel or antiparallel adjacent polypeptide chains of beta strand stabilized by hydrogen bonds form a beta sheet. For example, the proteins in silk have a beta-sheet structure. Those local structures are stabilized by hydrogen bonds and connected by tight turns and loose, flexible loops.



A B

A- The peptide chain curls into a helix so that each peptide carbonyl group is hydrogen-bonded to N¬H hydrogen on the next turn of the helix. Side chains are symbolized by green atoms in the space-filling structure.

B- The pleated sheet arrangement. Each peptide carbonyl group is hydrogen bonded to N¬H hydrogen on an adjacent peptide chain.

Tertiary structure: The spatial arrangement of secondary structure elements results in the formation of the tertiary structure or fold of a protein. The tertiary structure is held together by

non-covalent interactions (hydrogen bonding, ionic interactions, van der Waals forces, and hydrophobic packing), disulphide bonds and metal ion coordination.

An example of the tertiary structure is a single-domain globular protein. Globular proteins are sphere-like proteins that are more or less soluble in aqueous solutions (the other two protein classes are membrane and fibrous proteins).

Quaternary structure: Some proteins form assemblies (units) with other molecules, this is called the quaternary structure. Two examples are haemoglobin which is an assembly of four globular proteins and the actin microfilament, composed of many thousands actin molecules.

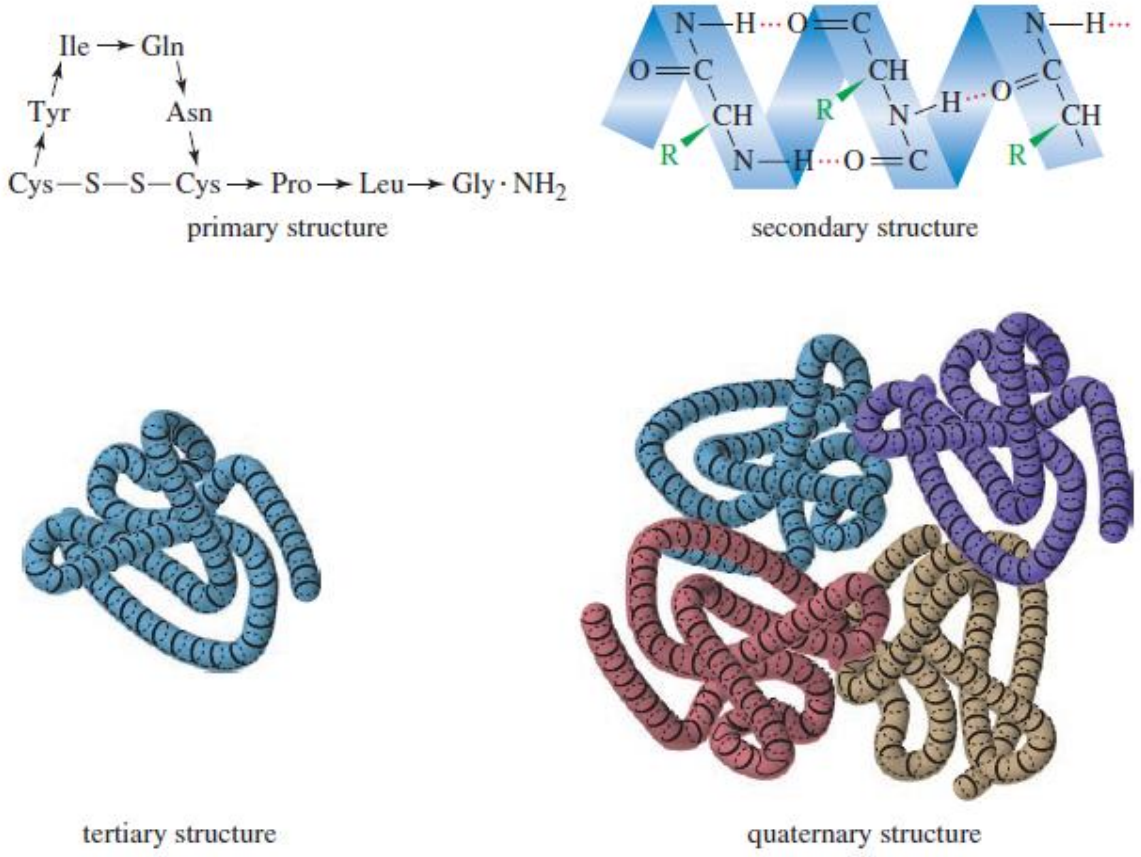


Figure 1: A schematic comparison of the levels of protein structure. Primary structure is the covalently bonded structure, including the amino acid sequence and any disulfide bridges. Secondary structure refers to the areas of helix, pleated sheet, or random coil. Tertiary structure refers to the overall conformation of the molecule. Quaternary structure refers to the association of two or more peptide chains in the active protein.

Amino acids sequences of proteins:

Protein sequencing denotes the process of finding the amino acid sequence, or primary structure of a protein. Sequencing plays a very vital role in Proteomics as the information obtained can be used to deduce function, structure, and location which in turn aids in identifying new or novel proteins as well as understanding of cellular processes. Better understanding of these processes allows for creation of drugs that target specific metabolic pathways.

The enzymes that achieve amino acids sequencing and break down proteins can be divided into exopeptidases and endopeptidases. Endopeptidases break peptide links (the bonds which hold amino acids together) in the middle of polypeptide chains. Exopeptidases break the peptide links of the amino acids at the ends of the chains.

It is often desirable to know the unordered amino acid composition of a protein prior to attempting to find the ordered sequence, as this knowledge can be used to facilitate the discovery of errors in the sequencing process or to distinguish between ambiguous results. Knowledge of the frequency of certain amino acids may also be used to choose which protease to use for digestion of the protein. A generalized method often referred to as amino acid analysis for determining amino acid frequency is as follows:

1. Hydrolyse a known quantity of protein into its constituent amino acids.
2. Separate and quantify the amino acids in some way.

Hydrolysis: Hydrolysis is done by heating a sample of the protein in 6 M hydrochloric acid to 100–110°C for 24 hours or longer.

Separation: The amino acids can be separated by ion-exchange chromatography or hydrophobic interaction chromatography.

Quantitative analysis: Once the amino acids have been separated, their respective quantities are determined by adding a reagent that will form a coloured derivative. If the amounts of amino acids are in excess of 10 nmol, ninhydrin can be used for this; it gives a yellow colour when reacted with proline, and a vivid purple with other amino acids. The concentration of amino acid is proportional to the absorbance of the resulting solution.

***N*-terminal Residue Identification:** *N*-terminal residue identification encompasses a technique which chemically determines which amino acid forms the *N*-terminus of a peptide chain. This information can be used to aid in ordering of individual peptide sequences that were generated using other sequencing techniques that fragment the peptide chain. The *N*-terminal residue identification can be achieved using the following reagents:

1- **Sanger reagent:** 1-fluoro-2,4-dinitrobenzene, or Sanger's reagent (commonly called dinitrofluorobenzene or DNFB) reacts with the amine group in amino acids to produce dinitrophenyl-amino acids. These DNP-amino acids are moderately stable under acid hydrolysis conditions that break peptide bonds. The DNP-amino acids can then be recovered, and the identity of those amino acids can be discovered through chromatography.

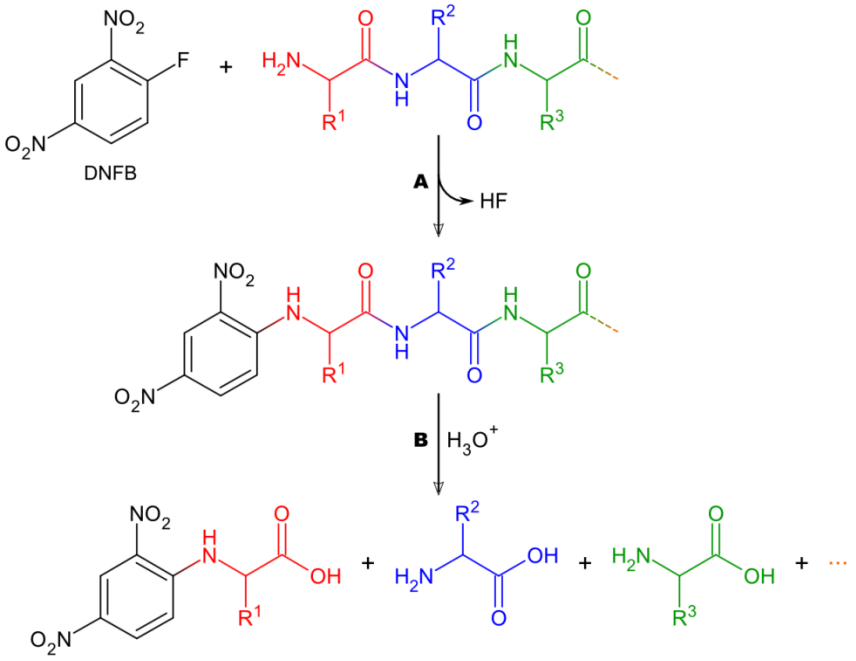


Figure 2: Sanger's method of peptide end-group analysis: A) complex of *N*-terminal end with Sanger's reagent (DNFB), B) total acid hydrolysis of the dinitrophenyl peptide.

2- **Dansyl chloride reagent:** Dansyl chloride or 5-(Dimethylamino) Naphthalene-1-Sulfonyl chloride is a reagent that reacts with the free amino groups of peptides and proteins to produce stable blue- or blue-green–fluorescent sulfonamide adducts.

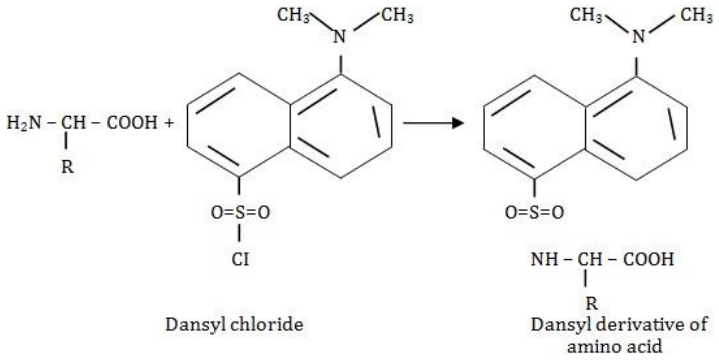


Figure 3: Dansyl chloride complex with amino acid.

3- **Edman reagent:** Edman degradation, developed by Pehr Edman, is a method of sequencing amino acids in a peptide. In this method, Phenyl isothiocyanate is reacted with an uncharged *N*-terminal amino group, under mildly alkaline conditions, to form a cyclical phenyl thiocarbamoyl derivative. Then, under acidic conditions, this derivative of the terminal amino acid is cleaved as a thiazolinone derivative. The thiazolinone amino acid is then selectively extracted into an organic solvent and treated with acid to form the more stable phenyl thiohydantoin (PTH)-amino acid derivative that can be identified using chromatography or electrophoresis.

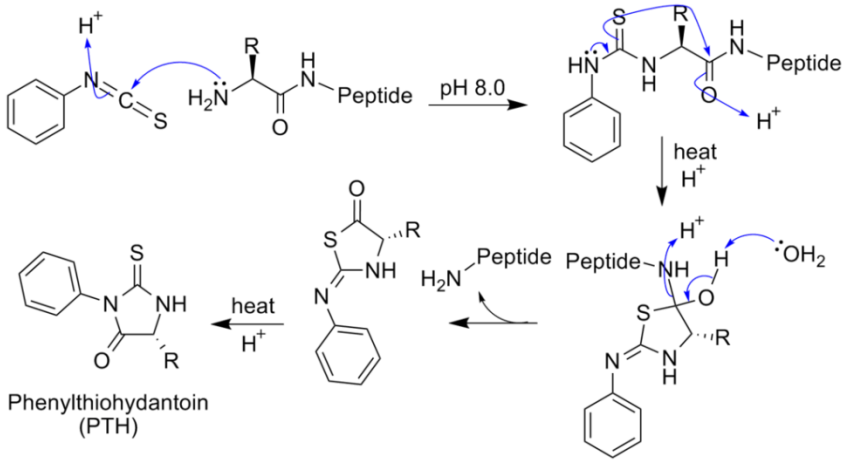


Figure 4: Formation of PTH from the reaction of Phenyl isothiocyanate with peptide.

***C*-terminal residue identification:** The number of methods available for *C*-terminal amino acid analysis is much smaller than the number of available methods of *N*-terminal analysis. The most common method is to add carboxypeptidases to a solution of the protein, take samples at regular intervals, and determine the terminal amino acid by analysing a plot of amino acid concentrations against time. *C*-terminal sequencing would greatly help in verifying the primary structures of proteins predicted from DNA.

Digestion into peptide fragments using endopeptidases: Peptides longer than about 50-70 amino acids long cannot be sequenced reliably by the Edman degradation. Because of this, long protein chains need to be broken up into small fragments that can then be sequenced individually. Digestion is done either by endopeptidases such as trypsin or pepsin or by chemical reagents such as cyanogen bromide. Different enzymes give different cleavage patterns, and the overlap between fragments can be used to construct an overall sequence. The most common endopeptidases are:

1- **Trypsin**: cuts after Arg or Lys, unless followed by Proline.
2- **Chymotrypsin**: cuts after Phe, Trp, or Tyr, unless followed by Pro.
3- **Pepsin:** cuts before Leu, Phe, Trp or Tyr, unless preceded by Pro.
4- **Cyanogen bromide (CNBr):** cuts after methionine residues.

Example: Unknown peptide was treated with trypsin first and again with CNBr. Giving the following sequences:

A) By trypsin (1) Asn-Thr-Trp-Met-Ile-Lys (2) Val-Leu-Gly-Met-Ser-Arg (3) Gly-Tyr-Met-Gln-Phe.
B) By CNBr (4) Gln-Phe (5) Val-Leu-Gly-Met (6) Ile-Lys-Gly-Tyr-Met (7) Ser-Arg-Asn-Thr-Trp-Met.

1- Deduce the sequence of the original peptide.
2- Give the sequence of the reaction of Phenyl isothiocyanate with peptide (5).
3- Give the products of peptide (1) if treated with chymotrypsin.

Denaturation of protein:

Protein structures have evolved to function in particular cellular environments. Conditions different from those in the cell can result in protein structural changes, large and small. A loss of three-dimensional structure sufficient to cause loss of function is called denaturation. The denatured state does not necessarily equate with complete unfolding of the protein and randomization of conformation.

Most proteins can be denatured by heat, which affects the weak interactions in a protein (primarily hydrogen bonds) in a complex manner. If the temperature is increased slowly, a protein's conformation generally remains intact until an abrupt loss of structure (and function) occurs over a narrow temperature range. The abruptness of the change suggests that unfolding is a cooperative process: loss of structure in one part of the protein destabilizes other parts. The effects of heat on proteins are not readily predictable.

Proteins can be denatured not only by heat but by extremes of pH, by certain miscible organic solvents such as alcohol or acetone, by certain solutes such as urea and guanidine hydrochloride, or by detergents. Each of these denaturing agents represents a relatively mild treatment in the sense that no covalent bonds in the polypeptide chain are broken. Organic solvents, urea, and detergents act primarily by disrupting the hydrophobic interactions that make up the stable core of globular proteins; extremes of pH alter the net charge on the protein, causing electrostatic repulsion and the disruption of some hydrogen bonding. The denatured states obtained with these various treatments need not be equivalent.

Plasma proteins:

Plasma is present in the blood and constitutes the liquid part of the blood. It is made up of 90% water which is required for hydration of body tissues. 7% of plasma is composed of proteins and looks like a pale yellow liquid. The total concentration of plasma protein in blood is 7-7.5 g/dl.

There are three main groups of plasma protein:

1. **Albumin:** It comprises of 60% of overall plasma protein and hence albumins are the highly abundant plasma proteins. These are produced by the liver and helps in transportation of different components in blood, along with drugs. Albumins also help to maintain water balance in the body and contribute to osmotic pressure.

2. **Fibrinogen:** These comprise of merely 4% of overall plasma proteins. This plasma protein is created by liver and its only function is to make clots and stop bleeding. It is very sticky and is a fibrous coagulant present in blood which produces thrombin. This thrombin then gets transformed into fibrin which is the major protein in blood clot.
3. **Globulin:** These comprise of 36% of overall plasma protein and include protein carriers, enzymes, gamma globulin and antibodies. Globulins are fractionized into alpha, beta and gamma and although all the globulins are manufactured in liver, gamma globulins are especially produced by lymphocytes also termed as plasma cells. Plasma globulins are categorized into four types based on its properties:
   i. **α1-Globulin:** This fraction includes several complex proteins containing carbohydrates and lipids. These are α1-glycoprotein and α-lipoproteins. The normal serum level of α1-globulin is 0.42 g/dl.
   ii. **α2-Globulins:** This fraction also contains complex proteins such as α2-glycoproteins, plasminogen, prothrombin, haptoglobulin, ceruloplasmin (transports Cu) and α2-macroglobulin. The normal serum value of this fraction is 0.67 g/dl.
   iii. **β-Globulins:** This fraction of plasma proteins contain these different β-lipoproteins which are very rich in lipid content. It also contains transferrin which transports non-heme iron in plasma. The normal serum value of β-globulins is 0.91 g/dl.
   iv. **γ-Globulins:** These are also called as Immunoglobulins and possess antibody activity. On the basis of their electrophoretic mobility they are classified as IgG, IgA and IgM.

Plasma protein functions:

The functions of plasma proteins include:

1. Osmotic or intravascular effect of plasma protein maintains fluid as well as electrolyte balance.
2. Viscosity of plasma is maintained by the plasma protein.
3. These are the protein reserves of our body.
4. Performs the important function of clotting.
5. Responds with inflammation in case of wound or injury.
6. The gamma globulins act as antibodies and protect our body from infection.
7. Plasma protein also maintains acid base balance.