

SBT1102 – BIOCHEMISTRY**UNIT 1 CARBOHYDRATES**

Introduction. Classification, Properties and Biological importance. Isomers, epimers, enantiomers, mutarotation, open chain and closed chain structures of glucose.

UNIT 2 AMINOACIDS AND PROTEINS

Amino acids: classification- essential and non-essential amino acids, protein and non-protein amino acids, Zwitter ions. Proteins: Classification- based on i) shape and solubility and ii) increasing complexity of structure. Structure of proteins: primary, secondary, tertiary and quaternary, biological significance. Concept of isoelectric point and its significance.

UNIT 3 LIPIDS

Introduction, Classification, Properties and Biological importance. Fatty acid nomenclature and structure, Lipids in cell membrane Cholesterol and Steroids, Hormones - structure and function

UNIT 4 NUCLEIC ACIDS

Introduction- Nitrogenous bases - Purines and Pyrimidines - Nucleosides and Nucleotides -- Structure of nucleic acids - DNA, RNA: m-RNA, t-RNA, r-RNA - Biological importance of nucleic acids. 16s rRNA and its significance.

UNIT 5 VITAMINS AND MINERALS

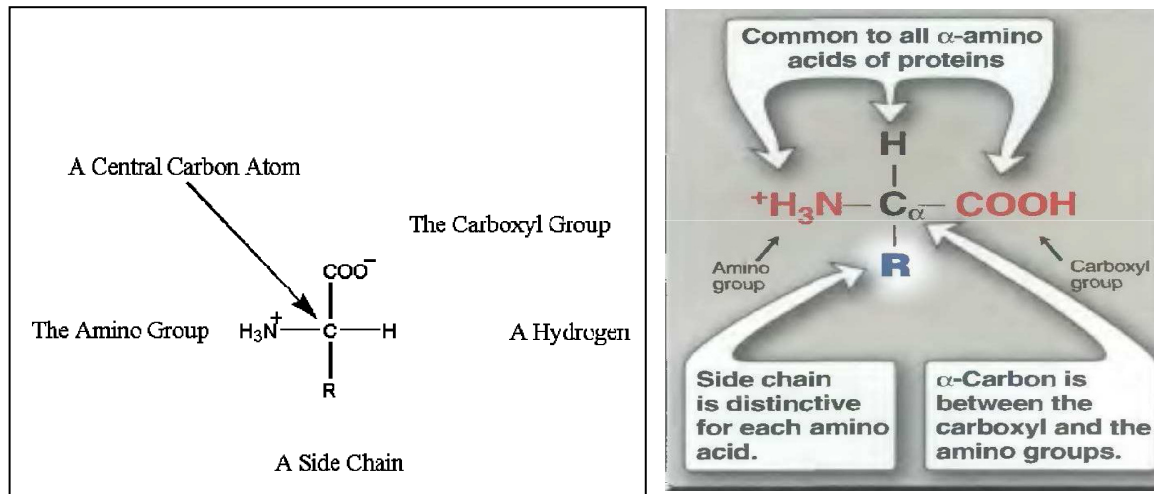
Vitamins: fat soluble and water soluble vitamins. Minerals: Micro and Macro minerals. Biological importance of vitamin and minerals, deficiency symptoms

Amino acids

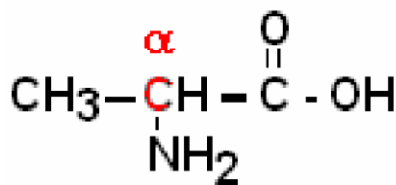
Amino acids are the building blocks of proteins. It has both an amino group ($-\text{NH}_2$) and an acid group ($-\text{COOH}$). There are more than 300 amino acids that occur in nature and many more yet to be characterized. Only 20 of the amino acids are found in the protein structure. The genetic code exists for only the 20 amino acids.

Structure of amino acids

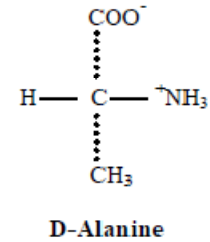
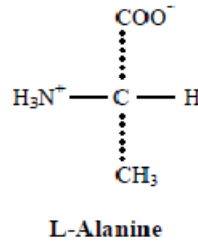
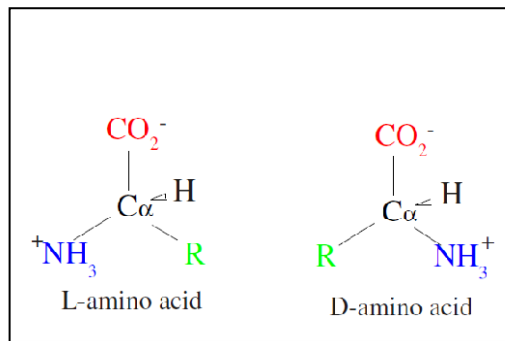
Each amino acid has 4 different groups attached to α -carbon (which is carbon atom next to carboxylic group – COOH).



The properties of each amino acid are determined by its specific side chain (R -groups). R -groups vary in structure, size, electric charge and solubility in water from one amino acid to other. Amino acids found in proteins are α -amino acids. The amino group is always found on the carbon adjacent to the carboxyl group.



Chirality – amino acids (except glycine) have a tetrahedral C_α bonded to four different chemical groups. As a result of this, amino acids are optically active or chiral. Common amino acids are all L stereoisomers. “CO-R-N” mnemonic is used for distinguishing L and D stereoisomers. Looking down the $\text{H}-\text{C}$ bond, CO-R-N spelled clockwise indicates the L stereoisomer.



There is no definitive answer on why the L isomer is found in proteins. Both D and L isomers have identical energies. Repetitive substructure in proteins (helices, sheets, turns) require all amino acids to have the same configuration. Apparently, living systems evolved from L amino acids based upon an initial random choice.

Amino acid names are often abbreviated as either 3 letters or single letter.

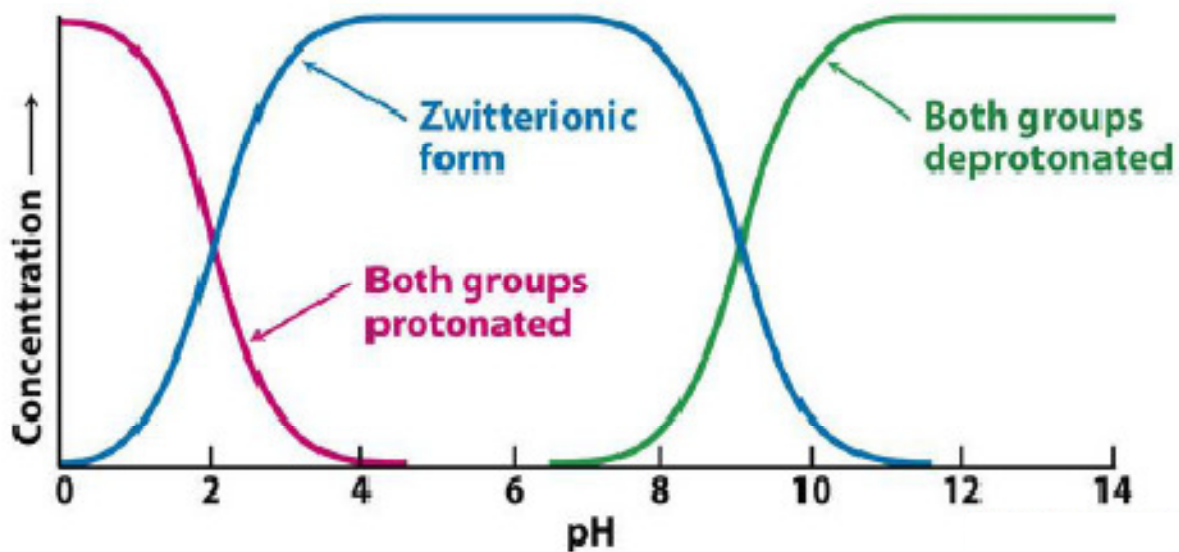
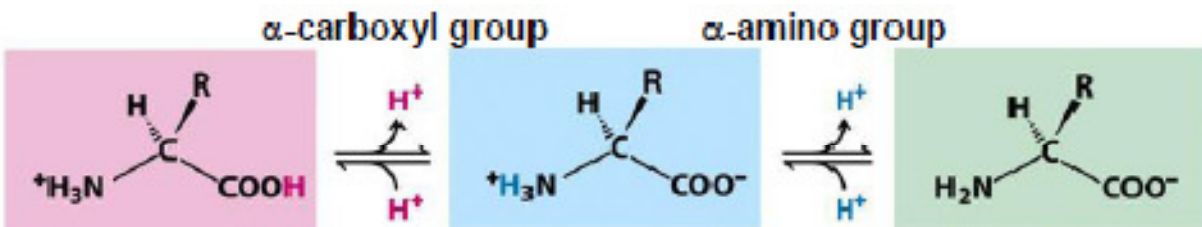
Names and Abbreviations of the Standard Amino Acids

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

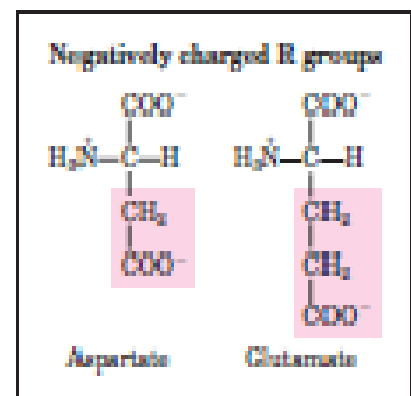
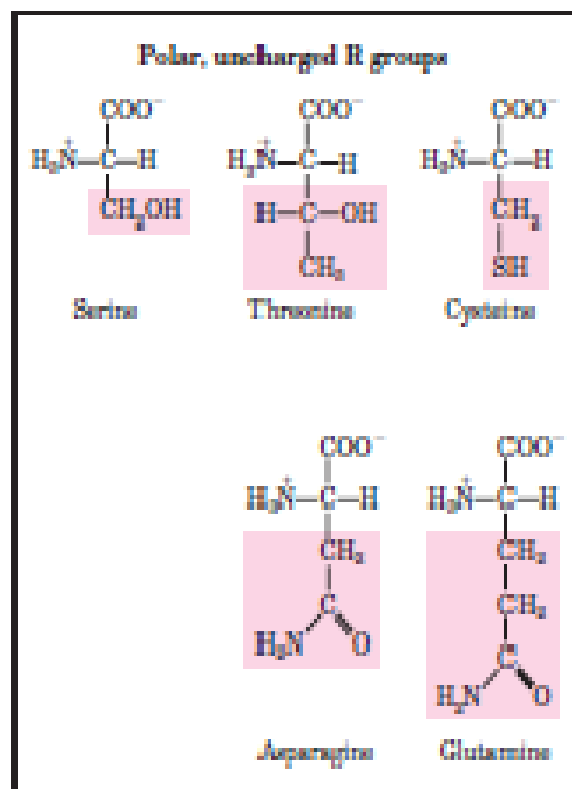
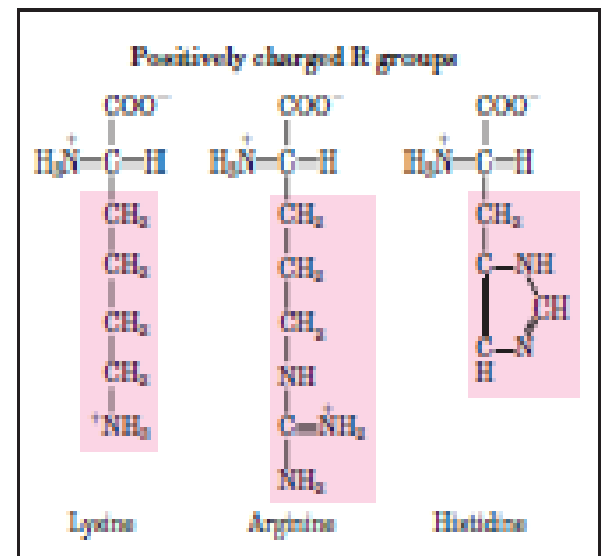
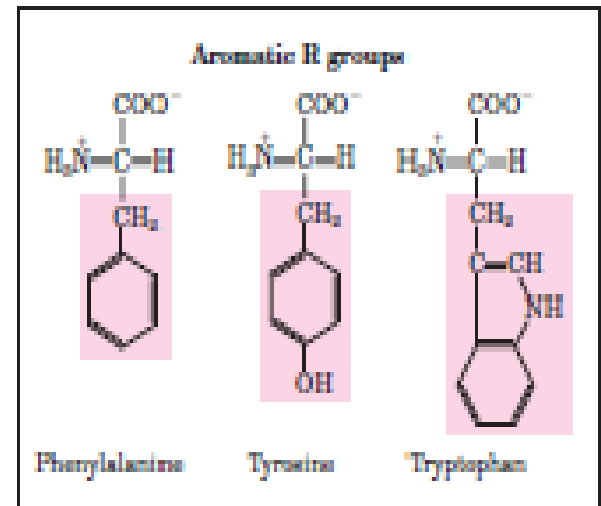
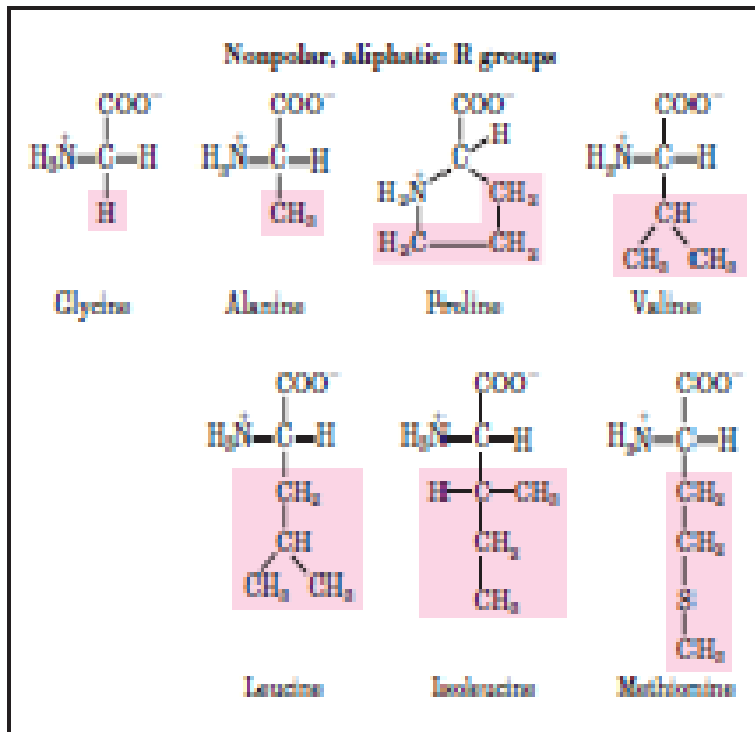
Zwitter Ions

At physiological pH of 7, the carboxyl group of an amino acid is in its conjugate base form ($-\text{COO}^-$) and the amino group is in its conjugate acid form ($-\text{NH}_3^+$). Thus each amino acid can behave as either an acid or a base. Such molecules which can behave both like an acid and a base are termed amphoteric molecules. Also molecules that bear both positive and negative charges are called zwitter ions.

Amino acids contain ionizable groups. The predominant ionic form of these molecules in solution therefore depends on the pH. At acidic pH ($\text{pH} < 7$) the carboxyl group ($-\text{COOH}$) is uncharged and the ammonium group ($-\text{NH}_3^+$) is protonated. Therefore the net charge on the amino acid is positive (+1). At basic pH ($\text{pH} > 7$) the carboxyl group ($-\text{COO}^-$) loses its proton and becomes charged and the amino group ($-\text{NH}_2$) becomes uncharged by losing the proton. Therefore the net charge on the amino acid is negative (-1). The pH at which the amino acid has no net charge and is electrically neutral is called as the isoelectric point (pI).



Structure of amino acids



Classification of amino acids

I) Nutritional classification – Based on the ability of the body to synthesize amino acids, they can be classified as essential and non-essential amino acids.

1. Essential amino acids – These amino acids cannot be formed (synthesized) in the body and so, it is essential to be included in the diet. Their deficiency in the body affects growth, health and protein synthesis. The following amino acids are essential;

- | | |
|---------------|-------------------|
| 1. Valine | 5. Methionine. |
| 2. Isoleucine | 6. Tryptophan |
| 3. Lysine | 7. Threonine |
| 4. Leucine | 8. Phenyl alanine |

2. Semi-essential amino acids – These amino acids are formed in the body but not in sufficient amount for body requirements especially in children. The semi-essential amino acids are;

1. Arginine
2. Histidine

3. Non-essential amino acids – The amino acids that can be synthesized in the body by regular metabolism in enough amounts are called as non-essential amino acids. They need not be included in the diet. They are;

- | | |
|-------------|--------------------|
| 1. Glycine | 6. Serine |
| 2. Alanine | 7. Asparagine |
| 3. Cysteine | 8. Glutamine |
| 4. Tyrosine | 9. Aspartic acid |
| 5. Proline | 10. Glutamic acid. |

II) Protein and Non-protein amino acids

1. Proteinogenic amino acids – The amino acids that are included in the genetic code are described as “proteinogenic”. With a few exceptions only these amino acids can be included in the protein structure by translation. These amino acids are also called as the standard amino acids. They are

- Alanine
- Glycine
- Proline
- Valine
- Leucine
- Isoleucine
- Tryptophan
- Phenylalanine
- Methionine
- Serine
- Threonine
- Cysteine
- Asparagine
- Glutamine
- Tyrosine
- Histidine
- Lysine
- Arginine
- Aspartic acid
- Glutamic acid

2. Non-protein amino acids – The amino acids that are not found in protein structures are termed non-protein amino acids. More than 700 amino acids have been detected in living systems which belong to this class. They are also called as non-standard amino acids. These amino acids are formed as metabolic intermediates (eg., ornithine and citrulline). Non-standard amino acids arise from post translational modification.

- Hydroxylysine
- Hydroxyproline
- Methylhistidine
- Methylarginine
- Phosphoserine
- Formylmethionine

Some amino acid derivatives also fall under these category (eg. Histamine, Catecholamine, Gamma amino butyric acid (GABA) and Dopamine).

Proteins – Introduction

Proteins are polypeptides, which are made up of many amino acids linked together as a linear chain. The structure of an amino acid contains a amino group, a carboxyl group, and a R group which is usually carbon based and gives the amino acid it's specific properties. These properties determine the interactions between atoms and molecules, which are: van der Waals force between temporary dipoles, ionic interactions between charged groups, and attractions between polar groups.

Proteins form the very basis of life. They regulate a variety of activities in all known organisms, from replication of the genetic code to transporting oxygen, and are generally responsible for regulating the cellular machinery and determining the phenotype of an organism. Proteins accomplish their tasks in the body by three-dimensional tertiary and quaternary interactions between various substrates. The functional properties depend upon the proteins three-dimensional structure. The (3D) structures arise because particular sequences of amino acids in a polypeptide chain fold to generate, from linear chains, compact domains with specific structures. The folded domains either serve as modules for larger assemblies or they provide specific catalytic or binding sites.

Protein Types and Functions

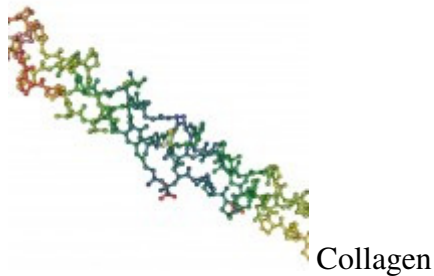
Role	Examples	Functions
Digestive enzyme	Amylase, lipase, pepsin	Break down nutrients in food into small pieces that can be readily absorbed
Transport	Hemoglobin	Carry substances throughout the body in blood or lymph
Structure	Actin, tubulin, keratin	Build different structures, like the cytoskeleton
Hormone signaling	Insulin, glucagon	Coordinate the activity of different body systems
Defense	Antibodies	Protect the body from foreign pathogens
Contraction	Myosin	Carry out muscle contraction
Storage	Legume storage proteins, egg white (albumin)	Provide food for the early development of the embryo or the seedling

Protein classification

Protein classification based on shape

On the basis of their shape, proteins may be divided into two classes: fibrous and globular.

Fibrous proteins



They have primarily mechanical and structural functions, providing support to the cells as well as the whole organism. These proteins are insoluble in water as they contain, both internally and on their surface, many hydrophobic amino acids. The presence on their surface of hydrophobic amino acids facilitates their packaging into very complex supramolecular structures. In this regard, it should be noted that their polypeptide chains form long filaments or sheets, where in most cases only one type of secondary structure, that repeats itself, is found. In vertebrates, these proteins provide external protection, support and shape; in fact, thanks to their structural properties, they ensure flexibility and/or strength. Some fibrous proteins, such as α -keratins, are only partially hydrolyzed in the intestine.

Here are some examples.

- **Fibroin**
It is produced by spiders and insects. An example is that produced by the silkworm, *Bombyx mori*.
- **Collagen**
The term “collagen” indicates not a single protein but a family of structurally related proteins (at least 29 different types), which constitute the main protein component of connective tissue, and more generally, the extracellular scaffolding of multicellular organisms. In vertebrates, they represent about 25-30% of all proteins. They are found in different tissues and organs, such as tendons and the organic matrix of bone, where they are present in very high percentages, but also in cartilage and in the cornea of the eye. In the different tissues, they form different structures, each capable of satisfying a particular need. For example, in the cornea, the molecules are arranged in an almost crystalline array, so that they are virtually transparent, while in the skin they form fibers not very intertwined and directed in all directions, which ensure the tensile strength of the skin itself. Note: the different types of collagen have low nutritional value as deficient in several amino acids (in fact, they contain no tryptophan and low amount of

the other essential amino acids). The gelatin used in food preparation is a derivative of collagen.

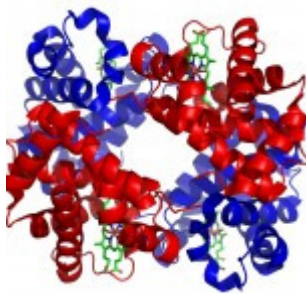
- α -Keratins

They constitute almost the entire dry weight of nails, claws, beak, hooves, horns, hair, wool, and a large part of the outer layer of the skin. The different stiffness and flexibility of these structures is a consequence of the number of disulfide bonds that contribute, together with other binding forces, to stabilize the protein structure. And this is the reason why wool keratins, which have a low number of disulfide bonds, are flexible, soft and extensible, unlike claw and beak keratins that are rich in disulfide bonds.

- Elastin

This protein provides elasticity to the skin and blood vessels, a consequence of its random coiled structure, that differs from the structures of the α -keratins and collagens.

Globular proteins



Haemoglobin

Most of the proteins belong to this class. They have a compact and more or less spherical structure, more complex than fibrous proteins. In this regard, motifs, domains, tertiary and quaternary structures are found, in addition to the secondary structures. They are generally soluble in water but can also be found inserted into biological membranes (transmembrane proteins), thus in a hydrophobic environment. Unlike fibrous proteins, that have structural and mechanical functions, they act as:

- enzymes;
- hormones;
- membrane transporters and receptors;
- transporters of triglycerides, fatty acids and oxygen in the blood;
- immunoglobulins or antibodies;
- grain and legume storage proteins.

Examples of globular proteins are myoglobin, hemoglobin, and cytochrome c. At the intestinal level, most of the globular proteins of animal origin are hydrolyzed almost entirely to amino acids.

Protein classification based on solubility and chemical composition

On the basis of their chemical composition, proteins may be divided into two classes: simple and complex.

Simple proteins

Also known as homoproteins, they are made up of only amino acids. Simple proteins yield only amino acids on hydrolysis. Examples are plasma albumin, collagen, and keratin. These proteins are further classified based on their solubility in different solvents as well as their heat coagulability.

Albumins

- Albumins are readily soluble in water, dilute acids and alkalies, coagulated by heat.
- Seed proteins contain albumin in lesser quantities.
- Albumins may be precipitated out from solution using high salt concentration, a process 'called **'salting out'**'.
- They are deficient in **glycine**.
- Serum albumin and ovalbumin (egg white) are examples.

Globulins

- Globulins are **insoluble or sparingly soluble in water**, but their solubility is greatly increased by the addition of neutral salts such as sodium chloride.
- These proteins are coagulated by heat.
- They are deficient in **methionine**.
- Serum globulin, fibrinogen, myosin of muscle and globulins of pulses are examples.

Prolamins

- Prolamins are insoluble in water but soluble in 70-80% aqueous alcohol.
- Upon hydrolysis they yield much proline and amide nitrogen, hence the name prolamin.
- They are deficient in **lysine**.
- Gliadin of wheat and zein of corn are examples of prolamins.

Glutelins

- Glutelins are insoluble in water and absolute alcohol but soluble in dilute alkalies and acids.
- They are plant proteins e.g., glutenin of wheat.

Histones

- Histones are small and stable basic proteins
- They contain fairly large amounts of basic amino acid, **histidine**.
- They are soluble in water, but insoluble in ammonium hydroxide.
- They are not readily coagulated by heat.
- They occur in **globin of hemoglobin and nucleoproteins**.

Protamines

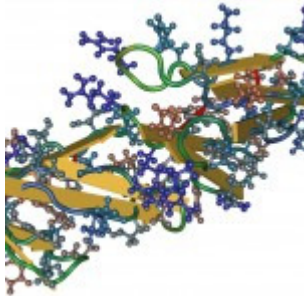
- Protamines are the simplest of the proteins.

- They are soluble in water and are not coagulated by heat.
- They are basic in nature due to the presence of large quantities of arginine.
- Protamines are found in association with nucleic acid in the sperm cells of certain fish.
- Tyrosine and tryptophan are usually absent in protamines.

Albuminoids

- These are characterized by great stability and insolubility in water and salt solutions.
- These are called albuminoids because they are essentially similar to albumin and globulins.
- They are highly resistant to proteolytic enzymes.
- They are fibrous in nature and form most of the supporting structures of animals.
- They occur as chief constituent of exoskeleton structure such as hair, horn and nails.

Conjugated proteins



Human Fibronectin

Sometimes also called heteroproteins, they contain in their structure a non-protein portion. These non-protein substances are known as prosthetic groups. The examples are glycoproteins, chromoproteins, nucleoproteins, mucoproteins, lipoproteins, metalloproteins and phosphoproteins.

Glycoproteins

They are proteins that covalently bind one or more carbohydrate units to the polypeptide backbone. Typically, the branches consist of not more than 15-20 carbohydrate units, where you can find arabinose, fucose (6-deoxygalactose), galactose, glucose, mannose, N-acetylglucosamine (GlcNAc, or NAG), and N-acetylneuraminic acid (Neu5Ac or NANA). Examples of glycoproteins are: glycophorin, the best known among erythrocyte membrane glycoproteins; fibronectin, that anchors cells to the extracellular matrix through interactions on one side with collagen or other fibrous proteins, while on the other side with cell membranes; all blood plasma proteins, except albumin; immunoglobulins or antibodies.

Chromoproteins

They are proteins that contain colored prosthetic groups. Typical examples are: hemoglobin and myoglobin, which bind, respectively, one and four heme groups; chlorophylls, which bind a porphyrin ring with a magnesium atom at its centre; rhodopsins, which bind retinal.

Phosphoproteins

They are proteins that bind phosphoric acid to serine and threonine residues. Generally, they have a structural function, such as tooth dentin, or reserve function, such as milk caseins (alpha, beta, gamma and delta), and egg yolk phosvitin.

Nucleoproteins

- Nucleoproteins are simple basic proteins (protamines or histones) in salt combination with nucleic acids as the prosthetic group.
- They are the important constituents of nuclei and chromatin.

Mucoproteins

- These proteins are composed of simple proteins in combination with carbohydrates like mucopolysaccharides, which include hyaluronic acid and chondroitin sulphates.
- On hydrolysis, mucopolysaccharides yield more than 4% of amino-sugars, hexosamine and uronic acid e.g., ovomucoid from egg white.
- Soluble mucoproteins are neither readily denatured by heat nor easily precipitated by common protein precipitants like trichloroacetic acid or picric acid.
- The term glycoproteins is restricted to those proteins that contain small amounts of carbohydrate usually less than 4% hexosamine.

Lipoproteins

These are proteins conjugated with lipids such as neutral fat, phospholipids and cholesterol

Metalloproteins

- These are **metal-binding proteins**.
- A γ -globulin, termed **transferrin** is capable of combining with **iron, copper and zinc**. This protein constitutes 3% of the total plasma protein.
- Another example is **ceruloplasmin**, which contains **copper**.

Derived proteins

These are proteins derived by partial to complete hydrolysis from the simple or conjugated proteins by the action of acids, alkalies or enzymes. They include two types of derivatives, primary-derived proteins and secondary-derived proteins.

Primary-derived proteins

These protein derivatives are formed by processes causing only slight changes in the protein molecule and its properties. There is little or no hydrolytic cleavage of peptide bonds.

Proteans

- Proteans are insoluble products formed by the action of water, dilute acids and enzymes.
- These are particularly formed from globulins but are insoluble in dilute salt solutions
- e.g., myosan from myosin, fibrin from fibrinogen.

Metaproteins

- These are formed by the action of acids and alkalies upon protein.
- They are insoluble in neutral solvents.

Coagulated proteins

Coagulated proteins are insoluble products formed by the action of heat or alcohol on natural proteins e.g., cooked meat and cooked albumin.

Secondary-derived proteins

- These proteins are formed in the progressive hydrolytic cleavage of the peptide bonds of protein molecule.
- They are roughly grouped into **proteoses, peptones and peptides according to average molecular weight.**
- Proteoses are hydrolytic products of proteins, which are soluble in water and are not coagulated by heat.
- Peptones are hydrolytic products, which have simpler structure than proteoses.
- They are soluble in water and are not coagulated by heat.
- Peptides are composed of relatively few amino acids.
- They are water-soluble and not coagulated by heat.
- The complete hydrolytic decomposition of the natural protein molecule into amino acids generally progresses through successive stages as follows:

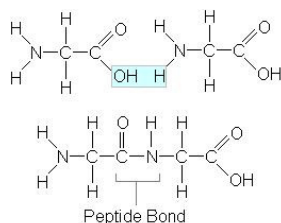
Protein -----> Protean -----> Metaprotein
Proteoses -----> Peptones -----> Peptides -----> amino acids

Protein Structures: Primary, Secondary, Tertiary, Quaternary

- Proteins are the largest and most varied class of biological molecules, and they show the greatest variety of structures. Many have intricate three-dimensional folding patterns that result in a compact form, but others do not fold up at all (“natively unstructured proteins”) and exist in random conformations. The function of proteins depends on their structure, and defining the structure of individual proteins is a large part of modern Biochemistry and Molecular Biology. To understand how proteins fold, we will start with the basics of structure, and progress through to structures of increasing complexity.

Peptide Bonds

- To make a protein, amino acids are connected together by a type of amide bond called a “peptide bond”. This bond is formed between the alpha amino group of one amino acid and the carboxyl group of another in a condensation reaction. When two amino acids join, the result is called a dipeptide, three gives a tripeptide, etc. Multiple amino acids result in a polypeptide (often shortened to “peptide”). Because water is lost in the course of creating the peptide bond, individual amino acids are referred to as “amino acid residues” once they are incorporated. Another property of peptides is polarity: the two ends are different. One end has a free amino group (called the “N-terminal”) and the other has a free carboxyl group (“C-terminal”).

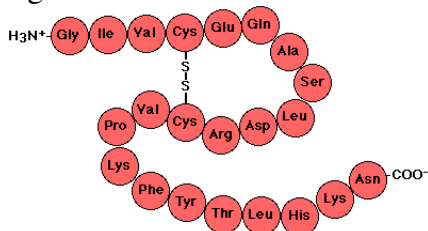


A molecule of water is removed from two glycine amino acids to form a peptide bond.

- In the natural course of making a protein, polypeptides are elongated by the addition of amino acids to the C-terminal end of the growing chain. Conventionally, peptides are written N-terminal first; therefore gly-ser is not the same as ser-gly or GS is not the same as SG. The connection gives rise to a repeating pattern of “NCC-NCC-NCC...” atoms along the length of the molecule. This is referred to as the “backbone” of the peptide. If stretched out, the side chains of the individual residues project outwards from this backbone.
- The peptide bond is written as a single bond, but it actually has some characteristics of a double bond because of the resonance between the C-O and C-N bonds:
- This means that the six atoms involved are coplanar, and that there is not free rotation around the C–N axis. This constrains the flexibility of the chain and prevents some folding patterns.

Primary Structure of Proteins

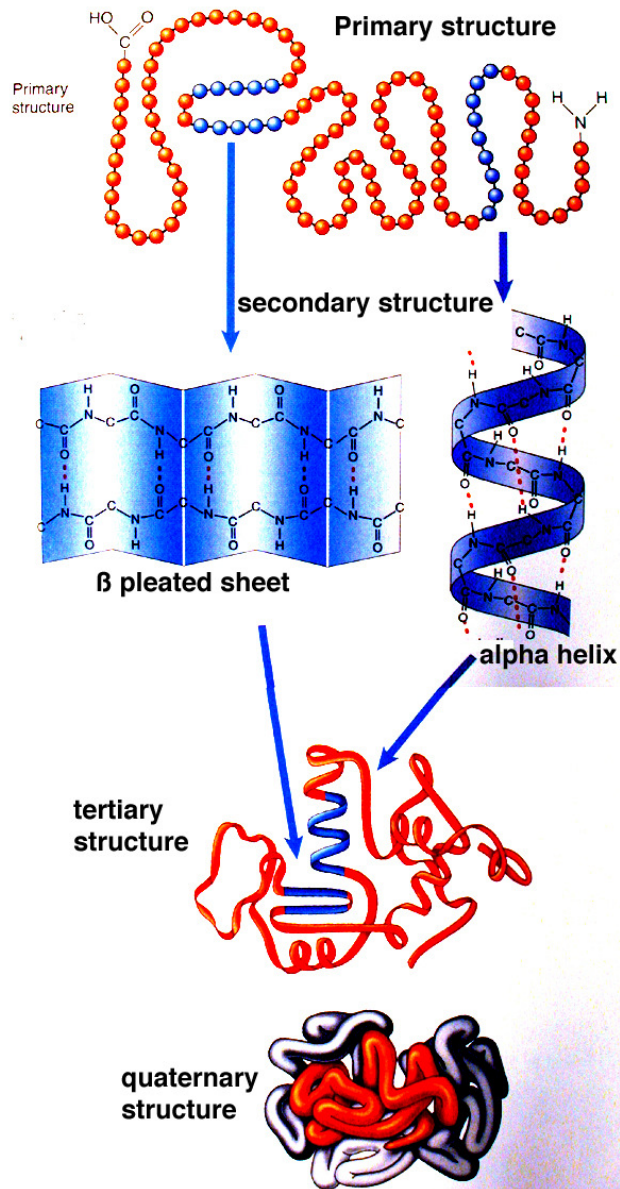
- It is convenient to discuss protein structure in terms of four levels (primary to quaternary) of increasing complexity. Primary structure is simply the sequence of residues making up the protein. Thus primary structure involves only the covalent bonds linking residues together.



- The minimum size of a protein is defined as about 50 residues; smaller chains are referred to simply as peptides. So the primary structure of a small protein would consist of a sequence of 50 or so residues. Even such small proteins contain hundreds of atoms and have molecular weights of over 5000 Daltons (Da). There is no theoretical maximum size, but the largest protein so far discovered has about 30,000 residues. Since the average molecular weight of a residue is about 110 Da, that single chain has a molecular weight of over 3 million Daltons.

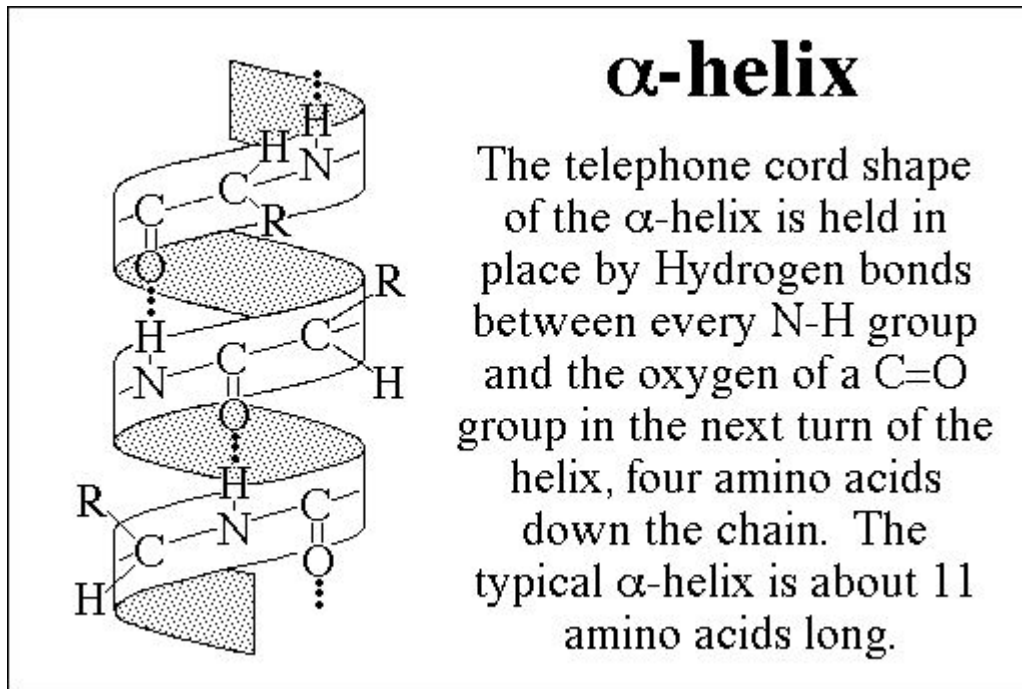
Secondary Structure

- This level of structure describes the local folding pattern of the polypeptide backbone and is stabilized by hydrogen bonds between N-H and C=O groups. Various types of secondary structure have been discovered, but by far the most common are the orderly repeating forms known as the α helix and the β sheet.



- An α helix, as the name implies, is a helical arrangement of a single polypeptide chain, like a coiled spring. In this conformation, the carbonyl and N-H groups are oriented parallel to the axis. Each carbonyl is linked by a hydrogen bond to the N-H of a residue located 4 residues further on in the sequence within the same chain. All C=O and N-H groups are involved in hydrogen bonds, making a fairly rigid cylinder. The α helix has precise dimensions: 3.6 residues per turn, 0.54 nm per turn. The side chains project outward and contact any solvent, producing a structure something like a bottle brush or a round hair brush. An example of a protein with many α helical structures is the keratin that makes up human hair.
- The structure of a β sheet is very different from the structure of an α helix. In a β sheet, the polypeptide chain folds back on itself so that polypeptide strands lie side by side, and are held together by hydrogen bonds, forming a very rigid structure. Again, the

polypeptide N-H and C=O groups form hydrogen bonds to stabilize the structure, but unlike the α helix, these bonds are formed between neighbouring polypeptide (β) strands. Generally the primary structure folds back on itself in either a parallel or antiparallel arrangement, producing a parallel or antiparallel β sheet. In this arrangement, side chains project alternately upward and downward from the sheet. The major constituent of silk (silk fibroin) consists mainly of layers of β sheet stacked on top of each another.

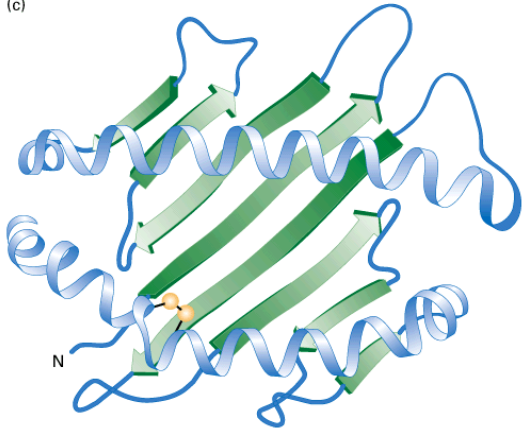


- Other types of secondary structure. While the α helix and β sheet are by far the most common types of structure, many others are possible. These include various loops, helices and irregular conformations. A single polypeptide chain may have different regions that take on different secondary structures. In fact, many proteins have a mixture of α helices, β sheets, and other types of folding patterns to form various overall shapes.
- What determines whether a particular part of a sequence will fold into one or the other of these structures? A major determinant is the interactions between side chains of the residues in the polypeptide. Several factors come into play: steric hindrance between nearby large side chains, charge repulsion between nearby similarly-charged side chains, and the presence of proline. Proline contains a ring that constrains bond angles so that it will not fit exactly into an α helix or β sheet. Further, there is no H on one peptide bond when proline is present, so a hydrogen bond cannot form. Another major factor is the presence of other chemical groups that interact with each other. This contributes to the next level of protein structure, the tertiary structure.

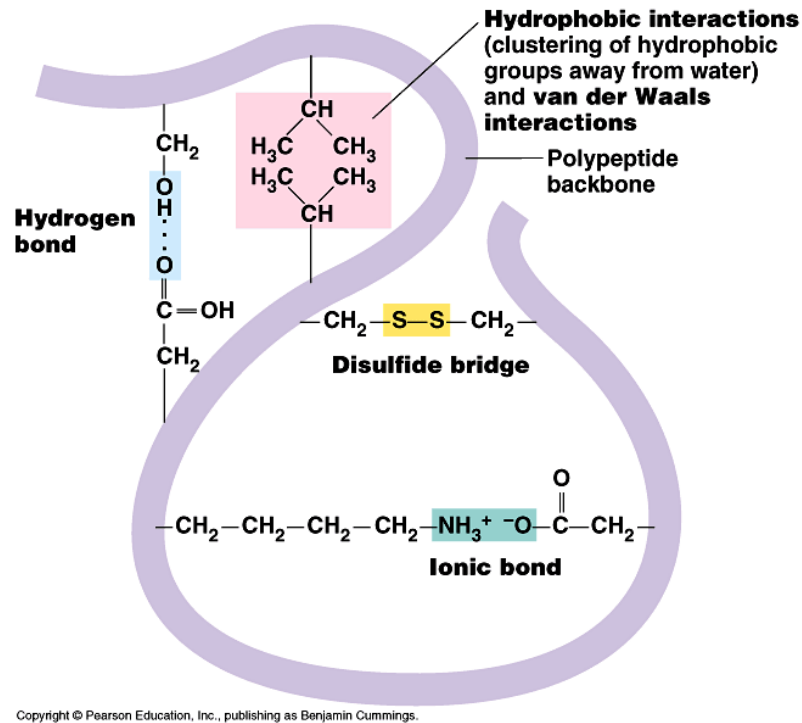
Tertiary Structure

- This level of structure describes how regions of *secondary* structure fold together – that is, the 3D arrangement of a polypeptide chain, including α helices, β sheets, and any other loops and folds. Tertiary structure results from interactions between side chains, or between side chains and the polypeptide backbone, which are often distant in sequence. Every protein has a particular pattern of folding and these can be quite complex.
- Whereas secondary structure is stabilized by H-bonding, all four “weak” forces contribute to tertiary structure. Usually, the most important force is hydrophobic interaction (or hydrophobic bonds). Polypeptide chains generally contain both hydrophobic and hydrophilic residues. Much like detergent micelles, proteins are most stable when their hydrophobic parts are buried, while hydrophilic parts are on the surface, exposed to water. Thus, more hydrophobic residues such as trp are often surrounded by other parts of the protein, excluding water, while charged residues such as asp are more often on the surface.

(c)



- Other forces that contribute to tertiary structure are ionic bonds between side chains, hydrogen bonds, and van der Waals forces. These bonds are far weaker than covalent bonds, and it takes multiple interactions to stabilize a structure.
- There is one covalent bond that is also involved in tertiary structure, and that is the disulfide bond that can form between cysteine residues. This bond is important only in non-cytoplasmic proteins since there are enzyme systems present in the cytoplasm to remove disulfide bonds.
- Visualization of protein structures Because the 3D structures of proteins involve thousands of atoms in complex arrangements, various ways of depicting them so they are understood visually have been developed, each emphasizing a different property of the protein. Software tools have been written to depict proteins in many different ways, and have become essential to understanding protein structure and function.

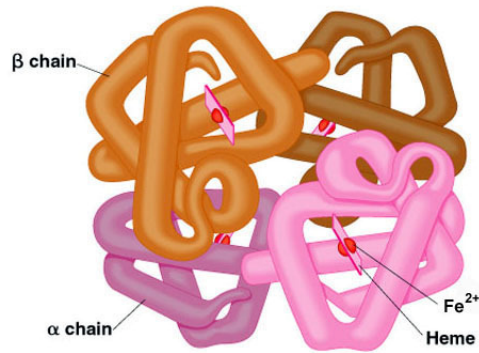


Structural Domains of Proteins

- Protein structure can also be described by a level of organization that is distinct from the ones we have just discussed. This organizational unit is the protein “domain,” and the concept of domains is extremely important for understanding tertiary structure. A domain is a distinct region (sequence of amino acids) of a protein, while a structural domain is an independently-folded part of a protein that folds into a stable structure. A protein may have many domains, or consist only of a single domain. Larger proteins generally consist of connected structural domains. Domains are often separated by a loosely folded region and may create clefts between them..

Quaternary Structure

- Some proteins are composed of more than one polypeptide chain. In such proteins, quaternary structure refers to the number and arrangement of the individual polypeptide chains. Each polypeptide is referred to as a subunit of the protein. The same forces and bonds that create tertiary structure also hold subunits together in a stable complex to form the complete protein.
- Individual chains may be identical, somewhat similar, or totally different. As examples, CAP protein is a dimer with two identical subunits, whereas hemoglobin is a tetramer containing two pairs of non-identical (but similar) subunits. It has 2 α subunits and 2 β subunits. Secreted proteins often have subunits that are held together by disulfide bonds. Examples include tetrameric antibody molecules that commonly have two larger subunits and two smaller subunits (“heavy chains” and “light chains”) connected by disulfide bonds and noncovalent forces.



- In some proteins, intertwined α helices hold subunits together; these are called coiled-coils. This structure is stabilized by a hydrophobic surface on each α helix that is created by a heptameric repeat pattern of hydrophilic/hydrophobic residues. The sequence of the protein can be represented as "abcdefgabcdefgabcdefg..." with positions "a" and "d" filled with hydrophobic residues such as A, V, L etc. Each α helix has a hydrophobic surface that therefore matches the other. When the two helices coil around each other, those surfaces come together, burying the hydrophobic side chains and forming a stable structure. An example of such a protein is myosin, the motor protein found in muscle that allows contraction.

Protein Folding

- How and why do proteins naturally form secondary, tertiary and quaternary structures? This question is a very active area of research and is certainly not completely understood. A folded, biologically-active protein is considered to be in its "native" state, which is generally thought to be the conformation with least free energy.
- Proteins can be unfolded or "denatured" by treatment with solvents that disrupt weak bonds. Thus organic solvents that disrupt hydrophobic interactions, high concentrations of urea or guanidine that interfere with H-bonding, extreme pH or even high temperatures, will all cause proteins to unfold. Denatured proteins have a random, flexible conformation and usually lack biological activity. Because of exposed hydrophobic groups, they often aggregate and precipitate. This is what happens when you fry an egg.
- If the denaturing condition is removed, some proteins will re-fold and regain activity. This process is called "renaturation." Therefore, all the information necessary for folding is present in the primary structure (sequence) of the protein. During renaturation, the polypeptide chain is thought to fold up into a loose globule by hydrophobic effects, after which small regions of secondary structure form into especially favorable sequences. These sequences then interact with each other to stabilize intermediate structures before the final conformation is attained.

- Many proteins have great difficulty renaturing, and proteins that assist other proteins to fold are called “molecular chaperones.” They are thought to act by reversibly masking exposed hydrophobic regions to prevent aggregation during the multi-step folding process. Proteins that must cross membranes (eg. mitochondrial proteins) must stay unfolded until they reach their destination, and molecular chaperones may protect and assist during this process.

Protein families/Types of proteins

- Proteins are classified in a number of ways, according to structure, function, location and/or properties. For example, many proteins combine tightly with other substances such as carbohydrates (“glycoproteins”), lipids (“lipoproteins”), or metal ions (“metalloproteins”). The diversity of proteins that form from the 20 amino acids is greatly increased by associations such as these. Proteins that are tightly bound to membranes are called “membrane proteins”. Proteins with similar activities are given functional classifications. For example, proteins that break down other proteins are called proteases.
- Because almost all proteins arise by an evolutionary process, ie. new ones are derived from old ones, they can be classified into families by their relatedness. Proteins that derive from the same ancestor are called “homologous proteins”. Studying the sequences of homologous proteins can give clues to the structure and function of the protein. Residues that are critical for function do not change on an evolutionary timescale; they are referred to as “conserved residues”. Identifying such residues by comparing amino acid sequences often helps clarify what a protein is doing or how it is folded. For example the proteases trypsin and chymotrypsin are members of the “serine protease” family; so-named because of a conserved serine residue that is essential to catalyze the reaction. Trypsin and chymotrypsin contain very similar folding patterns and reaction mechanisms. Recognizing a pattern of conserved residues in protein sequences often allows scientists to deduce the function of a protein.

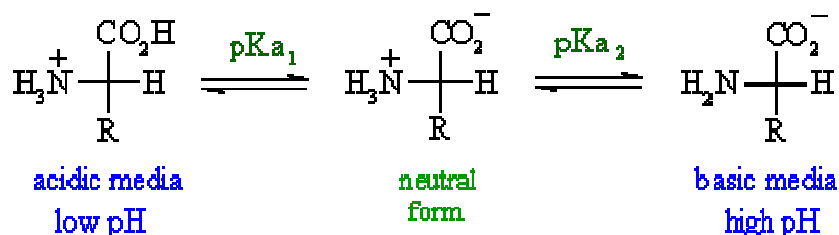
Isoelectric point, pI

- The isoelectric point or isoionic point is the pH at which the amino acid does not migrate in an electric field.
- This means it is the pH at which the amino acid is neutral, *i.e.* the zwitterion form is dominant.
- The pI is given by the average of the pK_as that involve the zwitterion, *i.e.* that give the boundaries to its existence.

There are 3 cases to consider....

- neutral side chains

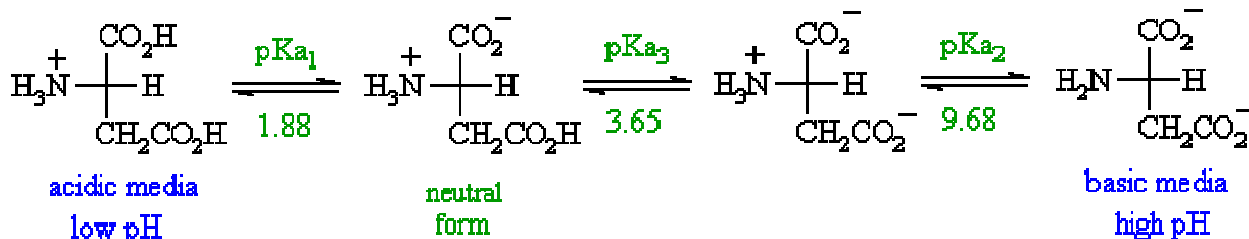
These amino acids are characterised by two pK_as : pK_{a1} and pK_{a2} for the carboxylic acid and the amine respectively. The isoelectric point will be halfway between, or the average of, these two pK_as, *i.e.* $pI = 1/2 (pK_{a1} + pK_{a2})$. This is most readily appreciated when you realise that at very acidic pH (below pK_{a1}) the amino acid will have an overall +ve charge and at very basic pH (above pK_{a2}) the amino acid will have an overall -ve charge. For the simplest amino acid, [glycine](#), pK_{a1} = 2.34 and pK_{a2} = 9.6, pI = 5.97.



The other two cases introduce other ionisable groups in the side chain "R" described by a third acid dissociation constant, pK_{a3}

- acidic side chains

The pI will be at a lower pH because the acidic side chain introduces an "extra" negative charge. So the neutral form exists under more acidic conditions when the extra -ve has been neutralised. For example, for [aspartic acid](#) shown below, the neutral form is dominant between pH 1.88 and 3.65, pI is halfway between these two values, *i.e.* $pI = 1/2 (pK_{a1} + pK_{a3})$, so pI = 2.77.



- basic side chains

The pI will be at a higher pH because the basic side chain introduces an "extra" positive charge. So the neutral form exists under more basic conditions when the extra +ve has been neutralised. For example, for [histidine](#), which was discussed on the [previous page](#), the neutral form is dominant between pH 6.00 and 9.17, pI is halfway between these two values, *i.e.* $pI = 1/2 (pK_{a2} + pK_{a3})$, so $pI = 7.59$.

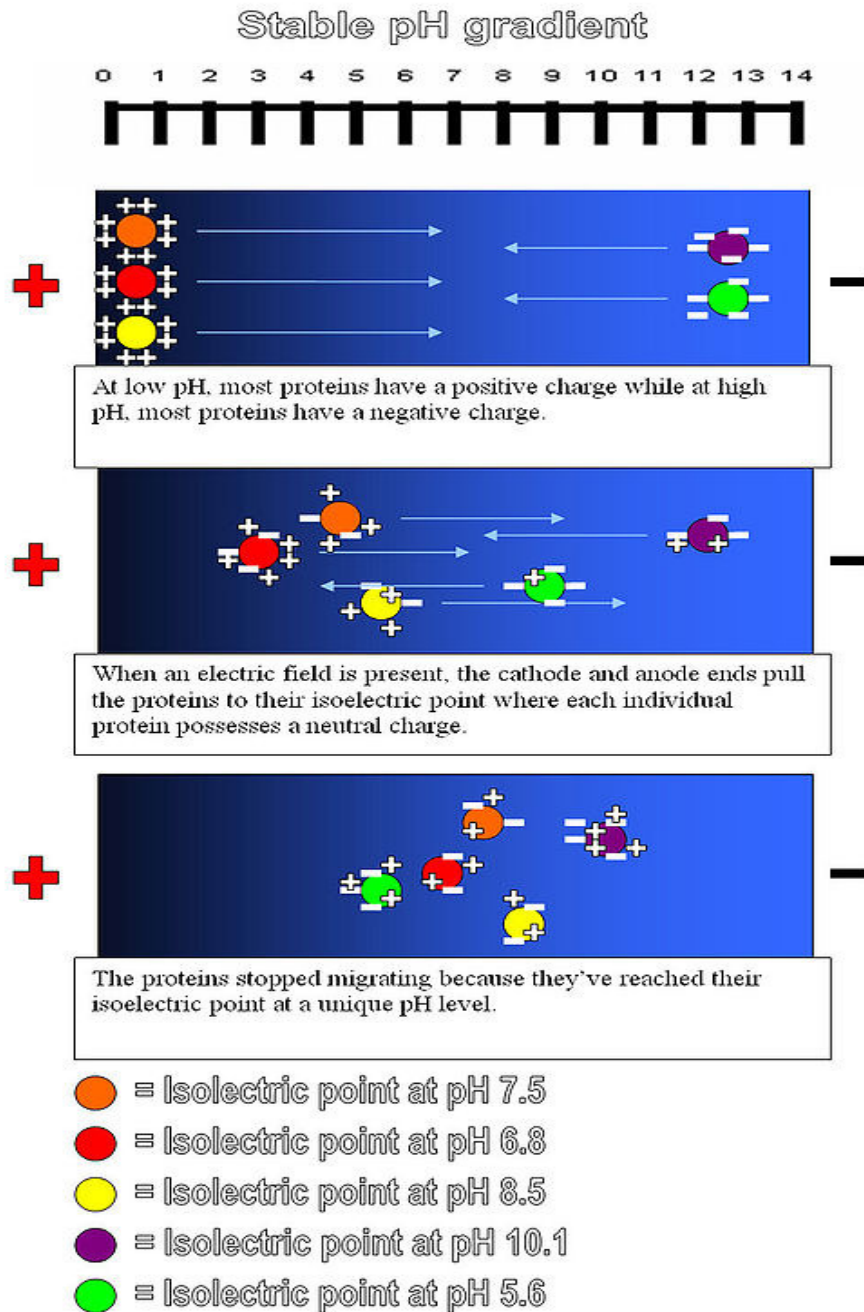
Isoelectric Focusing

Isoelectric focusing or also called the pI of the protein is the pH at which its net charge is zero. A separation technique which separates peptides according to how acidic and basic their residues are. A gel with a [pH](#) gradient is used as the medium. The pH gradient is made by adding polyampholytes, which are multi-charged polymers, with different pI into the gel. Then the sample is put onto the gel and a voltage is applied. The proteins will move along the gel until they reach their isoelectric points. In other words, each protein will move until it reaches a position in the gel at which the pH is equal to the pI of the protein. a protein band that forms at a given pH can then be removed and analyzed further. This process can successfully separate proteins that have a difference in net charge greater than or equal to 1.

Isoelectric point (pI): The pH at which the net charge on the [protein](#) is zero. For a protein with many basic [amino acid](#) pI will be high, while for an acidic protein the pI will be lower.

Isoelectric focusing is a type of zone electrophoresis, and it is usually performed in a gel, that takes advantage of the fact that a molecule's charge changes with the pH of its surroundings. A protein that is in a pH region below its isoelectric point (pI) will be positively charged and so will migrate towards the cathode. As it migrates, however, the charge will decrease until the protein reaches the pH region that corresponds to its pI. At this point it has no net charge and so migration ceases. As a result, the proteins become focused into the sharp stationary bands with each protein positioned at a point in the pH gradient corresponding to its pI. This technique is capable of extremely high resolution with proteins differing by a single charge being fractionated into separate bands.

Molecules to be focused are distributed over a medium that has a pH gradient (usually created by aliphatic ampholytes). An electric current is passed through the medium, creating a "positive" anode and "negative" cathode end. The negatively charged molecules migrate through the pH gradient in the medium toward the "positive" end while positively charged molecules move toward the "negative" end. As a particle moves towards the pole opposite of its charge it moves through the changing pH gradient until it reaches a point in which the pH of that molecules isoelectric point is reached. At this point the molecule no longer has a net electric charge (due to the protonation or deprotonation of the associated functional groups) and as such will not proceed any further within the gel. The gradient is initially established before adding the particles of interest by first subjecting a solution of small molecules such as polyampholytes with varying pI values to electrophoresis.



The method is applied in the study of proteins, which separate based on their relative content of acidic and basic residues, whose value is represented by the pI. Proteins are introduced into an immobilized pH gradient gel composed of polyacrylamide, starch, or agarose where a pH gradient has been established. Isoelectric focusing can resolve proteins that differ in pI value by as little as 0.01. Isoelectric focusing is the first step in two-dimensional gel electrophoresis, in which proteins are first separated by their pI and then further separated by molecular weight through SDS PAGE.