Detailed Guide to Biomolecules – Their Role, Definition and Types

An updated version of this ebook is available on:
https://gauravtiwari.org/biomolecules-role-definition-types/

You have always learned that a living system can grow, sustain, and reproduce itself. The most remarkable thing about a living system is that it itself is composed of non-living atoms and molecules. The scientific study of what goes on chemically within a living system falls in the domain of **biochemistry**.

Living systems are composed of various complex biomolecules such as carbohydrates, proteins, nucleic acids, lipids, and many more.

Proteins and carbohydrates are extremely crucial constituents of the food that we eat every day. These biomolecules regularly interact with each other and constitute the molecular logic of life processes. Moreover, some simple molecules such as vitamins and inorganic salts also play important roles in the biological functions of organisms.

In this section, we will learn more about the fascinating chemistry of life, what are the living organisms made of and what are the key types of biomolecules.

What are living organisms chemically made of?

We've already discussed the diversity of living organisms on earth. Miraculously, all of them are found to be made up of the same elements and compounds. Whether we chemically analyze a plant tissue, animal tissue, or microbial paste, we discover that they're composed of elements such as carbon, hydrogen, oxygen and many more. If we perform this same chemical analysis on a specimen of non-living matter, such as a piece of earth's crust, then we get another similar list of elements.

On a strictly technical basis, we can't really make out any significant differences between the two lists above. All the elements present in a sample of the earth's crust are also found in a sample of living tissue. However, upon more meticulous examination, we discover that the

relative abundance of carbon and hydrogen with respect to other elements is higher in any living tissue than in the earth's crust.

Element	% weight of the earth's crust	% weight of the human body
Hydrogen (H)	0.14	0.5
Carbon (C)	0.03	18.5
Oxygen (O)	46.6	65.0
Nitrogen (N)	Very little	3.3
Sulfur (S)	0.03	0.3
Sodium (Na)	2.8	0.2
Calcium (Ca)	3.6	1.5
Magnesium (Mg)	2.1	0.1
Silicon (Si)	27.7	Negligible

Analyzing the chemical composition of living cells

To discover the different kinds of organic compounds found in living organisms, we need to perform a chemical analysis of living tissues. After obtaining a sample of any living tissue, such as a slice of vegetable or a piece of liver, we grind it in trichloroacetic acid (CCl_3COOH) using a mortar and a pestle to obtain a thick slurry.

After straining this slurry through a cheesecloth or cotton, we obtain two different fractions. One is called the **filtrate** or, more technically, the **acid-soluble pool**. The second fraction is known as the **retentate** or the **acid-insoluble fraction**. Scientists have discovered thousands of organic compounds in the acid-soluble pool.

When we apply analytical techniques to a freshly isolated organic compound, we get an idea about the molecular formula and the probable structure of the said compound. All the carbon compounds that we obtain from living tissues are collectively known as **biomolecules**.

However, you should know that living organisms have also got inorganic elements and compounds within them. To prove that, we must perform a

slightly different experiment that is destructive in nature. First, we weigh a small amount of a living tissue (let's say a leaf or a piece of the liver; this is known as the *wet weight*) and dry it until all the water has evaporated. What's left behind is the *dry weight* of the given sample.

We now proceed to fully burn the tissue sample so that all the carbon compounds are oxidized to their gaseous form (CO_2 and water vapor) and are thus removed from the sample. The remaining material, known as the *ash*, contains inorganic elements such as calcium, magnesium, sodium, potassium, and so on. We also find inorganic compounds like sulfates and phosphates in the acid-soluble fraction.

Thus, performing an elemental analysis gives us the elemental composition of living tissues in the form of hydrogen, oxygen, chlorine, carbon, and so on. On the other hand, carrying out an analysis for compounds gives us an idea of the kind of organic and inorganic constituents that are present in living tissues.

From a chemist's point of view, you can identify functional groups such as aldehydes, ketones, aromatic compounds, etc. But as biologists, we classify these compounds like amino acids, nucleotide bases, fatty acids, and so on.

Biomacromolecules and micromolecules

All the compounds we find in the acid-soluble pool are found to have molecular weights ranging from 18 to about 800 Daltons. On the other hand, the acid-insoluble fraction has only four types of organic compounds - **proteins, nucleic acids, polysaccharides** and **lipids**. These classes of compounds, with the exception of lipids, happen to have molecular weights in the range of 10,000 Daltons and higher.

Because of this finding, scientists have classified biomolecules into two different types. The first type of biomolecules has molecular weights of less than one thousand Daltons; they are usually known as **micromolecules** or simply biomolecules. The second type of compound,

found in the acid-insoluble fraction, is known as macromolecule or **biomacromolecule**.

All the molecules found in the insoluble fraction are polymeric in nature, with the notable exception of lipids. Naturally, a question subsequently arises in our minds – why do we find lipids in the acid-insoluble fraction or macromolecular fraction if their molecular weights don't exceed 800 Daltons?

Compared to other compounds, lipids indeed have relatively low molecular weights. They're present not only in pure form but are also found arranged into the structure of cellular components such as the plasma membrane and other membranes.

When you grind a tissue sample, you are disrupting the structure of the cell. As a result, the plasma membrane and other membranes get disintegrated and form water-insoluble vesicles. These insoluble pieces of membrane get separated with the acid insoluble pool fragments in the form of vesicles and are subsequently found in the macromolecular fraction. Thus, we can say that lipids are *not strictly macromolecules*.

The acid soluble pool roughly represents the **cytoplasmic composition**, whereas the macromolecules from the cytoplasm and organelles form the acid insoluble fraction. Together, they represent the entire chemical composition of living organisms or tissues.

To sum up, if we happen to represent the chemical composition of a living tissue on the basis of the abundance of compounds and arrange them class-wise, we discover that **water** is the most abundant chemical compound in living organisms.

Primary and secondary metabolites

Most of the biomolecules we've discussed so far, such as amino acids, proteins and carbohydrates, are known as **primary metabolites**. We find them abundantly in animal cells. However, when we analyze plant, fungal and microbial cells, we notice thousands of compounds other than

primary metabolites such as alkaloids, flavonoids, rubber, essential oils, antibiotics, colored pigments, scents, gums, and spices. These compounds are known as **secondary metabolites**.

Primary metabolites have clearly identifiable functions and play well-known roles in normal physiological processes, while the role or functions of most of the secondary metabolites in host organisms are still not understood. However, many of these molecules such as rubber, drugs, and pigments are extensively used in human welfare.

Some secondary metabolites have ecological significance as well. For example, certain cyanobacterial secondary metabolites have toxic effects on living organisms. Many of these cyanotoxins have ecological roles as insecticides and herbicides.

Type of secondary metabolite	**Examples**
Pigments	Carotenoids, anthocyanins
Alkaloids	Morphine, codeine
Terpenoids	Monoterpenes, diterpenes
Essential oils	Lemongrass oil
Toxins	Abrin, ricin
Lectins	Concanavalin A
Drugs	Vinblastine, curcumin
Polymeric substances	Rubber, gums, cellulose

Carbohydrates

Primarily produced by plants, carbohydrates form a most extensive group of naturally occurring organic compounds. Some common examples of carbohydrates are cane sugar, glucose, and starch. They are primarily compounds of carbon, hydrogen and oxygen, and are also known as **saccharides** because their basic components are sugars.

Most carbohydrates have a general formula, $C_x(H_2O)_y$, and were considered to be hydrates of carbon (that's where the term "carbohydrate" comes from). For example, the molecular formula of glucose ($C_6H_{12}O_6$) readily fits into this general formula: $C_6(H_2O)_6$.

However, not all the compounds that fit into this formula may be classified as carbohydrates. For example, acetic acid (CH_3COOH) fits into the general formula $C_2(H_2O)_2$ but is not a carbohydrate. Similarly, rhamnose ($C_6H_{12}O_5$) is a carbohydrate but does not meet this criterion.

After observing a large number of their reactions, scientists have deduced that carbohydrates contain specific functional groups. Chemically, they may be defined as **optically active polyhydroxy aldehydes or ketones or the compounds that produce such units on hydrolysis**. Certain carbohydrates, which are sweet in taste, are also known as sugars. The common table sugar we use in our homes is called sucrose whereas the sugar present in milk is lactose.

On the basis of their behavior on hydrolysis, carbohydrates are broadly divided into three different groups, which we shall discuss in detail below.

Monosaccharides

A carbohydrate that cannot be hydrolyzed further to yield simpler units of polyhydroxy aldehydes or ketones is known as a monosaccharide. Around twenty monosaccharides are known to exist in nature. They are composed of 3-7 carbon atoms and are known as biomicromolecules. Examples of monosaccharides are *glucose, fructose,* and *ribose*.

Monosaccharides are further classified based on the number of carbon atoms and the functional group present in them. Monosaccharides containing an aldehyde group are known as **aldoses** and those containing a keto group are known as **ketoses**. The number of carbon atoms constituting the monosaccharide is subsequently introduced in its name too, as shown below:

Carbon atoms	General term	Aldehyde	Ketone
3	Triose	Aldotriose	Ketotriose
4	Tetrose	Aldotetrose	Ketotetrose
5	Pentose	Aldopentose	Ketopentose
6	Hexose	Aldohexose	Ketohexose

7	Heptose	Aldoheptose	Ketoheptose

Monosaccharides may be modified chemically to form several types of **derived monosaccharides**. Deoxy sugars are a good example; deoxygenation (removal of oxygen from the second carbon) of ribose produces deoxyribose, which is an important constituent of DNA. Other derived monosaccharides include amino sugars (*glucosamine, N-acetyl glucosamine*), sugar acids (*ascorbic acid, glucuronic acid*), and sugar alcohols (*mannitol*).

We'll now discuss certain monosaccharides in more detail.

(a) Glucose

Biologically and medically, glucose is arguably the most important monosaccharide in existence. It is found freely in nature as well as in combined form. It is found abundantly in honey and sweet fruits such as ripe grapes. It is also found in our blood and is thus known as "blood sugar".

Structure of glucose

Also known as dextrose, glucose is chemically an aldohexose. It serves as the monomeric unit of many larger carbohydrates such as starch and cellulose. Its structure was determined by chemists based on the following observations:

1. The molecular formula of glucose was deduced to be $\mathbf{C_6H_{12}O_6}$.
2. On prolonged heating with hydrogen iodide, it forms **n-hexane**. This observation infers that all the six carbon atoms in glucose are linked in a *straight chain*.
3. Glucose reacts with hydroxylamine to form an **oxime**. On the addition of a molecule of hydrogen cyanide, it produces cyanohydrin. These reactions confirm the presence of a *carbonyl group* (>C = O) in glucose.
4. Glucose gets oxidized to a carboxylic acid-containing six carbon atoms (**gluconic acid**) in reaction with mild oxidizing agents such as bromine water. This indicates that the carbonyl group present in glucose is an *aldehydic group*.

5. On acetylation with acetic anhydride, glucose produces **glucose pentaacetate**. This confirms the presence of *five –OH groups* in the structure of glucose. Since it exists as a stable compound, we may infer that these –OH groups are all attached to different carbon atoms.
6. On oxidation with nitric acid, both glucose and gluconic acid yield a dicarboxylic acid known as **saccharic acid**. This indicates the presence of a *primary alcoholic group* (–OH) in glucose.

The exact spatial arrangement of the different —OH groups in glucose was given by Emil Fischer after a careful study of many other properties of the sugar. Glucose is correctly named as D(+)-glucose; the 'D' here represents the configuration in three-dimensional space whereas the '(+)' represents the dextrorotatory nature of the molecule.

Cyclic structure of glucose

Certain facts could not be explained by the linear structure of glucose, as follows:

1. Although it has an aldehyde group, glucose does not give Schiff's test and also does not form the hydrogensulphite addition product with $NaHSO_3$.
2. The pentaacetate of glucose does not react with hydroxylamine. This indicates the absence of a free aldehyde group.
3. Glucose is found to exist in two different crystalline forms in solution, known as the α-form and the β-form.

The α-form of glucose (melting point: 419 K) is obtained by crystallization from a concentrated solution of glucose at 303 K. Similarly, the β-form (melting point: 423 K) is obtained by crystallization from a hot and saturated aqueous solution of glucose at 371 K.

Scientists were unable to explain this behavior by the open chain structure of glucose. As a result, they proposed that one of the —OH groups may possibly add to the —CHO group and form a cyclic hemiacetal structure.

Later, it was discovered that glucose forms a six-membered ring in which the —OH at C-5 is involved in the ring formation process. This explains the absence of a free —CHO group and also the existence of glucose in two different crystalline forms. Both of these cyclic hemiacetal forms exist in equilibrium with the open-chain structure of glucose.

The two cyclic forms of glucose differ only in the configuration of the hydroxyl group at C-1, known as the **anomeric carbon** (the aldehyde carbon just before cyclization). Such isomers are known as **anomers**. The six-membered cyclic structure of glucose is known as the *pyranose structure*, in analogy with a cyclic organic compound called pyran. Pyran is a cyclic organic compound with one oxygen atom. Diagrammatically, the cyclic form of glucose is accurately represented by the **Haworth structure**.

Biological significance

Glucose is the main source of energy for our body and is carried by the bloodstream to all our body parts. The normal concentration of glucose in the blood is 80–100 mg per 100 mL of blood. It is oxidized by living cells to produce energy in the form of ATP.

Since glucose requires no digestion, it may be given intravenously to patients who cannot take food orally. Glucose is found in the urine of patients who have diabetes mellitus. The condition in which glucose is excreted in urine is known as glycosuria.

(b) Fructose

Fructose is one of the most important ketohexoses in nature. It is obtained along with glucose during the hydrolysis of the disaccharide, sucrose. It is a natural monosaccharide abundantly found in fruits, honey and various vegetables. In its pure form, it is used as a sweetener in the food industry.

Like glucose, fructose has the molecular formula $C_6H_{12}O_6$. Scientists observed its reactions and found out that it contains a ketone group at C-2 and six carbon atoms in a straight chain, as in the case of glucose. It belongs to D-series as well, but unlike glucose, it is laevorotatory in nature. Thus, it is appropriately written as D-(–)-fructose.

Fructose also exists in two different cyclic forms that are obtained by the addition of —OH at C-5 to the carbonyl group. The ring thus formed is a

five membered ring and is known as the *furanose structure*, with analogy to a five-membered cyclic compound called furan.

Fructose is the major constituent of the polysaccharide inulin, a polysaccharide present in many plants such as dahlia tubers, chicory roots, and Jerusalem artichokes. It is the **sweetest of all natural sugars**, being about twice as sweet as glucose.

(c) Galactose

Galactose is an aldohexose and occurs, along with glucose, in lactose and in many other oligosaccharides and polysaccharides such as pectin, gums, and mucilage. Galactose is an **epimer** of glucose, differing only in the spatial arrangement of the -H and –OH groups around C-4.

Galactose is also synthesized in the mammary glands to produce the lactose in milk. It is also a constituent of glycolipids and glycoproteins in many cell membranes, such as those found in nervous tissues. It is less than half as sweet as glucose.

The inability of infants to metabolize galactose is an inherited condition known as galactosemia. In this disorder, the galactose concentration increases noticeably in the blood and also appears in the urine. Galactosemia results in vomiting, diarrhea, hepatomegaly, and often mental retardation. If not recognized within a few days after birth, it can even lead to death.

If the diagnosis is made early and lactose is removed from the diet, the symptoms disappear and the normal growth of the child may be resumed.

Oligosaccharides

They’re small carbohydrates that are formed by the condensation (the chemical combination of two molecules to form one molecule with the loss of a small molecule, usually water) of 2-9 monosaccharides. They’re

considered biomacromolecules. The monosaccharide units are joined together by a **glycosidic bond** between the aldehyde or ketone group of one monosaccharide and the alcohol group of another.

Depending on the number of monosaccharide molecules condensed to form an oligosaccharide, they can be:

- Disaccharides (*sucrose, lactose, maltose, trehalose)*
- Trisaccharides (*raffinose* – composed of glucose, fructose, and galactose)
- Tetrasaccharides (*stachyose*)

Let us study in detail about some important oligosaccharides.

(a) Sucrose

Sucrose is the common table sugar used in sweets, drinks, ice creams, cakes, and many more food items. It is commercially obtained from sugarcane stems and the roots of the sugar beet plant. On hydrolysis, it produces an equimolar mixture of D-(+)-glucose and D-(-) fructose.

These two monosaccharide units are held together by a glycosidic linkage between C-1 of α-D-glucose and C-2 of β-D-fructose. Sucrose is dextrorotatory in nature, but after hydrolysis it yields dextrorotatory glucose and laevorotatory fructose. Because the laevorotation of fructose (–92.4°) is much more than the dextrorotation of glucose (+ 52.5°), the resulting mixture is laevorotatory in nature.

Thus, we see that the hydrolysis of sucrose brings about a change in the sign of rotation from dextrorotatory (+) to laevorotatory (–). As a result, the solution is known as **invert sugar**.

(b) Maltose

Maltose or malt sugar is another disaccharide that is composed of two α-D-glucose units, in which C-1 of one glucose unit is linked to C-4 of another glucose unit.

(c) Lactose

Lactose is also known as *milk sugar* because of its prominent presence in milk. It is composed of β-D-galactose and β-D-glucose. The glycosidic linkage is between C-1 of galactose and C-4 of glucose.

Reducing sugars are sugars that can reduce Cu^{2+} ions to Cu^{+} ions. This property is found in all saccharides that possess free aldehyde or ketone groups, including all monosaccharides. On the basis of this property, Fehling's and Benedict's tests are used to detect the presence of glucose in urine.

When it comes to disaccharides, sucrose (which contains glucose and fructose) is a non-reducing sugar because both the aldehyde group of glucose and the ketone group of fructose participate in the formation of the glycosidic bond between the two of them. However, some other disaccharides such as maltose and lactose have reducing groups.

Polysaccharides

Polysaccharides are complex carbohydrates that contain a large number of monosaccharide units joined together by glycosidic linkages. These are by far the most commonly encountered carbohydrates in nature. They are mainly used by organisms for food storage or as structural components.

We'll now understand the structure of some important polysaccharides found in nature.

(a) Starch

Starch $[(C_6H_{10}O_5)_n]$ is the most important storage polysaccharide found in the plant kingdom. It is mainly concentrated in seeds, roots, and tubers. Corn, wheat, potatoes, rice, and cassava are the main dietary sources of starch. It is a polymer of α-glucose and consists of two components—**amylose** (13-20%) and **amylopectin** (80-87%).

Amylose is the water-soluble component of starch that constitutes about 15-20% of its bulk. It is a long, unbranched chain with 200-1000 a-D-(+)-glucose units held together by α-1, 4 glycosidic linkage.

Amylopectin is a water-insoluble and branched polymer of α-D-glucose units, in which the chain is formed by α-1, 4 glycosidic linkage whereas branching takes place by α-1, 6 glycosidic linkage. There are about 24-30 glucose resides between two branch points.

Despite the presence of numerous polar -OH groups, starch molecules are insoluble in water because of their very large size. However, it does form colloidal dispersions in hot water that give an intense blue-black color with free iodine because starch can hold I_2 molecules in its secondary helices. Thus, we may use a starch solution to detect free iodine or a dilute iodine solution to detect starch.

We can easily convert starch into glucose by heating it with water and a little acid (such as hydrochloric or sulphuric acid). It is also readily hydrolyzed at room temperature by certain digestive enzymes such as amylase and maltase.

(b) Glycogen

Glycogen is the main energy-storage carbohydrate of the animal kingdom. It is formed by the polymerization of glucose (a process known as **glycogenesis**) and is stored abundantly in the liver and in muscle tissues. Structurally, it is quite similar to the amylopectin fraction of starch, except that it is more highly branched. There are 12–18 glucose residues between two branch points (α-1, 6 glycosidic linkages) in a glycogen molecule.

Muscle glycogen granules (β-particles) are spherical in shape and contain up to 60,000 glucose residues. In the liver, there are similar granules and also rosettes of glycogen granules that seem to be aggregated β-particles.

(c) Cellulose

Cellulose is known for being the most abundant organic substance found in nature. It is the main structural component of plants and wood. Cotton fibers are composed of more or less pure cellulose, whereas dry wood

consists of around 50% cellulose. Cellulose plays a vital role in the textile and paper industries.

Like starch and glycogen, cellulose is a polymer of glucose. However, cellulose differs from the other two in that the glucose units in cellulose are joined by β-1,4 glycosidic linkages instead of α-1,4 glycosidic linkages. The stereochemistry of the β-anomer allows cellulose to form an extended chain that can form hydrogen bonds with adjacent cellulose molecules. The large number of hydrogen bonds so formed contributes to the strength of the resulting plant cell walls.

Cellulose has a much greater resistance to hydrolysis compared to starch or glycogen. It is not appreciably hydrolyzed when boiled in a 1% sulphuric acid solution. It does not show any color reaction with iodine. We cannot digest cellulose because our body has no enzymes capable of catalyzing its hydrolysis. There is only a little bacterial metabolism of cellulose in the human colon.

However, some microorganisms found in soil and in the digestive tracts of certain animals produce enzymes that do catalyze the breakdown of cellulose. The presence of these microorganisms explains why cows and other herbivorous animals live on grass and why termites feed on wood.

(d) Other polysaccharides

Inulin is a polysaccharide of fructose found in tubers and roots of dahlias, artichokes, and dandelions. It is highly soluble in water and is used to determine the glomerular filtration rate. However, since it is not hydrolyzed by intestinal enzymes, it has no nutritional value. **Dextrins** are intermediates formed during the hydrolysis of starch.

Chitin is a structural polysaccharide found in the exoskeleton of crustaceans, insects, and fungi. It consists of N-acetyl-D-glucosamine units joined by β-1, 4 glycosidic bonds. Pectin is a partially methylated polysaccharide found in fruits. It is a polymer of galacturonic acid

residues linked with α-1, 4 glycosidic bonds with some galactose and/or arabinose branches.

Glycosaminoglycans or **mucopolysaccharides** form a part of the connective tissue. These polymers act as shock absorbers between bones and prevent them from rubbing against each other when we move. Osteoarthritis, the most common form of arthritis, occurs due to the loss of glycosaminoglycans at joints.

Complex polysaccharides are found on the surfaces of almost all cells. They serve as “labels” or antigens that allow organisms to distinguish their own cells from foreign substances, materials or pathogens. Antigen recognition forms the basis of a very important biochemical principle – namely, molecular shape carries information that guides the reactions of life.

Polysaccharides found on the surface of our red blood cells give rise to different blood types, which are commonly classified by the ABO system.

Proteins

Proteins are physically and functionally complex macromolecules that are heteropolymers of amino acids. Their very name, derived from the Greek word *proteios*, literally means “holding the first place” and signifies the importance of these molecules.

Proteins play diverse roles in the biological world; for example, the spider-web protein is many times stronger than the toughest steel. Hair, feathers, and hooves are composed of keratin. Crystalin constitutes the transparent lens material found in our eyes and is required for vision. Very small quantities of certain proteins missing from the blood can signify that one’s metabolic processes are going out of control.

Similarly, juvenile-onset diabetes mellitus occurs due to a lack of insulin. Dwarfism can be due to deficiency of the growth hormone, which is again a protein. A unique “antifreeze” blood protein allows Antarctic fish

to survive at body temperatures below freezing. You can clearly see why proteins are considered to be so important in the field of biochemistry.

Amino acids

Amino acids contain amino ($–NH_2$) and carboxyl (–COOH) groups and, depending upon the relative position of amino group with respect to the carboxyl group, can be classified as α, β, γ, δ, and so on. We only obtain α-amino acids after the hydrolysis of proteins. They may also contain other functional groups.

All α-amino acids have trivial names that usually reflect the property of that compound or its source. For example, **glycine** gets its name from the fact that it is sweet in taste (Greek: *glykos*, meaning “sweet”) and **tyrosine** because it was first obtained from cheese (Greek: *tyros,* meaning “cheese”). Amino acids are generally represented by a three-letter symbol, though a one-letter symbol is also used at times.

Amino acids may be **acidic, basic** or **neutral** depending on the relative number of amino and carboxyl groups in their structure. For example, amino acids that have an equal number of amino and carboxyl groups are neutral in nature. While those with more amino than carboxyl groups are basic and those with more carboxyl groups than amino groups are acidic.

The amino acids that can be synthesized in the body are known as nonessential amino acids. On the other hand, those which cannot be synthesized in our body and must be obtained through diet are known as **essential amino acids**.

Properties of amino acids

Physically, amino acids are generally colorless, crystalline solids. They are water-soluble, have high melting points, and and behave like salts rather than simple amines or carboxylic acids. This behavior is because of the presence of both acidic (carboxyl group) and basic (amino group) groups in the same molecule. Amino acids are **amphoteric** (or

amphiprotic) in nature; that is, they can react either as an acid or as a base.

In an aqueous solution, the carboxyl group can lose a proton (H^+) and the amino group can accept a proton, producing a dipolar ion known as the **Zwitter ion**. Although the Zwitter ion itself is overall neutral, it contains both positive and negative charges. In Zwitter ionic form, amino acids demonstrate amphoteric behavior because they react both with acids and bases.

When there are equal positive and negative charges on an amino acid in solution, it is electrically neutral and does not migrate either towards the positive or the negative electrode when placed in an electrolytic cell. The pH at which this phenomenon takes place is known as the **isoelectric point**.

Except for glycine, all other naturally occurring α-amino acids are **optically active** because the α-carbon atom is asymmetric. Optically active amino acids exist both in 'D' and 'L' forms. Most naturally occurring amino acids are observed to possess the L-configuration. L-amino acids are represented by writing the $–NH_2$ group on left hand side while drawing their structure.

Polypeptides

We've already discussed that proteins are polymers of α-amino acids that are connected to each other by peptide bonds/linkages. Chemically, a peptide bond is an amide formed between a –COOH group and an $–NH^2$ group. The reaction between two molecules of similar or different amino acids takes place by the combination of the amino group of one molecule with the carboxyl group of the other.

This reaction results in the elimination of a water molecule and the formation of a peptide bond (–CO–NH–). The product of the reaction is known as a **dipeptide** because it is composed of two amino acids. For

example, when the carboxyl group of glycine combines with the amino group of alanine, it yields a dipeptide known as *glycylalanine*.

Similarly, if a third amino acid combines with a dipeptide, the resulting product is known as a **tripeptide**. It contains three amino acids linked by two peptide bonds. Likewise, when four, five or six amino acids are linked together, their respective products are known as **tetrapeptides, pentapeptides** or **hexapeptides** respectively.

When the number of the reacting amino acids is more than ten, then the products are known as **polypeptides**. A polypeptide having more than a hundred amino acid residues and a molecular mass higher than 10,000u is known as a **protein**. However, keep in mind that the difference between a polypeptide and a protein is not very clear.

Polypeptides with fewer amino acids are often referred to as proteins if they tend to have a well-defined conformation like a protein. A good example is **insulin**, which contains 51 amino acids.

Structure of proteins

Depending on their molecular shape, proteins can be classified into two types:

(a) Fibrous proteins

A fiber-like structure is formed when the polypeptide chains run parallel to each other and are held together by hydrogen bonds and disulfide bonds. Fibrous proteins are usually insoluble in water. Some common examples of such proteins are **keratin** (found in hair, wool, and silk) and **myosin** (found in muscles).

(b) Globular proteins

The structure of a globular protein is produced when the chains of polypeptides coil around to give rise to a spherical shape. These proteins are generally soluble in water. **Insulin** and **albumin** are common examples of globular proteins.

We may study the structure and shape of proteins at four different levels - the *primary, secondary, tertiary* and *quaternary* levels. Each of these levels is more complex than the one preceding it.

(a) Primary structure

Proteins may possess one or more polypeptide chains. Every polypeptide in a protein features amino acids linked to each other in a specific sequence; it is this sequence of amino acids that constitutes the primary structure of that protein. Any change in the primary structure leads to the formation of a different protein.

Scientists imagine the primary structure of a protein as a line, with the left end represented by the first amino acid and the right end represented by the last amino acid. The first amino acid is also known as the **N-terminal amino acid**, whereas the last amino acid is known as the **C-terminal** amino acid.

(b) Secondary structure

The protein thread does not exist throughout its length as an extended rigid rod. It is folded in the form of a helix, very similar to a revolving staircase. The secondary structure of proteins refers to the different shapes in which a long polypeptide chain can exist.

These chains are found to exist in two different kinds of structures: the α-helix structure and the β-pleated sheet structure. These structures arise due to the regular folding of the backbone of the polypeptide chain by virtue of hydrogen bonding between the C=O and the –NH– groups of the peptide bond.

In an α-Helix, the polypeptide chain is often found to form all possible hydrogen bonds by twisting into a **right handed screw (helix)** with the –NH– group of each amino acid residue that is hydrogen bonded to the C=O of an adjacent turn of the helix.

In the β-pleated sheet structure, all the peptide chains are stretched out to more or less the upper limit of extension and subsequently laid side by side and held together by intermolecular hydrogen bonds. The

structure appears very much like pleated folds of drapery, and thus earns its characteristic name.

In natural proteins, we only get to observe right-handed helices.

(c) Tertiary structure

The tertiary structure of proteins represents overall folding of the polypeptide chains arising due to the further folding of the secondary structure. Here, the long protein chain is folded upon itself like a hollow woollen ball, giving us a **three-dimensional view** of the protein.

It leads to the development of the two major molecular shapes we've already discussed above - fibrous and globular. The main forces that stabilize the secondary and tertiary structures of proteins are **hydrogen bonds, disulphide linkages, van der Waals forces** and **electrostatic forces of attraction**.

The tertiary structure has been found to be absolutely essential for the numerous biological activities of proteins.

(d) Quaternary structure

Some proteins are composed of two or more individually folded polypeptide chains known as **subunits**. The spatial arrangement of these subunits with respect to each other (for example, a linear string of spheres, spheres stacked upon each other in the form of a cube or a plate, and so on) is known as the **quaternary structure** of the protein.

Adult human haemoglobin consists of four subunits, two of which are identical to each other. Thus, two subunits of α-type and two subunits of β-type together constitute the human haemoglobin (Hb) molecule.

Denaturation of proteins

Proteins found in a natural biological system, possessing a unique three-dimensional structure and biological activity, are known as native proteins. When a protein in its native form is subjected to physical

change such as a change in temperature, or chemical change such as a change in pH, the hydrogen bonds in its structure are disturbed.

Due to disruption, the protein globules unfold, the helix gets uncoiled, and protein subsequently loses its biological activity. This phenomenon is known as the **denaturation** of the protein. During denaturation, the secondary and tertiary structures of the protein are destroyed but its primary structure remains intact.

A common example of denaturation is the coagulation of egg white on boiling. Another commonly seen example is curdling of milk that is caused due to the formation of lactic acid by the *Lactobacilli* present in the milk.

Lipids

All lipids are composed of carbon, hydrogen, and a little oxygen. They are insoluble in water but readily soluble in organic solvents such as benzene, acetone, and ether. Like we discussed before, lipids aren't polymeric substances but are assembled from smaller molecules by dehydration.

Lipids could be simple fatty acids or glycerol. Several lipids are composed of both fatty acids and glycerol, while some of them also contain phosphorus and a phosphorylated organic compound in their structure. Certain lipids possess more complex structures as well.

Classification of lipids

Scientists have classified lipids into subgroups as follows:

Simple lipids

These are esters of fatty acids with various alcohols. They may be of two types – neutral/true fats or waxes.

(a) Neutral or true fats

They are also known as glycerides. Every fat molecule is an ester of one molecule of glycerol and one to three molecules of the same or different long-chain fatty acids. Glycerol is chemically trihydroxy propane, whereas a fatty acid molecule is an unbranched chain of carbon atoms that has a carboxylic group attached to an R group.

The R group could be a methyl ($-CH_3$) group, ethyl ($-C_2H_5$) group, or have a higher number of $–CH_2$ groups (one carbon to 19 carbons). For example, palmitic acid ($C_{16}H_{32}O_2$) has 16 carbon atoms including the carboxyl carbon. Arachidonic acid possesses 20 carbon atoms including the carboxyl carbon. Fatty acids are of two types:

- *Saturated fatty acids* like palmitic acid and stearic acid don't contain double bonds.
- *Unsaturated fatty acids* such as oleic acid, linoleic acid, linolenic acid, and arachidonic acid possess one or more double bonds in their structure.

Oils containing p**olyunsaturated fatty acids (PUFA)** or fatty acids with more than one double bond are recommended by doctors to people having hypertension, high blood cholesterol, and other cardiovascular diseases because they have been found to lower blood cholesterol. Sunflower oil and safflower oil are rich in polyunsaturated fatty acids.

Depending on the number of fatty acid molecules attached to a molecule of glycerol, neutral or true fats may be **monoglycerides** (one molecule), **diglycerides** (two molecules), or **triglycerides** (three molecules).

Based on their melting point, triglycerides can be known as **fats** or **oils**. Fats (such as butter and ghee) have high melting points and remain in the solid-state at room temperature. On the other hand, oils (such as sunflower oil and groundnut oil) have lower melting points and exist in the liquid state at room temperature.

(b) Waxes

Waxes are esters of fatty acids with alcohols of higher molecular weight other than glycerol. They play an important protective role in living

organisms. They form water-insoluble coatings on the hair in skin in animals and stems, fruits, and leaves in plants.

Bee wax is chemically hexacosyl palmitate and is formed from palmitic acid and myricyl alcohol. It is secreted from the abdominal glands of worker bees for building their hives. **Lanolin** or wool fat is found in the form of a waterproof coating around animal furs. *Mycobacterium tuberculosis* and *Mycobacterium leprae* produce a harmful wax known as **wax-D** that contributes to their pathogenicity.

Compound or conjugated lipids

These are esters of fatty acids with alcohol that also contain some other chemical groups. They are classified into the following subgroups:

(a) Phospholipids

They are composed of a molecule of glycerol or another alcohol having:

- A phosphate group joined to one of its outer –OH groups
- Two fatty acid molecules linked to the other two –OH groups
- A nitrogen-containing chlorine molecule bound to the phosphate group

Phospholipids such as lecithin are found in cell membranes.

(b) Glycolipids

They contain fatty acids, the alcohol sphingosine, and galactose. One of the fatty acid molecules is replaced by the galactose residue. Glycolipids are found in the myelin sheath of nerve fibers and in the outer membrane of the cell membrane of chloroplasts.

(c) Lipoproteins

As their name suggests, lipoproteins are found to have lipids (mostly phospholipids) and proteins in their molecular structures. They are the main constituents of membranes.

(d) Chromolipids

They contain colored pigments such as carotenoids (carotene, vitamin A, and more).

Derived lipids

Although **steroids** don't have fatty acid residues, they are considered lipids because of their fat-like properties. Instead of a straight-chain, they possess four fused carbon rings. Steroids differ from each other in terms of the number and position of double bonds between carbon atoms and in the side groups linked to the ring.

Sterols are the most common steroids found in nature, with the most well-known example being **cholesterol**. It is the most abundant steroid in animal tissues and is found abundantly in foods containing animal fats. It is synthesized in the liver and is an essential constituent of the animal cell membrane.

Prostaglandins are a category of hormone-like unsaturated fatty acids that act as messenger molecules between cells. They are derived from arachidonic acid and other similar C_{20} fatty acids. **Diosgenin** is a steroidal compound found in yams that is used to manufacture antifertility pills.

Nucleic acids

Nucleic acids are polynucleotides that, together with polysaccharides and polypeptides, form the true macromolecular fraction of any living tissue or cell. Along with proteins, they are the main constituents of chromosomes and are responsible for the transmission of characters from one generation to the next (a phenomenon known as **heredity**).

The building blocks of nucleic acids are **nucleotides** that have three chemically distinct components - a heterocyclic compound, a monosaccharide, and a phosphoric acid or phosphate residue. The heterocyclic compounds found in nucleic acids are nitrogenous bases called **adenine, guanine, uracil, cytosine**, and **thymine**.

Adenine and guanine are substituted purines, whereas the others are substituted pyrimidines. The sugar found in polynucleotides is either β-D-ribose or β-D-2-deoxyribose, which we've already discussed before. A nucleic acid containing deoxyribose is known as deoxyribonucleic acid (DNA) while one that contains ribose is known as ribonucleic acid (RNA).

We'll now delve into the expansive world of nucleic acids and study about these fascinating molecules in detail.

Purines, pyrimidines, and nucleosides

To begin our study of nucleic acids, we must learn more about a crucial part of these molecules: two classes of heterocyclic bases known as the **purines** and the **pyrimidines**. Pyrimidines are six-membered heterocyclic rings whereas purines contain both a five-membered and a six-membered ring. Because of the presence of nitrogen atoms in the ring, these compounds are known as **heterocyclics** and also as bases.

In nature, we commonly find five important bases in nucleic acids—two purine bases (adenine and guanine) and three pyrimidine bases (cytosine, thymine, and uracil). Some "modified" versions of these bases, such as 5-fluorouracil, are used in cancer chemotherapy. The natural bases differ from each other in their ring substituents.

Every base has a lowermost nitrogen that is bonded to a hydrogen as well as two carbons. This particular –NH– group shares some chemical similarities with an alcohol (-OH) group. Just like two sugar molecules can be linked when an –OH group of one monosaccharide reacts with a second monosaccharide, a purine or pyrimidine molecule can be bonded to a sugar molecule by a reaction with the –NH– group.

A nucleoside is formed when either a purine or pyrimidine base is linked to a sugar molecule, usually β-D-ribose or β-D-2-deoxyribose. The base and sugar are bonded together between C-1 of the sugar and either the purine nitrogen at position 9 or the pyrimidine nitrogen at position 1 by eliminating a water molecule.

The nomenclature of each nucleoside underlines the importance of the base to the chemistry of the molecule. For example, adenine and

β-D-ribose react to yield **adenosine** whereas cytosine and β-D-2 -deoxyribose yield **deoxycytidine**.

Nucleotides

Nucleotides are chemically phosphate esters of nucleosides. They consist of a purine or a pyrimidine base linked to a sugar, which itself is bonded to at least one phosphate group. The ester linkage may be formed with the hydroxyl group at position 2, 3, or 5 of ribose or at position 3 or 5 of deoxyribose. The linkage of two or more phosphate residues results in the formation of a high-energy phosphate anhydride bond.

Nucleotides play a crucial role in the transfer of energy during many metabolic processes. **Adenosine diphosphate (ADP)** and **adenosine triphosphate (ATP)** are extremely important nucleotides that store and release energy to our body's cells and tissues. Part of the energy released from the biological oxidation of the food we eat is stored in the phosphate anhydride bonds of ADP and ATP.

Our body releases energy when required by reversibly hydrolysing the high-energy phosphate anhydride bonds in ADP and ATP. During the hydrolysis, there is a yield of about 35 kJ of energy per mole of ATP. All natural processes such as muscle movement, nerve impulse conduction, vision, and even the maintenance of our heartbeats are dependent on energy obtained from ATP.

Deoxyribonucleic acid (DNA)

DNA is the very molecule that started the dance of life on earth as we know it today. It is a long polymer of deoxyribonucleotides whose length is usually defined as number of nucleotides (or a pair of nucleotides referred to as base pairs) present in it. The haploid content of human DNA is **3.3×10^9 base pairs**.

DNA was first discovered by Friedrich Meischer (1869), who named it "nuclein" as he observed that it was an acidic substance present in the nucleus. In 1953, **James Watson and Francis Crick**, based on the X-ray diffraction data put forth by Maurice Wilkins and Rosalind Franklin, proposed a simple but world famous **double helical model** for the structure of DNA that subsequently earned them the Nobel Prize.

One of the most important features of their model was the base pairing present between the two strands of polynucleotide chains. This model was also based on the observation of Erwin Chargaff that for a double-stranded DNA molecule, the ratios between adenine and thymine and guanine and cytosine are constant and equals one (known as **Chargaff's rule**).

Important features of Watson and Crick's model of DNA are as follows:

- It is made of two polynucleotide chains whose backbone is composed of sugar and phosphate, and the nitrogenous bases project inside.
- The two chains have anti-parallel polarity, meaning if one chain has the polarity 5'→3', then the other chain has the polarity 3'→5' and vice versa.
- The nitrogenous bases in the two strands are paired via hydrogen bonds (H-bonds) and form base pairs (bp). Adenine forms two hydrogen bonds with thymine from the opposite strand, and vice-versa. Similarly, guanine is bonded with cytosine via three H-bonds. As a result, a purine is always seen opposite to a pyrimidine. This leads to the creation of an approximately uniform distance between the two strands of the helix.
- The two chains are coiled in a **right-handed fashion**. The pitch of the helix is 3.4 nm and there are about 10 bp in each turn. Consequently, the distance between a bp in a helix is around 0.34 nm.
- The plane of one base pair stacks over the other in the double helix. Apart from the H-bonds, this feature provides additional stability to the helical structure of DNA.

Packaging of the DNA molecule within the nucleus

The length of a DNA molecule is calculated to be approximately 2.2 metres – a length that is far greater than the dimension of a typical cell nucleus (approximately 10^{-6} m). Natually, you might be wondering how such a long polymer has been packaged in such a small space.

The answer, dear reader, is a true miracle of nature. Even in prokaryotes (which lack a well-defined nucleus), the DNA is not scattered throughout the cell. Being negatively charged, the DNA is organised in large loops

held by some positively charged proteins in a region known as the **nucleoid**.

Eukaryotes have a more complex organisation in the form of a set of positively charged, basic proteins known as **histones**. These proteins are rich in two basic amino acid residues, lysine and arginine, that carry positive charges in their side chains. They are organised to form a unit of eight molecules known as a **histone octamer**.

The negatively charged DNA molecule is wrapped around the positively charged histone octamer to form a structure known as the **nucleosome**. A typical nucleosome contains 200 base pairs of the DNA helix. Nucleosomes constitute the repeating unit of chromatin that we see in the nucleus, which have a characteristic **"beads-on-string"** appearance when viewed under the electron microscope.

This beads-on-string structure is packaged to form chromatin fibres that are further coiled and condensed at the metaphase stage of cell division to form chromosomes. The packaging of chromatin at higher level requires an additional set of proteins that are collectively referred to as **non-histone chromosomal (NHC) proteins**.

Different forms of DNA

Scientists have discovered more than a dozen forms of DNA with unique structural features in nature, named after various English alphabets. Some of them are:

- **B-DNA** – It is the regular DNA molecule with right-handed coiling and 10 base pairs per turn.
- **A-DNA** – It has 11 base pairs per turn. The base pairs aren't perpendicular to the axis but are tilted.
- **C-DNA** – It resembles B-DNA but has 9 base pairs per turn.
- **D-DNA** – It is similar to B-DNA but has 8 base pairs per turn.
- **Z-DNA** – Unlike the other examples above, it demonstrates **left-handed coiling**.

Ribonucleic acid (RNA)

RNA is usually single-stranded and occasionally double-stranded (like in the rice dwarf virus and reovirus). Scientists have enough evidence to prove that RNA was the first genetic material on earth; essential life processes such as metabolism, translation, and splicing evolved around RNA. It served as a genetic material as well as an enzymatic catalyst in primitive living systems.

However, RNA – being a catalyst – was reactive and thus unstable in nature. As a result, DNA subsequently evolved from RNA with chemical modifications (such as the presence of thymine instead of uracil) and special repair mechanisms that render it more stable and a better choice to serve as the genetic material in living organisms. Today, RNA serves as the genetic material in certain groups of viruses.

There are three types of non-genetic RNA found in living systems:

(a) Messenger RNA (mRNA)

Discovered by Jacob and Monod in 1961, mRNA is produced in the nucleus and harbors the genetic information for protein synthesis.

(b) Ribosomal RNA (rRNA)

It is the largest type of RNA and forms around 80% of the entire cellular RNA. It is found in ribosomes where protein synthesis takes place.

(c) Transfer RNA (tRNA)

Also known as **soluble RNA** or **adaptive RNA**, tRNA is the smallest type of RNA and forms around 10-15% of the total cellular RNA. tRNA molecules are found in the cytoplasm and are of as many types as the types of amino acids found in proteins (usually 20). They collect amino acids from the cytoplasm for protein synthesis during the process of translation.

Enzymes

Life is possible by virtue of the coordination of various chemical reactions that constantly take place in living organisms, such as the digestion of food, absorption of necessary molecules, and production of energy. This process involves a sequence of reactions that take place within the body under very mild conditions. All of this occurs with the help of certain biocatalysts known as **enzymes.**

The term *enzyme* was first coined by Wilhelm Kuhne in 1877, which roughly means “in yeast” in Greek (referring to the fermentation of sucrose to alcohol by zymase found in yeasts). In 1926, Sumner first crystallized the enzyme **urease** in pure form from jack beans (*Canavalia ensiformis*). For this work, he was awarded the Nobel Prize in Chemistry in 1946 along with Stanley and Northrop.

Almost all the enzymes have been found to be globular proteins in nature, except two recently discovered RNA enzymes – **ribozyme** and **ribonuclease-P**. These proteins, known as **ribozymes**, were discovered by Thomas Cech and Sidney Altman for which they were awarded the Nobel Prize in Chemistry (1989).

Structure of an enzyme

Like any other protein, an enzyme possesses the primary, secondary, and tertiary structures. When we observe the tertiary structure of an enzyme, we notice that the backbone of the protein chain folds upon itself, criss-crosses itself, and leads to the formation of many crevices or pockets.

One such pocket is the **active site** into which the substrate fits during a catalysis reaction. With the help of their active site, enzymes catalyse reactions at a high rate. Unlike inorganic catalysts that you must’ve read about in chemistry, enzymes cannot usually work efficiently at high temperatures; being proteins, they are denatured at high temperatures (on average, above 40°C).

However, special enzymes isolated from organisms who normally live under extremely high temperatures, such as hot vents and sulphur springs, are stable and retain their catalytic power even at high

temperatures (around 80° to 90°C). Thermal stability is an important virtue of such enzymes isolated from thermophilic organisms, and they are extensively used in biotechnological processes.

Properties of enzymes

Enzymes have certain distinct properties that have been enlisted below:

1. Most enzymes are chemically **proteins**. They may possess additional inorganic or organic groups known as **co-factors** that are necessary for their biological activity (discussed later).
2. Like inorganic catalysts, enzymes don't actually start a chemical reaction but increase the rate of an ongoing reaction by *decreasing the magnitude of the activation energy*. They don't change the equilibrium but help in the equilibrium state being attained faster.
3. The number of substrate molecules changed per minute by an enzyme molecule is known as the **turnover number**. The higher the turnover number, the higher is the enzyme's efficiency.
4. Being catalysts, enzymes aren't transformed or used up in the chemical reaction. They are released unchanged at the end of the reaction.
5. Unlike inorganic catalysts, enzymes are **highly specific in their action for a particular reaction and for a particular substrate**. For example the enzyme maltase acts only on the sugar maltose, not on sucrose or lactose.

How do enzymes catalyse biochemical reactions?

To understand how enzymes work, we must first take a closer look at chemical reactions in general. Chemical compounds can potentially undergo two types of changes –physical and chemical. A **physical change** simply refers to a change in physical shape without involving any breaking of chemical bonds. This also includes a change in state of matter, such as ice melting or water turning into vapour.

On the other hand, when bonds are broken and new bonds are formed during the transformation, a chemical reaction is said to take place. The hydrolysis of starch into glucose is an example of an organic chemical reaction. Given below is an example of an inorganic chemical reaction.

$Zu + CuSO_4 \rightarrow ZnSO_4 + Cu$

The **rate** of a physical or chemical process refers to the **amount of product formed per unit of time**. It can be expressed in the following manner:

$$\text{Rate} = \frac{\delta P}{\delta t}$$

If the direction happens to be specified, then the rate may also be known as the **velocity**. Several factors, including the temperature, influence the rates of physical and chemical processes. A general rule of thumb is that the rate gets doubled or halved for every change of 10°C in either direction.

Catalyzed reactions have been observed to proceed at much higher rates than the ones that aren't. When we study enzyme-catalyzed reactions, their rate is found to be much higher than the same reaction sans the catalyst.

Let's consider a relatively simple reaction – carbon dioxide getting dissolved in water to produce carbonic acid (H_2CO_3). In the absence of an enzyme, this reaction proceeds very slowly with only about 200 molecules of carbonic acid being formed every hour. However, in the presence of an enzyme named **carbonic anhydrase**, the reaction speeds up nearly 10 million times with about 600,000 molecules being formed every second!

In our body, there are thousands of different types of enzymes to be found. Each of these enzymes catalyzes a unique biochemical reaction. A multistep chemical reaction where each of the steps is catalyzed by the same enzyme complex or different enzymes is known as a **metabolic pathway**. For example,

The oxidation of glucose to pyruvic acid through ten different enzymes catalyzed metabolic reactions is a good example of a metabolic pathway. With one or two additional reactions, this same pathway can give rise to a range of diverse metabolic end products. For example, lactic acid is formed in our skeletal muscles under anaerobic conditions while ethyl alcohol is produced in yeast cells during fermentation. Hence, different products can possibly be formed in different conditions.

Mechanism of enzyme action

We have studied the concept of an “active site” in detail now. The chemical or metabolic conversion refers to a reaction, whereas the chemical being converted into a product is known as the **substrate**. Let us consider a hypothetical enzyme converting a substrate (S) into a product (P), in a reaction that may be depicted as:

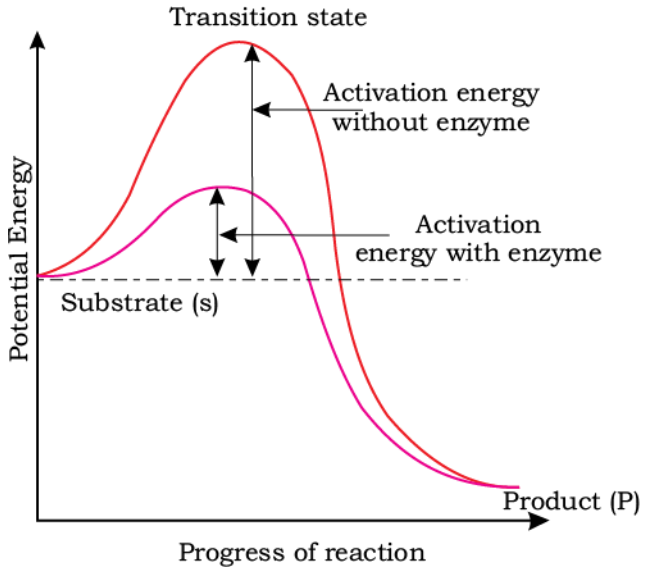
S➔P

Obviously, you must’ve understood that the substrate S needs to bind to the enzyme at its active site within a particular cleft or pocket. Because the substrate has to diffuse towards the active site, there is an inevitable formation of an **ES complex**. Here, “E” stands for the enzyme.

This complex formation is only a transient phenomenon. When the substrate is bound to the active site of the enzyme, a new structure of the substrate known as the **transition state** is formed. Soon afterward, the structure of the substrate gets transformed into that of the product and is released from the active site.

The pathway of this transformation must inevitably go through this transition state structure at some point or the other. There could be many more intermediate “altered structural states” between the stable substrate and the product, all of which are unstable.

Stability is depended greatly upon the energy status of the molecule or the structure that we’re dealing with. We can understand this concept better using the graph shown below:



Here, the y-axis represents the **potential energy content** whereas the x-axis represents the **progression of the reaction**. I would like you to take note of two things here: firstly, the energy level difference between S and P. If P is at a lower level than S, the reaction is an **exothermic reaction**. In this case, we don't need to supply energy (in the form of heat) in order to get the product.

However, in the case of both exothermic and endothermic reactions, the substrate S has to go through a much higher energy state or transition state. The difference in the average energy content of S from that of this transition state is known as the **activation energy**.

Enzymes work by bringing down this energy barrier and making the transition of S to P much easier. The ES complex is short-lived and soon dissociates into its product(s) P and the unaltered enzyme. The formation of the ES complex is absolutely essential for catalysis.

Models for the mode of enzyme action

Scientists have put forward three different models to explain the mode of enzyme action.

Lock and key hypothesis

It was proposed by **Emil Fischer** in 1894. According to this hypothesis, According to this model, the enzyme has a rigid structure or conformation. The substrate fits to the active site of the enzyme just like

a key only fits into the proper lock. The active site contains special groups having $–NH_2$, -COOH, and –SH moieties to establish contact with substrate molecules.

After the substrate molecules come in contact with the active site of the enzyme, they undergo chemical changes and form products. Because the product no longer fits into the active site, it is released to the surrounding medium and leaves the active site free to receive more substrate molecules. Essentially, the hypothesis states that active site of an enzyme is a rigid and pre-shaped template where only a specific substrate can bind.

This model demonstrates how a small concentration of an enzyme can act upon a large amount of the substrate. It explains why the enzyme remains unaffected at the end of the reaction. It also explains how competitive inhibitors possessing a structure similar to the substrate inhibit the enzyme.

However, the lock and key model does not adequately explain the flexible nature of enzymes. Thus, it fails to explain many features of enzymatic reactions such as the effect of allosteric modulators.

Induced fit hypothesis

In 1958, **Koshland** proposed a more acceptable and realistic model for enzyme-substrate complex formation. According to this hypothesis, the active site of the enzyme doesn't initially exist in a shape complementary to the substrate. Rather, it is induced to assume the said shape when the substrate is bound to the enzyme.

In other words, the active site is induced to assume a complementary shape much like a hand induces a change in the shape of a glove. Thus, the model asserts that enzymes or their active sites are flexible in nature. The active site of the enzyme possesses two groups –

(a) A buttressing group for supporting the substrate.
(b) A catalytic group for catalysing the reaction.

When the substrate comes in contact with the buttressing group, the active site becomes modified to bring the catalytic group opposite to the substrate bonds, which are then broken.

The induced fit model is supported by strong experimental evidence from X-ray diffraction studies. It also explains the action of allosteric modulators and competitive inhibitors on enzymes.

Substrate strain theory

In the substrate strain theory, the substrate is strained due to the induced conformation change in the enzyme. Also, it is possible that when a substrate binds to the preformed active site, the enzyme induces subsequently a strain to the substrate that leads to the formation of the product.

Some scientists feel that a combination of the induced fit model with the substrate strain is operative in the mechanism of enzymatic action.

Catalytic cycle of enzyme action

We may describe the catalytic cycle of enzyme action in the following steps:

1. First, the substrate binds to and fits into the active site of the enzyme.
2. The binding of the substrate stimulates the enzyme to change its shape and fit more tightly around the substrate.
3. The active site of the enzyme is now in close proximity of the substrate and breaks its chemical bonds of the substrate to form the new enzyme- product complex.
4. Finally, the enzyme releases the products of the reaction. The free enzyme is then ready to bind to another molecule of the substrate and run through the catalytic cycle once again.

Factors affecting enzyme activity

When we're speaking of enzyme activity, you must keep in mind that most enzymes are chemically proteins. Do you remember what we discussed about the tertiary structure of a protein is vital for its biological functions? With this concept in mind, it is easy to see why the activity of

an enzyme may be affected by a change in the conditions that can disrupt the tertiary structure of its protein chains.

These conditions include the temperature, pH, change in substrate concentration, or the binding of certain chemicals that regulate its activity. Let us understand more about these factors in detail.

Temperature and pH

Enzymes usually function in a narrow range of temperature and pH. Scientists have discovered that every enzyme demonstrates its highest level of activity at a particular temperature and pH known as the **optimum temperature** and **optimum pH**. The enzyme activity steadily declines both below and above the optimum value.

At low temperatures, the enzyme is preserved in a temporarily inactive state and regains its lost activity when the temperature is raised to normal. On the other hand, high temperature destroys enzymatic activity completely because the enzyme, being a protein, gets denatured by heat. Once the enzyme protein gets denatured, it remains inactive even if the temperature is brought down later.

Similarly, a rise or fall in pH above or below the pH reduces enzyme activity. Some enzymes act best in an acidic medium and others do so in an alkaline medium.

Concentration of substrate

With an increase in substrate concentration, the velocity of the enzymatic reaction rises at first. The reaction ultimately attains a peak value of velocity **(V_{max})** which is not exceeded by any further rise in the concentration of the substrate. The reason behind this phenomenon is that the enzyme molecules are fewer in number than the substrate molecules. After saturation of these molecules, there are no free enzyme molecules left to bind with the additional substrate molecules.

K_m or the **Michaelis constant** is a mathematical derivation or constant that indicates the substrate concentration at which the chemical reaction catalyzed by an enzyme attains half its maximum velocity. In other words,

it is the concentration of the substrate at which half of the maximum velocity of the enzyme action is attained.

Inhibition of enzyme activity

Any substance that can bring down the velocity of an enzyme-catalyzed reaction is known as an **inhibitor**. **Reversible inhibitors** bind to enzymes via non-covalent bonds. Subsequent dilution of the enzyme-inhibitor complex leads to the dissociation of the reversibly-bound inhibitor and the enzyme regains its activity.

On the other hand, irreversible inhibition takes place when an inhibited enzyme fails to regain its activity upon dilution of the enzyme-inhibitor complex. Certain **irreversible inhibitors** act by forming covalent bonds with specific groups of enzymes. For example, a class of insecticides known as organophosphates produces neurotoxic effects by irreversibly binding to the catalytic site of the enzyme **acetylcholinesterase**.

We'll now understand the concepts of reversible, irreversible, and allosteric inhibition in detail.

Reversible inhibition

In reversible inhibition, the inhibitor binds non-covalently to the enzyme. Furthermore, the enzyme inhibition can be reversed if the inhibitor is removed. Reversible inhibition is further sub-divided into the following categories:

(i) Competitive inhibition
(ii) Non-competitive inhibition

(a) Competitive inhibition

In this type of inhibition, the inhibitor closely resembles the real substrate structurally and is regarded as a **substrate analogue**. It binds reversibly to the same site on the enzyme that the substrate normally occupies by competing with it for the same. The effect of a competitive inhibitor is reversed by increasing the concentration of the substrate [S].

There is no effect on V_{max} during competitive inhibition; at a sufficiently high [S], V_{max} is attained in the absence of the inhibitor. However, competitive inhibitors increase the apparent K_m for the given substrate. Thus, in the presence of a competitive inhibitor, more substrate is required to achieve ½ V_{max}. Examples of competitive inhibition include:

- the inhibition of succinic dehydrogenase by malonate and oxaloacetate
- the inhibition of alcohol dehydrogenase by ethanol in methanol poisoning
- the control of bacterial pathogens by **sulpha drugs** that compete with the substrate P-amino benzoic acid (PABA) for the active site of the enzyme

(b) Non-competitive inhibition

Non-competitive inhibition takes place when the inhibitor and substrate bind at different sites of the enzyme. In here, the inhibitor can bind to either the free enzyme or the ES complex, thus preventing the reaction from taking place. The inhibitor has no structural resemblance to the substrate in this case.

Unlike competitive inhibitors, non-competitive inhibitors decrease the V_{max} of the reaction. Thus, we cannot overcome non-competitive inhibition by increasing the [S]. Since these inhibitors don't interfere with the binding of the substrate to the enzyme, the K_m remains unaltered.

Heavy metal ions such as Ag^{+}, Pb^{2+}, and Hg^{2+} can non-competitively inhibit enzymes by binding to cysteinyl sulfhydryl groups.

Irreversible inhibition

Here, the inhibitor binds covalently to the enzyme and inactivates it in an irreversible manner. These inhibitors are usually toxic substances that poison enzymes. For example, **iodoacetate** is an irreversible inhibitor of certain enzymes such as **papain** and **glyceraldehyde 3-phosphate dehydrogenase**.

Iodoacetate combines with sulfhydryl (-SH) groups at the active sites of these enzymes and renders them inactive. Another good example of non-competitive inhibition is **cyanide** that can kill animals by inhibiting the enzyme **cytochrome oxidase**.

Suicide inhibition is a specialized form of irreversible inhibition where the original inhibitor (the structural analogue or competitive inhibitor) is converted to a more potent form by the same enzyme it was supposed to inhibit. The new inhibitor formed as a result binds irreversibly with the enzyme (in contrast to the original inhibitor that binds reversibly).

A good example of suicide inhibition is **allopurinol**, a drug that inhibits the enzyme **xanthine** oxidase. It is used in the treatment of gout.

Allosteric inhibition

Some of the enzymes possess additional sites besides the active site, known as **allosteric sites**. Such enzymes are known as **allosteric enzymes**. Certain substances referred to as allosteric modulators (effectors or modifiers) bind to the allosteric site and regulate the enzyme activity.

When a positive (+) allosteric effector binds to the allosteric site (known as the *activator site*), the enzyme activity is increased. On the other hand, a negative (–) allosteric effector can inhibit the enzyme activity by binding to the allosteric site (called *inhibitor site* in this case).

For example, **glucose-6-phosphate** is an allosteric inhibitor of the enzyme **hexokinase**. It plays an important role in feedback regulation during glycolysis.

Isoenzymes

Isoenzymes are multiple molecular forms of an enzyme, synthesized by different genes, occurring in the same organism, and having a similar substrate activity. More than a hundred known enzymes have been found to have isoenzymes. For example:

- α-amylase of wheat endosperm has **16 isoenzymes**.
- Lactic acid dehydrogenase has **5 isoenzymes**.
- Alcohol dehydrogenase has **4 isoenzymes**.

Classification of enzymes

With years and years of dedicated effort, biochemists have discovered, isolated, and studied thousands of enzymes. Most of these enzymes have been classified into different groups based on the type of reactions they catalyse. Today, enzymes are divided into six different classes, with each class further being divided into 4-13 subclasses and named accordingly by a four-digit number, which we will discuss later.

(i) Oxidoreductases/dehydrogenases

These are enzymes that catalyse simultaneous oxidation and reduction (oxidoreduction) reactions between two substrates S and S' as shown below:

$S_{reduced} + S'_{oxidised} \rightarrow S_{oxidised} + S'_{reduced}$

(ii) Transferases

They are enzymes catalysing a transfer of a group G (other than hydrogen) between a pair of substrates S and S' as seen below:

$S - G + S' \rightarrow S + S' - G$

Examples of oxidoreductases include transaminases (transfer amino groups) and kinases (catalyse the phosphorylation of substrates by transferring phosphate groups, generally from ATP).

(iii) Hydrolases

They are enzymes that catalyse the hydrolysis of ester, ether, peptide, glycosidic, C-C, C-X or P-N bonds. They facilitate the breakdown of larger molecules into smaller molecules with the addition of water.

Examples of hydrolases include amylases, proteases, lipases, nucleases, maltase, invertase, and other digestive enzymes. They are abundantly found in lysosomes.

(iv) Lyases

These are enzymes that catalyze the cleavage of substrates into two parts without involving hydrolysis for the removal of groups. They leave behind a double bond at the place from where the group was removed.

Aldolase, carbonic anhydrase, and decarboxylase are examples of lyases.

(v) Isomerases

This class includes all enzymes that catalyze the rearrangement of the molecular structure of the substrate to form isomers. They facilitate the interconversion of optical, geometric, or positional isomers. Isomerase is a classic example of this class.

(vi) Ligases

Finally, ligases are enzymes that catalyze covalent bonding of two substrates to form a large molecule. They facilitate the joining of bonds such as C-O, C-S, C-N, and P-O with the help of energy obtained from ATP.

Examples of ligases are phosphoenolpyruvate carboxylase, RUBP carboxylase, and DNA ligase.

Exoenzymes and endoenzymes

Most enzymes remain and function within the cell, and are known as **intracellular enzymes** or **endoenzymes**. Some of them are dissolved in the cytoplasmic matrix, while others are bound to particles such as mitochondria, ribosomes, and chloroplasts. For example, respiratory enzymes that are required to convert lactic acid to CO_2 and water are present in the mitochondria.

However, certain enzymes leave the cell and function outside them. They are thus known as **extracellular enzymes** or **exoenzymes**. The main examples of such enzymes are digestive enzymes like salivary amylase, pepsin, and pancreatic lipase. It is worth noting that enzymes retain their catalytic activity even after they are extracted from the cells. This property has given them many commercial applications in various industries.

Nomenclature of enzymes

In 1961, the International Union of Biochemistry (IUB) appointed an Enzyme Commission (EC) to devise some basic principles for the

classification and nomenclature of enzymes. The IUB system has divided enzymes into six major classes that we've discussed above. Each of these classes is subdivided into many subclasses, which are further divided into sub-subclasses. Finally, a four-digit Enzyme Commission (EC) number is assigned to each enzyme.

In the each number, the significance of each digit is as follows:

1st digit – Class

2nd digit – Subclass

3rd digit – Sub-subclass

4th digit – Individual enzyme

Co-factors

Like other proteins, enzymes are composed of one or many polypeptide chains. However, scientists have observed that in a lot of instances, certain non-protein constituents known as **co-factors** are bound to the enzyme in order to render it catalytically active. In these cases , the protein portion of the enzymes is known as the **apoenzyme**

Biochemists have identified three different types of co-factors:

- **Prosthetic groups**
- **Co-enzymes**
- **Metal ions**

Prosthetic groups are organic compounds that, unlike the other two co-factors, are *tightly bound* to the apoenzyme. For example, in peroxidase and catalase (enzymes that catalyze the breakdown of H_2O_2 to water and oxygen), **haem** is the prosthetic group and it constitutes a part of the active site of the enzyme.

Co-enzymes are organic compounds like prosthetic groups, but their association with the apoenzyme is only transient. It usually takes place during the course of catalysis. Moreover, co-enzymes serve as co-factors in several different enzymes catalyzed reactions. The vital chemical components of many coenzymes are vitamins. For example, the

coenzymes nicotinamide adenine dinucleotide (**NAD**) and nicotinamide adenine dinucleotide phosphate (**NADP**) contain the vitamin niacin.

A large number of enzymes require **metal ions** for their biological activity. These ions form coordination bonds with side chains at the active site, and simultaneously also form one or more coordination bonds with the substrate. For example, **zinc** is an important cofactor for the proteolytic enzyme carboxypeptidase.

The enzyme's catalytic activity is found to be lost if the co-factor is removed from the enzyme. This proves that they play a most pivotal role in the catalytic activity of the enzyme.

Given below is a list of important metal ions and the metal-containing enzymes (metalloenzymes) they are found in.

Metal ion(s)	**Metalloenzymes**
Fe^{2+}, Fe^{3+}	Cytochrome oxidase, aconitase, catalase, peroxidase
Ca^{2+}	Succinate dehydrogenase, lipase
Mg^{+2}	Hexokinase, DNA polymerase, pyruvate kinase, phosphotransferase, enolase
Cu^{2+}	Cytochrome oxidase, tyrosinase
Co^{2+}	Peptidases, ascorbic acid oxidase
Mo	Nitrate reductase, dinitrogenase
Zn^{2+}	Carbonic anhydrase, Alcohol dehydrogenase, carboxypeptidase, LDH
Mn^{2+}	Arginase
Ni	Urease
Cl^{-}	Salivary amylase