

# BIOCHEMISTRY

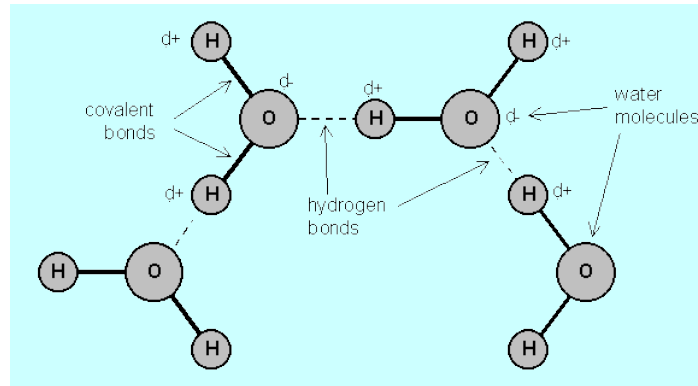
At least 80% of the mass of living organisms is water, and almost all the chemical reactions of life take place in aqueous solution. The other chemicals that make up living things are mostly organic macromolecules belonging to the 4 groups proteins, nucleic acids, carbohydrates or lipids. These macromolecules are made up from specific monomers as shown in the table below. Between them these four groups make up 93% of the dry mass of living organisms, the remaining 7% comprising small organic molecules (like vitamins) and inorganic ions.

GROUP NAME	MONOMERS	POLYMERS	% DRY MASS
Proteins	amino acids	polypeptides	50
nucleic acids	nucleotides	polynucleotides	18
carbohydrates	monosaccharides	polysaccharides	15

GROUP NAME	COMPONENTS	LARGEST UNIT	% DRY MASS
lipids	fatty acids + glycerol	Triglycerides	10

# WATER

Water molecules are charged, with the oxygen atom being slightly negative and the hydrogen atoms being slightly positive. These opposite charges attract each other, forming hydrogen bonds. These are weak, long distance bonds that are very common and very important in biology.



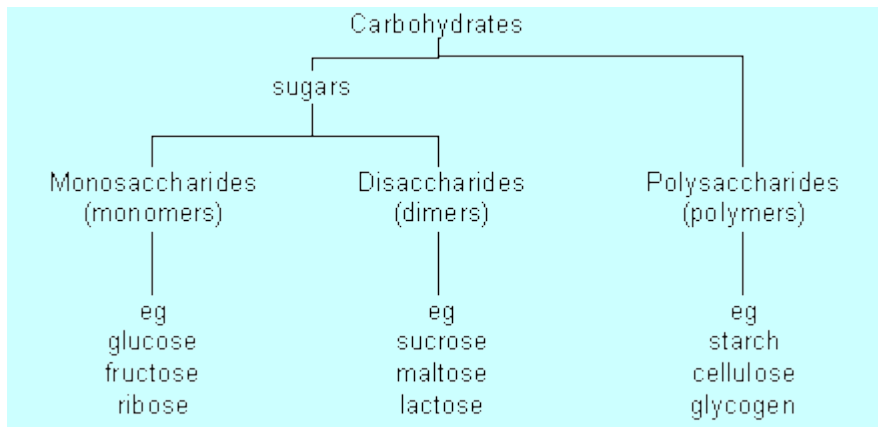
Water has a number of important properties essential for life. Many of the properties below are due to the hydrogen bonds in water.

- **Solvent.** Because it is charged, water is a very good solvent. Charged or polar molecules such as salts, sugars and amino acids dissolve readily in water and so are called hydrophilic ("water loving"). Uncharged or non-polar molecules such as lipids do not dissolve so well in water and are called hydrophobic ("water hating").
- **Specific heat capacity.** Water has a specific heat capacity of  $4.2 \text{ J g}^{-1} \text{ }^\circ\text{C}^{-1}$ , which means that it takes 4.2 joules of energy to heat 1 g of water by  $1^\circ\text{C}$ . This is unusually high and it means that water does not change temperature very easily. This minimizes fluctuations in temperature inside cells, and it also means that sea temperature is remarkably constant.
- **Latent heat of evaporation.** Water requires a lot of energy to change state from a liquid into a gas, and this is made use of as a cooling mechanism in animals (sweating and panting) and plants (transpiration). As water evaporates it extracts heat from around it, cooling the organism.
- **Density.** Water is unique in that the solid state (ice) is less dense than the liquid state, so ice floats on water. As the air temperature cools, bodies of water freeze from the surface, forming a layer of ice with liquid water underneath. This allows aquatic ecosystems to exist even in sub-zero temperatures.
- **Cohesion.** Water molecules "stick together" due to their hydrogen bonds, so water has high cohesion. This explains why long columns of water can be sucked up tall trees by transpiration without breaking. It also explains surface tension, which allows small animals to walk on water.
- **Ionization.** When many salts dissolve in water they ionize into discrete positive and negative ions (e.g.  $\text{NaCl} \rightarrow \text{Na}^+ + \text{Cl}^-$ ). Many important biological molecules are weak acids, which also ionize in solution (e.g. acetic acid  $\rightarrow$  acetate $^-$  +  $\text{H}^+$ ). The names of the acid and ionized forms (acetic acid and acetate in this example) are often used loosely and interchangeably, which can cause confusion. You will come across many examples of two names referring to the same substance, e.g.: phosphoric acid and phosphate, lactic acid and lactate, citric acid and citrate, pyruvic acid and pyruvate, aspartic acid and aspartate, etc. The ionized form is the one found in living cells.
- **pH.** Water itself is partly ionized ( $\text{H}_2\text{O} \rightleftharpoons \text{H}^+ + \text{OH}^-$ ), so it is a source of protons ( $\text{H}^+$  ions), and indeed many biochemical reactions are sensitive to pH ( $-\log[\text{H}^+]$ ). Pure water cannot buffer changes in  $\text{H}^+$  concentration, so is not a buffer and can easily be any pH, but the cytoplasm and tissue fluids of living organisms are usually well buffered at about neutral pH (pH 7-8).

# CARBOHYDRATES

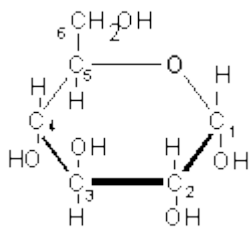


Carbohydrates contain only the elements carbon, hydrogen and oxygen. The group includes monomers, dimers and polymers, as shown in this diagram:

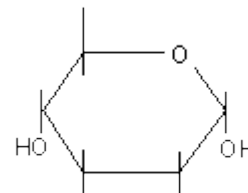


## Monosaccharides

All have the formula  $(\text{CH}_2\text{O})_n$ , where  $n$  is between 3 and 7. The most common & important monosaccharide is glucose, which is a six-carbon sugar. Its formula is  $\text{C}_6\text{H}_{12}\text{O}_6$  and its structure is shown below



or more simply



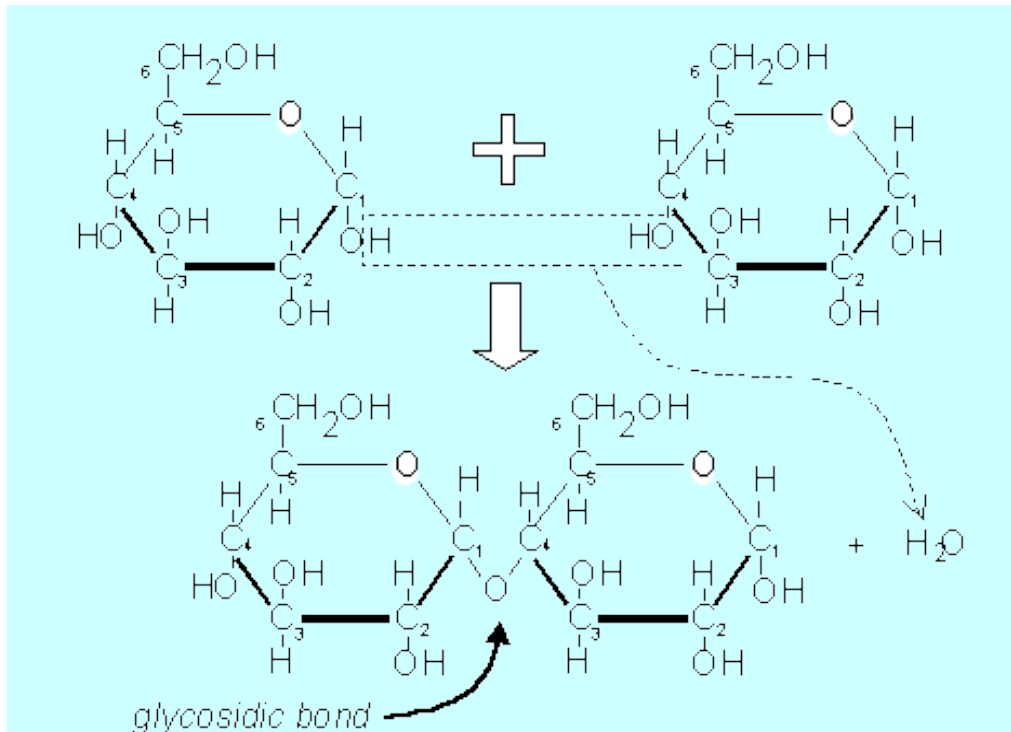
Glucose forms a six-sided ring. The six carbon atoms are numbered as shown, so we can refer to individual carbon atoms in the structure. In animals glucose is the main transport sugar in the blood, and its concentration in the blood is carefully controlled.

There are many monosaccharides, with the same chemical formula ( $\text{C}_6\text{H}_{12}\text{O}_6$ ), but different structural formulae. These include fructose and galactose.

Common five-carbon sugars (where  $n = 5$ ,  $\text{C}_5\text{H}_{10}\text{O}_5$ ) include ribose and deoxyribose (found in nucleic acids and ATP).

## Disaccharides

Disaccharides are formed when two monosaccharides are joined together by a glycosidic bond. The reaction involves the formation of a molecule of water (H<sub>2</sub>O):



This shows two glucose molecules joining together to form the disaccharide maltose. Because this bond is between carbon 1 of one molecule and carbon 4 of the other molecule it is called a 1-4 glycosidic bond. This kind of reaction, where water is formed, is called a condensation reaction. The reverse process, when bonds are broken by the addition of water (e.g. in digestion), is called a hydrolysis reaction.

- polymerisation reactions are condensation reactions
- breakdown reactions are hydrolysis reactions

There are three common disaccharides:

- Maltose (or malt sugar) is glucose & glucose. It is formed on digestion of starch by amylase, because this enzyme breaks starch down into two-glucose units. Brewing beer starts with malt, which is a maltose solution made from germinated barley. Maltose is the structure shown above.
- Sucrose (or cane sugar) is glucose & fructose. It is common in plants because it is less reactive than glucose, and it is their main transport sugar. It's the common table sugar that you put in tea.
- Lactose (or milk sugar) is galactose & glucose. It is found only in mammalian milk, and is the main source of energy for infant mammals.

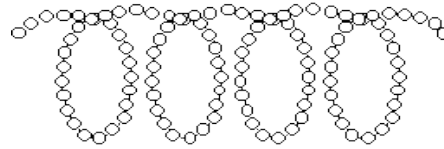
## Polysaccharides

Polysaccharides are long chains of many monosaccharides joined together by glycosidic bonds. There are three important polysaccharides:

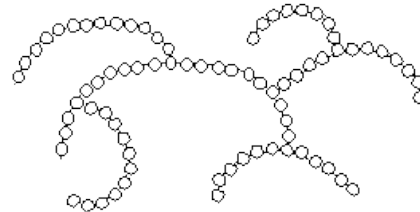
Starch is the plant storage polysaccharide. It is insoluble and forms starch granules inside many plant cells. Being insoluble means starch does not change the water potential of cells, so does not cause the

cells to take up water by osmosis (more on osmosis later). It is not a pure substance, but is a mixture of amylose and amylopectin.

Amylose is simply poly-(1-4) glucose, so is a straight chain. In fact the chain is floppy, and it tends to coil up into a helix.

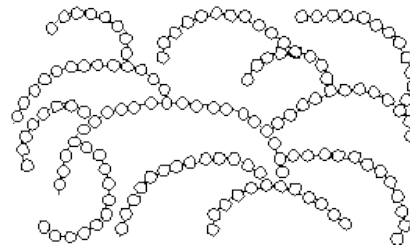


Amylopectin is poly(1-4) glucose with about 4% (1-6) branches. This gives it a more open molecular structure than amylose. Because it has more ends, it can be broken more quickly than amylose by amylase enzymes.

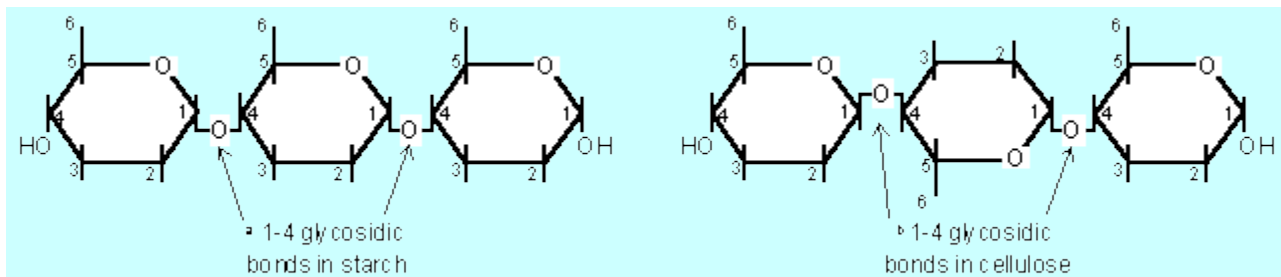


Both amylose and amylopectin are broken down by the enzyme amylase into maltose, though at different rates.

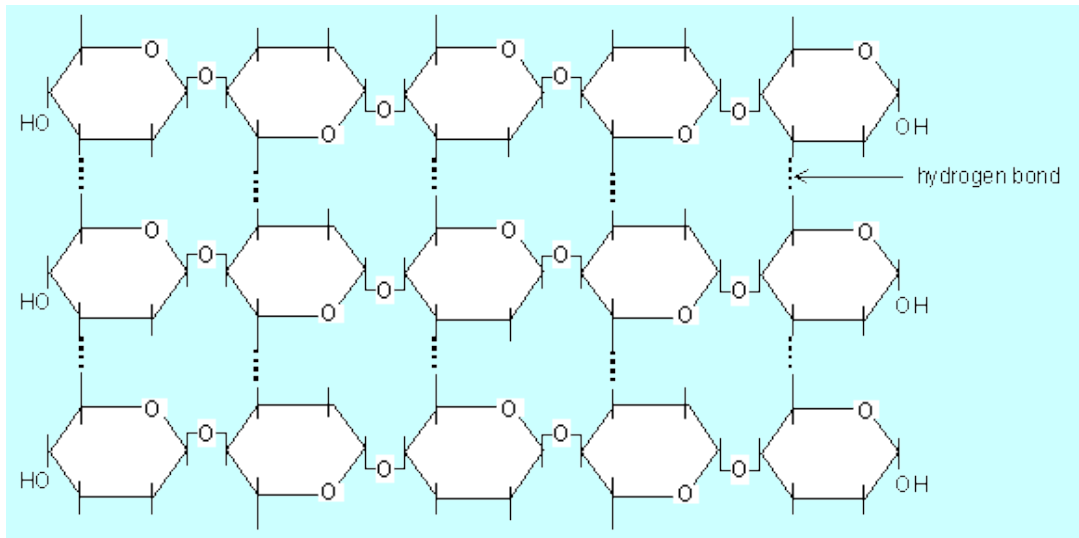
Glycogen is similar in structure to amylopectin. It is poly (1-4) glucose with 9% (1-6) branches. It is made by animals as their storage polysaccharide, and is found mainly in muscle and liver. Because it is so highly branched, it can be mobilised (broken down to glucose for energy) very quickly.



Cellulose is only found in plants, where it is the main component of cell walls. It is poly (1-4) glucose, but with a different isomer of glucose. Cellulose contains beta-glucose, in which the hydroxyl group on carbon 1 sticks up. This means that in a chain alternate glucose molecules are inverted.



This apparently tiny difference makes a huge difference in structure and properties. While the  $\alpha$ 1-4 glucose polymer in starch coils up to form granules, the  $\beta$ 1-4 glucose polymer in cellulose forms straight chains. Hundreds of these chains are linked together by hydrogen bonds to form cellulose microfibrils. These microfibrils are very strong and rigid, and give strength to plant cells, and therefore to young plants.



The beta-glycosidic bond cannot be broken by amylase, but requires a specific cellulase enzyme. The only organisms that possess a cellulase enzyme are bacteria, so herbivorous animals, like cows and termites whose diet is mainly cellulose, have mutualistic bacteria in their guts so that they can digest cellulose. Humans cannot digest cellulose, and it is referred to as fibre.

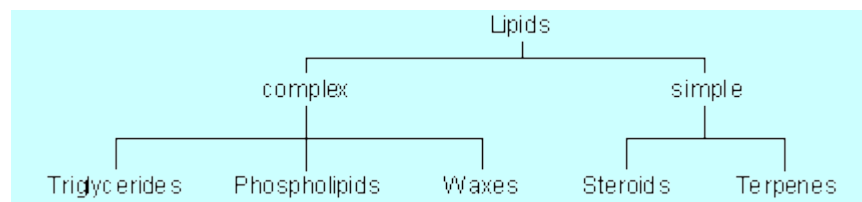
Other polysaccharides that you may come across include:

- Chitin (poly glucose amine), found in fungal cell walls and the exoskeletons of insects.
- Pectin (poly galactose uronate), found in plant cell walls.
- Agar (poly galactose sulphate), found in algae and used to make agar plates.
- Murein (a sugar-peptide polymer), found in bacterial cell walls.
- Lignin (a complex polymer), found in the walls of xylem cells, is the main component of wood.

# LIPIDS



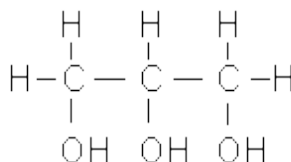
Lipids are a mixed group of hydrophobic compounds composed of the elements carbon, hydrogen and oxygen. They contain fats and oils (fats are solid at room temperature, whereas oils are liquid)



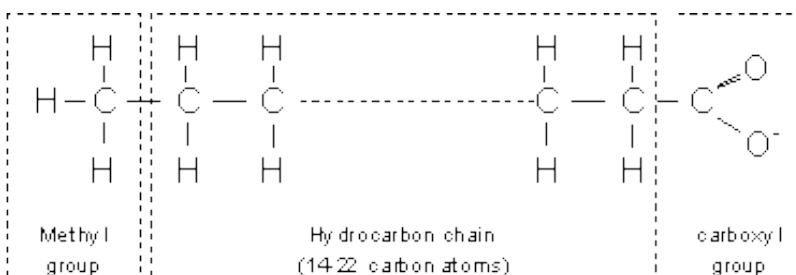
## Triglycerides

Triglycerides are commonly called fats or oils. They are made of glycerol and fatty acids.

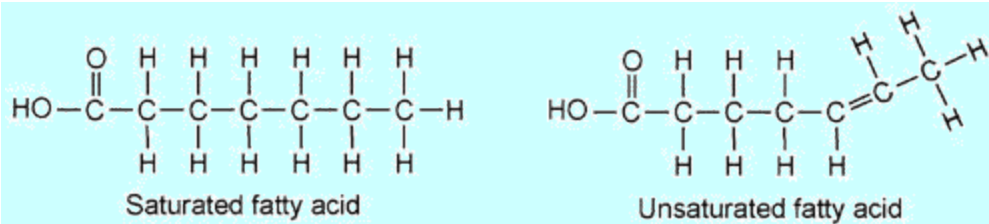
Glycerol is a small, 3-carbon molecule with three hydroxyl groups.



Fatty acids are long molecules with a polar, hydrophilic end and a non-polar, hydrophobic "tail". The hydrocarbon chain can be from 14 to 22  $\text{CH}_2$  units long. The hydrocarbon chain is sometimes called an R group, so the formula of a fatty acid can be written as  $\text{R}-\text{COOH}$ .



- If there are no  $\text{C}=\text{C}$  double bonds in the hydrocarbon chain, then it is a saturated fatty acid (i.e. saturated with hydrogen). These fatty acids form straight chains, and have a high melting point.
- If there are  $\text{C}=\text{C}$  double bonds in the hydrocarbon chain, then it is an unsaturated fatty acid (i.e. unsaturated with hydrogen). These fatty acids form bent chains, and have a low melting point. Fatty acids with more than one double bond are called poly-unsaturated fatty acids (PUFAs).

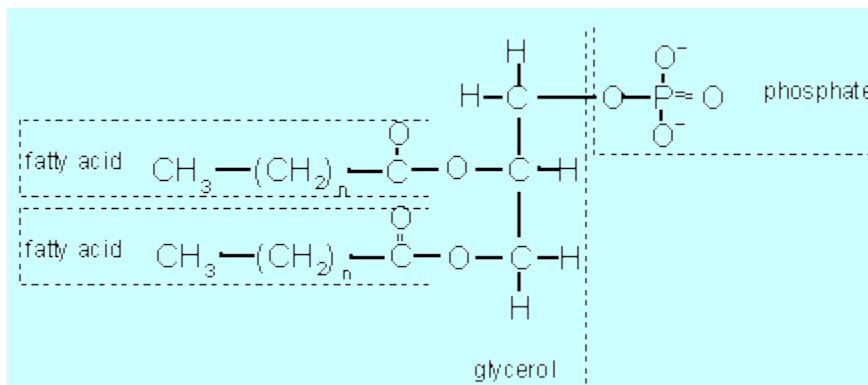


One molecule of glycerol joins together with three fatty acid molecules to form a triglyceride molecule, in another condensation polymerisation reaction:

Triglycerides are insoluble in water. They are used for storage, insulation and protection in fatty tissue (or adipose tissue) found under the skin (sub-cutaneous) or surrounding organs. They yield more energy per unit mass than other compounds so are good for energy storage. Carbohydrates can be mobilised more quickly, and glycogen is stored in muscles and liver for immediate energy requirements.

- Triglycerides containing saturated fatty acids have a high melting point and tend to be found in warm-blooded animals. At room temperature they are solids (fats), e.g. butter, lard.
- Triglycerides containing unsaturated fatty acids have a low melting point and tend to be found in cold-blooded animals and plants. At room temperature they are liquids (oils), e.g. fish oil, vegetable oils.

## Phospholipids



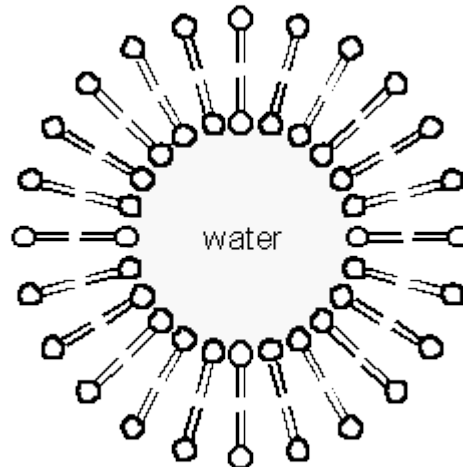
Phospholipids have a similar structure to triglycerides, but with a phosphate group in place of one fatty acid chain. There may also be other groups attached to the phosphate. Phospholipids have a polar hydrophilic "head" (the negatively-charged phosphate group) and two non-polar hydrophobic "tails" (the fatty acid chains). This mixture of properties is fundamental to biology, for phospholipids are the main components of cell membranes.

- When mixed with water, phospholipids form droplet spheres with the hydrophilic heads facing the water and the hydrophobic tails facing each other. This is called a micelle.





- Alternatively, they may form a double-layered phospholipid bilayer. This traps a compartment of water in the middle separated from the external water by the hydrophobic sphere. This naturally-occurring structure is called a liposome, and is similar to a membrane surrounding a cell.



## Waxes

Waxes are formed from fatty acids and long-chain alcohols. They are commonly found wherever waterproofing is needed, such as in leaf cuticles, insect exoskeletons, birds' feathers and mammals' fur.

## Steroids

Steroids are small hydrophobic molecules found mainly in animals. They include:

- cholesterol, which is found in animals cell membranes to increase stiffness
- bile salts, which help to emulsify dietary fats
- steroid hormones such as testosterone, oestrogen, progesterone and cortisol
- vitamin D, which aids  $\text{Ca}^{2+}$  uptake by bones.

# PROTEINS

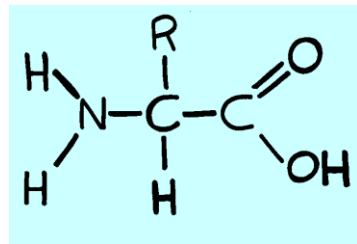


Proteins are the most complex and most diverse group of biological compounds. They have an astonishing range of different functions, as this list shows.

- structure e.g. collagen (bone, cartilage, tendon), keratin (hair), actin (muscle)
- enzymes e.g. amylase, pepsin, catalase, etc (>10,000 others)
- transport e.g. haemoglobin (oxygen), transferrin (iron)
- pumps e.g. Na<sup>+</sup>K<sup>+</sup> pump in cell membranes
- motors e.g. myosin (muscle), kinesin (cilia)
- hormones e.g. insulin, glucagon
- receptors e.g. rhodopsin (light receptor in retina)
- antibodies e.g. immunoglobulins
- storage e.g. albumins in eggs and blood, caesin in milk
- blood clotting e.g. thrombin, fibrin
- lubrication e.g. glycoproteins in synovial fluid
- toxins e.g. diphtheria toxin
- antifreeze e.g. glycoproteins in arctic flea
- and many more!

Proteins are made of amino acids. Amino acids are made of the five elements C H O N S. The general structure of an amino acid molecule is shown on the right. There is a central carbon atom (called the "alpha carbon"), with four different chemical groups attached to it:

- a hydrogen atom
- a basic amino group
- an acidic carboxyl group
- a variable "R" group (or side chain)

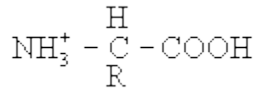


Amino acids are so-called because they have both amino groups and acid groups, which have opposite charges. At neutral pH (found in most living organisms), the groups are ionized as shown above, so there is a positive charge at one end of the molecule and a negative charge at the other end. The overall net charge on the molecule is therefore zero. A molecule like this, with both positive and negative charges is called a zwitterion. The charge on the amino acid changes with pH:

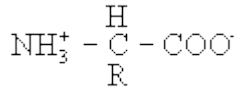
LOW PH (ACID)

NEUTRAL PH

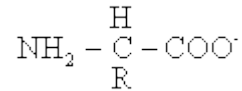
HIGH PH (ALKALI)



charge = +1

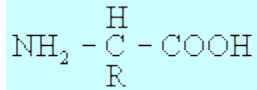


charge = 0



charge = -1

It is these changes in charge with pH that explain the effect of pH on enzymes. A solid, crystallised amino acid has the uncharged structure



however this form never exists in solution, and therefore doesn't exist in living things (although it is the form usually given in textbooks).

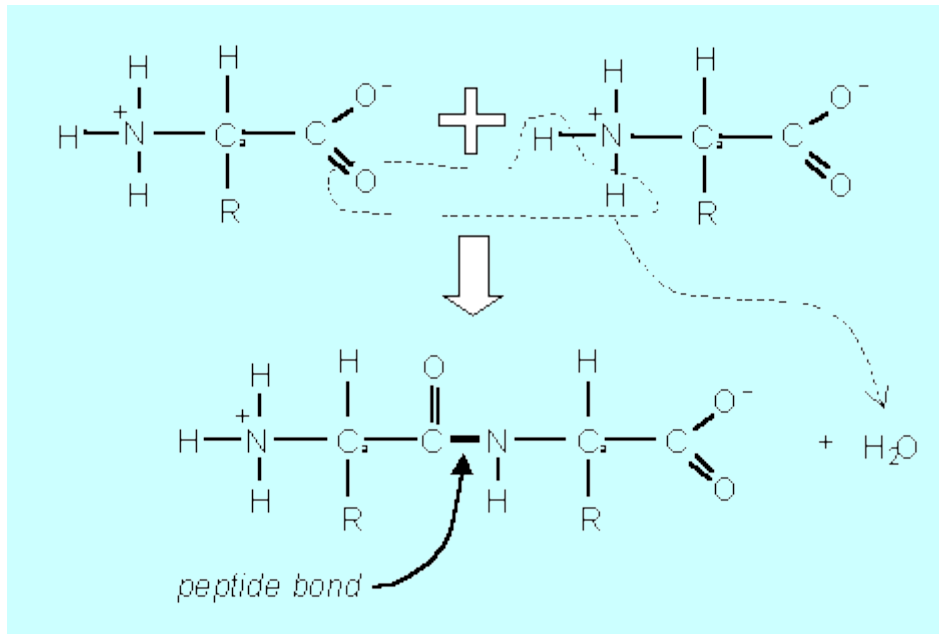
There are 20 different R groups, and so 20 different amino acids. Since each R group is slightly different, each amino acid has different properties, and this in turn means that proteins can have a wide range of properties. The following table shows the 20 different R groups, grouped by property, which gives an idea of the range of properties. You do not need to learn these, but it is interesting to see the different structures, and you should be familiar with the amino acid names. You may already have heard of some, such as the food additive monosodium glutamate, which is simply the sodium salt of the amino acid glutamate. Be careful not to confuse the names of amino acids with those of bases in DNA, such as cysteine (amino acid) and cytosine (base), threonine (amino acid) and thymine (base). There are 3-letter and 1-letter abbreviations for each amino acid.

THE TWENTY AMINO ACID R-GROUPS (FOR INTEREST ONLY NO KNOWLEDGE REQUIRED)			
	SIMPLE R GROUPS		BASIC R GROUPS
Glycine Gly G	— H	Lysine Lys K	$-\text{CH}_2-\text{CH}_2-\text{CH}_2-\text{CH}_2-\text{NH}_3^+$
Alanine Ala A	— CH <sub>3</sub>	Arginine Arg R	$-\text{CH}_2-\text{CH}_2-\text{CH}_2-\text{NH}-\text{C} \begin{array}{l} \text{NH}_2 \\ \text{NH}_2^+ \end{array}$
Valine Val V	$\begin{array}{l} \text{CH} \\ / \quad \backslash \\ \text{CH}_3 \quad \text{CH}_3 \end{array}$	Histidine His H	$-\text{CH}_2-\text{C} \begin{array}{l} \text{CH} \\ \text{NH} \\ \text{NH}^+ \\ \text{CH} \end{array}$
Leucine Leu L	$-\text{CH}_2-\text{CH} \begin{array}{l} \text{CH}_3 \\ \text{CH}_3 \end{array}$	Asparagine Asn N	$-\text{CH}_2-\text{C} \begin{array}{l} \text{O} \\ \text{NH}_2^+ \end{array}$

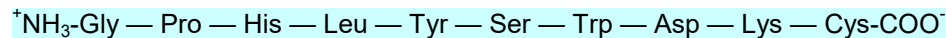
Isoleucine Ile I	$\begin{array}{c} \text{---CH} \text{ ---CH}_2\text{---CH}_3 \\   \\ \text{CH}_3 \end{array}$	Glutamine Gln Q	$\begin{array}{c} \text{O} \\    \\ \text{---CH}_2\text{---CH}_2\text{---C} \\   \\ \text{NH}_2^+ \end{array}$
	<b>HYDROXYL R GROUPS</b>		<b>ACIDIC R GROUPS</b>
Serine Ser S	$\text{---CH}_2\text{---OH}$	Aspartate Asp D	$\begin{array}{c} \text{O}^- \\    \\ \text{---CH}_2\text{---C} \\   \\ \text{O} \end{array}$
Threonine Thr T	$\begin{array}{c} \text{---CH} \text{ ---OH} \\   \\ \text{CH}_3 \end{array}$	Glutamate Glu E	$\begin{array}{c} \text{O}^- \\    \\ \text{---CH}_2\text{---CH}_2\text{---C} \\   \\ \text{O} \end{array}$
	<b>SULPHUR R GROUPS</b>		<b>RINGED R GROUPS</b>
Cysteine Cys C	$\text{---CH}_2\text{---SH}$	Phenylalanine Phe F	$\begin{array}{c} \text{CH---CH} \\ // \quad \backslash \\ \text{---CH}_2\text{---C} \quad \text{CH} \\ \backslash \quad // \\ \text{CH} \quad \text{---CH} \end{array}$
Methionine Met M	$\text{---CH}_2\text{---CH}_2\text{---S---CH}_3$	Tyrosine Tyr Y	$\begin{array}{c} \text{CH---CH} \\ // \quad \backslash \\ \text{---CH}_2\text{---C} \quad \text{CH---OH} \\ \backslash \quad // \\ \text{CH} \quad \text{---CH} \end{array}$
	<b>CYCLIC R GROUP</b>		
Proline Pro P	$\begin{array}{c} \text{C}_\alpha \text{---CH}_2 \\   \quad \backslash \\ \text{NH}_2\text{---CH}_2 \quad \text{CH}_2 \end{array}$	Tryptophan Trp W	$\begin{array}{c} \text{---CH}_2\text{---CH---CH} \\   \quad \backslash \quad   \\ \text{CH} \quad \text{---CH} \quad \text{NH} \\ // \quad \backslash \quad // \\ \text{C} \quad \text{CH} \quad \text{CH} \\ \backslash \quad // \\ \text{CH} \quad \text{=CH} \end{array}$

## Polypeptides

Amino acids are joined together by peptide bonds. The reaction involves the formation of a molecule of water in another condensation polymerisation reaction:



When two amino acids join together a dipeptide is formed. Three amino acids form a tripeptide. Many amino acids form a polypeptide. e.g.:



In a polypeptide there is always one end with a free amino ( $\text{NH}_2$ ) ( $\text{NH}_3$  in solution) group, called the N-terminus, and one end with a free carboxyl ( $\text{COOH}$ ) ( $\text{COO}$  in solution) group, called the C-terminus.

## Protein Structure

Polypeptides are just a string of amino acids, but they fold up to form the complex and well-defined three-dimensional structure of working proteins. To help to understand protein structure, it is broken down into four levels:

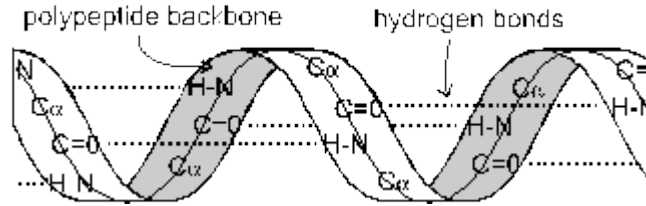
### 1. Primary Structure

- This is just the sequence of amino acids in the polypeptide chain, so is not really a structure at all. However, the primary structure does determine the rest of the protein structure. Finding the primary structure of a protein is called protein sequencing, and the first protein to be sequenced was the protein hormone insulin, by the Cambridge biochemist Fredrick Sanger, for which work he got the Nobel prize in 1958.

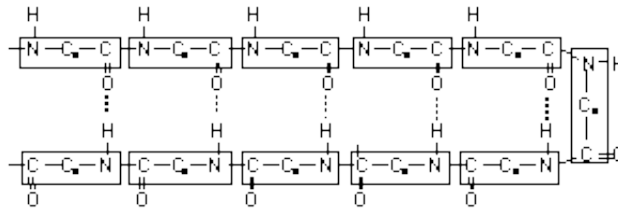
### 2. Secondary Structure

- This is the most basic level of protein folding, and consists of a few basic motifs that are found in all proteins. The secondary structure is held together by hydrogen bonds between the carboxyl groups and the amino groups in the polypeptide backbone. The two secondary structures are the  $\alpha$ -helix and the  $\beta$ -sheet.

**The  $\alpha$ -helix.** The polypeptide chain is wound round to form a helix. It is held together by hydrogen bonds running parallel with the long helical axis. There are so many hydrogen bonds that this is a very stable and strong structure. Helices are common structures throughout biology.



**The  $\beta$ -sheet.** The polypeptide chain zig-zags back and forward forming a sheet. Once again it is held together by hydrogen bonds.



### 3. Tertiary Structure

- This is the 3 dimensional structure formed by the folding up of a whole polypeptide chain. Every protein has a unique tertiary structure, which is responsible for its properties and function. For example the shape of the active site in an enzyme is due to its tertiary structure. The tertiary structure is held together by bonds between the R groups of the amino acids in the protein, and so depends on what the sequence of amino acids is. There are three kinds of bonds involved:

- hydrogen bonds, which are weak.
- ionic bonds between R-groups with positive or negative charges, which are quite strong.
- sulphur bridges - covalent S-S bonds between two cysteine amino acids, which are strong.

### 4. Quaternary Structure

- This structure is found only in proteins containing more than one polypeptide chain, and simply means how the different polypeptide chains are arranged together. The individual polypeptide chains are usually globular, but can arrange themselves into a variety of quaternary shapes. e.g.:

Haemoglobin, the oxygen-carrying protein in red blood cells, consists of four globular subunits arranged in a tetrahedral (pyramid) structure. Each subunit contains one iron atom and can bind one molecule of oxygen.



These four structures are not real stages in the formation of a protein, but are simply a convenient classification that scientists invented to help them to understand proteins. In fact proteins fold into all these structures at the same time, as they are synthesised.

The final three-dimensional shape of a protein can be classified as globular or fibrous.

globular structure

fibrous (or filamentous) structure



The vast majority of proteins are globular, including enzymes, membrane proteins, receptors, storage proteins, etc. Fibrous proteins look like ropes and tend to have structural roles such as collagen (bone), keratin (hair), tubulin (cytoskeleton) and actin (muscle). They are usually composed of many polypeptide chains. A few proteins have both structures: the muscle protein myosin has a long fibrous tail and a globular head, which acts as an enzyme.

This diagram shows a molecule of the enzyme dihydrofolate reductase, which comprises a single polypeptide chain. It has a globular shape

This diagram shows part of a molecule of collagen, which is found in bone and cartilage. It has a unique, very strong triple-helix structure. It is a fibrous protein

