Introduction to Chemistry for Biology

Biomolecules 2

Professor R Baxter

Course Synopsis see Course Booklet

Textbooks: J M^{c}Murry *Fundamentals of Organic Chemistry*, Chapters 14-16 (4th or subsequent editions); **and** appendices 11,12 and 13 in class booklet. Chapter 5 in *Lehninger's Principles of Biochemistry* (DL Nelson & MM Cox, 3rd ed.), although it covers the material in more depth, is worth consulting for explanations of amino acid properties.

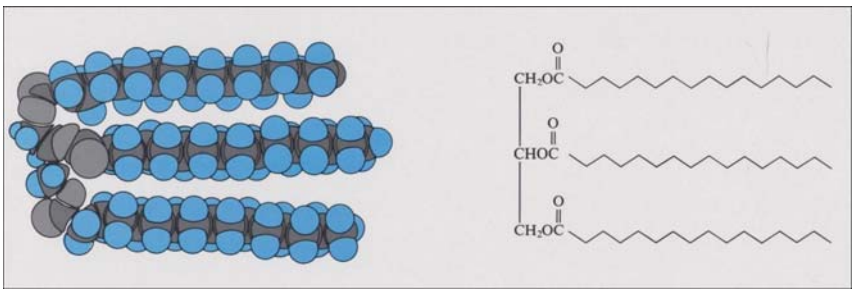
Introduction

In this part of the course we shall look in some detail at the common types of molecules found in biological systems – with emphasis on those of major importance in studies of cell biology – lipids, carbohydrates, proteins and nucleic acids. Each of these is built up of component entities, with some of which you should already be familiar.

I. Lipids

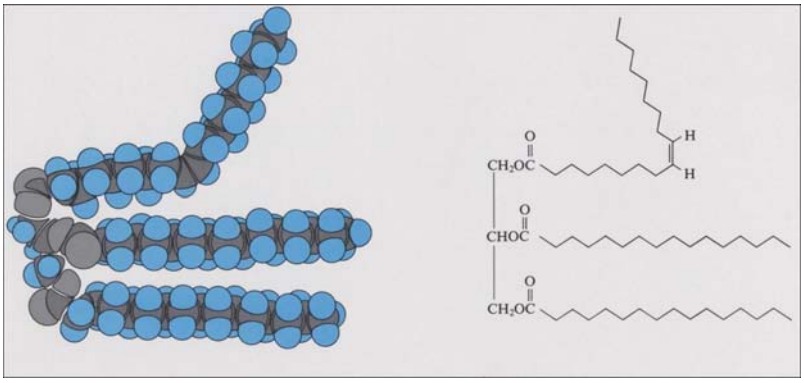
Lipids is a general name which is given to a group of biomolecules which are water insoluble (or of sparing solubility). These are really two quite different structural types of molecular species grouped together. The **Fats** which are (predominantly) long chain esters of the triol glycerol and the **Steroids** which are compounds having a pentacyclic cholesterol derived skeleton.

The commonest fats are **glycerol esters of long chain fatty acids** – and these can be divided into mono-, di- or triglycerides depending on the number of fatty acid side chains. One of the simplest is tributryrin – the triester of butyric acid and glycerol – which is found in milk. Triesters with one single saturated fatty acid are rare, examples are the triesters of stearic acid ($C_{17}H_{35}COOH$) and palmitic acid ($C_{15}H_{31}COOH$) which are common in mammalian fats. More common are mixed esters containing several different fatty acids.



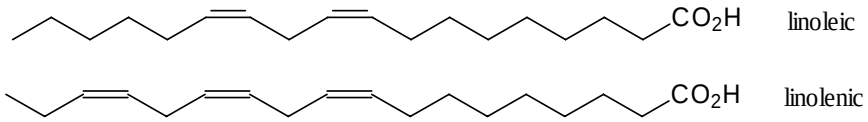
Glyceryl tripalmitate

The natural fatty acids belong to a homologous series differing by (-CH_2CH_2-) units. [That's because they are biosynthesised from acetyl CoA – which is a C_2 building block]. The names of ALL the fatty acids are not very important – it is a good idea to remember a common few like lauric, palmitic & stearic. Frequently unsaturated fatty acids (like oleic – ($C_{17}H_{33}COOH$)) are involved. Note that the presence of a double bond produces a 'kink' in the chain. This has an important consequence in packing of the fat molecules – unsaturated triglycerides have lower melting points. Vegetable fats contain more unsaturated fatty acids than mammalian fats. As a consequence vegetable fats are generally liquid at room temperature – human fats are liquid at >17°C. The 3D structures of solid fats are complex – most fats adopt more than one kind of packing in the solid phase. This results in different melting behaviour.



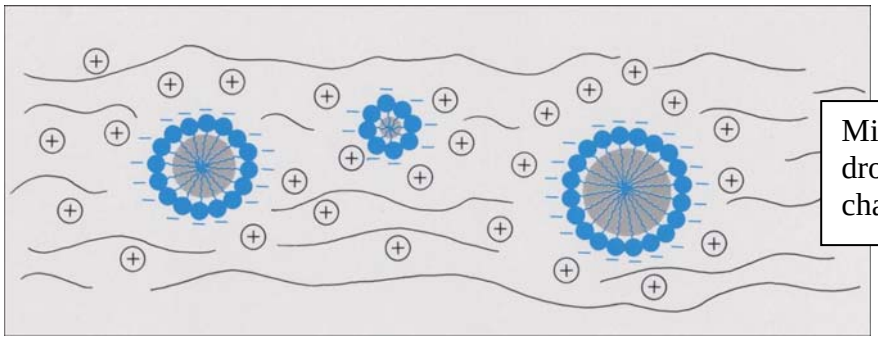
Glyceryl diipalmitooleate

The unsaturated fatty acids invariably have Z double bonds near the centre of the chain. Frequently vegetable and fish oils contain polyunsaturated fatty acids such as linoleic and linolenic acids. Margarine is produced by hydrogenating vegetable oils (olive, palm, etc) over nickel catalysts to give saturated fats with mammalian fat type characteristics.



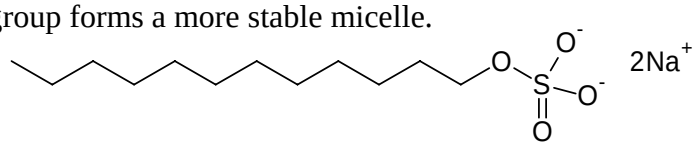
So-called 'trans fats' are triglycerides which have (a few) isolated E-double bonds in the fatty acids which arise from rearrangements during partial hydrogenation in margarine manufacture.

In Nature fats are hydrolysed to the component glycerol and the salts of the fatty acids by specialist enzymes called lipases. The same result can be effected by heating with alkali. The process is historically called saponification and this is the way soap was manufactured. Alkali metal salts of fatty acids are not very soluble in water because of their long hydrocarbon chains and form **micelles** – spherical structures where the acid groups point outwards to the water. Micelles can aggregate round other fatty molecules. Calcium salts of fatty acids are quite insoluble – hence 'natural' soaps form a scum in hard water.



Micelle formation –oil droplets emulsified by long chain fatty acids

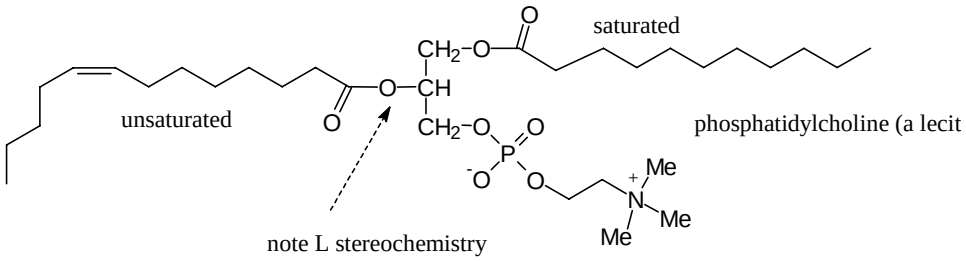
Modern soaps are based on synthetic sodium laurel sulfate. The superior solubility of the sulfate group forms a more stable micelle.

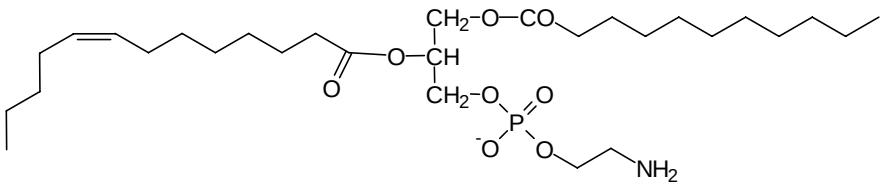


Fats as foodstuffs: Fats (as triglycerides) have the highest calorific content of all foods other than EtOH (9.3kcal/g = 39Kj/g). [Food calories aren't the same as the enthapy or the heat of combustion but are a measure of the energy the cell can derive from metabolism.] Fatty acid chains are broken down to acetyl CoA by a process called β-oxidation.

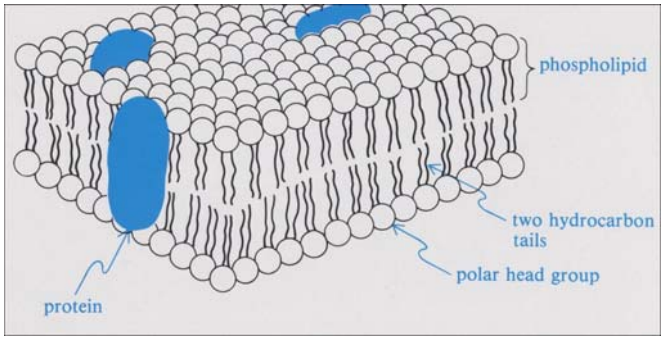
Waxes are esters of long chain fatty acids not with glycerol but with long chain alcohols (themselves derived from fatty acids by reduction) – beeswax is the straight chain palmitate ester $C_{15}H_{31}COOC_{30}H_{61}$.

The **phospholipids** (mainly as lecithins and cephalins) play a wide varied of structural roles in cell membranes and metabolic roles in nerve tissue. These are triglycerides in which one of the acids is phosphoric. In the membrane the charged groups (the polar heads groups) point to the outside of the cell and towards the cytoplasm.





phosphatidylethanolamine (a cephalin)

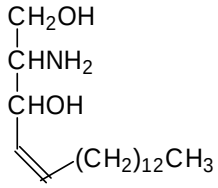


Schematic of a cell membrane

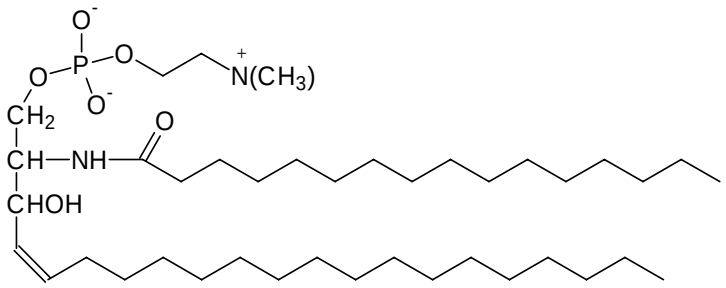
In lethicins the phosphate group is further esterified with choline ($HOCH_2CH_2N^+(CH_3)_3$) and in cephalins with ethanolamine. Lecithins lacking one

fatty acid cause haemolysis of red blood cells – and are components of many snake venoms.

Another group of major phospholipids are the **sphingolipids.** These are important components of cell membranes but do not contain glycerol. Instead the backbone is made of the C-18 compound **sphingosine.**



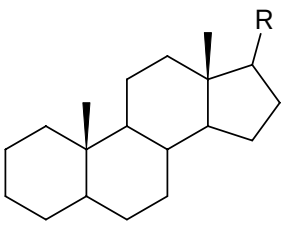
sphingosine
(a highly modified fatty acid)



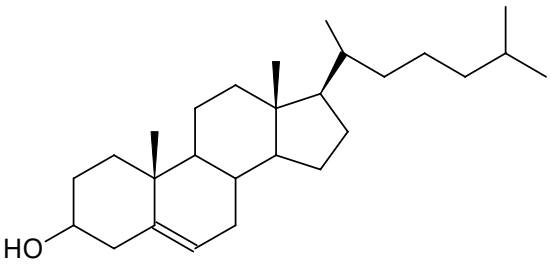
sphingomyelin (membrane lipid in nerve fibres)

Glycolipids are polysaccharides (see later) which are esterified with fatty acids. These have a variety of functions in brain and nerve cells. The myelin sheaths of axons contain large amounts of glycolipid.

Steroids are a separate class of lipids – all are derived from cholesterol. These include the bile acids and the steroid hormones. Cholesterol itself (and its fatty acid esters) play a structural role in (stiffening) membranes.

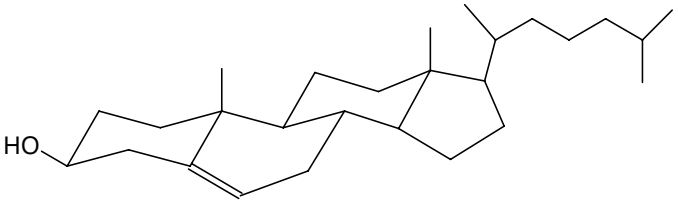


basic steroid stucture

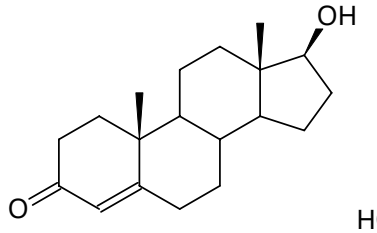


cholesterol

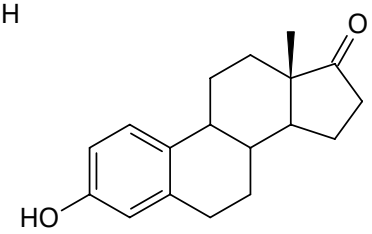
Note that cholesterol has a very flat (uncluttered) structure when drawn in 3D.



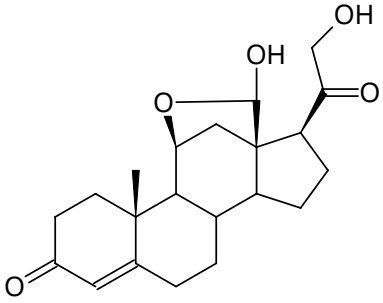
Cholesterol is the raw material for producing **the sex hormones –** the androgens (male) and the estrogens (female) and also **adrenocortical hormones** such as aldosterone (which regulates cellular salt balance).



testosterone



estrone

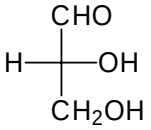


aldosterone

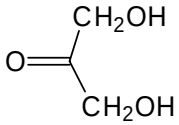
2. Carbohydrates

The carbohydrates are a class of molecules ranging from simple sugars (**monosaccharides**) to complex **polysaccharides** which contain several million monosaccharide monomers. These are important cell constituents – the principal energetic pathway of the cell is glycolysis – the breakdown (catabolism) of glucose. Glucose is stored in mammalian cells as the polysaccharide glycogen and in plant cells as starch. Other sugars play structural roles in cell matrices and membranes and addition of sugar units to proteins (glycoproteins) modulates the properties of the proteins.

<u>Monosaccharides</u> are the simplest sugar molecules. These are linear polyalcohols with an aldehyde or ketone group range in size from three carbon (trioses) to seven carbon (septoses) in chain length. Thus the trioses, glyceraldehyde and dihydroxyacetone are the simplest monosaccharides.



D-glyceraldehyde



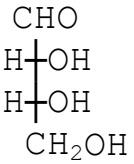
dihydroxyacetone

To begin with let's limit ourselves to the aldoses – monosaccharides with an aldehyde group. When we have a four carbon chain with a terminal aldehyde (and every other carbon hydroxylated) we have a tetrose. In fact we have two diastereomeric tetroses – erythrose and threose – each with two enantiomers (see below). This is where the Fischer formula are really useful.

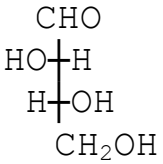
Fortunately Nature uses only sugars which have the R configuration at the penultimate carbon in the chain (2nd to the bottom as written in the Fischer). In the Fischer convention this centre is D so and monosaccharide with this configuration belongs to the D-series. (i.e. D-erythrose and D-glucose are natural but L-erythrose and L-glucose aren't).

However the stereochemistry at any other position can be variable so even limiting one stereochemistry allows whole families of diastereometric monosaccharides. Thus there are four natural D-pentoses, eight D-hexoses, etc. Appendix 12 in the course book describes these.

tetroses

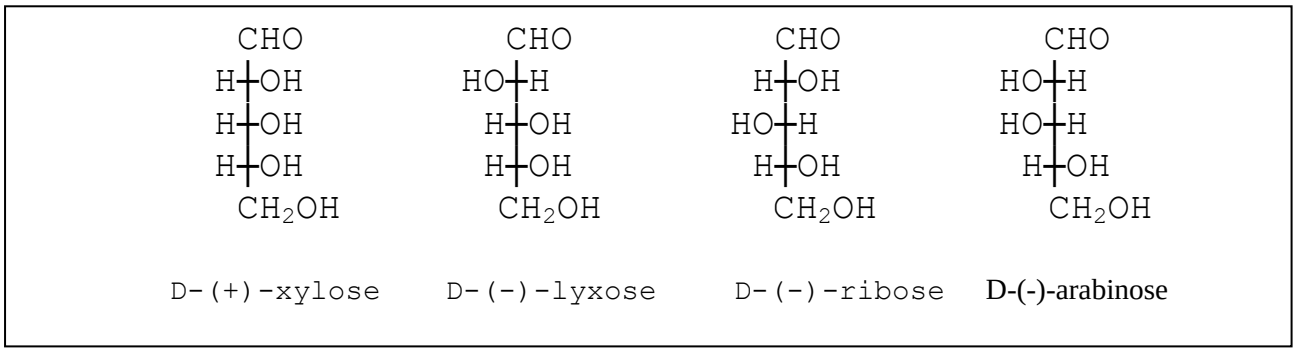


D(-)-erythrose

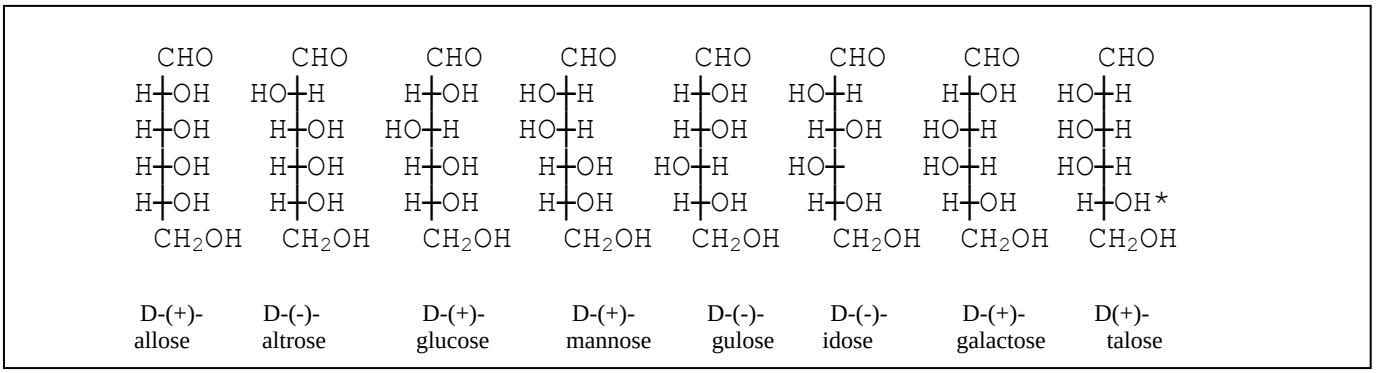


D-(-)-threose

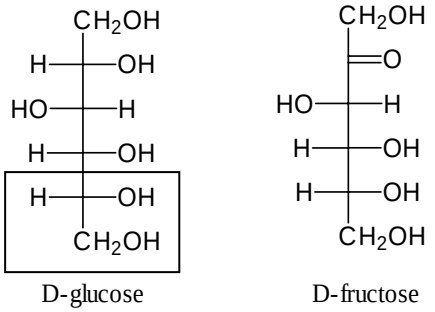
pentoses



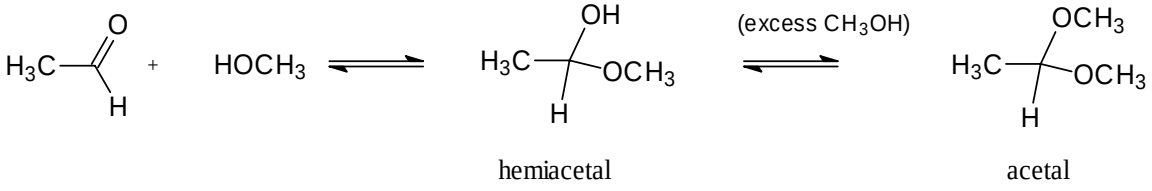
hexoses



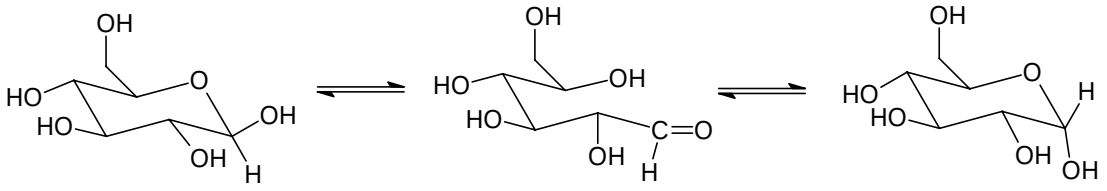
Ketoses, as the name implies, have a **ketone** instead of an aldehyde in their structure. Fortunately, although there are many more possible ketose than aldose structures, there are only a few ketoses's of biological significance. The most important is D-fructose which is metabolically closely related to D-glucose.



Pentoses and hexoses are not stable in their **open chain forms** – they rapidly cyclise with the elimination of water to form **cyclic hemiacetals**. A similar reaction occurs between **free** aldehydes and alcohols.

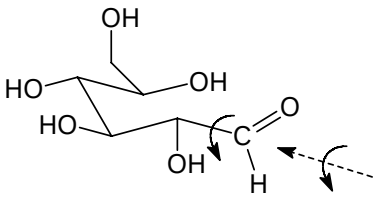


Internal (intramolecular) reactions, where they are sterically feasible, occur up to 100x faster than their intermolecular counterparts. With glucose the free aldehyde is in equilibrium with its ring closed **hemiacetals** and the linear form is only present in a few percent in solution.



β-anomer
$[\alpha]_D$ +112.2°

α-anomer
$[\alpha]_D$ +18.7°



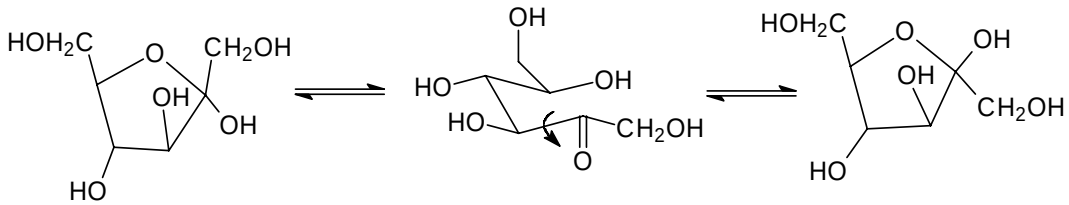
the open chain form can rotate round the C_2/C_3 bond and thus can close to give both anomers

The six-membered pyranose ring closed forms are diastereomers of each other and are referred to as the α- and β-anomers. The reaction can be followed to equilibrium by monitoring the change in optical rotation of the solution when the α-anomer (both anomers are stable in the solid state) is dissolved in water The α-anomer has $[\alpha]_D$ +18.7° and the β-anomer $[\alpha]_D$ +112.2°. When the pure α-anomer is allowed to equilibrate the optical rotation rises to +52° corresponding to a 60:40 mixture of α:β forms. This process is called **mutarotation.**

Note that glucose forms two pyranose rings – these are so called because they are (formally named as) derivatives of **pyran**. Similarly a five membered ring containing oxygen is named after **furan**.

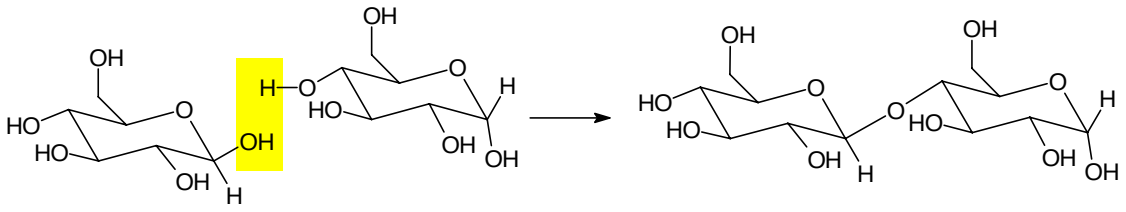
pyran furan

Fructose is also more stable as its internal **hemiketals** – these are called **furanoses** and formed in an analogous way to the glucose cyclisation to the hemiacetal structures.



mutarotation of D-fructose

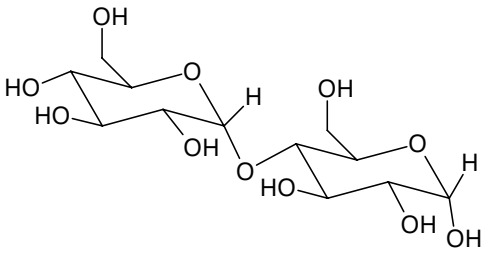
<u>Di-, tri- and polysaccharides</u> are oligomers of the monosaccharides and are formed by the hemiacetal → acetal chemistry we saw above. Reaction of a pyranose hemiacetal (formed from one monosaccharide) with the alcohol of a second pyranose (or furanose) sugar will give a disaccharide **in which the linkage is an acetal function**.



a (1-4) β-linked disaccharide

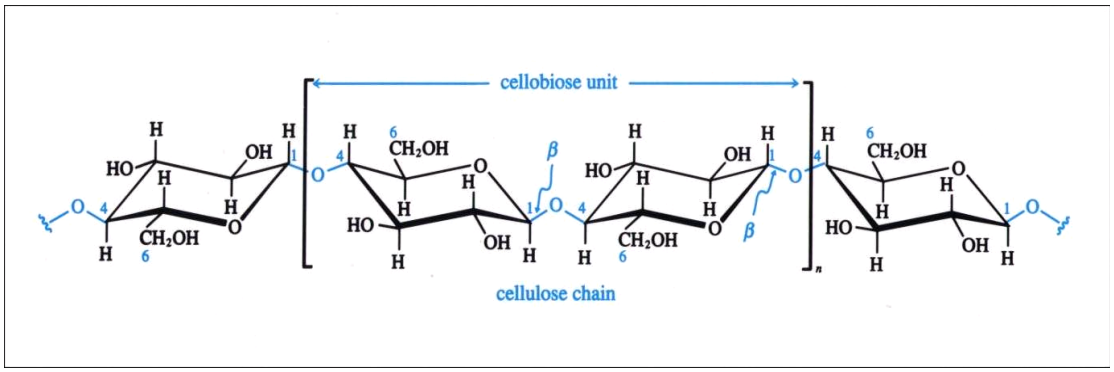
The linkages can be either β– (as above) or α–

An importance point here is that mammalian glycosidases can hydrolyse α-linkages but not β-linkages.

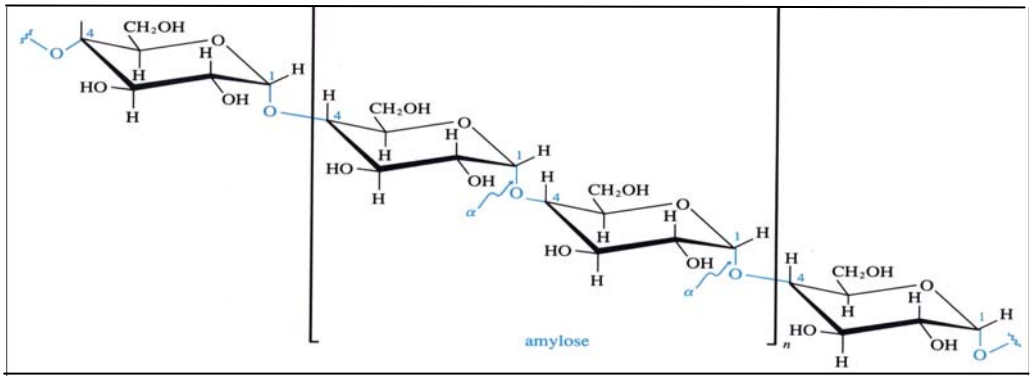


a (1-4) α-linked disaccharide

<u>Cellulose is a high molecular weight polymer of D-glucose with 1,4 β-linkages</u>



<u>Amylose is a high molecular weight polymer of D-glucose with 1,4 α-linkages</u>



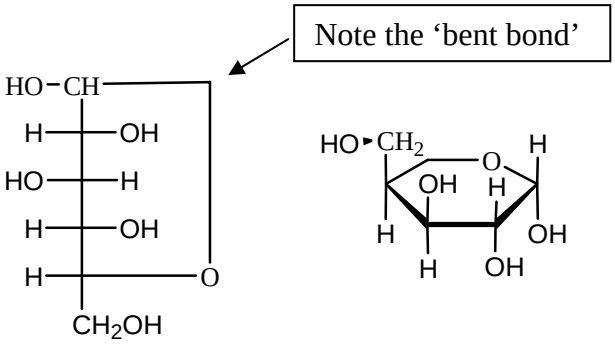
Amylopectin is a high molecular weight polymer of D-glucose similar to amylase with 1,4 α-linkages but with 1,6 α-linked branches every 25 units or so. The structure is very highly branched. (amylose and amylopectin are the components of starch)

Glycogen is the mammalian counterpart of amylopectin but a bit bigger (*ca* 100,000 units per molecule). Glycogen is the principal storage carbohydrate of mammalian cells.

[see the Chapter 14 in the McMurray for structures]

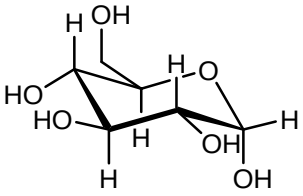
A note on drawing sugar structures

Fischer formulae tend to be a bit awkward to use when describing the cyclic forms. They also require drawing a 'bent bond'!! There are two alternatives – Haworth formulae (used a lot in traditional biochemical textbooks and particularly useful for extended polysaccharide structures) and perspective formulae. The latter two are pretty similar when used portray five membered rings but a bit different for six membered rings.



Fischer

Haworth

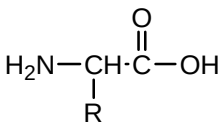


Stereochemical drawing

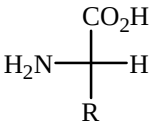
3. Amino Acids, Peptides and Proteins

Amino acids have already been introduced when we discussed chirality in **Biomolecules 1**. Essentially any molecule which contains both carboxylic acid and amino functional groups is by definition an amino acid. With natural amino acids we usually distinguish between proteinatious amino acids – those which occur as components of protein structures and are encoded by the DNA/RNA genetic code and non-proteinatious amino acids which are found in small non-mammalian peptides – such as the defence secretery peptides of fungi. Here we shall restrict ourselves to the amino acids (aa's) found in proteins.

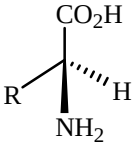
All protein aa's are α-amino acids, that is, the amino group is attached to C-2 (the alpha carbon) of the carbon skeleton. The stereochemistry at this position is also fixed – in the Fischer convention all are L. (Unfortunately, because of the priority rules in the CIP system all are 2R – except for cysteine). All but one of the compounds has a primary amino group. The common part of all protein aa's is therefore;



linear form



Fischer

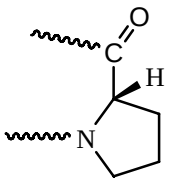


Stereoview

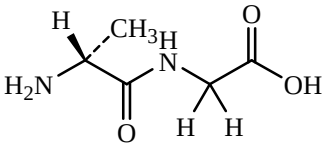
And these differ only in the R groups. The range of R groups found in Nature is limited. The 22 aa side chains are classified as alkyl (neutral), aryl (aromatic), acidic and basic. As we shall see the nature of the R groups is important in intramolecular interactions in protein structures. The table in the Appendix of the course book shows the structures. You should be make yourself familiar with a few of these structures.

In the literature each of the amino acids has its own three letter and one letter codes. The three letter codes are straightforward – e.g. ala for alanine, phe for phenylalanine, etc. One letter codes are more problematical – since several aa's start with the same letter. So alanine is A but phenylalanine is F. Don't worry about memorising more than a few as you will become very familiar with these later in your career.

The only secondary aa is proline – because of the strain imposed by its stereochemistry in a linear chain of aa's it introduces a kink.



Peptides are chains of amino acids produced by making an amide linkage (a peptide bond) between two aa's. Thus a dipeptide contains two aa residues, a tripeptide three, and so on. The term oligopeptide is frequently used to underline the fact that a peptide contains more than one different type of aa residue. The distinction between a peptide and a protein is largely semantic. Large natural oligopeptides (> 80 residues) with defined 3D structures are normally referred to as proteins.



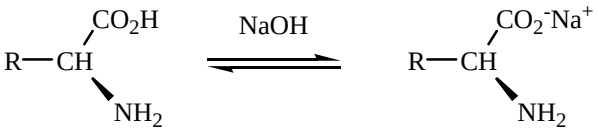
a dipeptide – alanylglycine (ala.gly)

Peptides are always named from the amino terminal end.

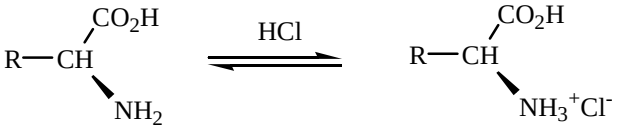
Properties of Amino Acids

Since the chemistry of proteins is largely an extension of the chemistry of their constituent aa's it is important that we learn a bit about these.

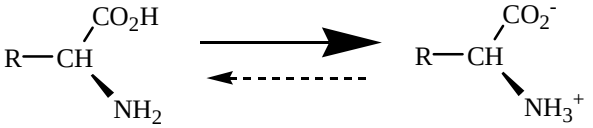
Because aa's have an acid group they obviously form salts with bases.



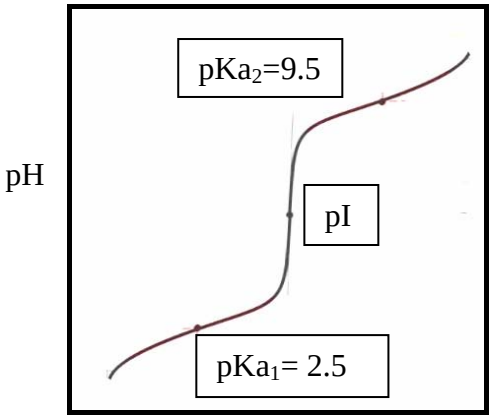
Because they have an amino group they also form salts with acids.



But since both functional groups are in the same molecule they also form internal salts. In fact this internal salt form is the natural form at neutral pH. This form is called a **zwitterion.** (a covalent ionic compound in which the charges are internally compensated) The uncharged form probably doesn't exist to any significant extent in solution.



The titration curve of a typical amino acid (alanine) is shown below: Note that there are two pKa's – that of the carboxylic acid group (pKa ~ 2.5) and that of the protonated amine group (pKa ~ 9.5). The mid-point, where the aa is in the zwitterionic form is called the pI – this is the pH where the overall structure is neutral – and the pH where it is least soluble in water.



When there are extra ionisable groups in the side chains (ie acids or amines) the picture gets more complicated. For example lysine has two amino groups each with a different pKa. This increases the number of forms we can have – but note that the form depends on pH. [Note that in a protein structure where the primary amine and the carboxylic acid are tied up in forming the backbone peptide bonds the acidic or basic properties of the side chain functionalities will play a major role in determining protein properties.]

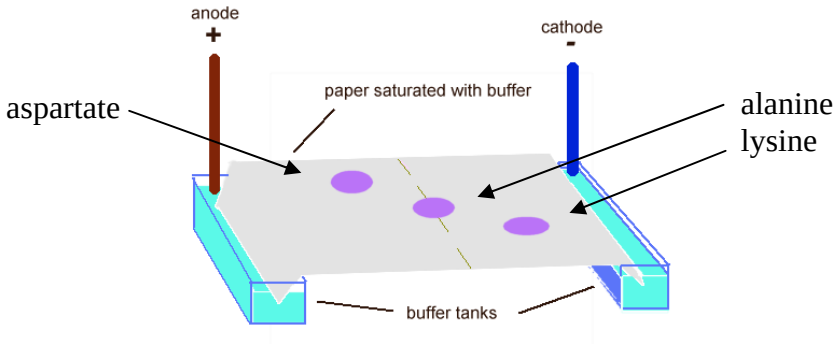
Separation of Amino acids

The charged states of amino acids (and proteins – since a protein can be considered as an extended aa structure) is exploited in separation. This may be for analytical purposes (to determine the aa content of a particular protein) or preparative (the aa glutamate, the flavouring in soy sauce, is prepared commercially on a multi-ton scale – some 2500 tonnes are produced per annum).

There are two methods of importance – electrophoresis and ion exchange chromatography.

Electrophoresis

We know that at any particular pH an amino acid is in a particular charge state. Take for example alanine and aspartic acid. At pH 6 (its pI) the alanine will be uncharged overall while aspartic acid will have one negative charge. If we put a mixture of these on a support (paper or a gel) saturated in a buffer at this pH and then apply a DC current (~ 100 milliamps) the uncharged alanine will remain stationary and the aspartate will migrate towards the anode (positive pole).



[Even small changes in the structures affect the pKa's of the acid and amino groups (eg inductive effects) so given a range electrophoresis experiments at different pH's we can separate all of the aa's in a mixture.]

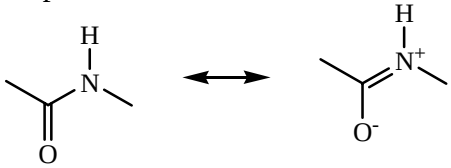
Ion Exchange

In ion exchange separations we use insoluble supports made of charged beads. These are made of polymers with acid or basic groups. A good example is a sulfonic acid resin – which has $-SO_3H_2$ groups. Under neutral conditions (buffers at pH 6-8) aa's will form amino salts [($RCH(COOH)NH_3^+)_2\ ^{2-}O_3S\sim\sim$] with the resin. If we slowly increase the acidity of the solution H+ will compete for the $\sim SO_3^{2-}$ groups driving the amino acid off the resin. Thus we can separate a mixture of aa's purely on the basis of the relative basicity (pKa) of the amino groups of the components.

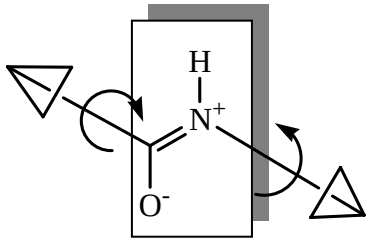
Peptides & Proteins

These are polymers of aa's made by coupling the acid and a-amino groups by an amide bond. The names 'peptide bond' or 'peptide linkage' are also used. Two properties of this new functional group are important.

1. The amide is uncharged
2. It is essentially rigid and planar.



The planar rigid structure is a consequence of resonance. We can draw two resonance forms for the amide – the bias is well towards the double bonded form. The consequence is that peptide structures are not are conformationally flexible as we might have imagined and can be considered as 'plates', formed by the amide functions, joined by bridges consisting of the Cα's of the aa residues. The fact that most of these are chiral and have relatively large groups on them really reduces the number of energetically favourable shapes that a large chain can adopt. So the ultimate 3D structure of a protein begins to emerge as a consequence of these factors.

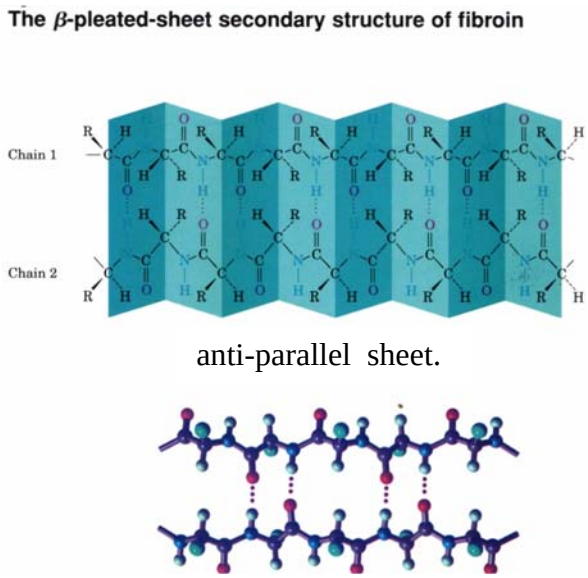


The amide group is essentially planar – the only degrees of freedom in a peptide are rotations round the Cα-N and Cα-CO bonds

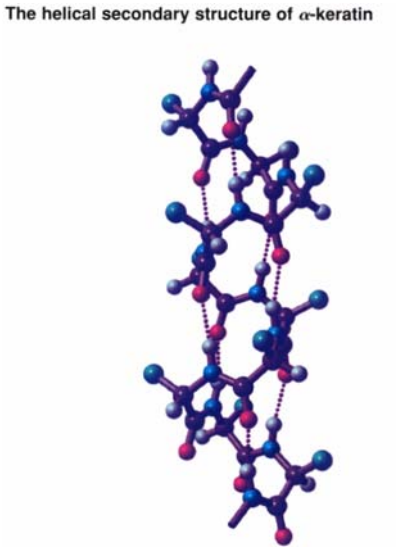
Primary, Secondary & Tertiary Stuctures

The primary structure of a protein or peptide is merely the amino acid sequence. This is normally written & named from the N-terminus.

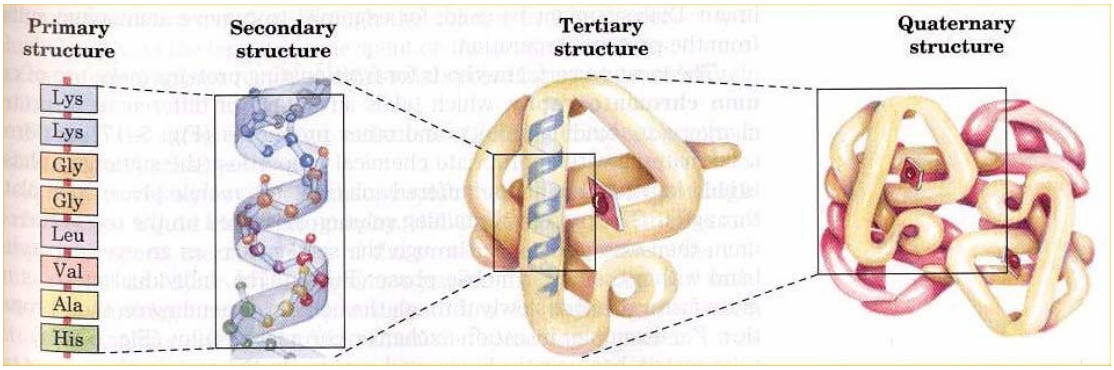
The distinct shapes taken up by certain sequences in a peptide chain are referred to as secondary structures. These are dictated by the types of amino acid side chains in the sequence. Here we have various bends and turns (eg b-βends, omega-loops, etc), helices (eg α-helix), and β-sheets (parallel and antiparallel).



anti-parallel sheet.

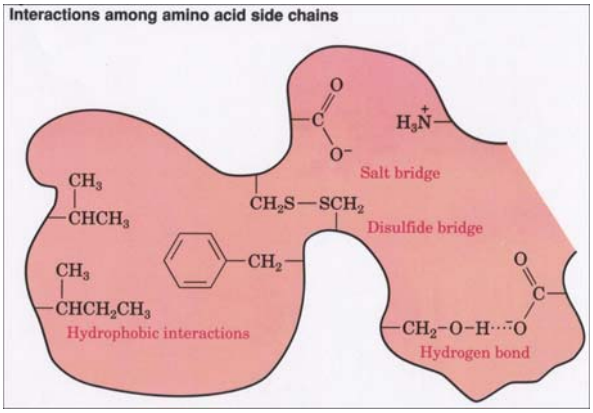


See slides in the lecture and Chapter 15 of McMurray's book for diagrams and a discussion.



Relationship between primary, secondary, tertiary and quaternary structures.

The tertiary structure of a protein refers to the folding of these secondary structural elements in the overall architecture. This is mediated by forces (ionic attraction, hydrogen bonding and hydrophobic interactions) acting between side chains of individual amino acids. Disulfide bridges can provide covalent stabilisation. Quaternary structure refers to interactions between different protein monomers.



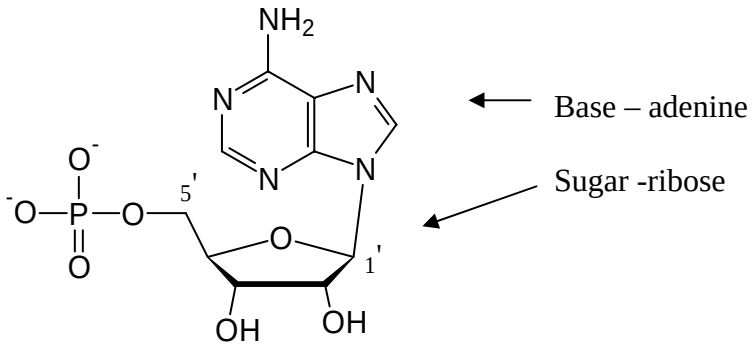
Nucleosides, Nucleotides and Nucleic Acids

The nucleic acids, **deoxyribonucleic acid** (DNA) and **ribonucleic acid** (RNA) are the carriers and processors of the cells genetic information. DNA codes all the information (in the genome) – while messenger RNA (mRNA) carries the sequence information of an individual gene and acts as a 'template' for its translation into a protein sequence. (There are other types of RNA's involved in other cellular functions – such as tRNA which is involved in the activation of amino acids for protein synthesis).

The building blocks (monomers) of nucleic acid structures are the **nucleotides**.

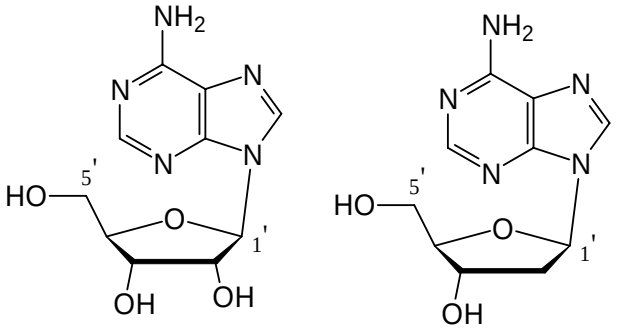
Nucleotides themselves are built up of three structural elements – a heterocyclic aromatic base, a sugar unit (either ribose or 2-deoxyribose) and a phosphate.

e.g. Adenosine 5'-monophosphate is a **nucleotide**.



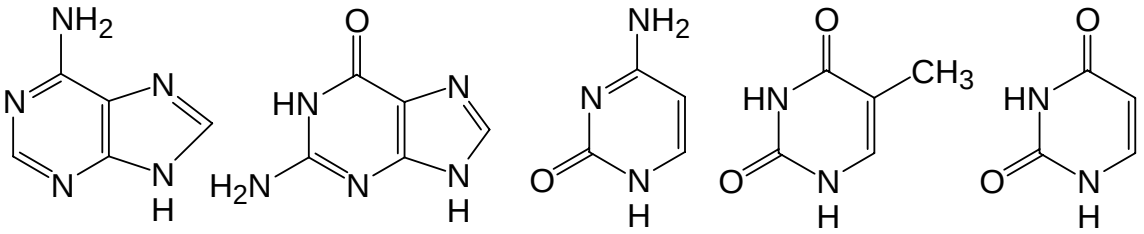
When there is no phosphate the compound is referred to as a **Nucleoside**.

e.g. Adenosine and Deoxyadenosine are nucleosides.



Ribose is found in RNA nucleosides and 2-deoxyribose in DNA nucleosides.

DNA contains four nucleotides – each with a different heterocyclic base component. Two of these are **purines** and two are **pyrimidines**.



adenine guanine cytosine thymine uracil

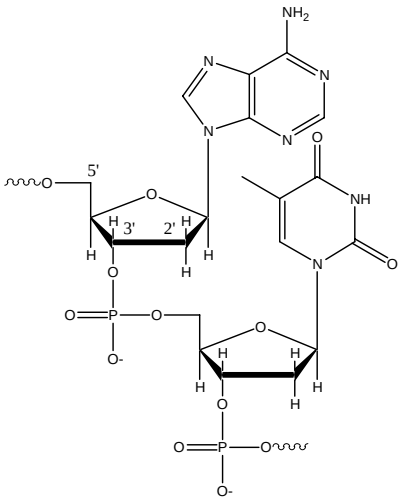
purines **pyrimidines**

Adenine, guanine and cytosine also occur in RNA **but thymine is replaced by uracil**.

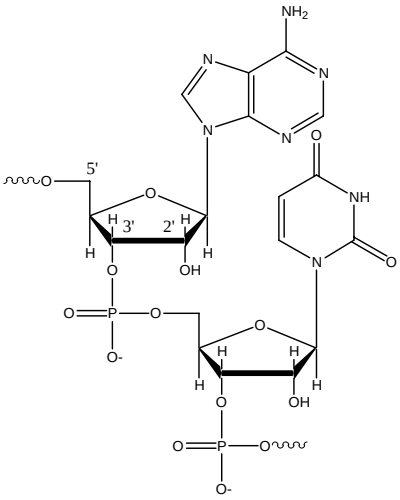
The bases are given single letter codes – A,G,C, T and U.

<u>The primary structure of Nucleic acids</u>.

DNA and RNA are polymers of 2'-deoxynucleotides and nucleotides, respectively. In both cases the monomeric units are linked by 3'- 5'- phosphate groups. (sometimes refered to as the phosphate backbone)

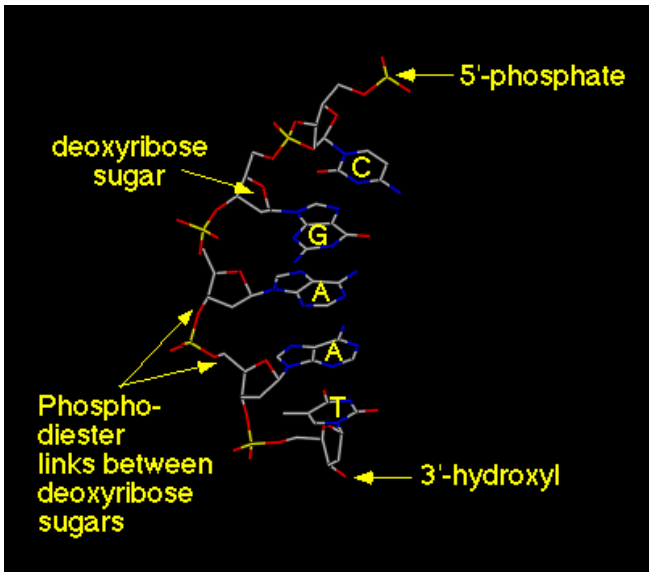


DNA fragment



RNA fragment

Example of a DNA Backbone: single stranded 5'-d(CGAAT) sequence



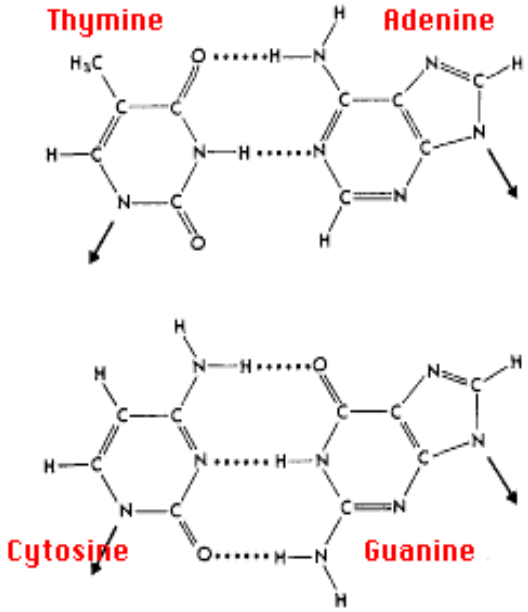
The DNA backbone is a polymer with an alternating sugar-phosphate sequence. The deoxyribose sugars are joined at both the 3'-hydroxyl and 5'-hydroxyl groups to phosphate groups in ester links, also known as "phosphodiester" bonds.

NOTE

- The alternating backbone of deoxyribose and phosphodiester groups
- Chain has a direction (known as polarity), 5'- to 3'- from top to bottom
- Oxygens (red atoms) of phosphates are polar and negatively charged
- A, G, C, and T bases can extend away from chain, and stack atop each other
- Bases are hydrophobic

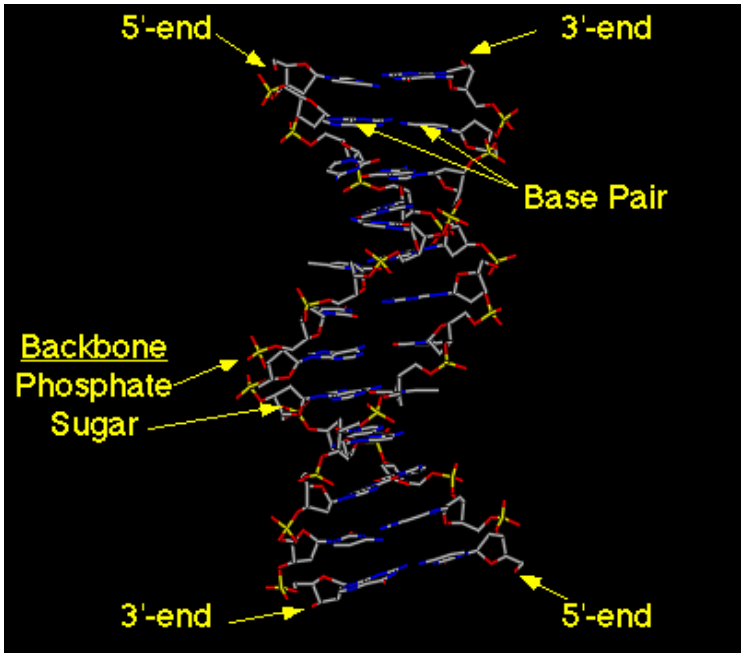
The DNA Double Helix

DNA is a normally a double stranded macromolecule. Two polynucleotide chains, held together by hydrogen bonding between the bases, form a DNA molecule.



The four bases form specific H-bonded pairs

- A only bonds with T
- G only bonds with C.



Features of the DNA Double Helix

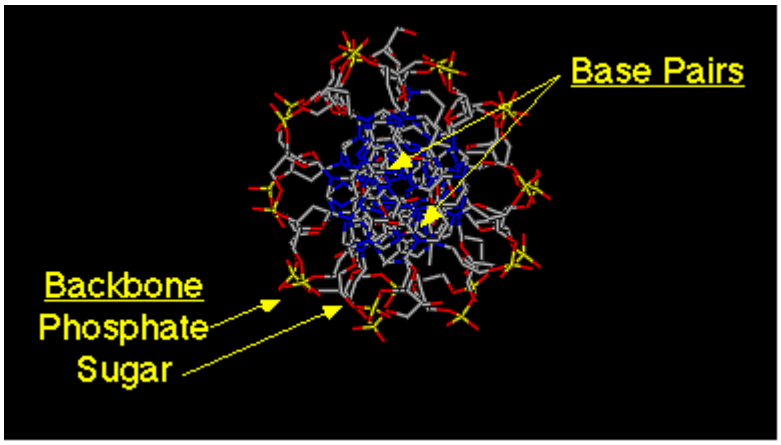
Two DNA strands form a helical spiral, winding around a helix axis in a right-handed spiral

The two polynucleotide chains run in opposite directions

The sugar-phosphate backbones of the two DNA strands wind around the helix axis like the railing of a spiral staircase

The bases of the individual nucleotides are on the inside of the helix, stacked on top of each other like the steps of a spiral staircase.

Two grooves – major and minor, run along the helix.



View down the helix axis

RNA Structure

Unlike DNA, RNA's do not form a helix. Although single stranded, some RNA molecules can adopt a number of defined shapes. Different types of sequences, which have different functions in the cell, adopt different conformations. The structures can be partially single stranded and may adopt stabilised loops or 'hairpin' bends with the same strand.

An RNA molecule – note the H-bonding between bases (the green shaded region) forming a short helix in part of the structure.