Edexcel AS Biology Teacher 1
Contents

Neil C Millar (nmillar@ntlworld.co.uk)
Head of Biology, Heckmondwike Grammar School
High Street, Heckmondwike, WF16 0AH
July 2015

Biology Teacher 1 Specification

1.01 Water
The importance of the dipole nature of water leading to hydrogen bonding and the significance of the following to organisms: high specific heat capacity; polar solvent; surface tension; incompressibility; maximum density at 4 °C.

1.02 Carbohydrates.
The difference between monosaccharides, disaccharides and polysaccharides. The structure of the hexose glucose (alpha and beta) and the pentose ribose. How monosaccharides (glucose, fructose, galactose) join to form disaccharides (sucrose, lactose and maltose) and polysaccharides (starch formed from amylose and amylopectin; glycogen) through condensation reactions forming glycosidic bonds, and how these can be split through hydrolysis reactions. How the structure of glucose, starch, glycogen and cellulose relates to their function.

1.03 Lipids
How a triglyceride is synthesised, including the formation of ester bonds during condensation reactions between glycerol and three fatty acids. The differences between saturated and unsaturated lipids. How the structure of lipids relates to their role in energy storage, waterproofing and insulation. How the structure and properties of phospholipids relate to their function in cell membranes.

1.04 Proteins
The structure of an amino acid (structures of specific amino acids are not required). The formation of polypeptides and proteins (as amino acid monomers linked by peptide bonds in condensation reactions). The role of ionic, hydrogen and disulphide bonding in the structure of proteins. The significance of the primary, secondary, tertiary and quaternary structure of a protein in determining the properties of fibrous and globular proteins, including collagen and haemoglobin. How the structure of collagen and haemoglobin are related to their function.

1.05 Enzymes
Enzymes are catalysts that reduce activation energy. Enzymes catalyse a wide range of intracellular reactions as well as extracellular ones. The structure of enzymes as globular proteins. The concepts of specificity and the induced fit hypothesis.

How the initial rate of enzyme activity can be measured and why this is important. Temperature, pH, substrate and enzyme concentration affect the rate of enzyme activity. Enzymes can be affected by competitive, non-competitive and end-product inhibition.

1.06 DNA
The structure of DNA, including the structure of the nucleotides (purines and pyrimidines), base pairing, the two sugar-phosphate backbones, phosphodiester bonds and hydrogen bonds. How DNA is replicated semi-conservatively, including the role of DNA helicase, polymerase and ligase.

1.07 Gene Expression
A gene is a sequence of bases on a DNA molecule coding for a sequence of amino acids in a polypeptide chain. The structure of mRNA including nucleotides, the sugar phosphate backbone and the role of hydrogen bonds. The structure of tRNA, including nucleotides, the role of hydrogen bonds and the anticodon.

The nature of the genetic code, including triplets coding for amino acids, start and stop codons, degenerate and non-overlapping nature, and that not all the genome codes for proteins. The processes of transcription in the nucleus and translation at the ribosome, including the role of sense and anti-sense DNA, mRNA, tRNA and the ribosomes.

1.08 Gene Mutations
The term gene mutation as illustrated by base deletions, insertions and substitutions. The effect of point mutations on amino acid sequences, as illustrated by sickle cell anaemia in humans.

1.09 Viruses
The classification of viruses is based on structure and nucleic acid types as illustrated by λ (lambda) phage (DNA), tobacco mosaic virus and Ebola (RNA) and human immunodeficiency virus (RNA retrovirus). The lytic cycle of a virus and latency.

1.10 Viral Diseases
Viruses are not living cells and so antivirals must work by inhibiting virus replication. as viruses can be difficult to treat once infection has occurred, the focus of disease control should be on preventing the spread, as exemplified by the 2014 Ebola outbreak in West Africa. Be able to evaluate the ethical implications of using untested drugs during epidemics.

1.11 Cell Cycle and Mitosis
The cell cycle is a regulated process in which cells divide into two identical daughter cells, and that this process consists of three main stages: interphase, mitosis and cytokinesis. What happens to genetic material during the cell cycle, including the stages of mitosis. Mitosis contributes to growth, repair and asexual reproduction.

1.12 Meiosis
Meiosis results in haploid gametes, including the stages of meiosis. Meiosis results in genetic variation through

recombination of alleles, including independent assortment and crossing over.

1.13 Chromosome Mutations
What chromosome mutations are, as illustrated by translocations. How non-disjunction can lead to polysomy, including Down's syndrome, and monosomy, including Turner's syndrome.

1.14 Sexual Reproduction in Mammals
The processes of oogenesis and spermatogenesis. The events of fertilisation from the first contact between the gametes to the fusion of nuclei. The early development of the embryo to blastocyst stage.

1.15 Sexual Reproduction in Plants
How a pollen grain forms in the anther and the embryo sac forms in the ovule. How the male nuclei formed by division of the generative nucleus in the pollen grain reach the embryo sac, including the roles of the tube nucleus, pollen tube and enzymes. The process of double fertilisation inside the embryo sac to form a triploid endosperm and a zygote.

1.16 Classification
The limitations of the definition of a species as a group of organisms with similar characteristics that interbreed to produce fertile offspring. Why it is often difficult to assign organisms to any one species or to identify new species.

The classification system consists of a hierarchy of domain, kingdom, phylum, class, order, family, genus and species. The evidence for the three-domain model of classification as an alternative to the five-kingdom model and the role of the scientific community in validating this evidence.

DNA sequencing and bioinformatics can be used to distinguish between species and determine evolutionary relationships. How gel electrophoresis can be used to distinguish between species and determine evolutionary relationships.

Organisms occupy niches according to physiological, behavioural and anatomical adaptations.

1.17 Natural selection
Evolution can come about through natural selection acting on variation bringing about adaptations. Reproductive isolation can lead to allopatric and sympatric speciation. The role of scientific journals, the peer review process and scientific conferences in validating new evidence supporting the accepted scientific theory of evolution. There is an evolutionary race between pathogens and the development of medicines to treat the diseases they cause.

1.18 Biodiversity
Biodiversity can be assessed at different scales:
- within a habitat at the species level using a formula to calculate an index of diversity
- within a species at the genetic level by looking at the variety of alleles in the gene pool of a population.

1.19 Conservation
The ethical and economic reasons (ecosystem services) for the maintenance of biodiversity. The principles of ex-situ (zoos and seed banks) and in-situ conservation (protected habitats), and the issues surrounding each method.

BLANK PAGE

Biological Molecules

Living things are made up of thousands and thousands of different chemicals. These chemicals are called organic because they contain the element carbon. In science organic compounds contain carbon–carbon bonds, while inorganic compounds don't. There are four important types of organic molecules found in living organisms: carbohydrates, lipids, proteins, and nucleic acids (DNA). These molecules are mostly polymers, very large molecules made up from very many small molecules, called monomers. 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	Elements	Monomers	Polymers	% dry mass of a cell
Carbohydrates	CHO	monosaccharides	polysaccharides	15
Lipids	CHOP	fatty acids + glycerol*	triglycerides*	10
Proteins	CHONS	amino acids	polypeptides	50
Nucleic acids	CHONP	nucleotides	polynucleotides	18

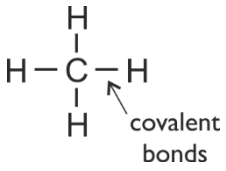
* Triglycerides are not polymers, since they are formed from just four molecules, not many (see p12).

We'll study each of these groups in turn.

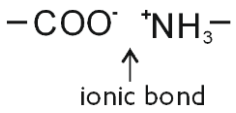
Chemical Bonds

In biochemistry there are three important types of chemical bond.

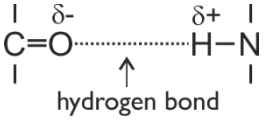
Covalent bonds are strong. They are the main bonds holding the atoms together in the organic molecules in living organisms. Because they are strong, covalent bonds don't break or form spontaneously at the temperatures found in living cells. So in biology covalent bonds are always made or broken by the action of enzymes. Covalent bonds are represented by solid lines in chemical structures.



Ionic Bonds are fairly strong. They are formed between a positive ion (such as NH_3^+) and a negative ion (such as COO^-). They are not common in biology since ionic compounds dissociate in solution, but ionic bonds are sometimes found inside protein molecules.

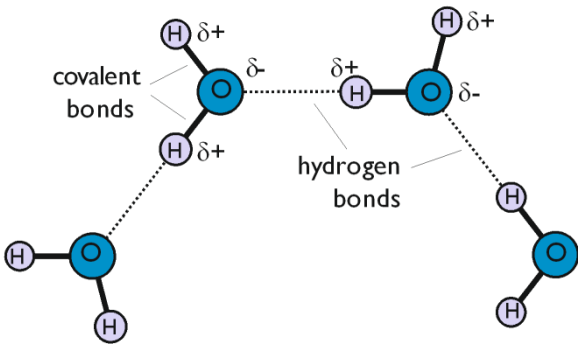


Hydrogen bonds are much weaker. They are formed between an atom (usually hydrogen) with a slight positive charge (denoted $\delta+$) and an atom (usually oxygen or nitrogen) with a slight negative charge (denoted $\delta-$). Because hydrogen bonds are weak they can break and form spontaneously at the temperatures found in living cells without needing enzymes. Hydrogen bonds are represented by dotted lines in chemical structures.



Water

Life on Earth evolved in the water, and all life still depends on water. At least 80% of the total mass of living organisms is water. Water molecules are a charged dipole, with the oxygen atom being slightly negative (δ-) and the hydrogen atoms being slightly positive (δ+). These opposite charges attract each other, forming hydrogen bonds that bind water molecules loosely together.



This dipole property of water gives it many specific properties that have important implications in biology.

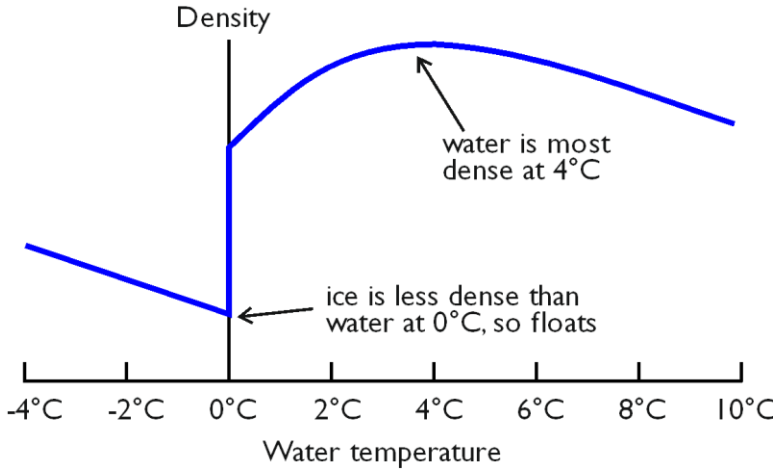
1. **Water is an extremely good solvent.** The water dipoles will stick to the atoms in almost all crystalline solids, causing them to dissolve. Substances are often transported around living organisms as solutes in aqueous solution (e.g. in blood or sap) and almost all the chemical reactions of life take place in solution.
   - Charged or polar molecules such as salts, sugars, 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").

Many important biological molecules ionise when they dissolve (e.g. acetic acid $\rightleftharpoons$ acetate$^-$ + H^+), so the names of the acid and ionised 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 ionised form is the one found in living cells.

2. **Water has a High Specific Heat.** Water has a high specific heat capacity, which means that it takes a lot of energy to heat, so water does not change temperature very easily. This minimises fluctuations in temperature inside cells, and it also means that sea temperature is remarkably constant.
3. **Water has a High Latent Heat.** Water requires a lot of energy to change state from a liquid into a gas, since so many hydrogen bonds have to be broken. So as water evaporates it extracts heat from around it, and this is used to cool animals (sweating and panting) and plants (transpiration). Water also

must lose a lot of heat to change state from a liquid to a solid. This means it is difficult to freeze water, so ice crystals are less likely to form inside cells.

4. **Water is cohesive and adhesive.**
   - Cohesion means that water molecules "stick together" due to their hydrogen bonds. 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.
   - Adhesion means that water molecules stick to other surfaces, such as xylem vessels. This explains capillary action (where water will be drawn along a narrow tube) and the meniscus on test tube walls.
5. **Water is most dense at 4°C.** Most substances get denser as they cool down, and the solid form is denser than the liquid form. Water is unique in that the solid state (ice) is less dense that the liquid state, and in fact water is most dense at 4°C. This property causes several important effects:

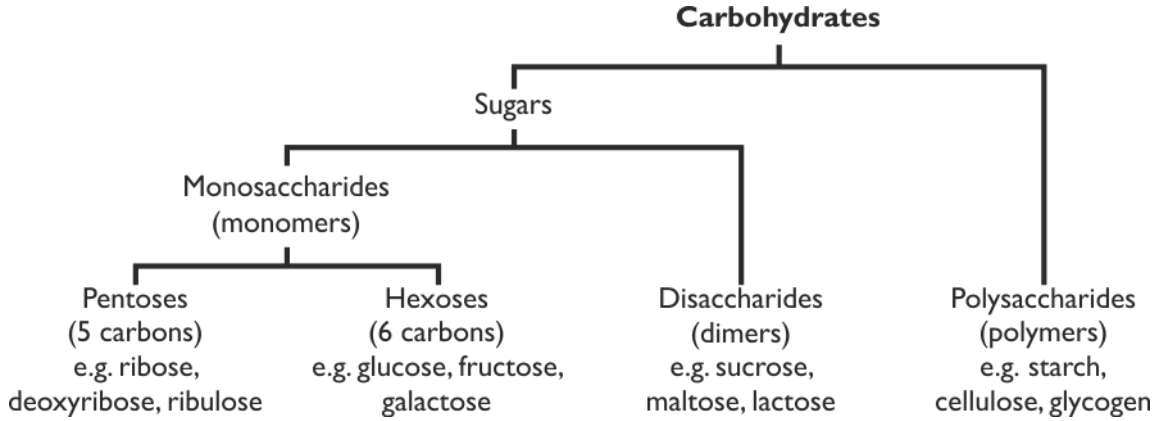


   - Ice floats on water, so as the air temperature cools, bodies of water freeze from the surface, forming a layer of ice with liquid water underneath. This allows aquatic ecosystems to exist in sub-zero temperatures, and even throughout long ice ages.
   - The expansion of water as it freezes causes freeze-thaw erosion of rocks, which results in the formation of soil, without which there could be no terrestrial plant life.
   - Cold water sinks below warm water, and warm water rises above cold water, which gives rise to many ocean currents.

6. **Water is incompressible.** The hydrogen bonds hold water molecules closer together than other liquids, so water is very incompressible, since the molecules can't be pushed any closer. So if a force is applied to water, the water will move rather than squash, which allows blood to be pumped round a body. The incompressibility is also used to make plant cells turgid and give eyes their shape.

Carbohydrates

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

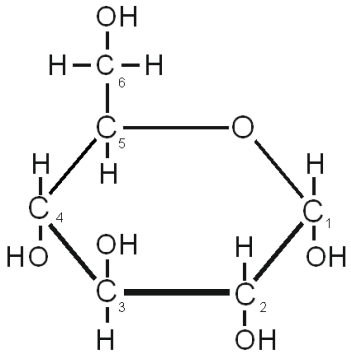


Monosaccharides

Monosaccharides all have the formula $(CH_2O)_n$, where n can be 3-7.

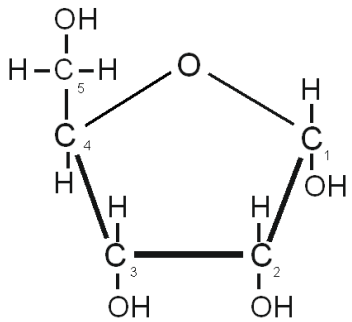
- **Hexose sugars** have six carbon atoms, so have the formula $C_6H_{12}O_6$. Hexose sugars include glucose, galactose and fructose. These are isomers, with the same chemical formula ($C_6H_{12}O_6$), but different structural formulae. In animals glucose is the main transport sugar in the blood, and its concentration in the blood is carefully controlled.
- **Pentose sugars** have five carbon atoms, so have the formula $C_5H_{10}O_5$. Pentose sugars include ribose and deoxyribose (found in nucleic acids and ATP) and ribulose (which occurs in photosynthesis).
- **Triose sugars** have three carbon atoms, so have the formula $C_3H_6O_3$. Triose sugars are found in respiration and photosynthesis.

Structural formula for α-Glucose
($C_6H_{12}O_6$)



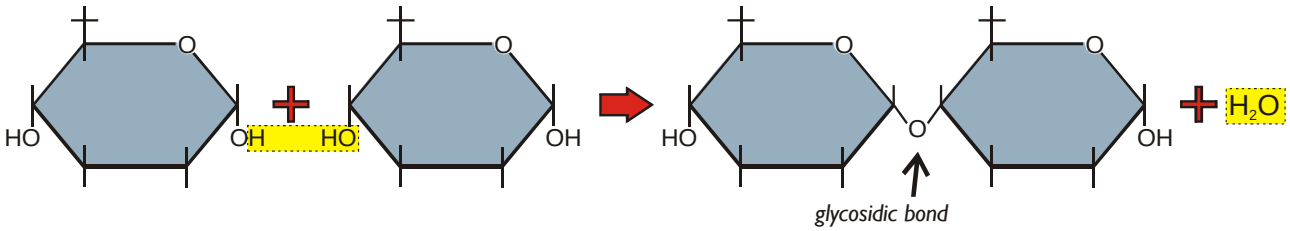
You need to know these formulae!

Structural formula for Ribose
($C_5H_{10}O_5$)



Disaccharides

Disaccharides are formed when two monosaccharides are joined together by a glycosidic bond (C–O–C). The reaction involves the formation of a molecule of water (H_2O):



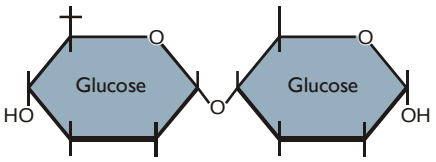
This shows two glucose molecules joining together to form the disaccharide maltose. This kind of reaction, where two molecules combine into one bigger molecule, is called a condensation reaction. The reverse process, where a large molecule is broken into smaller ones by reacting with water, is called a hydrolysis reaction.

In general:

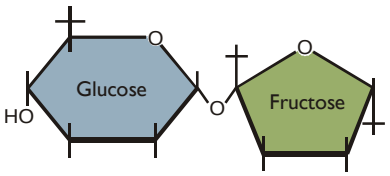
- polymerisation reactions are condensations
- breakdown reactions are hydrolyses

There are three common disaccharides:

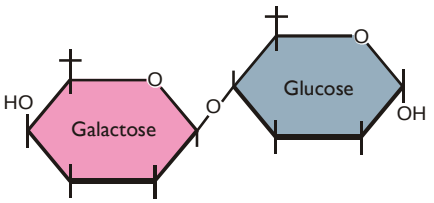
Maltose (or malt sugar) is glucose–glucose. It is formed on digestion of starch by amylase, because this enzyme breaks starch down into two-glucose units. Brewing beer starts with malt, which is a maltose solution made from germinated barley.



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 is the common table sugar that you put in your tea.



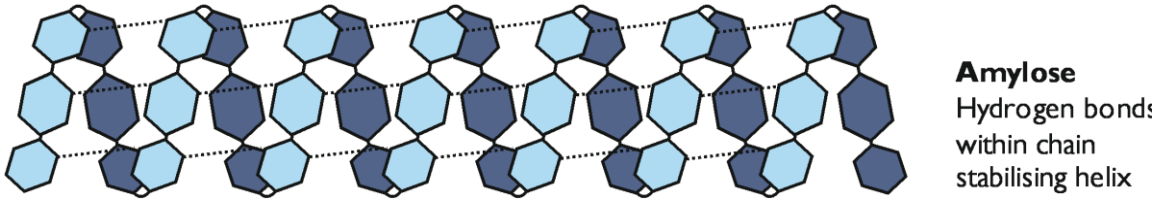
Lactose (or milk sugar) is galactose–glucose. It is found only in mammalian milk, and is the main source of energy for infant mammals.



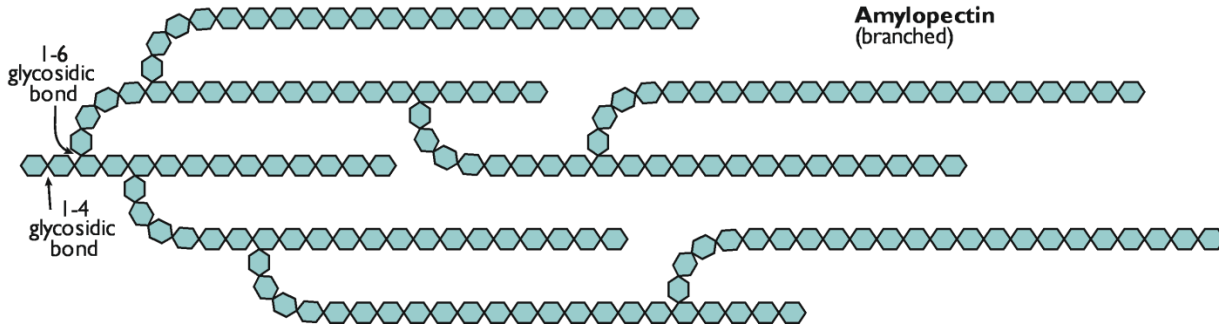
Polysaccharides

Polysaccharides are chains of many glucose monomers (often 1000s) joined together by glycosidic bonds. The main polysaccharides are starch, glycogen and cellulose.

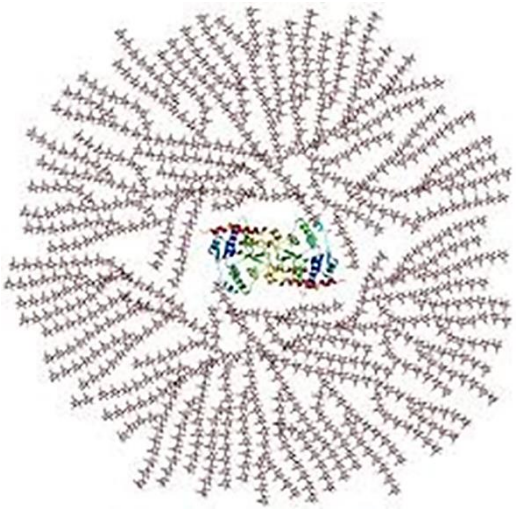
1. **Starch** is the plant storage polysaccharide. It is insoluble and forms starch granules inside many plant cells. Being insoluble means starch does not change the water potential of cells, so does not cause the cells to take up water by osmosis. It is not a pure substance, but is a mixture of amylose and amylopectin. Amylose is poly-(1-4) glucose, so is a long glucose chain that coils up into a helix held together by hydrogen bonds.



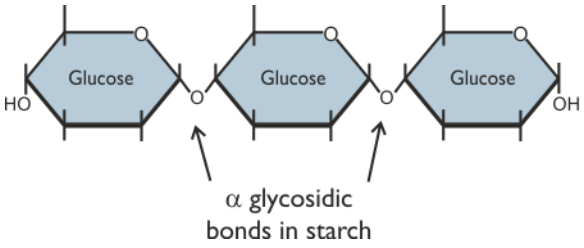
Amylopectin is poly(1-4) glucose with about 4% (1-6) branches. This gives it a more open molecular structure than amylose. Because it has more ends, it can be broken more quickly than amylose by amylase enzymes. Both amylose and amylopectin are broken down by the enzyme amylase into maltose, though at different rates.

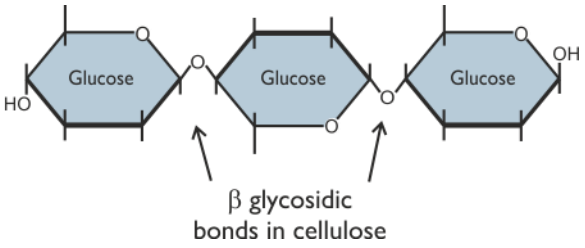


2. **Glycogen** is the animal storage polysaccharide and is found mainly in muscle and liver cells. It is similar in structure to amylopectin: poly (1-4) glucose with 9% (1-6) branches, and is sometimes called animal starch. Because it is so highly branched, it can be mobilised (broken down to glucose for energy) very quickly. It is broken down to glucose by the enzyme glycogen phosphorylase.

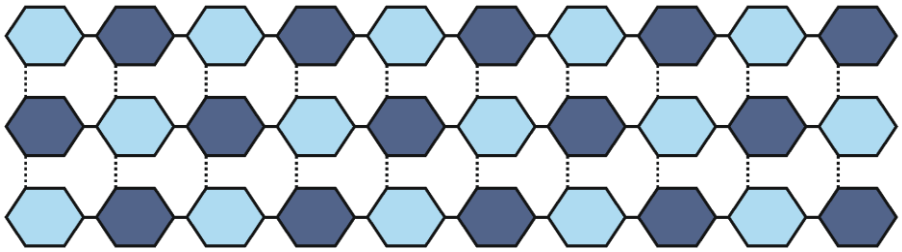


3. **Cellulose** is only found in plants, where it is the main component of cell walls. It is poly (1-4) glucose, but with a different isomer of glucose. Starch and glycogen contain α-glucose, while cellulose contains β-glucose, with a different position of the hydroxyl group on carbon 1. This means that in a cellulose chain alternate glucose molecules are inverted.





This apparently tiny difference makes a huge difference in structure and properties. The α bond is flexible so starch molecules can coil up, but the β bond is rigid, so cellulose molecules form straight chains. Hundreds of these chains are linked together by hydrogen bonds between the chains to form cellulose microfibrils. These microfibrils are very strong and rigid, and give strength to plant cells, and therefore to young plants and also to materials such as paper, cotton and sellotape.



Cellulose
Hydrogen bonds between chains forming microfibril

The β-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. Carnivores and omnivores cannot digest cellulose, and in humans it is referred to as fibre.

Starch and Glycogen	Cellulose
α glycosidic bonds	β glycosidic bonds
flexible chains	straight chains
H bonds within each chain, forming helix	H bonds between chains, forming microfibrils
Can form H-bonds with water, so can be soluble	Can't form H bonds with water, so insoluble
Reacts with iodine to form blue-black complex	Doesn't react with iodine
Easy to digest	Difficult to digest
Storage role	Structural role

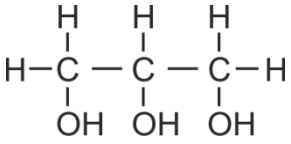
Lipids

Lipids are a mixed group of hydrophobic compounds composed of the elements carbon, hydrogen, oxygen and sometime phosphorus (CHOP). The most common lipids are triglycerides and phospholipids.

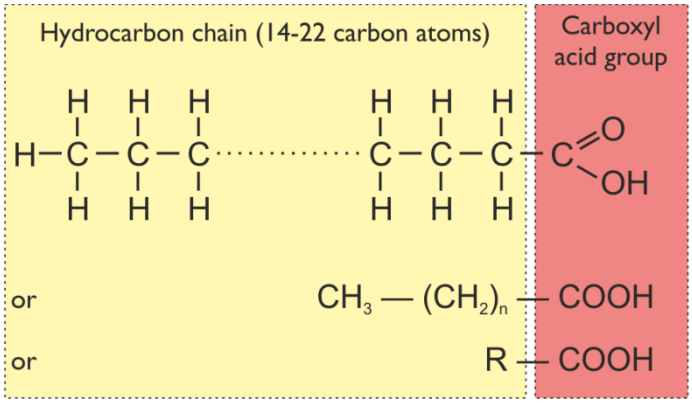
Triglycerides

Triglycerides, or triacylglycerols, are commonly known as fats or oils. They are made of glycerol and fatty acids.

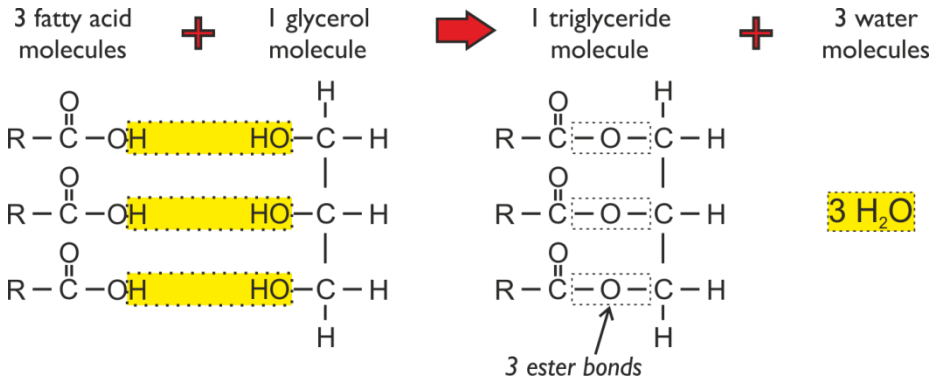
Glycerol is a small, 3-carbon molecule with three alcohol (OH) groups.



Fatty acids are long molecules made of a non-polar hydrocarbon chain with a polar carboxyl acid group at one end. The hydrocarbon chain can be from 14 to 22 CH_2 units long. Because the length of the hydrocarbon chain can vary it is sometimes called an R group, so the formula of a fatty acid can be written as R-COOH.



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

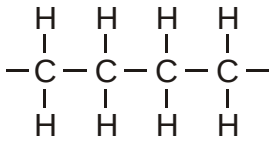


Triglycerides are found in fatty (or adipose) tissue. They are used for:

- **Energy storage**. Triglyceride respiration yields more energy per unit mass than other compounds, so adipose tissue is used a long-term energy store. However, triglycerides can't be mobilised quickly since they are so insoluble, so are no good for quick energy requirements. Tissues that need energy quickly (like muscles) instead use glycogen.
- **Insulation**. Adipose tissue under the skin (sub-cutaneous) is used to insulate warm-blooded mammals against heat loss e.g. blubber in whales. A fatty myelin sheath electrically insulates nerve cells so the electrical impulses travel faster.
- **Waterproofing**. Mammals' fur and birds' feathers contain the lipid lanolin for waterproofing. Insect exoskeletons contain waxy lipids to stop water loss, and plants have a lipid waxy cuticle to reduce water loss.

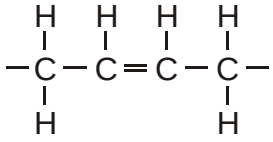
Saturated and Unsaturated Fats

- If the fatty acid chains in a triglyceride have no C=C double bonds, then they are called saturated fatty acids (i.e. saturated with hydrogen). Triglycerides with 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.



saturated

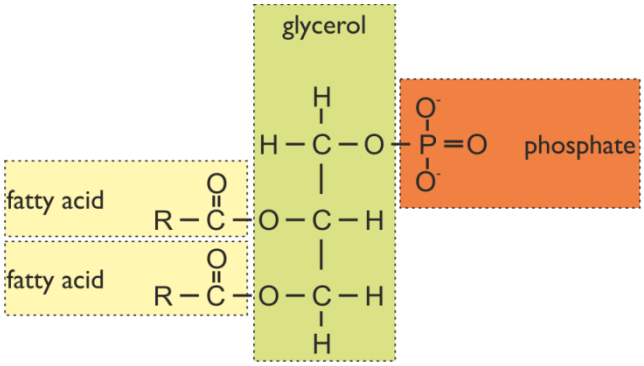
- If the fatty acid chains in a triglyceride do have C=C double bonds they are called unsaturated fatty acids (i.e. unsaturated with hydrogen). Fatty acids with more than one double bond are called poly-unsaturated fatty acids (PUFAs). Triglycerides with unsaturated fatty acids have a low melting point and tend to be found in cold-blooded animals and plants. At room temperature they are liquids (oils), e.g. fish oil, vegetable oils. An "omega number" is sometimes used to denote the position of a double bond, e.g. omega-3 fatty acids.



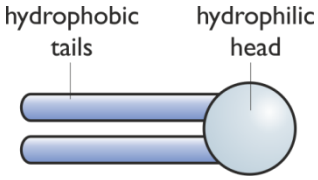
unsaturated

Phospholipids

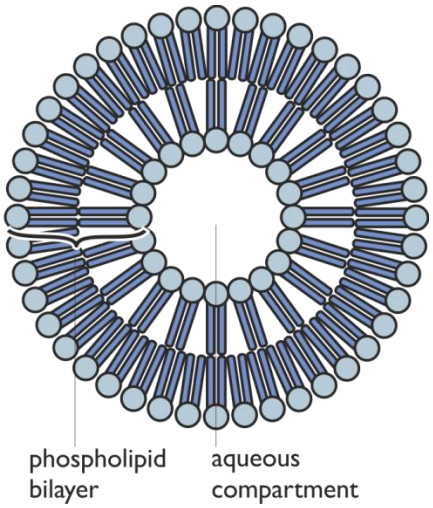
Lipids have a very low density, so the body fat of water mammals helps them to float easily Phospholipids have a similar structure to triglycerides, but with a phosphate group in place of one fatty acid chain. There may also be other groups attached to the phosphate. Phospholipids have a polar hydrophilic "head" (the negatively-charged phosphate group) and two non-polar hydrophobic "tails" (the fatty acid chains).



or



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 a double-layered phospholipid bilayer. The hydrophilic heads facing the water and the hydrophobic tails facing each other. 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 (see pxx).



Proteins

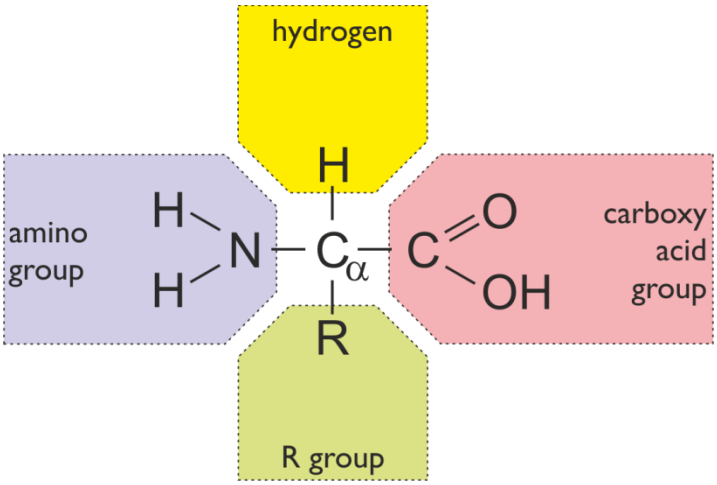
Proteins are the most complex and most diverse group of biological compounds. They have an astonishing range of different functions, as this list shows.

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^+K^+ 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. cholera toxin
antifreeze	e.g. glycoproteins in arctic flea

and many more!

Amino Acids

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

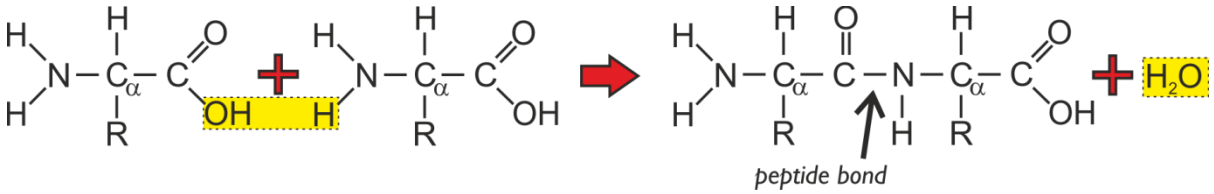


1. a hydrogen atom
2. a basic amino group (NH_2 or NH_3^+)
3. an acidic carboxyl group (COOH or COO^-)
4. a variable "R" group (or side chain)

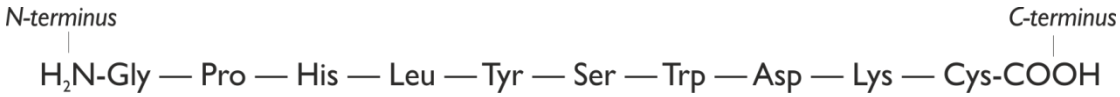
There are 20 different R groups, and so 20 different amino acids. Since each R group is slightly different, each amino acid has different properties, and this in turn means that proteins can have a wide range of properties. The table on the next page shows the 20 different R groups, grouped by property, 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. There are 3-letter and 1-letter abbreviations for each amino acid.

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 (NH_2) group, called the N-terminus, and one end with a free carboxyl (COOH) group, called the C-terminus.

In a protein the polypeptide chain may be many hundreds of amino acids long. Amino acid polymerisation to form polypeptides is part of protein synthesis. It takes place in ribosomes, and is special because it requires an RNA template. The sequence of amino acids in a polypeptide chain is determined by the sequence of the bases in DNA. Protein synthesis is studied in detail on pxx.

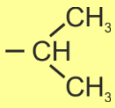
The Twenty Amino Acid R-Groups

Simple R groups

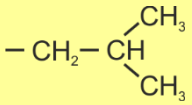
Glycine
Gly G

Alanine
Ala A

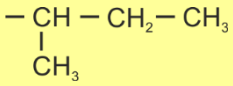
Valine
Val V



Leucine
Leu L

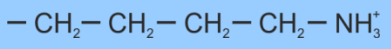


Isoleucine
Ile I

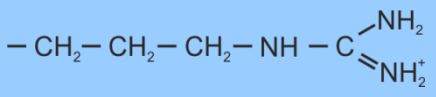


Basic R groups

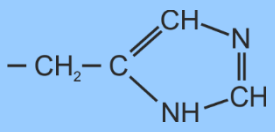
Lysine
Lys K



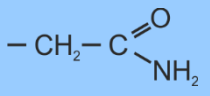
Arginine
Arg R



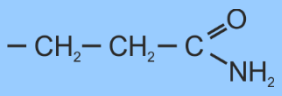
Histidine
His H



Asparagine
Asn N

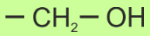


Glutamine
Gln Q

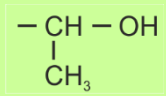


Hydroxyl R groups

Serine
Ser S

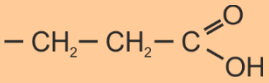


Threonine
Thr T

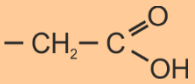


Acidic R groups

Aspartate
Asp D

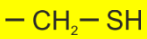


Glutamate
Glu E

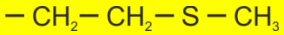


Sulphur R groups

Cysteine
Cys C

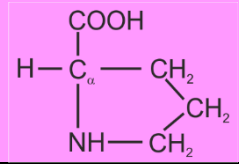


Methionine
Met M



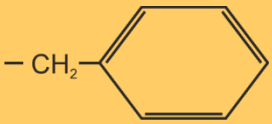
Cyclic R group

Proline
Pro P

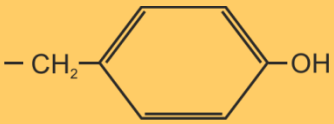


Ringed R groups

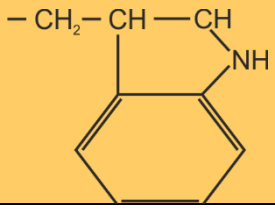
Phenylalanine
Phe F



Tyrosine
Tyr Y



Tryptophan
Trp W



Protein Structure

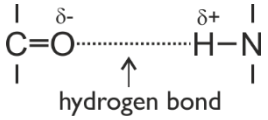
Polypeptides are just strings of amino acids, but they fold up and combine to form the complex and well-defined three-dimensional structure of working proteins. To help to understand protein structure, it is broken down into four levels:

1. Primary Structure

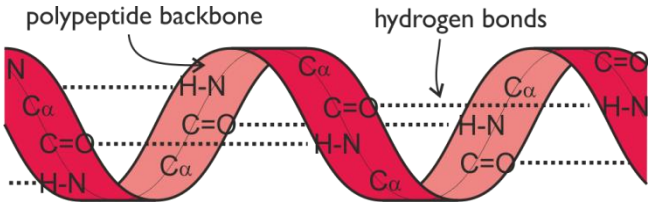
This is just the sequence of amino acids in the polypeptide chain, so is not really a structure at all. However, the primary structure does determine the rest of the protein structure. Most polypeptide chains contain hundreds or even thousands of amino acids.

2. Secondary Structure

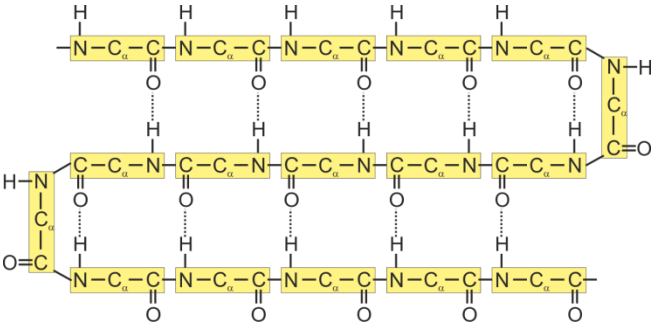
This is the most basic level of protein folding, and consists of a few basic motifs that are found in almost all proteins. The secondary structure is held together by hydrogen bonds between the carboxyl groups and the amino groups in the polypeptide backbone. The two most common secondary structure motifs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



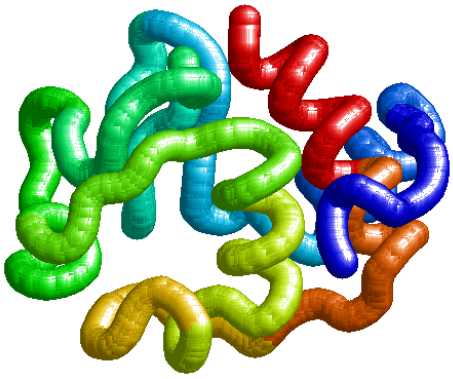
The α-helix. The polypeptide chain is wound round to form a helix. It is held together by hydrogen bonds running parallel with the long helical axis. There are so many hydrogen bonds that this is a very stable and strong structure. Do not confuse the α-helix of proteins with the famous double helix of DNA – helices are common structures throughout biology.



The β-sheet. The polypeptide chain zig-zags back and forward forming a sheet of antiparallel strands. Once again it is held together by hydrogen bonds.



3. Tertiary Structure

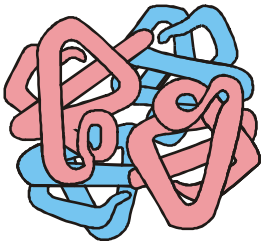


This is the complete structure formed by the folding up of a polypeptide chain. Every protein has a unique tertiary structure, which is responsible for its properties and function. For example the shape of the active site in an enzyme is due to its tertiary structure. The tertiary structure is held together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

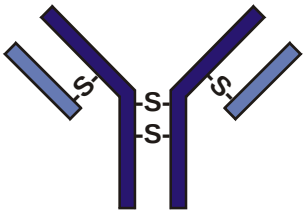
- hydrogen bonds, which are weak but numerous.
- Ionic bonds (or salt bridges) between oppositely-charged R-groups e.g. NH_3^+ in lysine or arginine with COO^- in aspartate or glutamate. These ionic bonds are stronger than hydrogen bonds but weaker than covalent bonds.
- covalent S–S bonds called <u>sulphur bridges</u> between two cysteine amino acids, which are much stronger.

4. Quaternary Structure

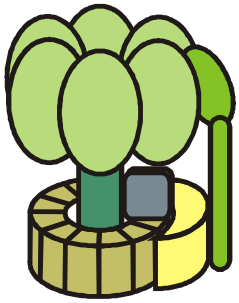
Almost all working proteins are actually composed of more than one polypeptide chain, and the quaternary structure is the arrangement of the different chains. There are a huge variety of quaternary structures e.g.:



Haemoglobin consists of four chains arranged in a tetrahedral (pyramid) structure.



Antibodies comprise four chains arranged in a Y-shape.



The enzyme ATP synthase is composed of 22 chains forming a rotating motor.

Collagen consists of three chains in a triple helix structure.

Actin consists of hundreds of globular chains arranged in a long double helix.

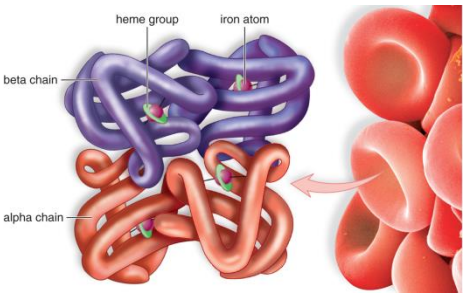
These four structures are not real stages in the formation of a protein, but are simply a convenient classification that scientists invented to help them to understand proteins. In fact proteins fold into all these structures at the same time, as they are synthesised.

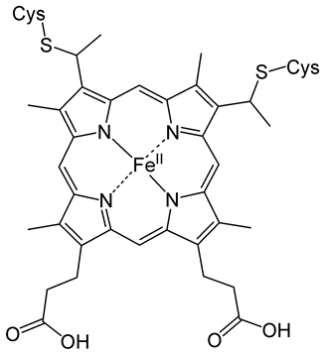
The final three-dimensional shape of a protein can be classified as globular or fibrous.

Globular Proteins e.g. Haemoglobin

The vast majority of proteins are globular, i.e. they have a compact, roughly spherical structure. This group includes enzymes, membrane proteins, receptors, transport proteins and storage proteins.

Haemoglobin is a globular protein found in red blood cells. One molecule is composed of four globular polypeptide chains called globins. There are two α chains with 141 amino acids each and two β chains with 146 amino acids each, giving a total of 574 amino acids. Each chain contains a small non-polypeptide group called haem, which has an iron atom at its centre. The haem groups are attached to the globin polypeptide chains by covalent sulphur bridges.



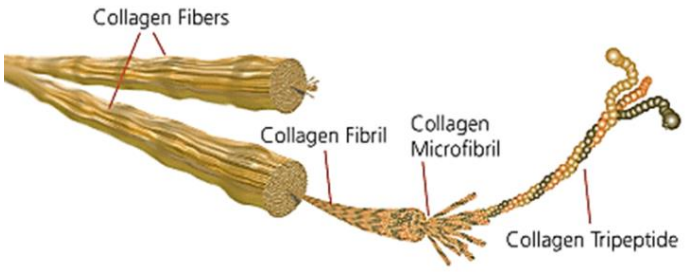


One oxygen molecule (O_2) can bind to each iron atom, so a haemoglobin molecule can bind up to four O_2 molecules. The polypeptide chains provide a suitable environment for the oxygen molecules to bind reversibly, so haemoglobin acts as an effective oxygen transport protein.

Fibrous Proteins e.g. Collagen

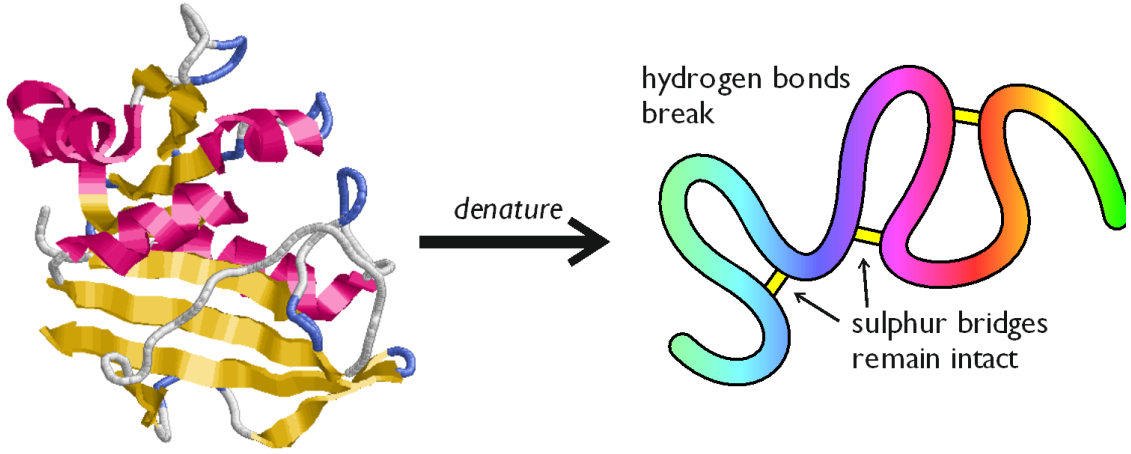
Fibrous proteins are long and thin, like ropes. They tend to have structural roles, such as collagen (bone), keratin (hair), tubulin (cytoskeleton), actin (muscle), fibrin (blood clots) and fibroin (silk). They are always composed of many polypeptide chains.

Collagen is a fibrous protein found in bone and cartilage. A single molecule consists of three long polypeptide chains linked by numerous hydrogen bonds and wrapped round each other in a triple helix. Each polypeptide chain is about 10,000 amino acids long and contains a simple repeating sequence of just three amino acids (e.g. Gly-Pro-Ala). Many collagen molecules bind together to form a fibril, and many of these fibrils link to form fibres. These collagen fibres have high tensile strength and give strength and flexibility to cartilage, ligaments, tendons, bone, skin and blood vessels. Collagen is also found in the cornea and lens of the eye. The synthesis of collagen requires vitamin C, which is why vitamin C deficiency causes the disease scurvy, where connective tissue breaks down.

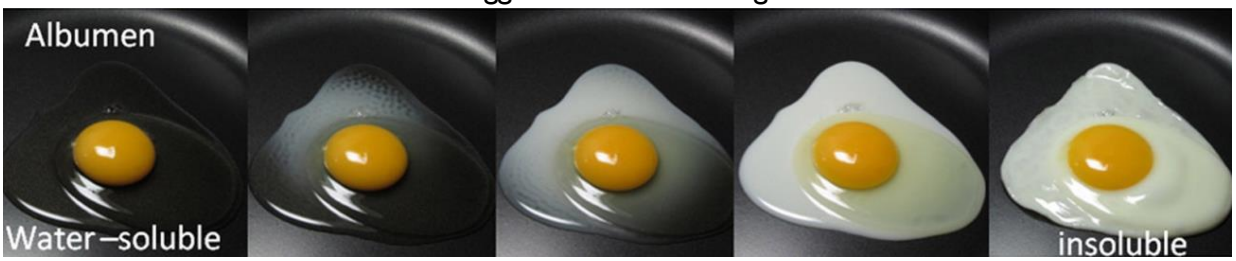


Protein Denaturing

Since the secondary, tertiary and quaternary structures are largely held together by hydrogen bonds, the three-dimensional structure of proteins is lost if the hydrogen bonds break. The polypeptide chain just folds up into a random coil and the protein loses its function. This is called denaturing, and happens at temperatures above about 50°C or at very low or high pH. Covalent bonds are not broken under these conditions, so the primary structure is maintained (as are sulphur bridges).



Egg albumin denaturing



Enzymes

Enzymes are biological catalysts. There are about 40,000 different enzymes in living cells, each controlling a different chemical reaction. They increase the rate of reactions by a factor of between 10^6 to 10^{12} times, allowing the chemical reactions that make life possible to take place at normal temperatures. They were discovered in fermenting yeast in 1900 by Buchner, and the name enzyme means "in yeast".

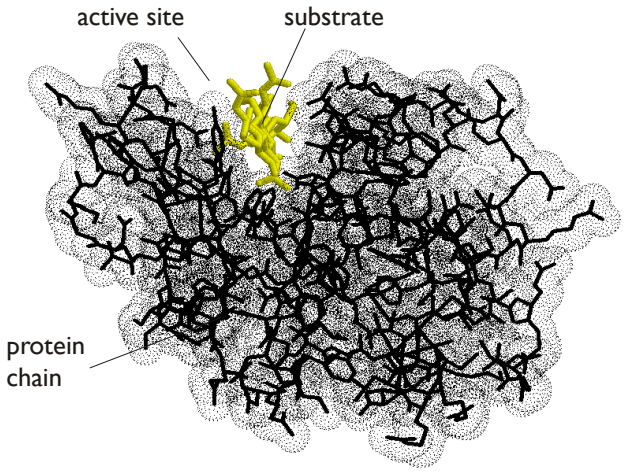
- **Intracellular enzymes** catalyse all the metabolic reactions inside cells and organelles (such as respiration, photosynthesis, DNA replication and protein synthesis). They also act as motors, membrane pumps and receptors.
- **Extracellular enzymes** are synthesised inside cells but then exported out of cells and catalyse reactions outside the cell. Digestive enzymes are the best example: they are synthesised in the cells of the pancreas, but are secreted out of these cells and work in the lumen of the intestine.

How do enzymes work?

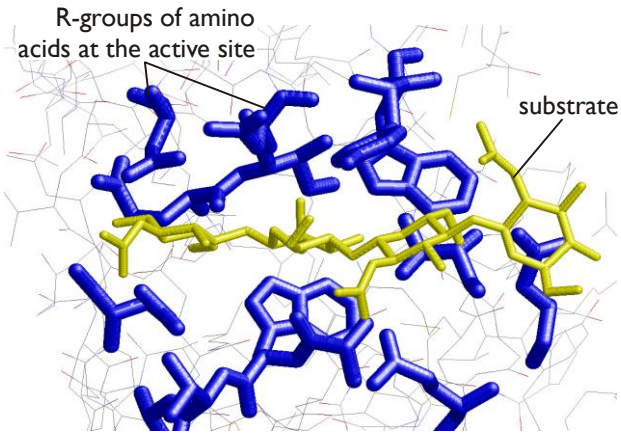
There are three ways of thinking about enzyme catalysis. They all describe the same process, though in different ways, and you should know about each of them.

1. Enzymes Manipulate the Substrate in the Active Site

Enzymes are proteins, and their function is determined by their complex 3-dimentional structure. The reaction takes place in a small part of the enzyme called the active site, while the rest of the protein acts as "scaffolding". The substrate molecule binds to the active site and the product is released.

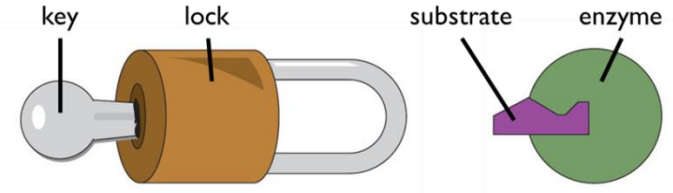


Lysozyme – whole molecule



Close-up of substrate binding to amino acids in the active site

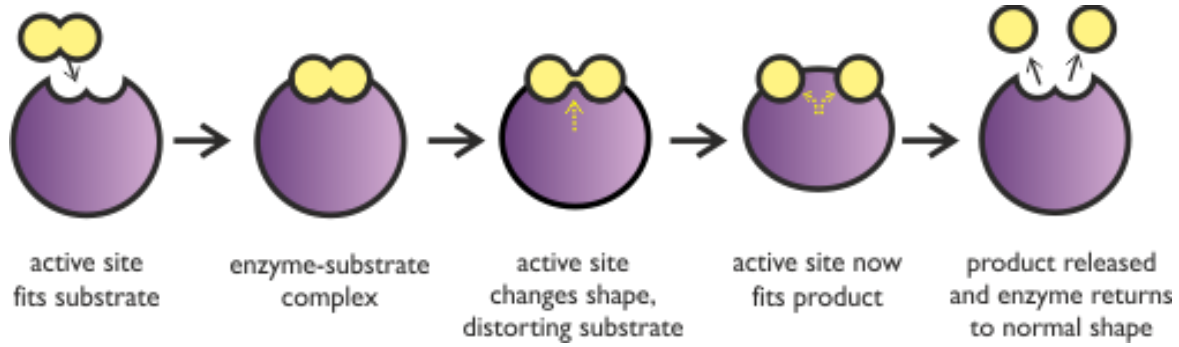
The shape of an enzyme's active site is complementary to the shape of the substrate molecule. This is called the lock and key model, since the active site is like a lock and the substrate is like a key fitting into the lock.



The shape and properties of the active site are given by the amino acids around it, which form weak hydrogen and ionic bonds with the substrate molecule. This means the active site can bind one substrate only, in other words the enzyme is specific for that one reaction. Molecules with a different shape won't fit the active site, just as the wrong key won't fit into a lock.

The lock and key analogy is quite good, but it doesn't reflect the fact that enzyme molecules are flexible (unlike locks). When a substrate binds, the whole enzyme changes shape, distorting the substrate molecule in the active site and turning it into the product. For example if a bond in the substrate is to be broken, that bond might be stretched by the enzyme, making it more likely to break. Alternatively if a bond is to be made between two molecules, the two molecules can be held in exactly the right position and orientation

and "pushed" together, making the bond more likely to form. The enzyme can also make the local conditions inside the active site quite different from those outside (such as pH, water concentration, charge), so that the reaction is more likely to happen. This flexible model is called the induced fit model, and is a more accurate description of enzyme action.



Many enzymes also have small non-protein molecules called coenzymes at their active sites to help bind to the substrate. Many of these are derived from dietary vitamins, which is why vitamins are so important.

2. Enzymes Take an Alternative Reaction Pathway

In any chemical reaction, a substrate (S) is converted into a product (P):

$$\mathbf{S \rightleftharpoons P}$$

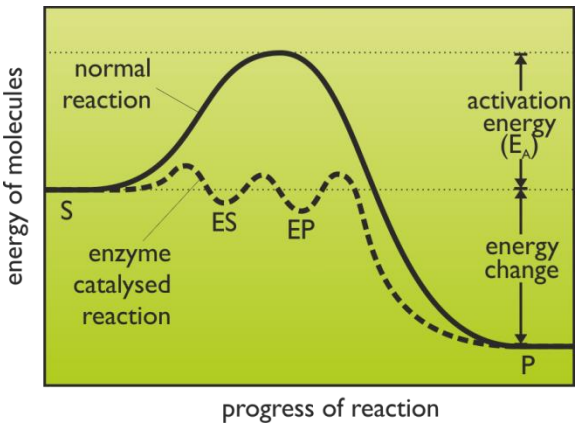
(There may be more than one substrate and more than one product, but that doesn't matter here.) In an enzyme-catalysed reaction, the substrate first binds to the active site of the enzyme to form an enzyme-substrate (ES) complex, then the substrate is converted into product while attached to the enzyme, and finally the product is released. This mechanism can be shown as:

$$\mathbf{E + S \rightleftharpoons ES \rightleftharpoons EP \rightleftharpoons E + P}$$

The enzyme is then free to start again. The end result is the same ($S \rightleftharpoons P$), but a different route is taken, so that the $S \rightleftharpoons P$ reaction as such never takes place. In by-passing this step, and splitting the reaction up into many small steps rather than one big step, the reaction can be made to happen much more quickly.

3. Enzymes Lower the Activation Energy

The way enzymes work can also be shown by considering the energy changes that take place during a chemical reaction. We shall consider a reaction where the product has a lower energy than the substrate, so the substrate naturally turns into product (in other words the equilibrium lies in the direction of the product). Before it can change into product, the substrate must overcome an "energy barrier" called the activation energy (E_A). The larger the activation energy, the slower the reaction will be because only a few substrate molecules will by chance have sufficient energy to overcome the activation



energy barrier. Imagine pushing boulders over a hump before they can roll downhill, and you have the idea. Most physiological reactions have large activation energies, so they simply don't happen on a useful time scale. Enzymes dramatically reduce the activation energy of a reaction, so that most molecules can easily get over the activation energy barrier and quickly turn into product.

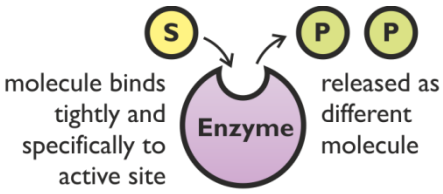
For example for the breakdown of hydrogen peroxide ($2H_2O_2 \rightleftharpoons 2H_2O + O_2$):

- E_A = 86 kJ mol^{-1} with no catalyst
- E_A = 62 kJ mol^{-1} with an inorganic catalyst of iron filings
- E_A = 1 kJ mol^{-1} in the presence of the enzyme peroxidase (catalase).

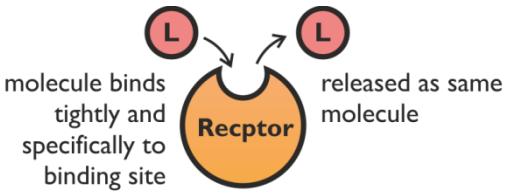
Active sites and binding sites

Enzymes and receptors are both protein molecules that work in similar ways. They have specific three-dimensional shapes with a site where another molecule can bind.

Enzymes have an active site. The molecule that binds (the substrate) is changed and released as a different molecule (the product).

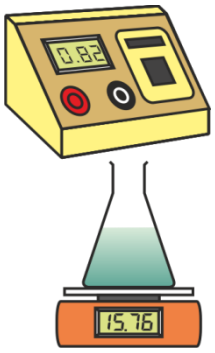


Receptors have a binding site. The molecule that binds (the ligand) is released unchanged.

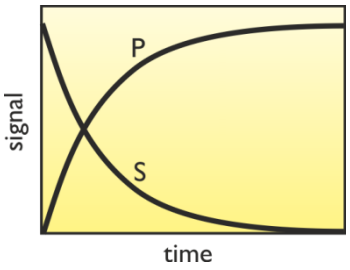


Measuring the Rate of Enzyme Reactions

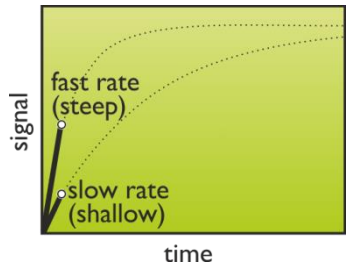
1. Firstly you need a signal to measure that shows the progress of the reaction. The signal should change with either substrate or product concentration, and it should preferably be something that can be measured continuously. Typical signals include colour changes, pH changes, mass changes, gas production, volume changes or turbidity changes. If the reaction has none of these properties, it can sometimes be linked to a second reaction that does generate one of these changes.



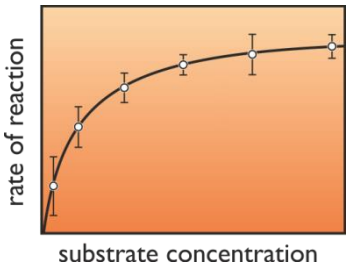
2. If you mix the substrate with enzyme and measure the signal, you will obtain a time-course. If the signal is proportional to substrate concentration it will start high and decrease, while if the signal is proportional to product it will start low and increase. In both cases the time-course will be curved (actually an exponential curve).



3. How do you obtain a rate from this time-course? One thing that is **not** a good idea is to measure the time taken for the reaction, for as the time-course shows it is very difficult to say when the reaction actually ends: it just gradually approaches the end-point. The rate is in fact the slope (or gradient) of the time-course, so we can see that the rate (and slope) decreases as the reaction proceeds. The best measurement is the initial rate - that is the initial slope of the time-course. This also means you don't need to record the whole time-course, but simply take one measurement a short time after mixing.



4. Repeat this initial rate measurement under different conditions (such as different temperatures or substrate concentrations) and then plot a graph of rate vs. the factor. Each point on this second graph is taken from a separate initial rate measurement (or better still is an average of several initial rate measurements under the same conditions). Draw a smooth curve through the points.

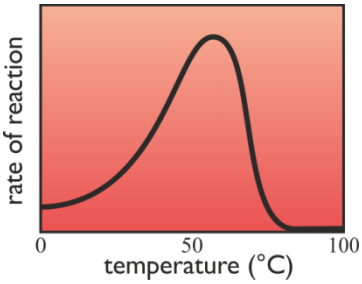


Be careful not to confuse the two kinds of graph (the time-course and rate graphs) when interpreting data.

Factors that Affect the Rate of Enzyme Reactions

1. Temperature

All chemical reactions get faster as the temperature increases, but with enzyme reactions this is only true up to a certain temperature, above which the rate slows down again. This optimum temperature is about 40°C for mammalian enzymes but there are enzymes that work best at very different temperatures, e.g. enzymes from the arctic snow flea work at -10°C, and enzymes from thermophilic bacteria work at 90°C.

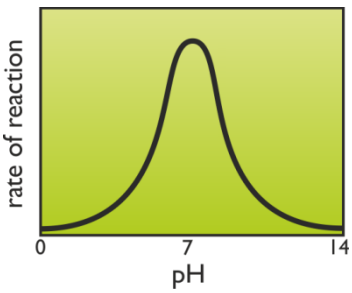


Up to the optimum temperature the rate increases geometrically with temperature (i.e. it's a curve, not a straight line). The rate increases because the enzyme and substrate molecules both have more kinetic energy so collide more often, and also because more molecules have sufficient energy to overcome the (greatly reduced) activation energy. The rate is not zero at 0°C, so enzymes still work in the fridge (and food still goes off), but they work slowly. Enzymes can even work in ice, though the rate is extremely slow due to the very slow diffusion of enzyme and substrate molecules through the ice lattice.

This increase in rate with temperature would continue indefinitely except that the enzyme molecule itself is affected by temperature. Above about 40°C there is enough thermal energy to break the weak hydrogen bonds holding the secondary, tertiary and quaternary structures of the enzyme together, so the enzyme (and especially the active site) loses its specific shape to become a random coil. The substrate can no longer bind, and the reaction is no longer catalysed. This denaturation is usually irreversible. The optimum temperature of enzymes is normally about 40°C because that is the temperature at which hydrogen bonds break. This is also the reason why mammals and birds maintain their body temperature at around 40°C. Remember that only the weak hydrogen bonds not peptide bonds are broken at these mild temperatures; to break strong covalent bonds you need to boil in concentrated acid for many hours.

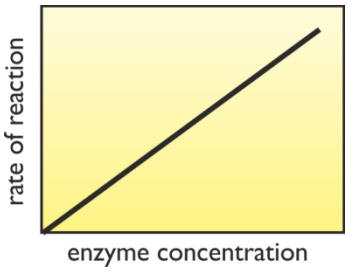
2. pH

Enzymes have an optimum pH at which they work fastest. For most enzymes this is about pH 7-8 (physiological pH of most cells), but a few enzymes can work at extreme pH, such as protease enzymes in animal stomachs, which have an optimum of pH 1. The pH affects the charge of the R-groups of the amino acids at the active site. For example carboxyl R-groups are uncharged (COOH) in acid pH but negatively charged (COO^-) in alkali pH. Similarly amino R-groups are positively charged (NH_3^+) in acidic pH but uncharged (NH_2) in alkali pH. These changes can affect the shape as well as the charge of the active site, so the substrate can no longer bind and the reaction isn't catalysed.



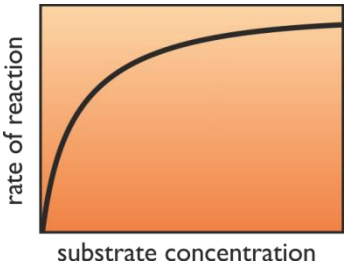
3. Enzyme concentration

As the enzyme concentration increases the rate of the reaction increases linearly, because there are more enzyme molecules available to catalyse the reaction. At very high enzyme concentration the substrate concentration may become rate-limiting, so the rate stops increasing. Normally enzymes are present in cells in rather low concentrations.



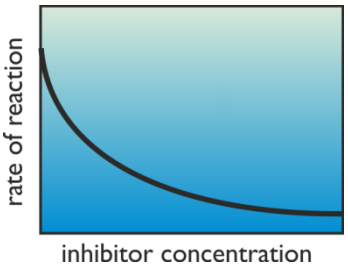
4. Substrate concentration

The rate of an enzyme-catalysed reaction shows a curved dependence on substrate concentration. As the substrate concentration increases, the rate increases because more substrate molecules can collide with enzyme molecules, so more reactions will take place. At higher concentrations the enzyme active sites become saturated with substrate, so there are few free enzyme molecules, so adding more substrate doesn't make much difference (though it will increase the rate of E–S collisions).



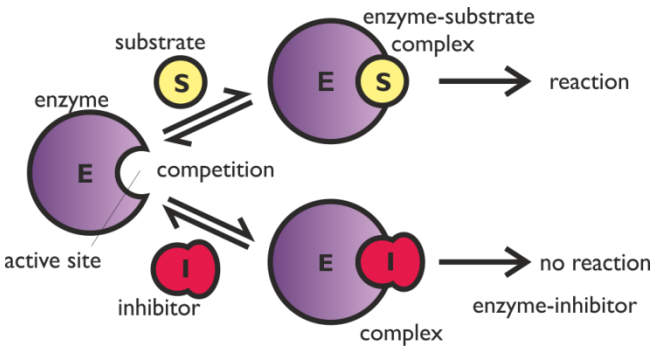
5. Inhibitors

Inhibitors inhibit the activity of enzymes, reducing the rate of their reactions. They are found naturally but are also used artificially as drugs, pesticides and research tools. Inhibitors that bind fairly weakly and can be washed out are called reversible inhibitors, while those that bind tightly and cannot be washed out are called irreversible inhibitors.



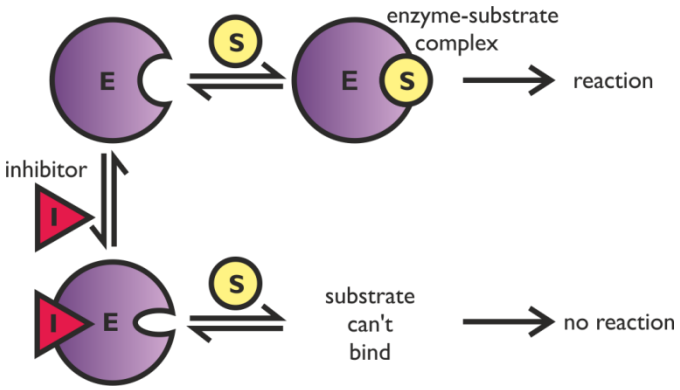
There are two kinds of reversible inhibitors:

- Competitive Inhibitors are molecules with a similar structure to the normal substrate molecule, and can fit into the active site of the enzyme. They therefore compete with the substrate for the active site, so the reaction is slower. However, if the substrate concentration is increased high enough the substrate will out-compete the inhibitor and the rate can approach a normal rate. The sulphonamide anti-bacterial drugs are competitive inhibitors.

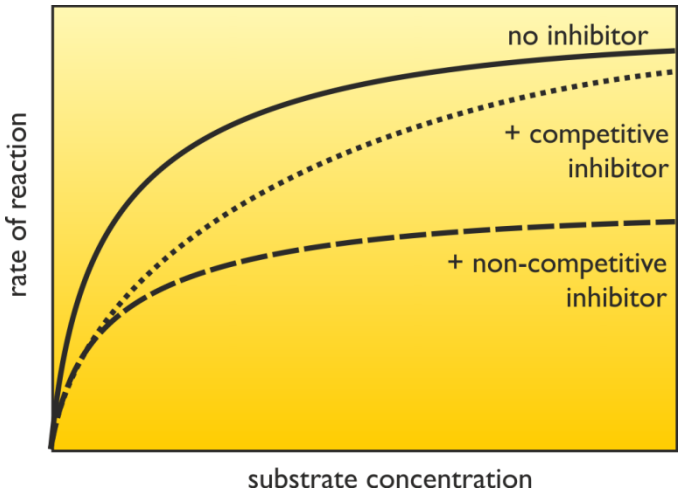


- Non-competitive Inhibitors are molecules with a quite different in structure from the substrate molecule and do not fit into the active site. They bind to another part of the enzyme molecule, changing

the shape of the whole enzyme, including the active site, so that it can no longer bind substrate molecules. Non-competitive inhibitors therefore simply reduce the amount of active enzyme (just like decreasing the enzyme concentration). Poisons like cyanide, heavy metal ions and some insecticides are all non-competitive inhibitors.

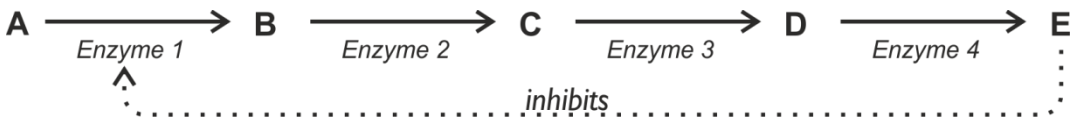


The two types of inhibitor can be distinguished experimentally by carrying out a substrate vs. rate experiment in the presence and absence of the inhibitor. If the inhibition is reduced at high substrate concentration then the inhibitor is a competitive one.



Metabolic Pathways and End-Product Inhibition

The thousands of reactions taking place in every cell are arranged into metabolic pathways. The product of one reaction is the substrate for the next, and each step has its own enzyme. These pathways only operate when needed, so are under tight control. One form of control is end-product inhibition. If the final product builds up then it inhibits the first enzyme in the pathways (usually by non-competitive inhibition), thus reducing its own production. If the concentration of the final product is too low, then the first enzyme is no longer inhibited, the pathway runs again, and more end product is made. This is a form of negative feedback.

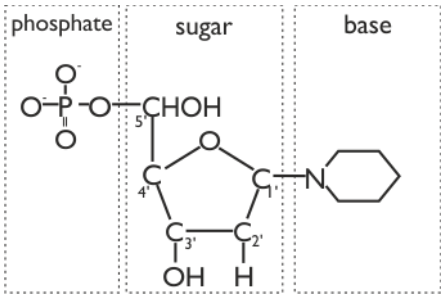


DNA

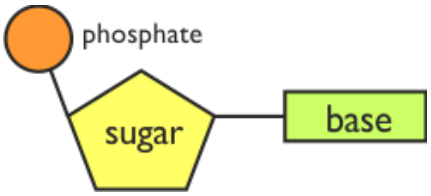
DNA and its close relative RNA are perhaps the most important molecules in biology. They contain the instructions that make every single living organism on the planet, and yet it is only in the past 50 years that we have begun to understand them. DNA stands for deoxyribonucleic acid and RNA for ribonucleic acid, and they are called nucleic acids because they are weak acids, first found in the nuclei of cells. They are polymers, composed of monomers called nucleotides.

Nucleotides

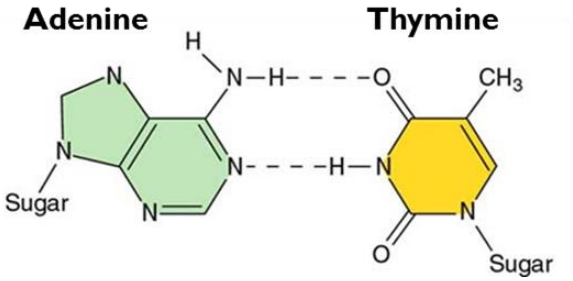
Nucleotides contain the elements CHONP, and have three parts to them:

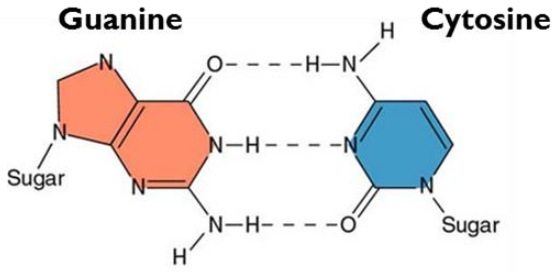


or more simply:



- A phosphate group (PO_4^{2-}), which is negatively charged, and gives nucleic acids their acidic properties.
- A pentose sugar. If carbon 2' has a hydroxyl group attached then the sugar is ribose, found in RNA. If the carbon 2' just has a hydrogen atom attached instead (as shown), then the sugar is deoxyribose, found in DNA.
- A nitrogenous base. These are small organic basic groups that contain the elements CHON, so are often called nitrogenous bases. There are four different bases in DNA nucleotides, called Adenine (A), Cytosine (C), Guanine (G) and Thymine (T). RNA nucleotides also have four bases, but instead of thymine RNA has Uracil (U). Adenine and Guanine have a double ring and are called purines while thymine and cytosine have single rings and are called pyrimidines (you don't need to know their structures).





The diagrams show that each base can form hydrogen bonds with another specific base to form complementary base pairs. The pairs are always formed between a purine and pyrimidine (A with T and G

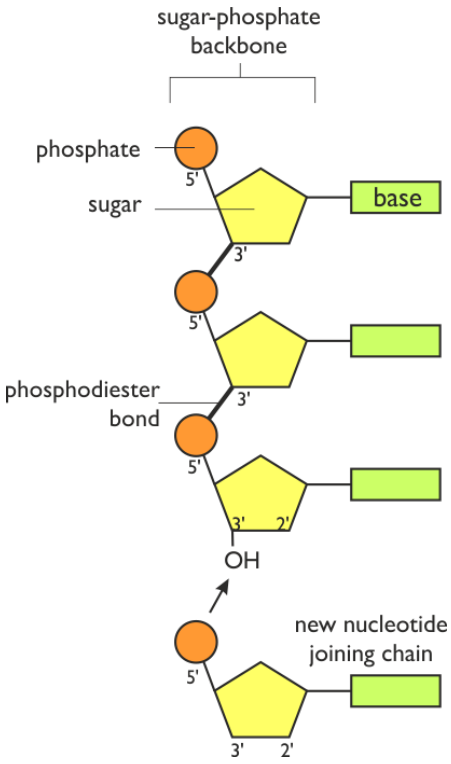
with C), so the two base pairs are the same size. Since there are four bases, there are four different nucleotides:

Base:	Adenine (A)	Cytosine (C)	Guanine (G)	Thymine (T)
Nucleotide:	Adenosine	Cytidine	Guanosine	Thymidine

The nucleotide above is shown with a single phosphate group, but in fact nucleotides can have one, two or three phosphate groups. So for instance you can have adenosine monophosphate (AMP), adenosine diphosphate (ADP) and adenosine triphosphate (ATP). These nucleotides are very common in cells and have many roles other than just part of DNA. For example, ATP is used as the energy transfer molecule, while AMP and GTP are used as messenger chemicals.

Nucleotide Polymerisation

Nucleotides polymerise by forming phosphodiester bonds between carbon 3' of the sugar and an oxygen atom of the phosphate. This is a condensation reaction. The bases do not take part in the polymerisation, so there is a sugar-phosphate backbone with the bases extending off it. This means that the nucleotides can join together in any order along the chain. Two nucleotides form a dinucleotide, three form a trinucleotide, a few form an oligonucleotide, and many form a polynucleotide.

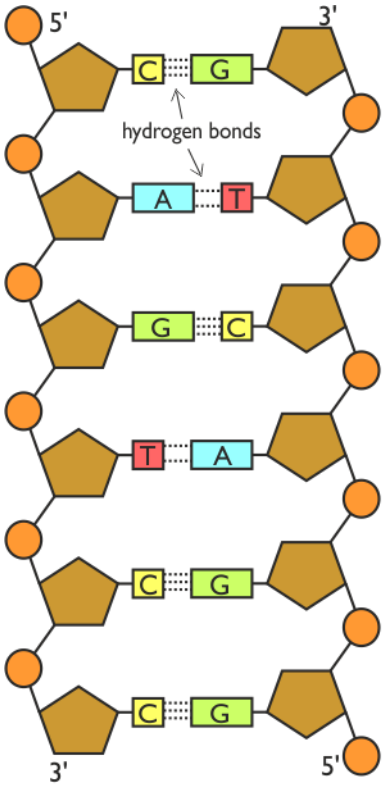


A polynucleotide has a free phosphate group at one end, called the 5' end because the phosphate is attached to carbon 5' of the sugar, and a free OH group at the other end, called the 3' end because it's on carbon 3' of the sugar. The terms 3' and 5' are often used to denote the different ends of a DNA molecule.

Structure of DNA

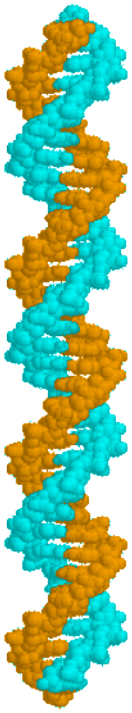
The three-dimensional structure of DNA was discovered in 1953 by Watson and Crick in Cambridge, using experimental data of Wilkins and Franklin in London, for which work they won a Nobel Prize. The main features of the structure are:

- DNA is double-stranded, so there are two polynucleotide stands alongside each other. The strands are antiparallel, i.e. they run in opposite directions.
- The two strands are wound round each other to form a double helix (not a spiral, despite what some textbooks say).
- The two strands are joined together by hydrogen bonds between the bases. The bases therefore form base pairs, which are like rungs of a ladder.
- The base pairs are specific. A only binds to T (and T with A), and C only binds to G (and G with C). These are called complementary base pairs. This means that whatever the sequence of bases along one strand, the sequence of bases on the other stand must be complementary to it. (Incidentally, complementary, which means matching, is different from complimentary, which means being nice.)
- DNA is a very long molecule, so it can store a great deal of information (a single human DNA molecule has 3 billion base-pairs and is 1m long). The bases are protected on the inside of the molecule and the billions of hydrogen bonds holding the two strands together make DNA a very stable molecule.



DNA showing the complementary base pairing between antiparallel strands

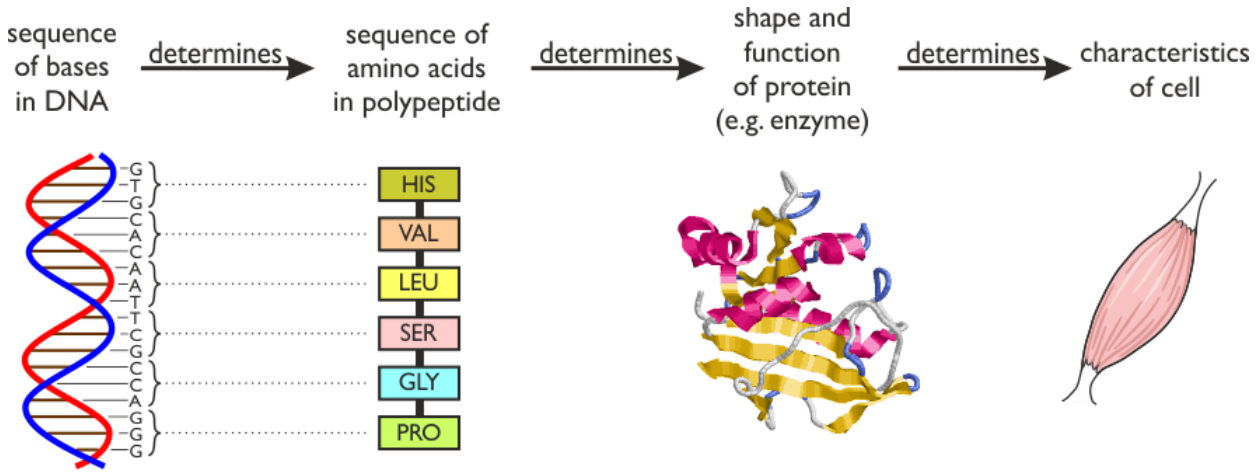
DNA showing the double helix



spacefilling model of the double helix

Function of DNA

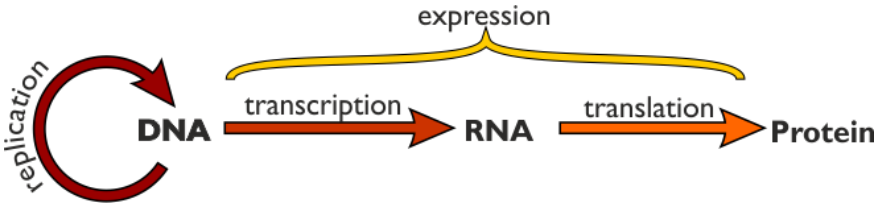
DNA contains genes, and genes control characteristics. A gene is simply a section of DNA around 1000-2000 base pairs long. The four bases can appear in any order along the DNA molecule, so their sequence can encode information, like writing with a 4-letter alphabet. DNA doesn't control characteristics directly, but instead DNA controls characteristics by controlling protein synthesis. In fact a gene is defined as ***a sequence of bases on a DNA molecule coding for a sequence of amino acids in a polypeptide chain***. It is the numerous proteins in a cell (mostly enzymes) that control what the cell does, and therefore the characteristics of the organism.



So there is an old and a new definition of a gene that say the same thing:

A gene is an inherited factor that controls a particular characteristic.	A gene is a section of DNA that codes for a particular polypeptide.

This process of making proteins and so controlling characteristics is called gene expression (because the gene "expresses" itself). Expression is split into two steps: transcription and translation, so this diagram summarises what DNA does:



The Genome

No one knows exactly how many genes we humans have to control all our characteristics, but the current best estimate is around 20 thousand. The sum of all the genes in an organism is called the genome, and this table shows the estimated number of genes in different organisms:

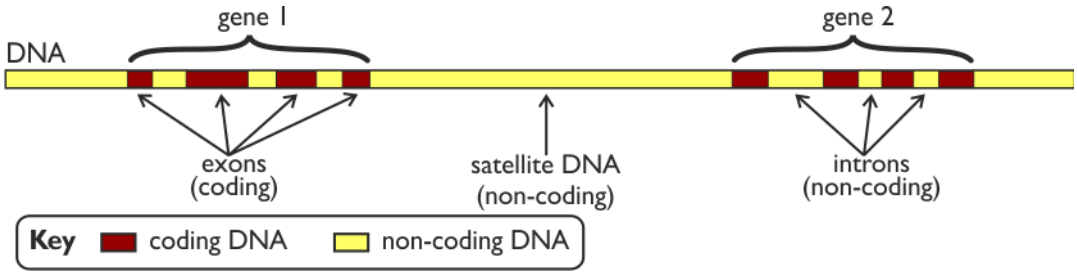
Species	Common name	length of DNA (kbp)*	no of genes
phage λ	virus	48	60
Eschericia coli	bacterium	4 639	4 000
Saccharomyces cerevisiae	Yeast	13 500	6 000
Caenorhabditis elegans	nematode worm	90 000	~10 000
Drosophila melaogaster	fruit fly	165 000	~10 000
Homo sapiens	human	3 150 000	~20 000

*kbp = kilo base pairs, i.e. thousands of nucleotide monomers.

Coding and Non-Coding DNA

Surprisingly, a lot of the DNA in eukaryotes does not code for polypeptides. In fact, only about 2% of the DNA in a eukaryotic cell is coding DNA. The rest, called non-coding DNA, does not form genes. There are two kinds of non-coding DNA:

- Non-coding regions of DNA within a gene are called introns (for interruption sequences), while the coding parts of DNA are called exons (for expressed sequences). All eukaryotic genes have introns, and they are usually longer than the exons, so genes are often much longer than they really need to be!
- Non-coding regions of DNA between genes are called satellite DNA. Satellite DNA often contains simple base sequences repeated many times (sometime thousands of times).

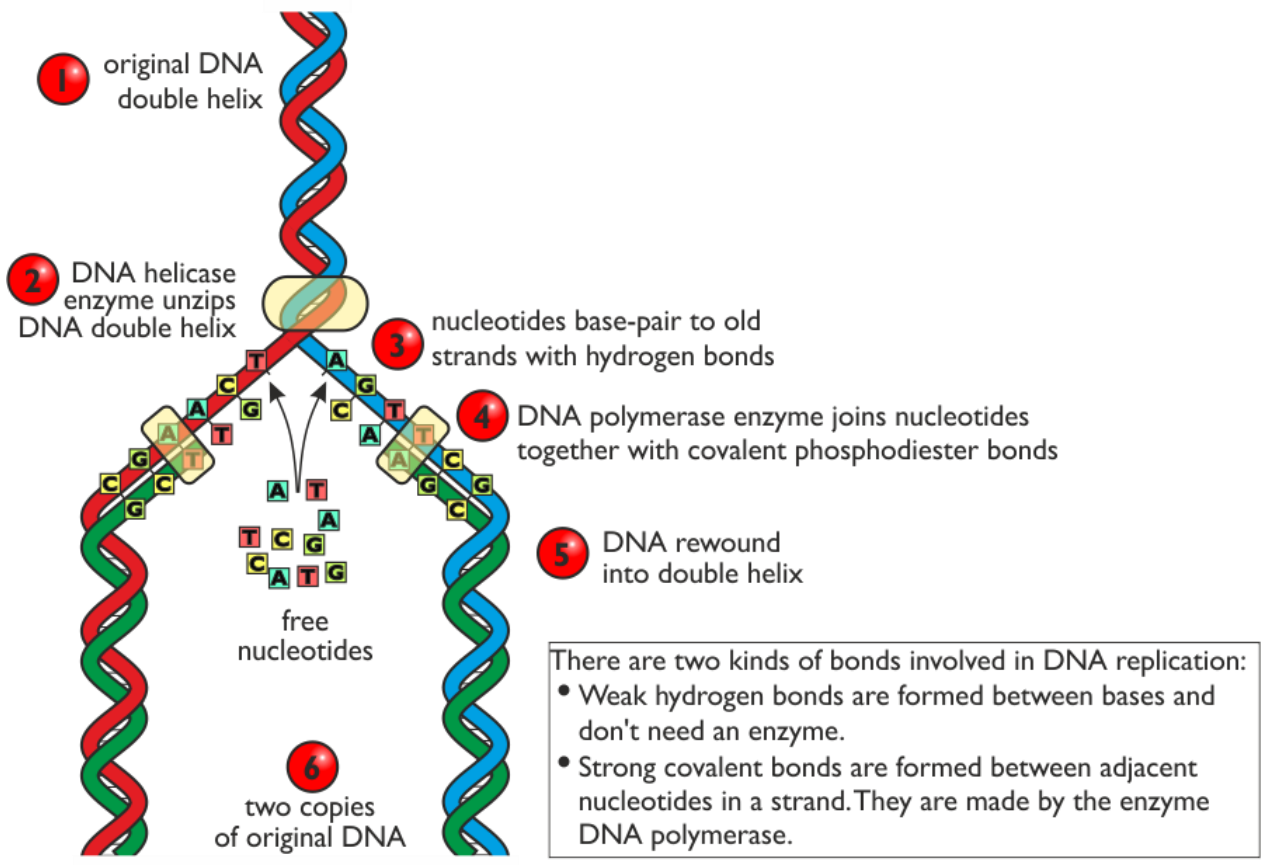


Non-coding DNA was originally termed junk DNA, but in fact it probably serves many different functions.

- Some non-coding DNA is structural, helping to coil the DNA molecule into chromosomes.
- Some non-coding DNA has a control function, regulating when genes are expressed.
- Some non-coding DNA is involved in DNA replication.
- Some non-coding DNA contains unused copies of genes (pseudogenes).

Replication of DNA

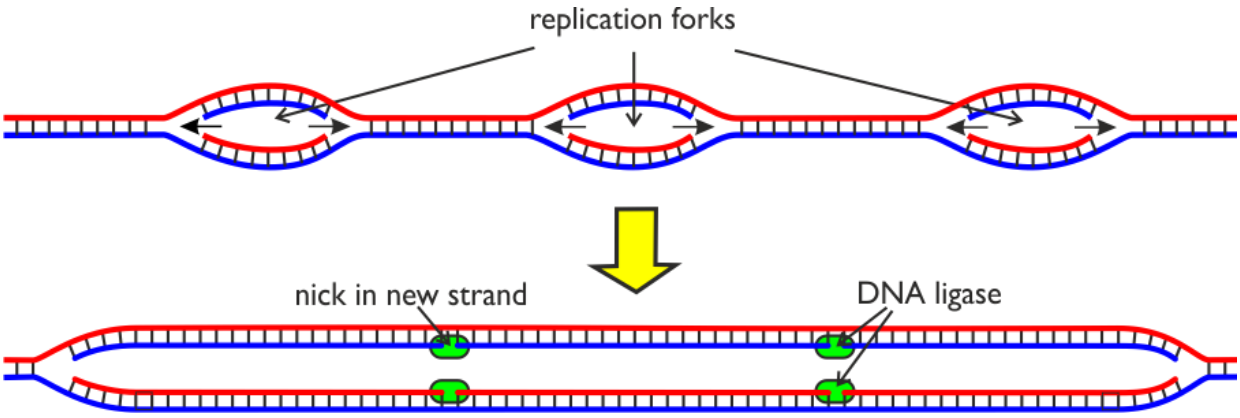
DNA is copied, or replicated, before every cell division, so that one identical copy can go to each daughter cell. The method of DNA replication is obvious from its structure: the double helix unzips and two new strands are built up by complementary base-pairing onto the two old strands.



1. Replication starts at a specific sequence on the DNA molecule called the replication origin.
2. The enzyme DNA helicase unwinds and separates the two strands of DNA, breaking the hydrogen bonds between the base pairs.
3. The new DNA is built up from the four nucleotides (A, C, G and T) that are present in the nucleoplasm. These nucleotides attach themselves to the bases on the old strands by complementary base pairing. Where there is a T base, only an A nucleotide will bind, and so on.
4. The enzyme DNA polymerase joins the new nucleotides to each other by strong covalent phosphodiester bonds, forming the sugar-phosphate backbone. This enzyme is enormously complex and contains 18 subunits.
5. The two strands wind up to form a double helix.
6. The two new DNA molecules are identical to the old molecule. Each new DNA molecule contains one "new" strand and one "old" strand.

Replication Forks and DNA Ligase

In eukaryotes replication is speeded up by taking place at many sites along the DNA simultaneously. These multiple sites of replication lead to structures called replication forks:



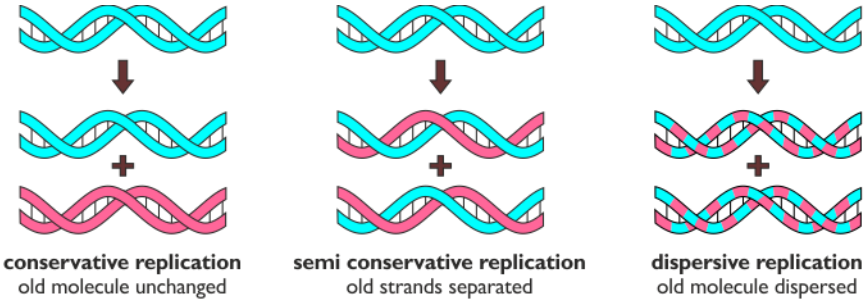
As the polymerase enzymes from two forks meet, there is always a gap, or nick, formed, where adjacent nucleotides cannot be joined by polymerase. These nicks are joined (or ligated) by another enzyme, DNA ligase, which makes phosphodiester bonds to complete the sugar-phosphate backbone of the new stands.

Occasionally, mistakes are made in replication, i.e. an incorrect base pair (such as A-C) is formed. To allow for this, there are repair enzymes that follow DNA polymerase along the new DNA molecules, checking for mismatched base pairs. Incorrect nucleotides are cut out and replaced, and repaired with DNA ligase again.

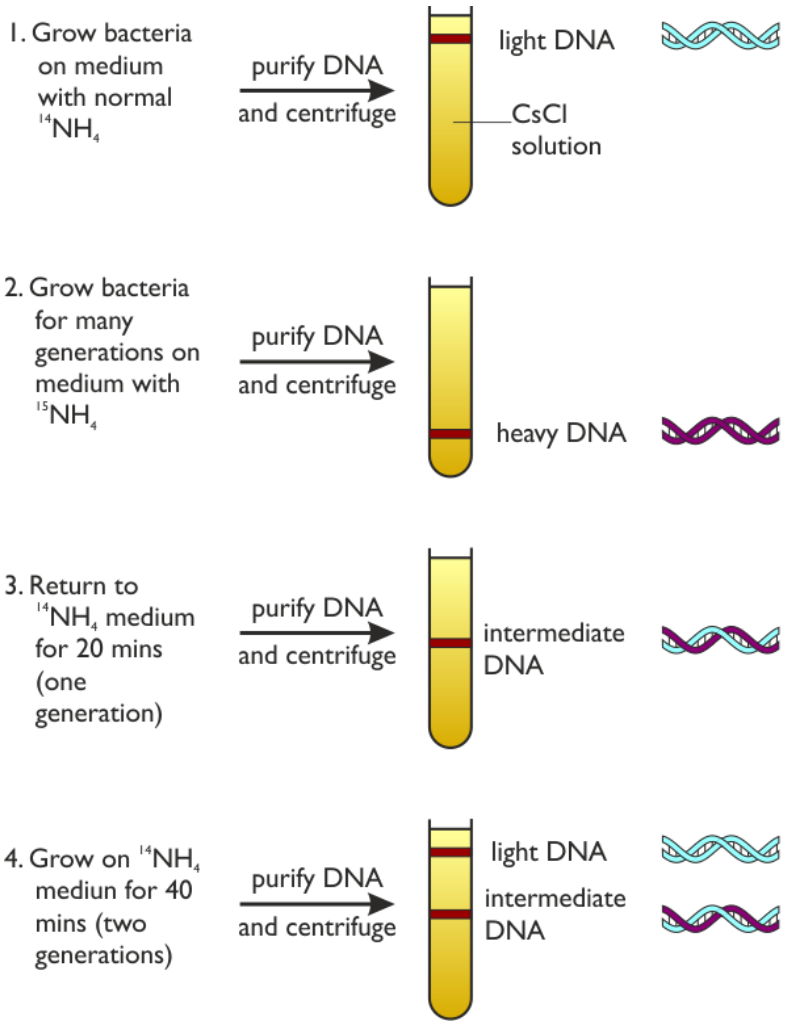
Even with multiple origins, DNA replication can take a few hours, and in fact this limits the speed of cell division. One reason bacteria can reproduce so fast is that they have a relatively small amount of DNA.

The Meselson-Stahl Experiment

This replication mechanism is called semi-conservative replication, because *each new DNA molecule contains one new strand and one old strand*. This need not be the case, and alternative theories included conservative replication and dispersive replication:



The evidence for the semi-conservative method came from an elegant experiment performed in 1958 by Matthew Meselson and Franklin Stahl. They used the bacterium *E. coli* together with the technique of density gradient centrifugation, which separates molecules on the basis of their density.



RNA

As we have seen, gene expression is split into two parts – transcription and translation, and it involves the other nucleic acid – RNA (ribonucleic acid). RNA is a nucleic acid like DNA, but with 4 differences:

- RNA is made of ribose nucleotides instead of deoxyribose nucleotides
- RNA has the base uracil instead of thymine (so the four bases in RNA are A, U, C and G).
- RNA is single stranded (though it can fold into 3-dimentional structures)
- RNA is shorter than DNA

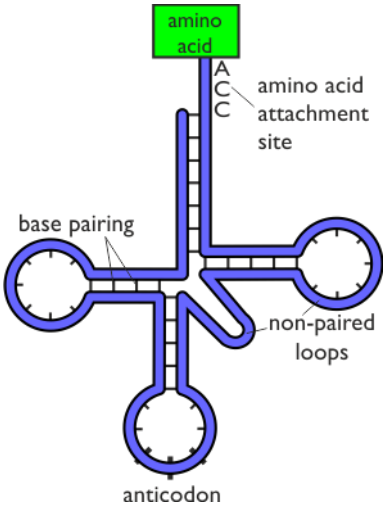
There are three kinds of RNA, with three different jobs:

Messenger RNA (mRNA)

mRNA carries the "message" that codes for a particular protein from the nucleus (where the DNA master copy is) to the cytoplasm (where proteins are synthesised). It is single stranded and just long enough to contain one gene only (about 1000 nucleotides). It has a short lifetime and is degraded soon after it is used.

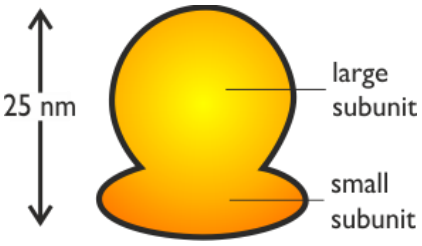
Transfer RNA (tRNA)

tRNA is an "adapter" that matches amino acids to their codon. tRNA is only about 80 nucleotides long, and it folds up by complementary base pairing to form a looped clover-leaf structure. At one end of the molecule there is always the base sequence ACC, where the amino acid binds. On the middle loop there is a triplet nucleotide sequence called the anticodon. There are 64 different tRNA molecules, each with a different anticodon sequence complementary to the 64 different codons. The amino acids are attached to their tRNA molecule by specific aminoacyl tRNA synthase enzymes. These are highly specific, so that each amino acid is attached to a tRNA adapter with the appropriate anticodon.



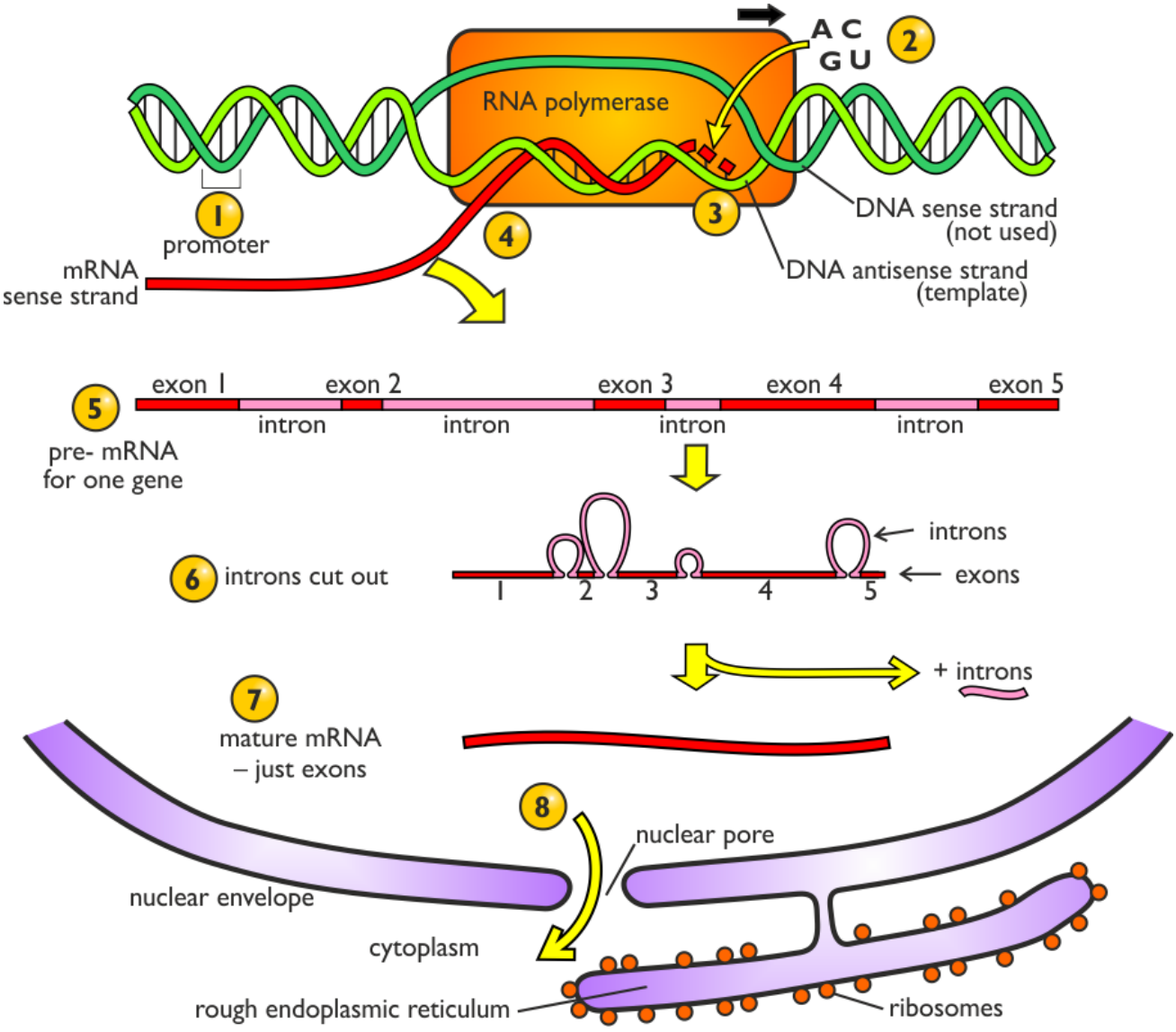
Ribosomal RNA (rRNA)

rRNA together with proteins forms ribosomes, which are the site of mRNA translation and protein synthesis. Ribosomes have two subunits, small and large, and are assembled in the nucleolus of the nucleus and exported into the cytoplasm. rRNA is coded for by numerous genes in many different chromosomes. Ribosomes free in the cytoplasm make proteins for use in the cell, while those attached to the RER make proteins for export.



Transcription - RNA Synthesis

DNA never leaves the nucleus, but proteins are synthesised in the cytoplasm, so a copy of each gene is made to carry the "message" from the nucleus to the cytoplasm. This copy is mRNA, and the process of copying is called transcription. As we have seen, eukaryotic genes contain non-coding sequences (introns) as well as the coding sequences (exons). The introns need to be removed before the mRNA can be translated into protein. This removal is called post-transcriptional modification.



1. The start of each gene on DNA is marked by a special sequence of bases called the promoter.
2. The RNA molecule is built up from the four ribose nucleotides (A, C, G and U) in the nucleoplasm. The ribose nucleotides attach themselves to the bases on the DNA by complementary base pairing, just as in DNA replication. However, only one strand of RNA is made.

> The mRNA strand will be used to code for a polypeptide, so the sequence of bases on mRNA is called the sense sequence. The DNA strand that is copied is therefore called the DNA antisense strand, since its sequence is complementary to the RNA sequence. The other DNA strand is not used in transcription. Since its sequence is complementary to the antisense DNA strand (and the same as the RNA strand), it is called the DNA sense strand.

3. The new nucleotides are joined to each other by strong covalent phosphodiester bonds by the enzyme RNA polymerase.
4. Only about 8 base pairs remain attached at a time, since the mRNA molecule peels off from the DNA as it is made. The DNA rewinds as it leaves the polymerase enzyme. At the end of the gene the transcription stops, so the mRNA molecule is just the length of the gene.
5. The initial mRNA that is transcribed is called the primary transcript or pre-mRNA. Pre-mRNA is an exact copy of the gene on the DNA, so it contains exons and introns.
6. The introns in the mRNA are cut out and the exons are joined together by enzymes in a process called splicing. Some of this splicing is done by the RNA intron itself, acting as an RNA enzyme. The recent discovery of these RNA enzymes, or ribozymes, illustrates what a diverse and important molecule RNA is. Other splicing is performed by RNA/protein complexes called snurps.
7. The result is a shorter mature RNA containing only exons. The introns are broken down. This completes the post-transcriptional modification.
8. The mRNA diffuses out of the nucleus through a nuclear pore into the cytoplasm. There, it attaches to ribosomes for translation. It usually doesn't have far to go to find a ribosome, as many are attached to the rough endoplasmic reticulum, which is contiguous with the nuclear envelope.

Prokaryotic DNA does not have introns, so mRNA processing is not needed.

The Genetic Code

There are 20 different amino acids and only 4 different bases, so the bases are read in groups of three. This gives 4^3 or 64 combinations, more than enough to code for 20 amino acids. A triplet of three bases coding for an amino acid is called a codon, and the meaning of each of the 64 codons is called the genetic code. Because there are more codons than amino acids, most amino acids are coded for by more than codon. For example CCA, CCT, CCC and CCG all code for the amino acid glycine. Some codons also mark the beginning and end of a gene.

The Genetic Code (mRNA codons)

UUU phe UUC UUA leu UUG	CUU CUC CUA CUG leu	AUU AUC AUA ile AUG start/met	GUU GUC GUA GUG val
UCU UCC UCA UCG ser	CCU CCC CCA CCG pro	ACU ACC ACA ACG thr	GCU GCC GCA GCG ala
UAU tyr UAC UAA stop UAG	CAU his CAC CAA gln CAG	AAU asn AAC AAA lys AAG	GAU asp GAC GAA glu GAG
UGU cys UGC UGA stop UGG trp	CGU CGC CGA CGG arg	AGU ser AGC AGA arg AGG	GGU GGC GGA GGG gly

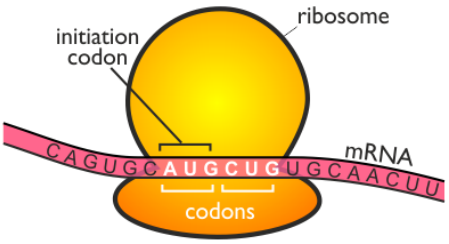
There are several interesting points from this code:

- The code is degenerate, i.e. there is often more than one codon for an amino acid.
- The degeneracy is on the third base of the codon, which is therefore less important than the others.
- One codon means "start" i.e. the start of the gene sequence. It is AUG, which also codes for methionine. Thus all proteins start with methionine (although it may be removed later). AUG in the middle of a gene simply codes for methionine.
- Three codons mean "stop" i.e. the end of the gene sequence. They do not code for amino acids.
- The code is non-overlapping, e.g. the sequence UGCACG contains only the two codons UGC and ACG. It can never be read as the overlapping codons UGC, GCA, CAC, ACG.

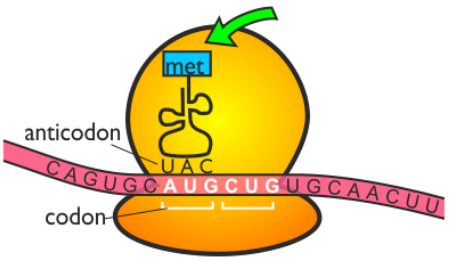
The process of synthesising proteins is called translation because it translates from the 4-base DNA language to the 20-amino acid protein language.

Translation - Protein Synthesis

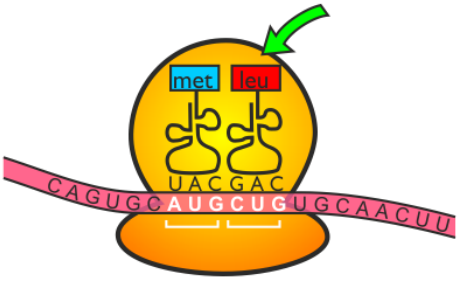
1. A ribosome attaches to the mRNA at an initiation codon (AUG). The ribosome encloses two codons.



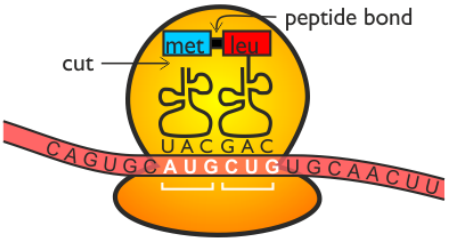
2. The first tRNA molecule with an amino acid attached (met-tRNA) diffuses to the ribosome. Its anticodon attaches to the first mRNA codon by complementary base pairing.



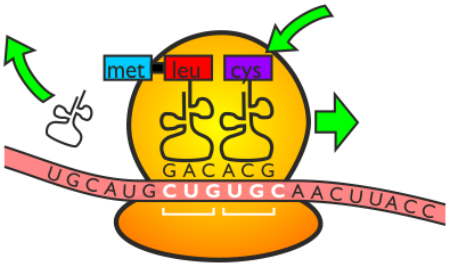
3. The next amino acid-tRNA attaches to the adjacent mRNA codon (CUG, leu in this case) by complementary base pairing.



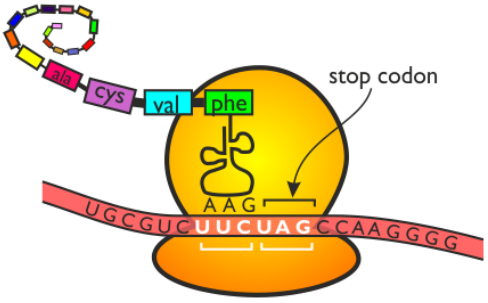
4. The bond between the amino acid and the tRNA is cut and a peptide bond is formed between the two amino acids. These operations are catalysed by enzymes in the ribosome called ribozymes.



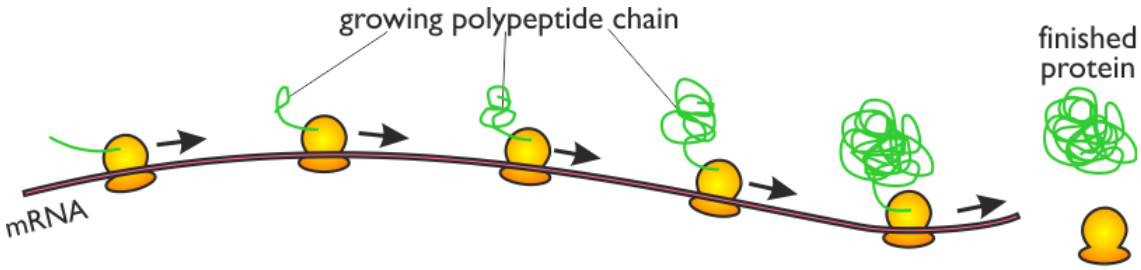
5. The ribosome moves along one codon so that a new amino acid-tRNA can attach. The free tRNA molecule leaves to collect another amino acid. The cycle repeats from step 3.



6. The polypeptide chain elongates one amino acid at a time, and peels away from the ribosome, folding up into a protein as it goes. This continues for hundreds of amino acids until a stop codon is reached, when the ribosome falls apart, releasing the finished protein.



A single piece of mRNA can be translated by many ribosomes simultaneously, so many protein molecules can be made from one mRNA molecule. A group of ribosomes all attached to one piece of mRNA is called a polyribosome, or a polysome.

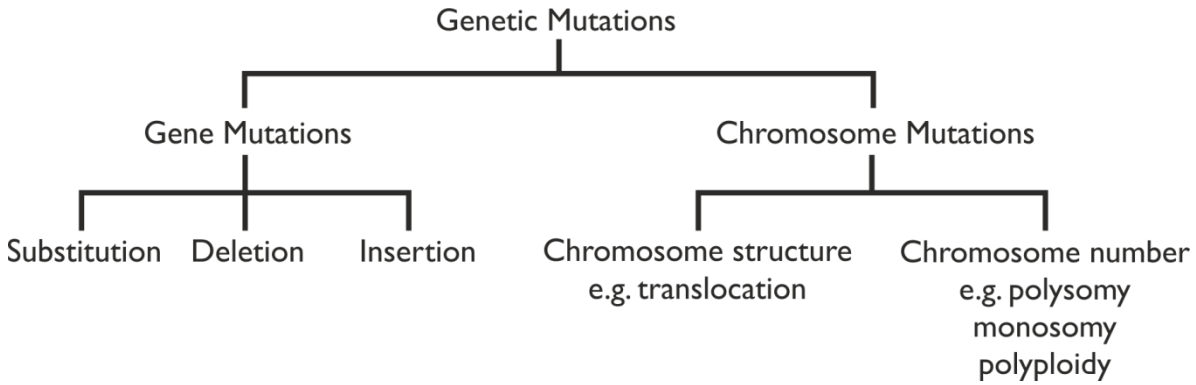


Post-Translational Modification

In eukaryotes, proteins often need to be altered before they become fully functional. Because this happens after translation, it is called post-translational modification. Modifications are carried out by other enzymes and include: chain cutting, adding methyl or phosphate groups to amino acids, adding sugars (to make glycoproteins) or lipids (to make lipoproteins).

Mutations

A mutation is a change in DNA in a cell (a change in genotype). A mutation may have no effect on phenotype or it may may a major effect, including death. There are many different types of mutation, categorised into two groups: gene mutations and chromosome mutations:

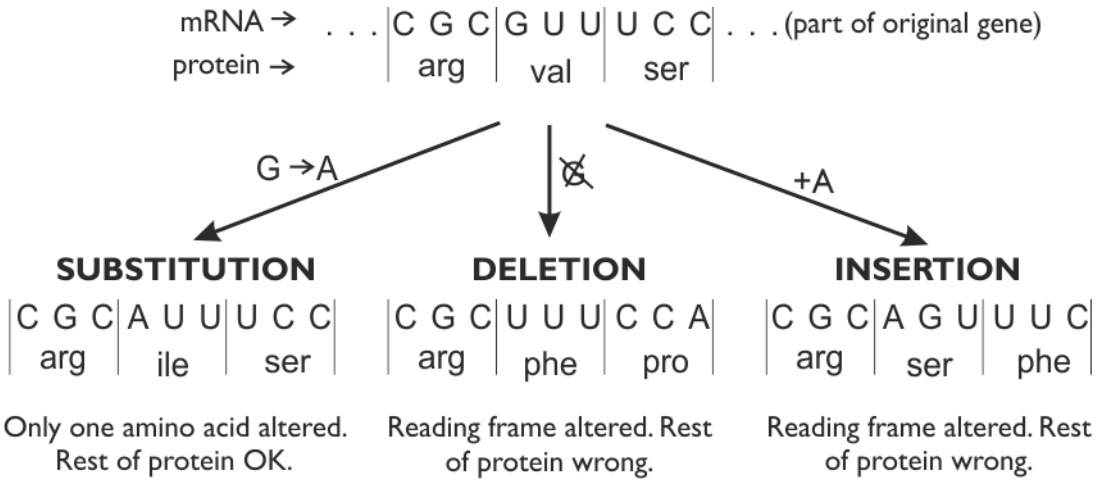


Here, we shall look at gene mutations. On p xx we shall look at chromosome mutations.

Gene Mutations

Gene mutations are changes in DNA base sequences due to a base-pairing error during DNA replication. DNA is a very stable molecule, and it doesn't suddenly change without reason, but bases can change when DNA is being replicated. Normally replication is extremely accurate, and there are even error-checking procedures in place to ensure accuracy, but very occasionally mistakes do occur (such as a T–C base pair).

There are three kinds of gene mutation, shown in this diagram:



- Substitution mutations only affect one amino acid, so tend to have less severe effects. In fact if the substitution is on the third base of a codon it may have no effect at all, because the third base often doesn't affect the amino acid coded for (e.g. all codons beginning with CC code for proline). These are

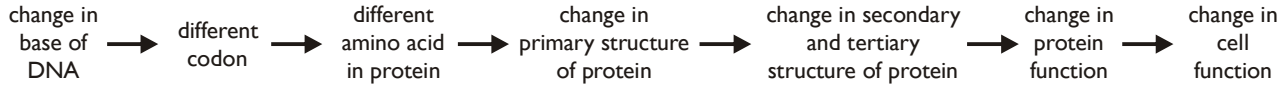
called silent mutations. However, if a mutation leads to a premature stop codon the protein will be incomplete and certainly non-functional. This is called a nonsense mutation.

- Deletion and insertion mutations have more serious effects because they are frame shift mutations i.e. they change the codon reading frame even though they don't change the actual sequence of bases. So all amino acids "downstream" of the mutation are wrong, and the protein is completely wrong and non-functional. However, the effect of a deletion can be cancelled out by a near-by insertion, or by two more deletions, because these will restore the reading frame. A similar argument holds for a substitution.

Impact on phenotype

Most mutations have no phenotypic effect because they occur in non-coding DNA, or the mutated gene is not expressed in this cell, or the mutation doesn't change the amino acid. These are called silent mutations, and we all have a few of these.

Mutations in coding DNA that is expressed will change the cell's phenotype like this:



Many of the proteins in cells are enzymes, and most changes in enzymes will stop them working (because there are far more ways of making an inactive enzyme than there are of making a working one). When an enzyme stops working a reaction in a cell doesn't happen, so the cell's function is changed. It's just possible (though unlikely) that a mutation could make a modified enzyme that actually worked faster than the original enzyme. This means cell's function could be improved.

Since mutations change genes, they give rise to new alleles (i.e. different versions of genes). A cell with the original, functional gene has one allele, while a cell with a mutated, non-functional version of the same gene has a different allele. For example in a flower a "red" allele might encode a functional enzyme that makes a red pigment, while a "white" allele might encode a non-functional enzyme so the flower stays white.

Sickle Cell Anaemia

We saw on px that haemoglobin is a protein made of four polypeptide chains. A single-base mutation in the gene for one of these chains causes the disease sickle cell anaemia.

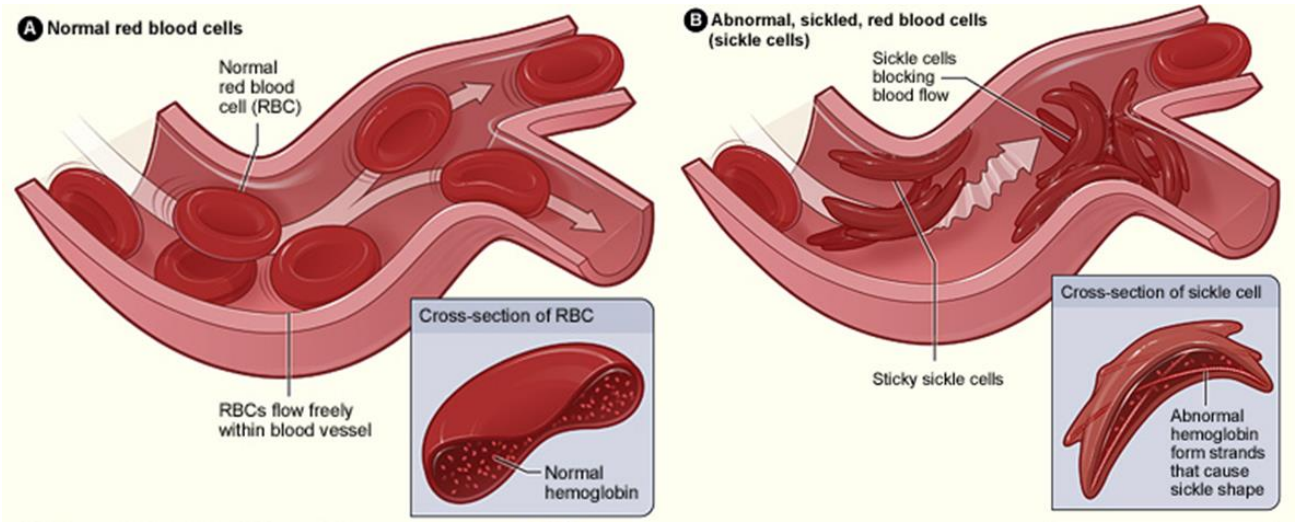
Normal haemoglobin

DNA	ACT	CCT	GAG	GAG	TCT
mRNA	UGA	GGA	CUC	CUC	AGA
Polyeptide	Thr	Pro	Glu	Glu	Ser

Sickle cell haemoglobin

DNA	ACT	CCT	GTG	GAG	TCT
mRNA	UGA	GGA	CAC	CUC	AGA
Polyeptide	Thr	Pro	Val	Glu	Ser

Although this substitution mutation only changes a single amino acid, this slightly changes the shape of the whole haemoglobin molecule, and causes them to link together to form long chains, distorting the red blood cells into sickle shapes. These sickled red blood cells are less flexible than normal cells, so can block capillaries and arterioles, causing cell death and sever pain. Sickle cells are also destroyed by the spleen faster than they can be made, so not enough oxygen can be carried in the blood (anaemia). Without treatment this phenotype is fatal in early childhood, though modern medical intervention can extend life expectancy to 50.



Viruses

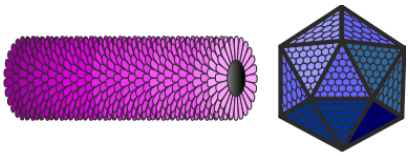
Viruses are the smallest and most abundant of biological entities. They are considered biological since they contain DNA, can reproduce themselves and evolve by natural selection. However they are not considered to be living organisms because they are not made of cells, and cannot do any of the basic characteristics of life on their own. They are in fact more like complex chemicals than simple living organisms. Viruses are obligate parasites that can only reproduce inside host cells, which get damaged in the process, leading to disease. Viruses are thought to have arisen from lengths of DNA that became separated from their cells.

Virus Structure

Viruses have very simple structures, usually made of just three components:

- Nucleic acid, which can be DNA or RNA, and single or double-stranded. The nucleic acid typically codes for 5-100 proteins (by comparison, the bacterium *E. coli* has about 4000 genes).

- A protein coat called a capsid, made of subunits called capsomeres. Because capsids are composed of many repeating subunits, they tend to have simple geometrical shapes, such as a helix or icosahedron (20 triangular faces).



- A lipid envelope, not made by the virus itself, but derived from a host cell membrane. The lipid envelope may contain matrix proteins, which attach the capsid to the envelope, and glycoproteins, which attach the virus to host cells.

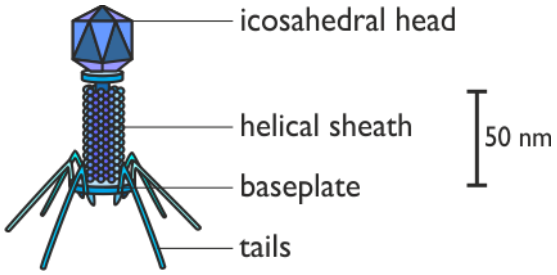
A complete virus particle outside a cell is called a virion. Virions are too small to see with a light microscope and were first seen in the 1930s using the electron microscope. Once inside a host cell virions dismantle into their separate parts so that the virus can be reproduced.

We need to know about four viruses:

	Structure	Nucleic Acid	Genome size	Host cell	Disease	Life Cycle
Phage λ	complex	ss DNA	6.4 kb 70 genes	E.coli bacteria	-	lysogenic and lytic
TMV	Helix	ss RNA (+ve sense)	6.4 kb 3 genes	Tobacco plants	Tobacco mosaic disease	
Ebola virus	Enveloped helix	ss RNA (-ve sense)	19 kb 7 genes	Human cells	haemorrhagic fever	budding
HIV	enveloped icosahedron	ss RNA (retrovirus)	10 kb 9 genes	Human white blood cells	AIDS	latency and budding

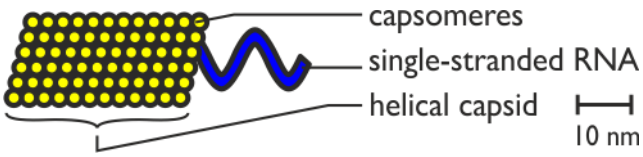
Phage Lambda

Bacterial viruses are known as bacteriophages (or just phages), which means "bacteria eaters". Phage Lambda (λ) infects the bacterium *E.coli*, which lives in human intestines. Like many phages, it has a complex structure combining icosahedral and helical capsids. The head contains double-stranded DNA.



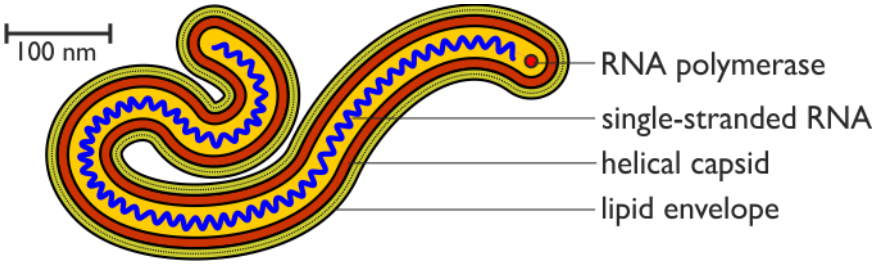
Tobacco Mosaic Virus (TMV)

Tobacco mosaic virus (TMV) was the first virus to be discovered, and is one of the smallest. It infects a wide variety plants (not just tobacco), where it causes a mosaic-like mottling of the leaves and causes a great deal of economic damage. TMV has a very simple structure: just a coil of single-stranded RNA surrounded by a helical capsid made of 2130 capsomeres.



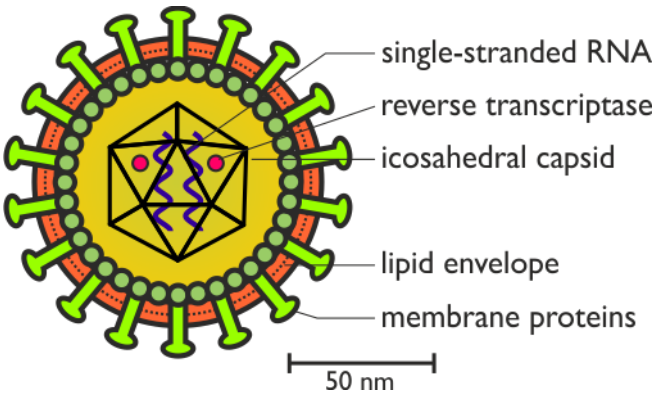
Ebola Virus

Ebola virus is a human virus that causes severe and often fatal hemorrhagic fever. It is a very large virion consisting of a negative-sense single-stranded RNA surrounded by a large flexible helical capsid 80nm in dameter, which in turn is enclosed in a lipid envelope. The virus usually coils up.



Human Immunodeficiency Virus (HIV)

The human immunodeficiency virus (HIV) infects the white blood cells of the immune system and causes the fatal disease AIDS. HIV comprises single-stranded RNA, surrounded by an icosahedral capsid, which is in turn surrounded by a sphere of proteins attached to a lipid envelope.



Viral Nucleic Acid and Reproductive Cycles

Unlike cells (which all contain double-stranded DNA), viruses contain a variety of nucleic acid types, including single and double-stranded DNA and RNA. Indeed viruses are usually classified by their type of nucleic acid, as this table shows:

Classification Group	Nucleic Acid Type	examples
1	Double-stranded DNA	**Phage** λ, smallpox, chicken pox, herpes
2	Single-stranded DNA	Parvo-virus
3	Double-stranded RNA	Reovirus
4	Single-stranded RNA (positive sense)	**TMV**, rhinovirus (common cold), polio
5	Single-stranded RNA (negative sense)	**Ebola**, measles, mumps, flu
6	Single-stranded RNA (retrovirus)	**HIV**, leukemia

70% of viruses contain RNA, but there are different ways in which this RNA is used to make proteins:

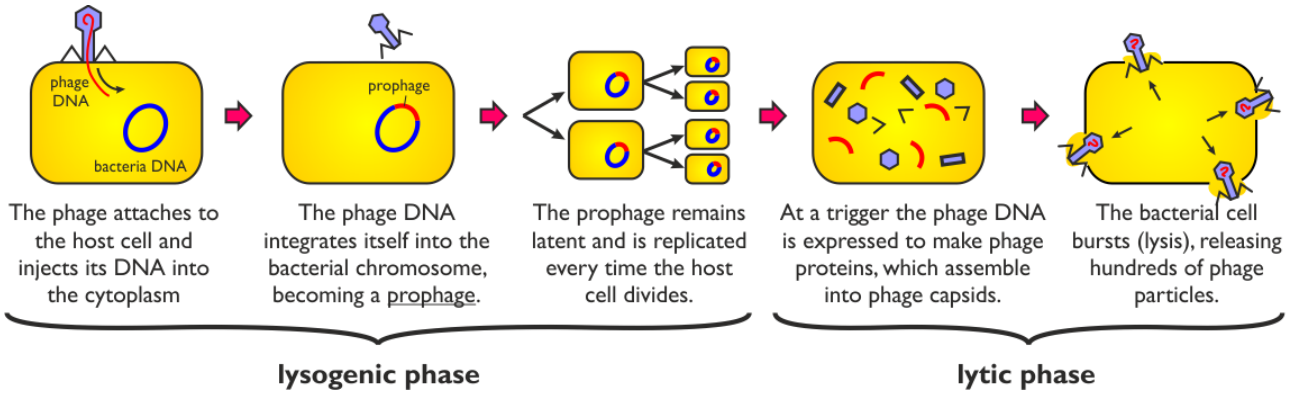
- Positive sense RNA has a sequence that can be translated directly by ribosomes to synthesise proteins, in other words the viral RNA is mRNA.
- Negative sense RNA has the complementary sequence to mRNA, so a complementary mRNA copy has to be made using a special RNA-directed RNA polymerase enzyme.
- Retroviruses have a unique enzyme called reverse transcriptase, which makes double-stranded DNA from a single-stranded RNA template (in other words the reverse of transcription). This DNA can then be integrated into the host cell's genomic DNA.

Viruses can only reproduce inside a host cell. They rely on the host cell's ribosomes, enzymes, nucleotides, amino acids, tRNA and even ATP to synthesise new viral nucleic acids and proteins. Different viruses have different reproductive cycles depending on their type of nucleic acid. The different ways these nucleic acids are used in the reproductive cycles of the four viruses are shown below.

Phage λ Reproductive Cycle

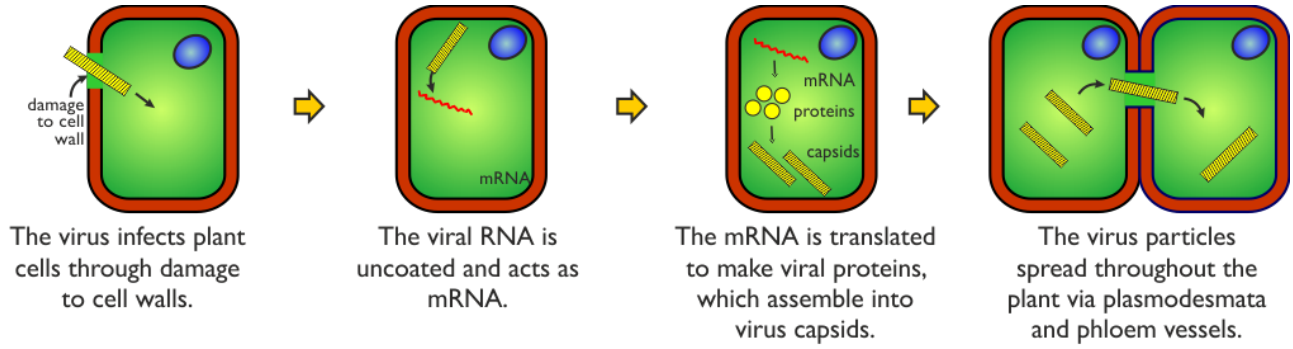
When phage λ infects bacterial cells the capsid remains outside the cell and injects the DNA into the cytoplasm. There are now two different pathways the virus can follow.

- In the lysogenic pathway the phage DNA is integrated into the bacterium's own DNA, where it becomes a latent prophage. In this latent state the virus doesn't cause any harm, but is replicated along with the bacterial DNA every time the cell divides. At some environmental signal the phage switches to the lytic pathway.
- In the lytic pathway the prophage separates from the bacterial DNA. The viral DNA is replicated, transcribed and translated using bacterial enzymes and ribosomes, to make thousands of phage capsid proteins. These proteins spontaneously assemble to form hundreds of capsids, which burst out of the cell. This causes cell lysis and kills the cell.



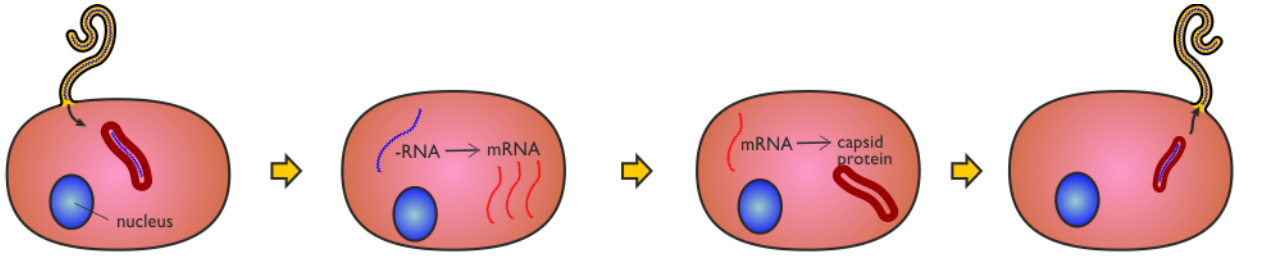
TMV Reproductive Cycle

TMV first infects plants through damage to cell walls and the capsid is disassembled. Because the viral single-stranded RNA is positive sense it can be used directly as mRNA. This mRNA is translated into capsid proteins using plant cell ribosomes and amino acids, and the RNA is also replicated. The virus re-assembles and spreads throughout the plant through plasmodesmata between cells, and through the phloem vessels.



Ebola Reproductive Cycle

The Ebola virus can infect a variety of human cells, including blood and liver cells. The viral envelope fuses with the cell membrane releasing the capsid into the cytoplasm. The negative-sense single-stranded RNA molecule is used a template for the viral enzyme RNA-directed RNA polymerase to make many copies of mRNA. This mRNA is translated using human cell ribosomes and amino acids to make viral capsid proteins. The virus reassembles and exits the cell through budding, collecting the envelope of host cell membrane as it leaves. This kills the cell.



The virus attaches to the host cell and the two membranes fuse, releasing the capsid into the cytoplasm

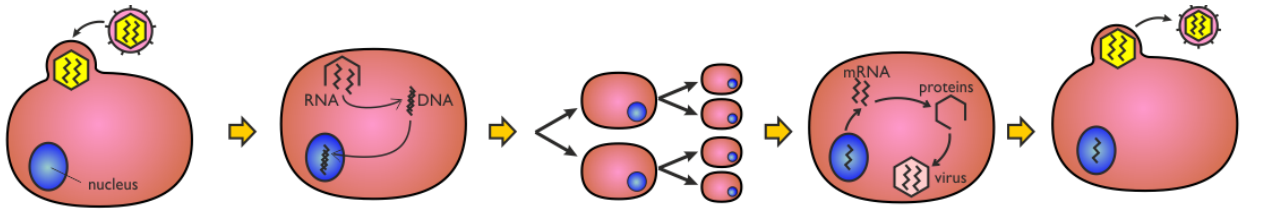
The viral RNA is uncoated and the viral enzyme RNA polymerase synthesises many copies of mRNA from the -RNA.

The mRNA is translated to make viral proteins, which assemble into virus capsids.

The virus particles are released by budding, collecting their lipid envelope from the host cell membrane.

HIV Reproductive Cycle

HIV infects T-lymphocyte cells of the human immune system. Virus particles cells attach to the lymphocyte membrane then the envelope and cell membrane fuse together, releasing the capsid inside the cytoplasm. HIV is a retrovirus, which means it can copy its RNA to make double-stranded DNA using the viral enzyme reverse transcriptase. This DNA is integrated into the nuclear DNA, where it can remain latent for many years. This is another example of latency, like phage λ. At an environmental trigger the viral DNA is suddenly expressed, making viral proteins, which assemble to form virions, which exit the cell by budding, killing the cell.



The virus attaches to the host cell and the two membranes fuse, releasing the capsid into the cytoplasm

The viral enzyme reverse transcriptase synthesises ds DNA from the ss RNA. The DNA is incorporated into the host's nuclear DNA, where it is called a provirus.

The provirus remains latent for years and is replicated every time the host cell divides.

At a trigger the viral DNA is expressed to make viral proteins, which assemble into virus capsids.

The virus particles are released by budding, collecting their lipid envelope from the host cell membrane.

Viral Diseases

All viruses cause disease. Since viruses cannot reproduce on their own, they must infect host cells to reproduce, and this reproduction inevitably harms or kills the host cell. Viruses cause disease in plants, animals, fungi and even bacteria. Human viral diseases are common and vary from harmless to fatal.

Disease	Virus
common cold	Rhinovirus
influenza	Myovirus
measles	Paramyxovirus
Rubella (German measles)	Rubella virus
Mumps	paramyxovirus
Chickenpox	Varicella zoster virus
hepatitis B	DNA virus
Ebola haemorrhagic fever	Ebola virus
AIDS	HIV

Treating Viral Diseases

It is easy to kill organisms (for example using heat, acid, high-energy radiation or bleach), but none of these methods can be used once a pathogen has already infected a human host. Treating diseases requires selectively toxic drugs that kill the pathogen but not the host cells. For example bacteria can be killed using antibiotics that target enzymes only found in bacteria, and fungi can be killed using antifungals that target enzymes only found in fungi.

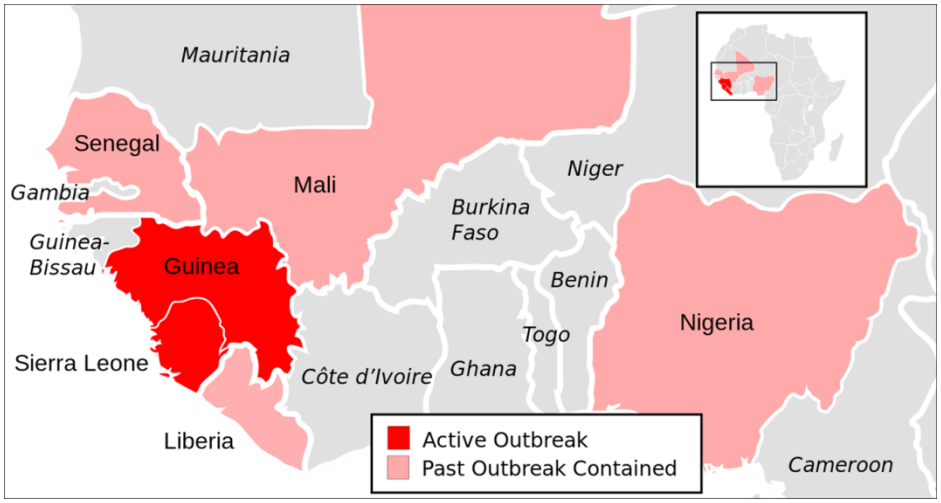
But viruses are not living cells and do not carry out metabolic reactions, which could be targeted by antiviral drugs. The only time viruses are active is when they reproduce inside a host cell. Viral reproduction mostly uses host cell enzymes, but there are a few viral enzymes that can be targeted. For example (you don't need to learn these, just appreciate them):

- AZT and acyclovir are nucleotide analogues that competitively inhibit the enzyme reverse transcriptase, only found in retroviruses like HIV. These drugs are therefore effective treatments for AIDS.
- Raltegravir inhibits the viral enzyme integrase, used by retroviruses to integrate their DNA into the host DNA.
- Tamiflu inhibits the viral enzyme neuraminidase that many viruses use to escape from cells. So this drug stops the viruses reproducing and spreading.
- Ribavirin is a RNA nucleotide analogue that inserts itself into viral RNA, thus stopping its translation.

Even with these drugs, some viral diseases are almost impossible to treat and rely on combination therapy, where three of four different antiviral drugs are used together. These have shown great success, especially with AIDS, but are very expensive and so not available in many countries.

Viral Disease Epidemics

An epidemic is when an infectious disease spreads far more quickly and widely than usual. If an epidemic spreads across continents it is called a pandemic. So the most important strategy for controlling epidemics is to limit the spread of the disease, especially for viral diseases, which are so difficult to treat with drugs. A good example is the 2014 outbreak of Ebola disease in Guinea, West Africa. The outbreak quickly spread to neighbouring countries Sierra Leone, Liberia, Mali and Sengal. The epidemic remains today (2015), with a total of 11,000 deaths so far.



There are three stages in preventing the spread of an epidemic:

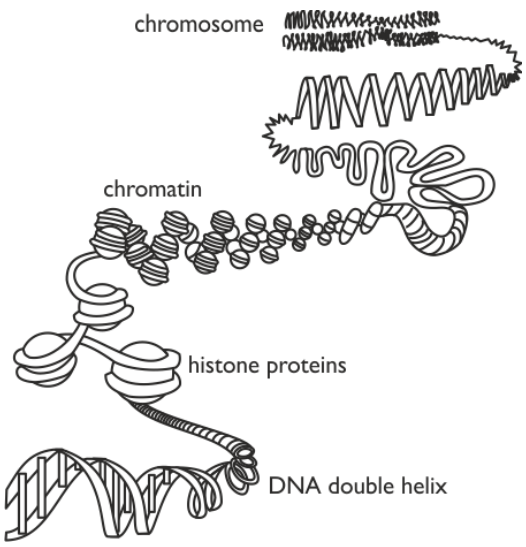
1. Identify the outbreak and respond quickly. In the 2013 Philippines Earthquake the health implications were immediately obvious and emergency health teams were on site within days. But in the 2014 West Africa Ebola outbreak, it took six months before the World Health Organisation (WHO) coordinated an international response.
2. Reduce spread by enforcing simple cultural and hygiene practices, such as hand-washing by healthcare workers and others in contact with patients and victims. About two-thirds of the Ebola cases in Guinea were caused by the local practice washing dead bodies, which are still infectious. This practice has now been stopped.
3. Impose travel restrictions to and from the region to prevent the infection from spreading to other countries. Many countries (including the USA) still impose restrictions on those travelling to and from the Ebola region. On 8 August 2014 a *cordon sanitaire* (or quarantine area) was established round a triangular area between Guinea, Liberia, and Sierra Leone, where 70% of the known cases had been found. No one was allowed to enter or leave, in order to stop further spread. The quarantine has since been reduced to a series of checkpoints for hand-washing and measuring body temperature on major roads throughout the region.

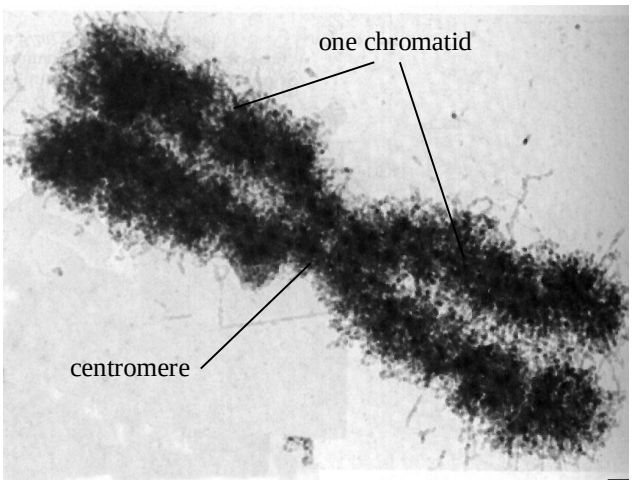
BLANK PAGE

DNA and Chromosomes

The DNA molecule in a single human cell is 1 m long, so is 10 000 times longer than the cell in which it resides (< 100μm). (Since an adult human has about 10^{14} cells, all the DNA is one human would stretch about 10^{14} m, which is a thousand times the distance between the Earth and the Sun.) In order to fit into the cell nucleus the DNA in eukaryotes is cut into shorter lengths and each length is tightly wrapped up with histone proteins to form a complex called chromatin.

Just before cell division the DNA is replicated, and more histone proteins are synthesised, so there is temporarily twice the normal amount of chromatin. Following replication the chromatin then coils up even tighter to form short fat bundles called chromosomes. These are about 100 000 times shorter than fully stretched DNA, and therefore 100 000 times thicker, so are thick enough to be seen with the light microscope. Each chromosome is roughly X-shaped because it contains two replicated copies of the DNA. The two arms of the X are therefore identical. They are called chromatids, and are joined at the centromere. (Do not confuse the two chromatids with the two strands of DNA.) The complex folding of DNA into chromosomes is shown below.

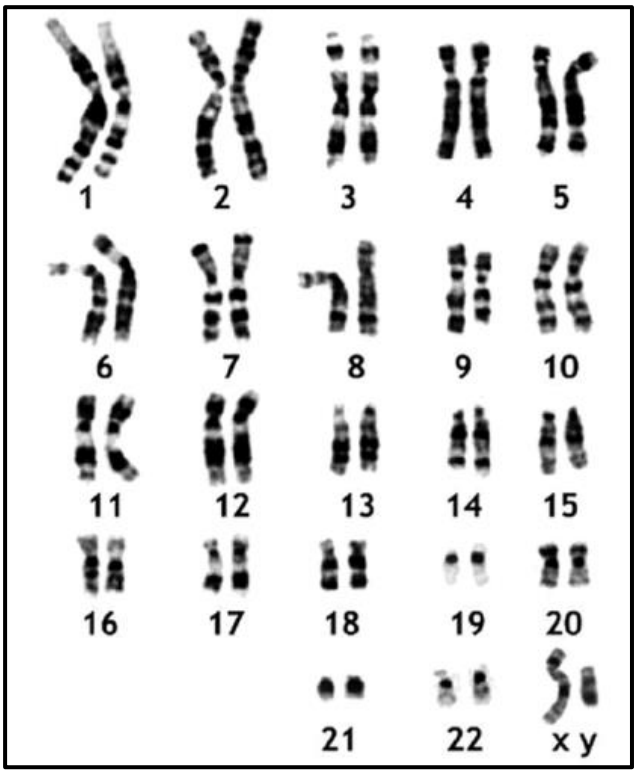




micrograph of a single chromosome

Karyotypes and Homologous Chromosomes

If a dividing cell is stained with a special fluorescent dye and examined under a microscope during cell division, the individual chromosomes can be distinguished. They can then be photographed and studied. This is a difficult and skilled procedure, and it often helps if the chromosomes are cut out and arranged in order of size.

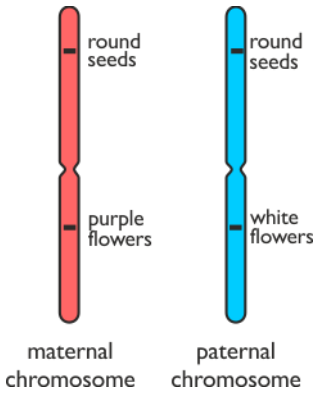


This display is called a karyotype, and it shows several features:

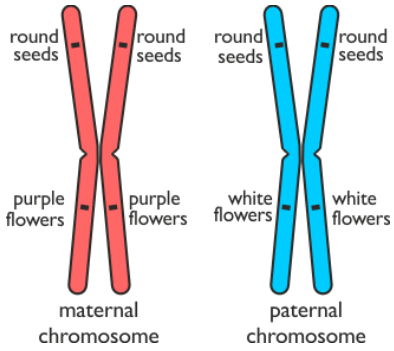
- Different species have different number of chromosomes, but all members of the same species have the same number. Humans have 46 (this was not known until 1956), chickens have 78, goldfish have 94, fruit flies have 8, potatoes have 48, onions have 16, and so on. The number of chromosomes does not appear to be related to the number of genes or amount of DNA.
- Each chromosome has a characteristic size, shape and banding pattern, which allows it to be identified and numbered. This is always the same within a species. The chromosomes are numbered from largest to smallest.
- Chromosomes come in pairs, with the same size, shape and banding pattern, called homologous pairs ("same shaped"). So there are two chromosome number 1s, two chromosome number 2s, etc., and humans really have 23 pairs of chromosomes.
- One pair of chromosomes is different in males and females. These are called the sex chromosomes, and are non-homologous in one of the sexes. In humans the sex chromosomes are homologous in females (XX) and non-homologous in males (XY). In other species it is the other way round. The non-sex chromosomes are called autosomes, so humans have 22 pairs of autosomes, and 1 pair of sex chromosomes.

It is important to understand exactly what homologous chromosomes are. We have two copies of each chromosome because we inherit one copy from each parent, so each homologous pair consists of a maternal and paternal version of the same chromosome. Since the homologous chromosomes contain the same genes, this also means we have two copies of each gene (again, one from each parent). This is why we write two letters for each gene in a genetic cross. The two homologous chromosomes may have the same versions (or alleles) of the gene (e.g. AA), or they may have different alleles, because one copy is a mutation (Aa).

Sometimes the chromosomes in a cell nucleus are represented by rods called ideograms, although these structures never actually exist because the chromatin is usually uncoiled. Each ideogram represents the long coiled DNA molecule in one chromosome. This diagram shows a pair of homologous chromosomes with two genes marked. The plant cell containing these chromosomes is homozygous for the seed shape gene (RR) and heterozygous for the flower colour gene (Pp).



The only time chromosomes can actually be seen is during cell division. At this point in the cell cycle each chromosome is made of two identical chromatids, because each DNA molecule has now been replicated. This diagram shows the same pair of homologous chromosomes during mitosis. The two chromatids in each chromosome contain the same alleles because they're exact replicas of each other. But again the two homologous chromosomes contain the same genes but different alleles.



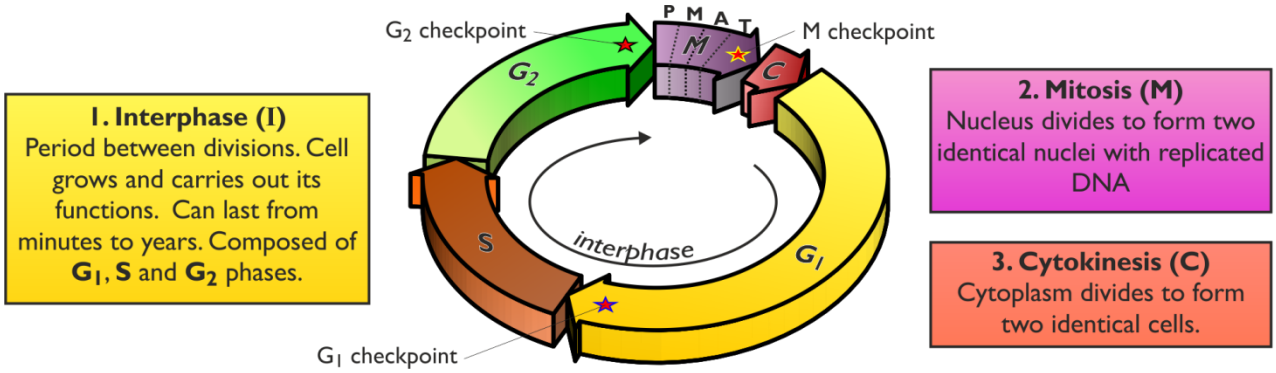
Chromatin	DNA + histone complex during interphase
Chromosome	compact X-shaped form of chromatin formed (and visible) during mitosis
Chromatids	the two arms of an X-shaped chromosome. The two chromatids are identical since they are formed by DNA replication.
Homologous chromosomes	two chromosome of the same size and shape, one originating from each parent. They contain the same genes, but different alleles.

The Cell Cycle

Cell theory states that new cells are always formed by division of old cells. The normal form of cell division in eukaryotes is mitosis, which forms two genetically-identical "daughter cells". Mitosis is used to make new cells for:

- **Growth**, when an organism is growing in mass. A human's 10^{14} body cells are made by mitosis from one original zygote.
- **Replacement**, to replace cells that are lost. Each day humans replace 10^{7} gut epithelial cells in faeces; 10^{7} skin epidermal cells in house dust; and 10^{11} red blood cells that are recycled.
- **Repair**, to replace cells that are damaged by injury e.g. to mend a broken bone or a cut in the skin.
- **Asexual Reproduction**, in some eukaryotic organisms (e.g. yeast, strawberry, starfish).
- **Artificial Cloning** of plants and animals in biotechnology.

The life of a cell from one division to the next is called the cell cycle, and has three main phases:



1. **Interphase.** This is when the cell grows and carries out its normal functions (e.g. respires, synthesises molecules, secretes hormones, contracts, transmits nerve impulses, etc.). Typically 90% of the cell cycle is spent in interphase. Interphase is sub-divided into:
   - **Gap phase G_1**, where the cell grows back to its original size. Genes are expressed into whatever proteins are needed by this cell, and organelles are replicated.
   - **Synthesis phase S**, where DNA and histones are replicated in preparation for mitosis. This can take a few hours.
   - **Gap phase G_2**, where spindle proteins are synthesised, ready for mitosis.
2. **Mitosis Phase.** This is where the nucleus divides to make two nuclei with identical copies of the DNA that was replicated during the previous S phase. The mitotic phase can be sub-divided into the four phases: prophase, metaphase, anaphase and telophase (details below).
3. **Cytokinesis.** This is division of the cytoplasm, to form two daughter cells with one nucleus each.

Control of the Cell Cycle

The cell cycle is tightly controlled, so that cells only divide when they need to. The cycle is controlled through checkpoints, where the cell is checked to see if it is safe to proceed to the next phase of the cycle:

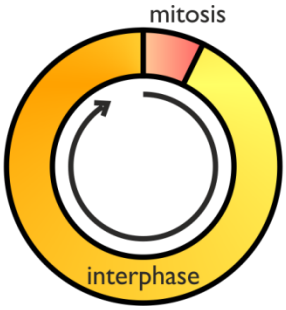
- The **G_1** checkpoint is only passed if there is a need for more cells, signalled by an external factor such as a growth hormone. This prevents excessive cell division.
- The **G_2** checkpoint is only passed if DNA has been replicated correctly and there is the correct number of chromosomes.
- The **M** checkpoint is only passed if mitosis has been completed successfully.

Sometimes the checkpoints can fail, perhaps due to mutations or viral infection. This failure allows cells to divide uncontrollably, leading to cancer.

The cell cycle can last from hours to years in different cell types. For example embryonic cells divide very quickly with no growth in between, so they have very short interphases, just long enough to replicate the DNA. Skin cells divide about every 12 hours on average, liver cells every 2 years, and muscle cells never divide at all after maturing, so remain in the G_1 phase for decades.

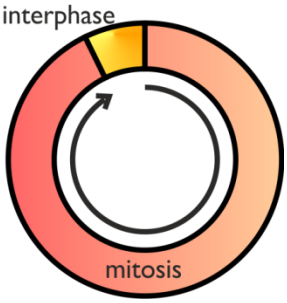
Slow cell cycle in a mature tissue.

mitosis:	1h
interphase	100h
total cycle	101h



Fast cell cycle in a growing tissue.

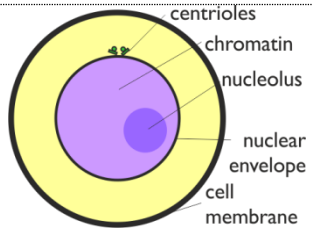
mitosis:	60min
interphase	10min
total cycle	70min



Mitosis

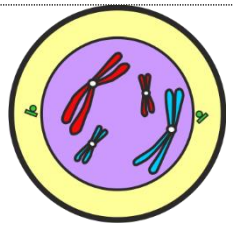
The point of mitosis is to make two genetically-identical cells, and the purpose of the stages shown below is to ensure that the two daughter cells each receive a replicated copy of the DNA. This cell has just two pairs of homologous chromosomes.

Interphase



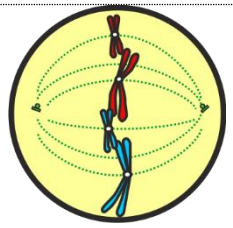
- no chromosomes visible
- DNA, histones and centrioles all replicated

Prophase



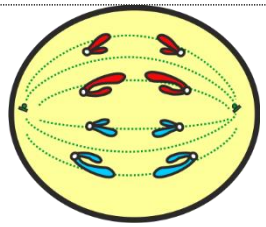
- chromosomes condensed and visible
- centrioles at opposite poles of cell
- nucleolus disappears

Metaphase



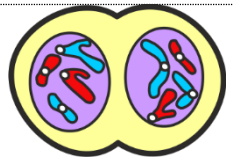
- nuclear envelope disappears
- chromosomes align along equator of cell
- spindle fibres (microtubules) connect centrioles to chromosomes

Anaphase



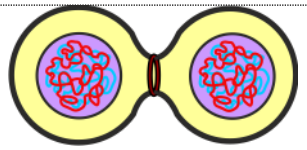
- centromeres split, allowing chromatids to separate
- microtubules shorten, pulling chromatids towards poles, centromeres first
- Other microtubules lengthen, pushing poles apart and elongating cell.

Telophase

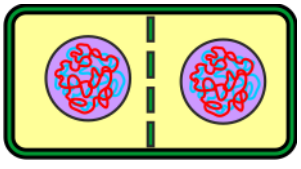


- spindle fibres disperse
- nuclear envelopes form
- chromatids uncoil and become too thin to see

Cytokinesis



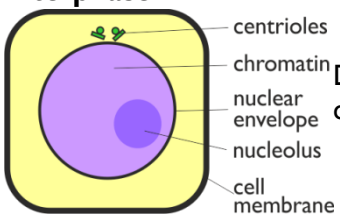
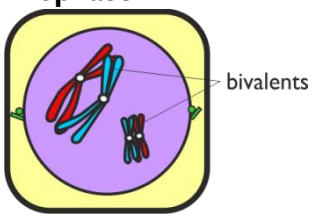
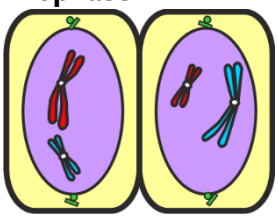
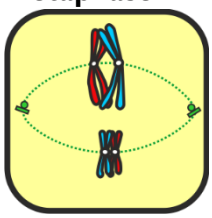
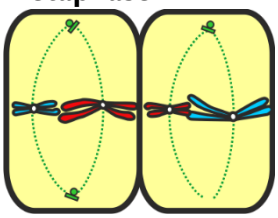
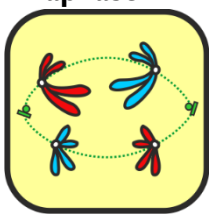
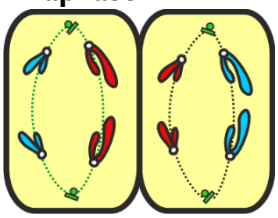
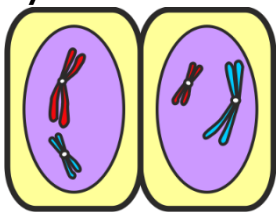
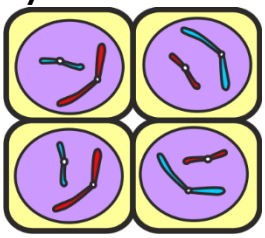
- In animal cells a ring of actin filaments forms round the equator of the cell, and then tightens to form a cleavage furrow, which splits the cell in two.



- In plant cells vesicles move to the equator, line up and fuse to form two membranes called the cell plate. A new cell wall is laid down between the membranes, which fuses with the existing cell wall.

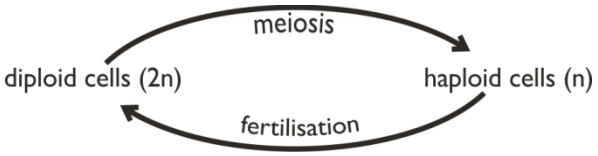
Meiosis

Meiosis is the special cell division used by sexually-reproducing organisms to make gametes. It starts with DNA replication, like mitosis, but then proceeds with two divisions one immediately after the other. Meiosis therefore results in four daughter cells rather than the two cells formed by mitosis. Each division has four phases, like mitosis, but, while meiosis II is just like mitosis, meiosis I has a number of significant differences.

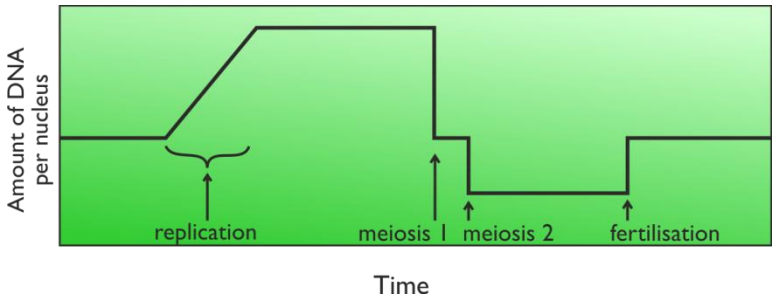
Division I		Division II	
Interphase 	DNA, histones and centrioles all replicated		Meiosis division II proceeds immediately without any interphase. There is no replication of DNA.
Prophase I 	Chromosomes condense and homologous chromosomes join together to form bivalents.	**Prophase II** 	Centrioles move to form new poles at 90° to original poles. Chromosomes already coiled.
Metaphase I 	Bivalents line up on the equator.	**Metaphase II** 	Chromosomes line up on new equator.
Anaphase I 	Bivalents separate and homologous chromosomes move to opposite poles. Chromatids do not separate.	**Anaphase II** 	Centromeres finally split, separating chromosomes. Chromatids move to opposite poles.
Telophase I and Cytokinesis I 	Two nuclei with replicated chromosomes. Chromosomes don't uncoil. Cells divide, but remain joined.	**Telophase II and Cytokinesis II** 	Nuclear envelopes form around each set of chromosomes. Cells divide, but remain joined to form a tetrad.

Meiosis differs from mitosis in two important aspects (apart from the two divisions):

Firstly, in meiosis the chromosome number is halved from the normal diploid number (2n) to the haploid (half) number (n). This is necessary so that the chromosome number remains constant from generation to generation, so in sexual reproduction meiosis is always followed by fertilisation at some point in the life cycle:



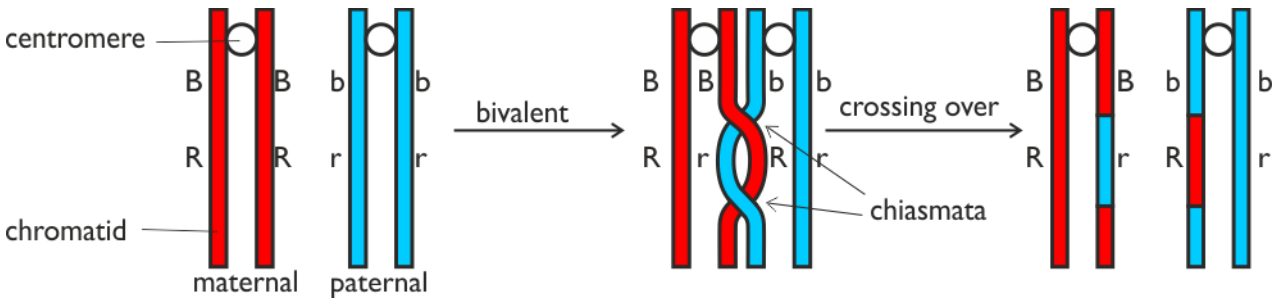
The halving is done in a particular way: meiosis ensures that each haploid cell has one of each homologous pair of chromosomes. So for example human gametes have 23 chromosomes: one of each homologous pair. Remember that other species have different haploid numbers. The changes in chromosome number can also be shown as a graph of DNA mass over time:



Secondly, in meiosis the chromosomes are re-arranged during meiosis to form new combinations of alleles. This genetic recombination is vitally important and is a major source of genetic variation. It means for example that of all the millions of sperm produced by a single human male, the probability is that no two will be identical. The whole point of meiosis and sex is to introduce genetic variation, which allows species to adapt to their environment and so to evolve. There are two sources of genetic variation in meiosis: crossing over and Independent assortment. We'll look at each of these in turn.

Crossing Over

This happens at prophase I of meiosis, when the bivalents first form. While the two homologous chromosomes are joined in a bivalent, bits of one chromosome are swapped (crossed over) with the corresponding bits of the other chromosome.

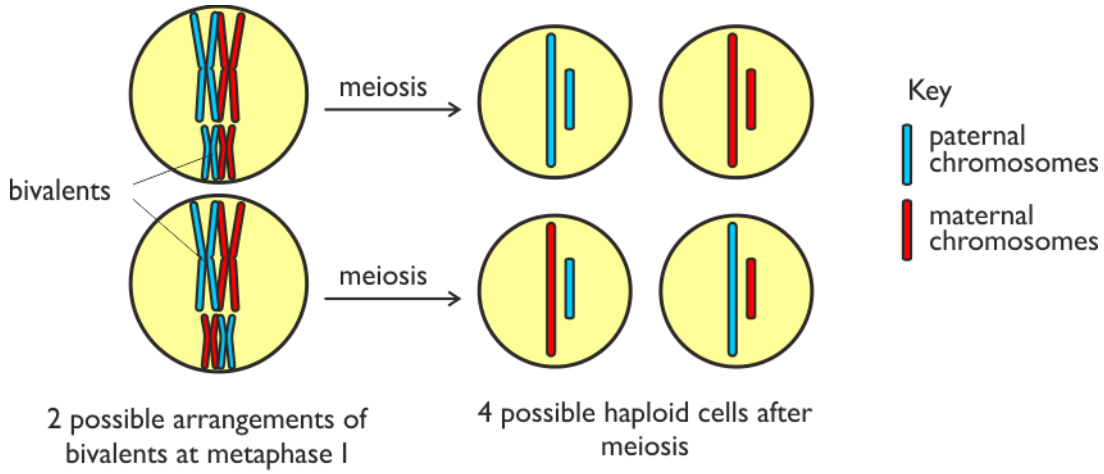


The points at which the chromosomes actually cross over are called chiasmata (singular chiasma), and they involve large, multi-enzyme complexes that cut and join the DNA. There is always at least one chiasma in a bivalent, but there are usually many, and it is the chiasmata that actually hold the bivalent together. The chiasmata can be seen under the microscope and they can give the bivalents some characteristic strange shapes. There are always equal amounts crossed over, so the chromosomes stay the same length.

Crossing over means that maternal and paternal alleles can be combined even though they are on physically different chromosomes. In the example in the diagram some gametes will have the new genotype combinations Br or bR. This potentially allows any combination of alleles to form and, since there are around 20 000 genes in humans, there is the potential to make an astronomically large number of combinations.

Independent Assortment

This happens during metaphase I of meiosis, when the homologous chromosomes join together to form bivalents that line up on the equator. Each bivalent is made up of two homologous chromosomes, which originally came from two different parents (they're often called maternal and paternal chromosomes). Since they can line up in any orientation on the equator, the maternal and paternal versions of the different chromosomes can be mixed up in the final gametes.



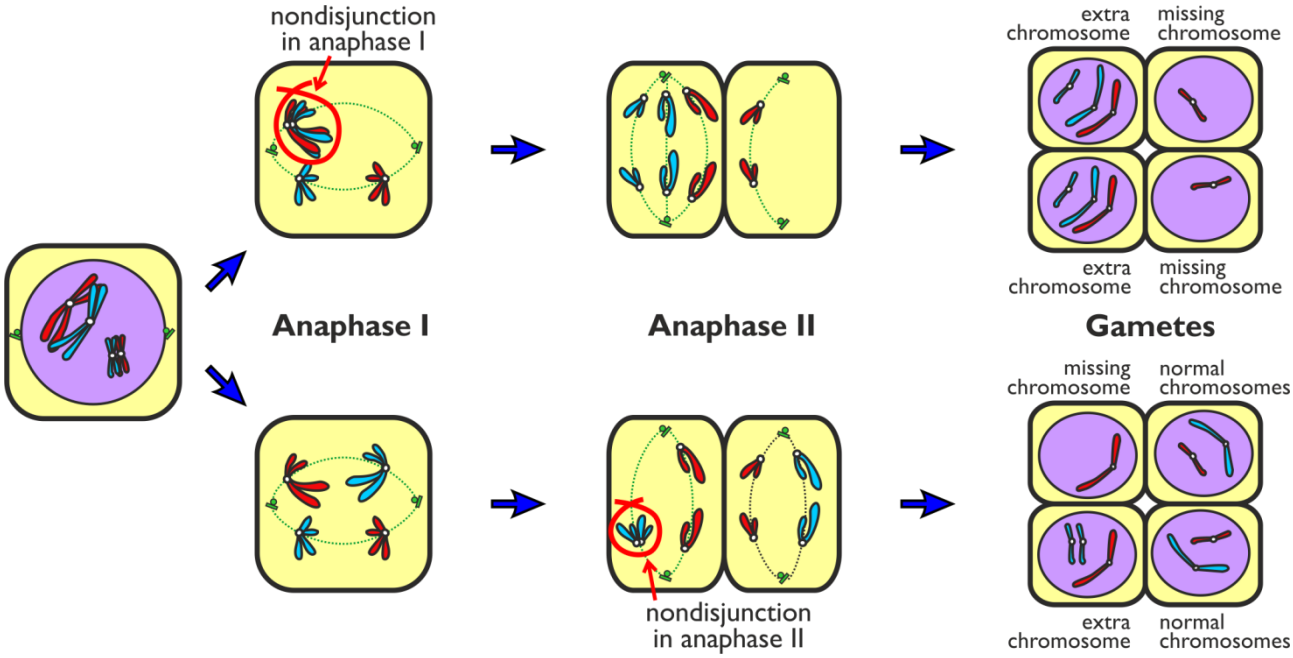
In this simple example with 2 homologous chromosomes (n=2) there are 4 possible different gametes (2^2). In humans with n=23 there are over 8 million possible different gametes (2^{23}). This is just the number of different arrangements of chromosomes. When we consider all the different rearrangements of genes within chromosomes due to crossing over we can understand why every gamete produced by an individual is genetically unique.

Chromosome Mutations

Earlier we looked at gene mutations, where a few bases in DNA were changed, leading to new alleles. But mutations can also happen on a larger scale, when large regions of a chromosome, perhaps containing hundreds of genes, can be changed. These large mutations are known as chromosome mutations, and include nondisjunction (which changes the number of chromosomes) and translocation, which changes the structure of chromosomes.

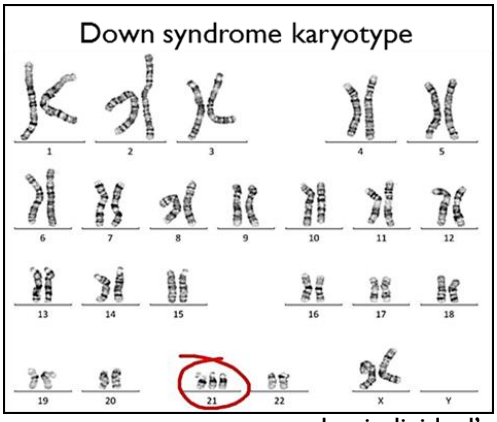
Nondisjunction

Nondisjunction is a chromosome mutation that changes the number of chromosomes in a cell. It is caused by a fault in anaphase of meiosis, where the chromatids or chromosomes normally separate and move to opposite poles. In nondisjunction the chromatids or chromosomes remain stuck together and move together to one pole of the cell. This nondisjunction can happen at anaphase I or anaphase II, but in both cases it results in some gametes with an extra chromosome, and some with a missing chromosome.



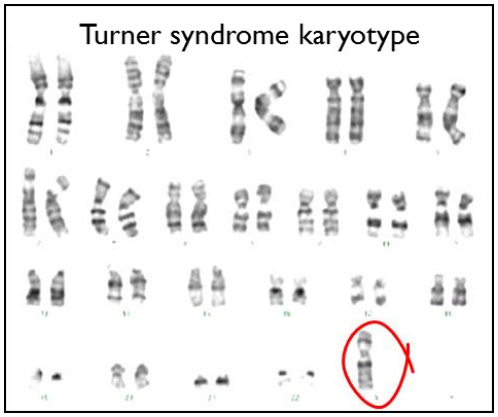
If an abnormal gamete with too many or too few chromosomes is fertilised then the resulting zygote will show polysomy in that chromosome, since there will be the wrong number of chromosomes. Some examples of polysomy are shown below.

Trisomy, e.g. Down syndrome



If a gamete with an extra chromosome is fertilised by a normal gamete then the resulting zygote will show trisomy in that chromosome, since there will be three copies instead of two. Trisomy is almost always fatal early in embryo development, as it appears that the extra chromosome presents too great a burden on cell division and gene expression. The only common exception is trisomy in the smallest chromosome – chromosome 21 (21 is actually smaller than 22). Trisomy 21 causes Down syndrome, which, although not fatal, does have severe consequences on the individual's phenotype. Symptoms of Down syndrome include particular recognisable facial features, short height, heart defects, poor vision, severe learning difficulties and a shorter life expectancy.

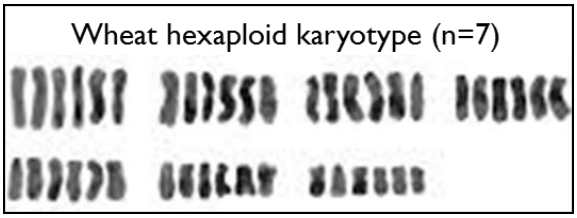
Monosomy e.g. Turner syndrome



If a gamete with a missing chromosome is fertilised by a normal gamete then the resulting zygote will show monosomy in that chromosome, since there will only be one copy instead of two. Monosomy is almost always fatal early in embryo development, since a full complement of genes is required for normal cell function. The only exception is the X chromosome: it is possible to survive with a single copy of the X chromosome (after all that's all males have!). Monosomy X (often written as X0 in contrast to the normal XX) is called Turner syndrome and occurs about once in every 5000 births. X0 individuals appear female but their sex organs do not mature at adolescence, and they are sterile. They have normal intelligence.

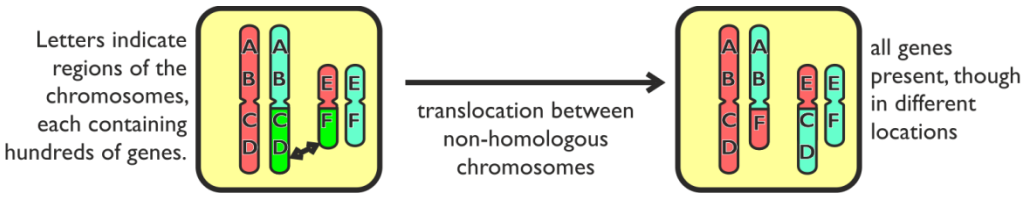
Polyploidy

Polyploidy happens when nondisjunction causes all the chromosomes to remain together, forming a gamete with two copies of every chromosome (in other words a diploid gamete). If this gamete is fertilised by a normal haploid gamete the resulting zygote has three copies of the entire genome – it's triploid. Surprisingly, polyploidy has little adverse effect on phenotype and is quite common in nature, especially in the plant kingdom. For example apples are triploid (3n) and wheat is hexaploid (6n). It seems that carrying an extra set of chromosomes (polyploidy) is easier than carrying just one extra chromosome (polysomy).



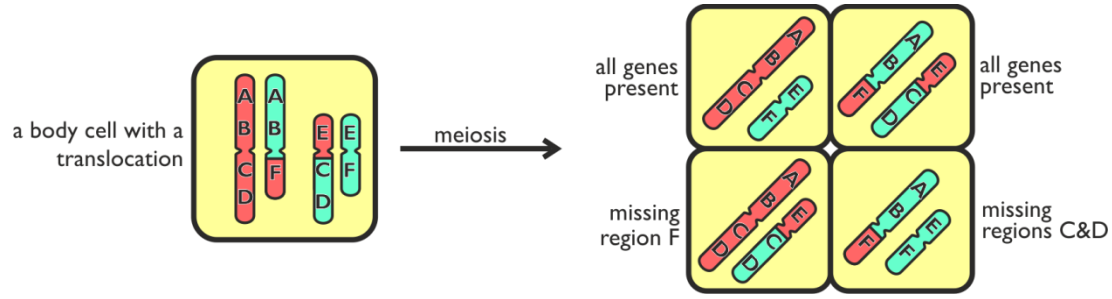
Translocation

A translocation is a chromosome mutation where one part of a chromosome is swapped with a part of a different chromosome. This sounds a bit like crossing over in meiosis, but in crossing over equal parts are swapped between homologous chromosomes, whereas in translocation the swapped parts can be unequal in size and they are swapped between non-homologous chromosomes. The result is different-sized chromosomes, and indeed translocations can be identified by their appearance in a karyotype.



In this example the bottom part of a chromosome 1 (regions labelled C and D) is swapped with the bottom part of a chromosome 2 (region labelled F). This is an unequal translocation, so chromosome 1 gets smaller and that chromosome 2 gets bigger. Exactly the same genes and alleles are still present, albeit some are now located on different chromosomes. Translocation mutations are therefore usually harmless, with little or no effect on the phenotype. Around 1 in 500 people are thought to contain translocation mutations. There are however two possible complications caused by a translocation:

1. Cells with a translocation can undergo mitosis without any problems, but meiosis doesn't work properly since some of the homologous chromosomes are no longer truly homologous and so cannot form bivalents properly. Even if meiosis completes, some of the gametes produced will be missing certain regions of some chromosomes and therefore hundreds of genes (see diagram below). If these gametes are fertilised the resulting embryo will be unviable. So individuals with a translocation may be perfectly normal but are unsuccessful in having children, often resulting in miscarriage. This difficulty in having children can be the first time someone may realise they have a translocation.



2. When the translocated regions are cut and swapped into their new locations they may disrupt a gene. This disrupted gene will no longer function. Several different kinds of cancer have been found to be caused by translocations in genes that regulate the cell cycle (pxx). Alternatively a new "fusion gene" may be formed by joining two parts of genes together, which can cause unpredictable effects of the phenotype.