

2

Biochemistry and cell biology

Patricia Revest

Introduction	15	•	Specialised plasma membrane structures	58
Basic chemistry	16	•	Cellular transport processes	60
Chemical bonds	16	•	Cell-to-cell communication	65
Basic chemical reactions	19	•	Types of signalling	65
Inorganic molecules and ions	19	•	Ligands and receptors	66
Biological compounds	20	•	Cell recognition	68
Simple organic compounds	20	•	Cell adhesion molecules	68
Macromolecules	20	•	Cell division and DNA replication	68
Proteins	20	•	Stem cells	68
Carbohydrates	32	•	Phases of the cell cycle	69
Lipids	35	•	Mitosis	69
Nucleic acids	39	•	Meiosis	70
Basic cell structure	44	•	DNA replication	71
Histology	44	•	Cellular ageing and cell death	72
Organelles, structure and function	44	•	Key metabolic pathways	74
Intracellular and extracellular fluids	50	•	Glycolysis	75
Protein synthesis and processing	51	•	The metabolism of fatty acids	76
Genes and the genome	51	•	TCA cycle	77
The genetic code	52	•	Amino acid metabolism	78
Transcription	52	•	Electron transport chain and oxidative phosphorylation	78
Translation	54	•	Tissues and organs	80
Processing pathways for proteins	56	•	Origins of cells	80
Protein production in the cytosol	56	•	Basic tissue types	80
Membrane structure and functions	57	•	Organs and systems	88
Fluid mosaic model of membranes	57	•		

INTRODUCTION

The human body is mainly water plus a wide variety of biologically active chemicals which subserve all the functions of the body. **Biochemistry** describes how these molecules are made and the interactions between them at molecular level.

Cell biology then goes on to describe how the biochemicals are organised into cells and cellular components, which then form the tissues of the body. Normal processes within the body are called **physiological**, whereas processes that cause disease are called **pathological**.

Metabolism

All the processes of replication, growth, repair and so on that are vital to the survival of a cell require energy. Much of the economy of any cell is taken over by the chemistry required, firstly, to generate that energy, and, secondly, to harness it to drive crucial processes or to build molecules needed by the cell to maintain itself. All this chemistry is termed **metabolism**. We can think of cells as being extremely complex chemical machines. At any time a cell may have many thousands of individual chemical reactions going on simultaneously, each a part of a **metabolic pathway** that may result in the breakdown of larger molecules, or the assembly of

large molecules from smaller ones, or simply the conversion of one molecule into another. All of these pathways need to be controlled so that molecules are produced in the right place in the right amounts.

Catabolism and anabolism

In general, metabolism that involves degrading complex molecules to simpler ones, called **catabolism**, is the route by which chemical energy is generated. By contrast, the metabolism that results in the synthesis of more complex molecules from simpler components, **anabolism**, has a net requirement for chemical energy. It is these synthetic metabolic pathways that produce the components needed for cell growth, repair and replication.

Cells

The **cell** is the basic unit of living organisms. All living things consist of cells and although there are many organisms which exist as single cells, such as bacteria and amoeba, the largest living organisms on Earth, such as the great blue whale, also consist of aggregates of the same basic cellular units, although in rather large numbers. An adult human has somewhere in the region of 10^{15} cells.

All living organisms are made up of one of two types of cell: **prokaryotes** and **eukaryotes**.

Table 2.1 Eukaryotic kingdoms

Kingdom	Examples	Examples of pathogens
Protists	Protozoans, algae, slime moulds	<i>Plasmodium</i> spp. (malaria)
Fungi	Yeasts, moulds, mushrooms	<i>Candida albicans</i> (candidiasis or thrush)
Plants	Mosses, ferns, conifers, flowering plants	None, although many plants are poisonous to humans
Animals	Sponges, corals, arthropods, amphibians, reptiles, birds, mammals	<i>Taenia solium</i> (tapeworm)

Table 2.2 Chemical elements of the human body

	Element (atomic symbol)	Abundance (% wet weight)
Major elements	Oxygen (O)	65.0
	Carbon (C)	18.5
	Hydrogen (H)	9.5
	Nitrogen (N)	3.3
	Total	96.3
Minor elements (in order of abundance)	Calcium (Ca), phosphorus (P), potassium (K), sulphur (S), sodium (Na), chlorine (Cl), magnesium (Mg)	
	Total	3.7
Trace elements (in alphabetical order)	Aluminium (Al), boron (B), chromium (Cr), cobalt (Co), copper (Cu), fluorine (F), iron (Fe), manganese (Mn), molybdenum (Mo), selenium (Se), silicon (Si), tin (Sn), vanadium (V), zinc (Zn)	

Prokaryotes

Two of the major kingdoms into which organisms are classified are the Eubacteria and the Archaeobacteria. These are all prokaryotes. Prokaryotic cells lack a nucleus and other internal structures. All prokaryotic organisms are single celled; they are extremely numerous and can be found in almost all environments, including some which seem too hostile to support living things. The earliest organisms on Earth were prokaryotes, and they were the only living things for more than 2 billion years. From the point of view of medicine, the bacteria are important in that they cause many of the infectious diseases.

Many chemicals and chemical processes which occur in the human body are similar to those found in many other organisms, including prokaryotes. However, there are also significant differences and sometimes these can be exploited, for example in the use of antibiotics that can kill bacterial cells that have invaded the body, without killing the host, by exploiting differences in their biochemistry.

Eukaryotes

The other four kingdoms are made up of organisms with **eukaryotic** cells (Table 2.1). These cells have a nucleus and other internal structures which allow compartmentation of chemical processes within a single cell. The Protista contains both unicellular and multicellular organisms and it is thought that the organisms in the other three kingdoms evolved from ancient protists.

Viruses

Although **viruses** are not cells and as such are not strictly living organisms, they are dependent on the host's cells to reproduce, i.e. they are parasites. The effects of this parasitism are the symptoms of many viral infections. However, as with bacteria, differences in chemical processes used by viruses can be exploited to produce antiviral drugs.

Cells and viruses exist on different scales. Viruses are very small (60–80 nm), which is smaller than some of the intracellular components of eukaryotic cells. Most bacteria

are between 1 μm and 10 μm and many eukaryotic cells are 10–100 μm .

BASIC CHEMISTRY

Chemical composition of the human body

The chemicals which make up the human body are mainly compounds consisting of combinations of carbon, hydrogen and oxygen atoms (Table 2.2) with small quantities of other elements. These are known as **organic** compounds because they contain carbon. Together with nitrogen, carbon, hydrogen and oxygen make up 96.3% of the body weight. The remaining 3.7% is made up of seven **minor elements**. As well as these major and minor elements there are a number of **trace elements** which are required in very small amounts (less than 0.01%). Many of these are needed as special components of proteins called enzymes and are required for their activation (see Regulation of enzyme activity).

CHEMICAL BONDS

In order for atoms to combine to form molecules, chemical bonds must be formed between the individual atoms (Fig. 2.1). There are three main types of chemical bonds:

- Ionic
- Covalent
- Hydrogen.

The strength of the different types of bond varies enormously (Table 2.3). Stronger bonds are more stable but, often, large numbers of weaker bonds stabilise molecular interactions.

Atomic structure

In order to understand how chemical bonds are formed it is necessary to know something about the structure of atoms. All atoms of a particular element have a nucleus of **protons**

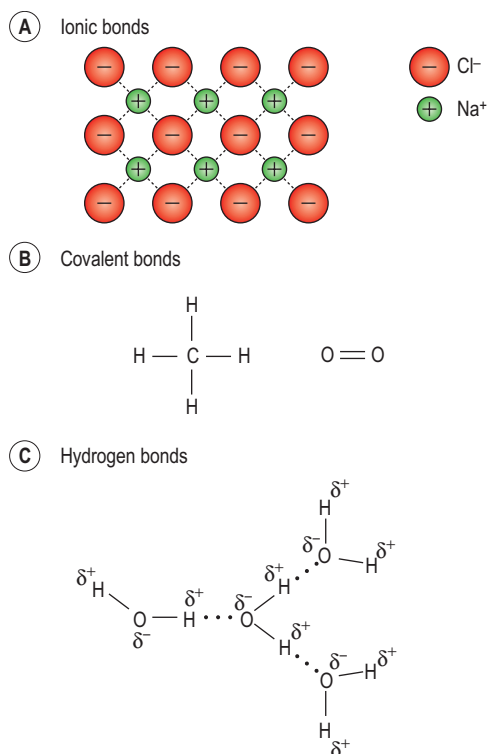


Fig. 2.1 Different types of chemical bond.

Table 2.3 Energy of some typical chemical bonds in biological systems

Type of bond	Energy (kJ/mol)
Ionic	12.6–29.3
Covalent (single)	210–460
Covalent (double)	500–710
Covalent (triple)	815
Hydrogen	4.2–8.4
Van der Waals interaction	4.2

and **neutrons** (except ordinary hydrogen which has a single proton as a nucleus) surrounded by a cloud of electrons. The total number of protons in an element gives its **atomic number**, and the total of protons and neutrons (which may vary between different forms or **isotopes** of an element) gives the **atomic mass**. Carbon has six protons and an atomic number of 6, but it can have six, seven or eight neutrons. Its atomic mass is thus the average of the mass of the different isotopes: 12.01. This reflects the much greater abundance of carbon-12 (where 12 is the atomic mass) than the other isotopes, carbon-13 and carbon-14.

Atomic shells

The electrons which surround the nucleus are found in fixed energy levels called shells and there are strict rules about how many electrons can occupy each shell:

- The innermost shell can contain a maximum of two electrons
- The next shell can contain eight electrons
- The third shell can hold up to 18.

Atoms are at their most stable when their outermost shell is either full or contains eight electrons. The most stable

elements which exist are the noble gases such as helium or neon, which, with two and 10 electrons, respectively, have full outer shells. These atoms are chemically inert, their full outer shells rendering them non-reactive, and they rarely form compounds with other atoms. Atoms with unfilled outer shells are more reactive.

Ionic bonds

When dissolved in water many atoms are able to form charged particles called **ions** by either losing or gaining electrons from their outer shell:

- Atoms which gain an electron become negatively charged **anions**.
- Atoms which lose an electron become positively charged **cations**.

These ions will be attracted to each other by the opposite electrical charges, resulting in the formation of **ionic bonds**. The tendency of an atom to lose or gain electrons depends on the number of electrons already present.

Sodium chloride: an example of an ionic bond

Common salt, or sodium chloride, is a compound formed by the combination of sodium and chlorine atoms. Sodium has 11 electrons: both its inner two shells are full and it has a single electron in its outer shell. Chlorine has 17 electrons, again with full inner shells but with seven electrons in its outer shell. In water, sodium (Na) atoms lose a single electron to become sodium (Na^+) cations with a single positive charge. Chlorine (Cl) atoms gain the electron lost by the sodium to become chloride (Cl^-) anions with a single negative charge. When sodium chloride solution is dehydrated, then the ions form a lattice of interlocking anions and cations held together by the ionic bonds between the ions to give white salt crystals (Fig. 2.1A). The resulting sodium chloride has completely different properties from its component parts. Sodium is a highly reactive metal and chlorine is a poisonous green gas, but sodium chloride is the main ionic constituent of seawater, where life was thought to have evolved, and forms the major ions in the extracellular fluid which bathe all the cells in the body.

Ionic bonds are frequently formed between molecules such as potassium (K) or calcium (Ca), which have either one or two electrons in their outer shell – forming K^+ and Ca^{2+} , respectively – and fluorine or iodine, which like chlorine require only one electron to fill their outer shell.

Hydrogen ions

Hydrogen ions are formed by losing its single electron to form a cation (H^+). This ion is stable because its nucleus with its lone proton has an empty electron shell. However, hydrogen can also form an anion (H^-) with more reactive metals, called a hydride, with two electrons in its single shell. Water forms a small number of ions when it dissociates into H^+ and OH^- (hydroxyl) ions. This is depicted as $\text{H}_2\text{O} \rightleftharpoons \text{H}^+ + \text{OH}^-$ and is reversible, as indicated by the double arrows. These ions are important in many chemical reactions.

Covalent bonds

It is not necessary for an atom to completely lose or gain an electron in order to participate in chemical bonding. Atoms can share their electrons with other atoms to fill their outer

electron shells part of the time. Up to three pairs of electrons can be shared, producing increasingly strong links known as **covalent bonds**. These bonds can form between atoms of the same element, as in the formation of O_2 , where two oxygen atoms, each with six electrons in the second shell, share two pairs of electrons forming a double bond ($O=O$). Two hydrogen atoms form molecules of H_2 by sharing their electrons with a single bond ($H-H$). When a single atom of oxygen shares electrons with two atoms of hydrogen the result is the most abundant biological molecule of them all, water or H_2O .

Carbon bonds

Carbon can form a wide variety of covalent bonds with other atoms by sharing four of its electrons to form pairs (Fig. 2.1B). When a single carbon atom combines with four hydrogen atoms the simplest organic molecule, methane (CH_4), is formed. However, carbon can form long chains of atoms both with other carbon atoms and with other atoms to form complex biochemicals.

Stereoisomers

The four covalent bonds around a carbon atom are arranged in a tetrahedral pattern. When the four chemical groups are different this means that the molecule can exist as two mirror images, which cannot be superimposed. The asymmetric carbon atom is said to be **chiral** and the two molecules are called **stereoisomers**. Stereoisomers rotate polarised light in either a clockwise, or dextrorotatory (+), direction or in an anticlockwise, or levorotatory (-), direction.

Stereoisomers can only be converted, one to the other, by breaking and re-forming the covalent bonds. A molecule with one chiral centre has two stereoisomers, and one with n chiral centres has 2^n possible stereoisomers. This means that many organic molecules with identical formulae and chemical groups can exist in different forms depending on the number of asymmetrical carbon atoms. This is particularly important in the structures of macromolecules such as proteins and carbohydrates, in which the building blocks consist of specific stereoisomers.

Carbon-carbon double bonds

Carbon atoms can also share two pairs of electrons to form carbon-carbon double bonds. These bonds are planar and rigid and do not rotate. When they occur in carbon chains they produce either **cis** or **trans isomers** depending on whether the next carbon atoms are:

- On the same side of the double bond (*cis*)
- On opposite sides (*trans*).

This isomerism is important in the structures of fatty acids, which are components of many lipids. Most important dietary fatty acids are in the *cis* form but commercially processed fats in foods such as some margarines and peanut butter contain *trans* forms, which are known to raise blood cholesterol. Recent nutritional guidelines recommend that consumption of these types of fatty acids be limited.

Polar and non-polar covalent bonds

In covalent bonds between atoms, the electrons are not always equally shared between the atoms. When the electrons are more attracted to one of the atomic nuclei, then the bond has two poles of charge called a **dipole** and the bond is called a **polar** covalent bond. This occurs in H_2O , where

the electrons are closer to the oxygen atom than the hydrogen atoms. The oxygen then has a partial negative charge (δ^-) and the hydrogen atoms partial positive charges (δ^+). These partial charges affect the way these molecules react, and are particularly important in the way water molecules interact. In compounds such as methane, where the electrons are shared equally, the bonds are called **non-polar** covalent bonds.

Hydrogen bonds

The partial positive charges caused by the unequal sharing of electrons by hydrogen atoms are attracted to the partial negative charges on atoms such as oxygen and nitrogen. This forms a weak bond called a **hydrogen bond**. Although this type of bond is only about 1/20 as strong as a covalent bond, the presence of a large number of hydrogen bonds can produce a significant cohesive force. Many large molecules contain numerous hydrogen bonds between different parts of the molecule and these help to maintain the structure.

Hydrogen bonds can be disrupted easily, by heat and by excess acidity, causing a molecule to lose its shape and possibly its function. This is shown clearly in the formation of ice crystals where the regular lattice of the frozen water molecules is held together by the hydrogen bonds between the molecules (see Fig. 2.1c). When ice is heated, the hydrogen bonds are disrupted and ice becomes liquid water.

Other molecular interactions

Two types of interaction between non-polar molecules produce attractive forces between molecules:

- Van der Waals interactions
- Hydrophobic interactions.

Van der Waals interactions happen when transient dipoles occur in the electron clouds of uncharged molecules in close proximity. The resultant force depends on the distance between the molecules, as these interactions can only occur when the atoms are 0.3–0.4 nm apart.

Van der Waals interactions are weaker than hydrogen bonding and are very dependent on the precise molecular shapes of the interacting molecules. However, two interlocking molecules can form very many of these contacts and, because of this, van der Waals interactions are particularly important in specific interactions between biological molecules, such as enzymes and their substrates.

Hydrophobic interactions take place between non-polar molecules, which do not contain either ions or dipoles and so do not interact with water molecules and are almost completely insoluble in aqueous solution. These molecules will aggregate together to exclude water. Such hydrophobic interactions are due to the fact that the insertion of a non-polar molecule into water requires energy because it involves distorting the interactions between the polar water molecules. Thus the most stable conformation of non-polar molecules is to exclude water and to form structures that have the least possible surface area exposed to water.

This exclusion of water by hydrophobic forces is the basis of cell membrane structure. Cell membranes are largely made up of molecules that are polar at one end, where they interact with water, attached to long, non-polar carbon and hydrogen chains which aggregate together and do not react with water.

BASIC CHEMICAL REACTIONS

In order for molecules to be formed, chemical bonds must be made or broken between the constituent atoms. There are three basic types of chemical reaction:

- **Synthetic** or **combination** reactions in which bonds form between two molecules or atoms to form a larger molecule, e.g. $X + Y \rightarrow XY$.
- **Decomposition** reactions where a molecule is broken down into smaller parts, e.g. $XY \rightarrow X + Y$.
- **Exchange** reactions where atoms are exchanged between molecules, e.g. $XY + AB \rightarrow XA + YB$.

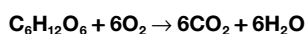
Some reactions are thermodynamically irreversible; that is they will only proceed in a single direction. However, many reactions can be reversed, although sometimes this may require special conditions.

Oxidation–reduction (redox) reactions

A commonly occurring biochemical reaction is the **oxidation–reduction** or **redox** reaction. In this reaction one molecule loses electron(s) in a process called oxidation, with the lost electron(s) being transferred to another molecule in a process called reduction. The name ‘redox’ derives from the fact that, in biological systems, the molecule which often receives the transferred electron is oxygen.

Many redox reactions also involve the loss of two hydrogen atoms in the form of a hydrogen ion (H^+) and a hydride ion (H^-). Oxidation–reduction reactions allow energy, in the form of chemical bonds, to be transferred from the molecule that is oxidised to the molecule that is reduced. The process is best thought of as two tightly linked processes in which there is a reduction in the potential energy of the oxidised molecule, whereas the potential energy of the reduced molecule is increased.

For example, one of the basic reactions used to provide energy to cells by the metabolism of glucose is summarised as:



glucose + oxygen → **carbon dioxide + water**

This is an oxidation–reduction reaction: glucose is oxidised to carbon dioxide, losing hydrogen, and oxygen accepts the hydrogen and is reduced to water. However, in cellular metabolism it does not occur in a single step. In a series of linked reactions, the energy liberated in the oxidation of the glucose is transferred to other molecules which can then be used to drive other reactions.

Energy in chemical reactions

The breaking of chemical bonds requires energy input, whereas the formation of new bonds releases energy.

Thus chemical reactions produce:

- Either a net release of energy (**exergonic** reactions)
- Or a net absorption of energy (**endergonic** reactions).

In exergonic reactions the energy required to break existing bonds is less than that released by the newly formed bonds. The products have a lower potential energy than the reactants. The reverse occurs in endergonic reactions, where the products have a higher potential energy than the reactants. Many biochemical reactions (e.g. the redox reactions)

couple exergonic and endergonic reactions to transfer energy between molecules.

Activation energy

The energy initially required to break the chemical bonds in the reactants is the **activation energy**. In order to react, molecules and atoms must collide with sufficient kinetic energy to overcome the repulsive force between their electron clouds, and this kinetic energy can be increased by raising the temperature. Increasing the concentration of the reactants also increases the chance of favourable collisions. Some reactions also require energy input to alter certain properties of the covalent bonds and others may require the excitation of electrons from inner electron shells into the outermost shell before the reaction can occur.

Catalysts

Special transport systems and mechanisms of compartmentation within cells can help to increase the effective concentration of a compound. However, large increases in temperature are undesirable in biological systems and many biochemical reactions would not occur without the presence of a **catalyst**. Reactions between molecules occur more readily if the reactants collide in the correct orientation. Catalysts act to lower the activation energy of the reaction by helping the reactant molecules to collide in a more favourable direction. The catalysts themselves are unchanged by the process so they can be used again and again. Biological catalysts called enzymes allow a large number of chemical reactions to occur at an appropriate rate and under normal physiological conditions, such as at normal body temperature. (see Enzymes, below)

INORGANIC MOLECULES AND IONS

We tend to think of life as being carbon-based. However, as well as the large number of carbon-containing organic molecules, there are large numbers of molecules which do not contain carbon but are also essential for life. These include water and many salts, acids and bases.

Water

Water makes up about 60% of the human body by weight. The large amount of water which is present in and around the cells has a number of functions. The most important of these is as a **solvent** for all the molecules, the **solutes**, which need to be in **solution** in order to participate in chemical reactions. Water is also an important **substrate** in many of these reactions. Many chemicals are transported around the body in solution and, without water, the gases oxygen and carbon dioxide could not be moved between the lungs and the blood.

A 70 kg human has approximately 42 L of total body water divided into two main compartments:

- **Intracellular fluid** contained within cells: about 28 L (approximately 40% of body weight).
- **Extracellular fluid:** about 14 L (approximately 20% of body weight), divided between the interstitial fluid of about 10.5 L and the blood plasma volume of about 3.5 L.

Water acts as a **lubricant**, preventing solid structures rubbing against one another and causing damage. It also

cushions the brain within the solid skull, reducing the pressure applied by the brain's own weight to the base of the brain. This cushioning prevents some of the damage which could occur due to sudden movements of the head banging the brain against the inside of the skull (see Ch. 8).

The high thermal capacity of water also allows it to be used to dissipate heat easily from areas of high temperature, preventing cell and tissue damage, and the large amount of heat required to evaporate water released as sweat is a very efficient cooling mechanism.

Acids, bases and salts

Acids, bases and salts are organic or inorganic compounds which, when dissociated in water, form ions:

- Acids form hydrogen ions (H^+) and anions
- Bases form hydroxyl ions (OH^-) and cations
- Salts form anions and cations.

Acids and bases are crucially important in the control of the acidity (pH) of the body, which is maintained within very strict limits close to the neutral pH of 7 (see Ch. 1).

The salts sodium chloride (NaCl) and potassium chloride (KCl) are the main ionic constituents of the fluids outside and inside cells, respectively. Sodium and potassium ions are essential for the transmission of nerve impulses, and calcium is required for muscle contraction. However, the commonest salts present in the body are the calcium phosphates which make up the bones and teeth (see Ch. 9).

BIOLOGICAL COMPOUNDS

SIMPLE ORGANIC COMPOUNDS

The simplest types of organic molecules are the **hydrocarbons**, which are composed of carbon and hydrogen. They are divided into two groups:

- Those that do not contain double bonds and are called **saturated**
- Those containing double bonds and are called **unsaturated**.

Hydrocarbons are **non-polar** and **hydrophobic**. Long hydrocarbon chains form the fats that make up cell membranes and form dense energy stores. Another distinctive type of hydrocarbon is produced when carbon atoms form a ring structure with six carbon and six hydrogen atoms, known as benzene. These ring structures, called aromatic rings, may also contain nitrogen.

Common chemical groups

Apart from hydrocarbons, all of the other biochemicals in the body consist of molecules in which hydrogen has been replaced by other atoms or groups of atoms. Potentially, although there is a huge range of possible groupings of atoms that can be formed, there are a limited number of distinctive functional groups found on biological molecules (Table 2.4). The reactivity of these groups determines the roles played by the molecules that contain them and determine their interactions with other molecules.

Table 2.4 Common functional groups and classes of compounds (R represents the rest of the molecule)

Group	Formula	Class
Hydroxyl	R-OH	Alcohols
Aldehyde	R-COH	Aldehydes
Carbonyl	R-CO-R'	Ketones
Carboxylate	R-COOH	Carboxylic acids
Ester	R-COO-R'	Esters
Amino	R-NH ₂	Amines
Imino	R-NH	Imines
Amide	R-CONH ₂	Amides
Phosphate	R-PO ₄	Organic phosphates
Sulphydryl	R-SH	Thiols

Table 2.5 Macromolecules

Macromolecule	Monomer	Major functions
Proteins (15%)*	Amino acids	Structural material, biological catalysts
Carbohydrates (2%)*	Sugars	Energy source, cell surface markers
Nucleic acids	Nucleotides	Genetic information (DNA and RNA)
Lipids (20%)*	(Many lipids are triacylglycerols)	Energy storage, cell membranes

*Note the percentages quoted are very variable due to the difference in fat and protein content in different individuals. The values shown here are a guide to their relative abundance.

MACROMOLECULES

Although a large number of small molecules are involved in biochemical processes, it is only in the formation of large molecules that the complexity of life becomes evident. There are three major groups of **macromolecules** present in the body. Each of these groups is formed from a number of related small molecules, **monomers**, which can be linked together by covalent bonds into polymers. The three groups of macromolecules are:

- Proteins
- Carbohydrates
- Nucleic acids.

A fourth group of molecules, which are not macromolecules in the same sense, are the **lipids**. These form macromolecular complexes by hydrophobic bonding.

Together, these four groups of molecules make up most of the remaining 40% of the human body, after water (Table 2.5). Although each of these groups has a particular type of building block, they often contain other molecules in varying amounts. For example, proteins often have carbohydrates covalently linked to them.

PROTEINS

Proteins have an enormous diversity of roles in the body and have the most complex structures of all of the macromolecules. Although all proteins are made in a similar manner and formed from similar building blocks, they have a wide variety of structures and functions, some examples of which are described below.

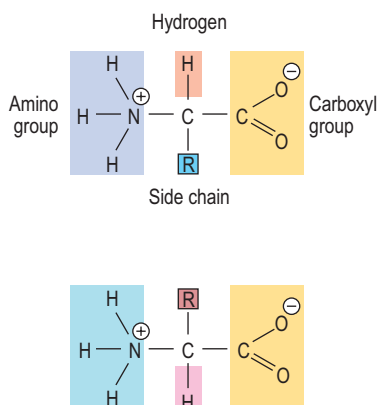


Fig. 2.2 Generalised amino acid with ionised NH_3^+ and COO^- groups in its two stereoisomers. The two molecules are mirror images of each other and cannot be superimposed.

Amino acids

Amino acids are the building blocks from which proteins are made. They also have other roles as neurotransmitters, as a source of energy, and as key intermediates in the maintenance of nitrogen balance.

Amino acid structure

Amino acids are a group of molecules with a common structure, which consists of a central carbon atom, attached to which are four different groups:

- A hydrogen atom ($-\text{H}$)
- An **amino** group ($-\text{NH}_2$)
- A **carboxyl** group ($-\text{COOH}$)
- A **side chain** ($-\text{R}$), which varies between the different amino acids.

At the normal pH of body fluids the amino and carboxyl groups of amino acids are ionised to NH_3^+ and COO^- , respectively (Fig. 2.2). This produces a molecule with positive and negative charges at opposite ends called a **zwitterion**. However, when amino acids bond together to form proteins, these charged groups combine. Thus it is only the side chains (and the terminal groups) that remain charged in proteins.

A wide range of amino acids are found in living things, but only 20 of these are made into human proteins (Table 2.6). The side chains vary in size and complexity, from a single hydrogen atom in glycine to the large side chains of tryptophan and arginine (Fig. 2.3). Each amino acid has a three-letter and a single-letter abbreviation.

Essential amino acids

There is a continual turnover of proteins in the body. This varies from minutes, in the case of some enzymes involved in metabolic control, to months, in the case of the main structural protein, collagen. Although some of the amino acids are reused many are metabolised, which means that even adults (who are not growing) require a daily dietary intake of protein. Eight of the 20 amino acids found in human proteins are called **essential amino acids** and must be included as components of the diet because they cannot be manufactured from metabolic intermediates (see Ch. 16). Dietary protein must be broken down in the gut to its constituent amino acids, which are then absorbed into the blood.

Table 2.6 Amino acids found in human proteins

Classification	Amino acid	Abbreviations	
Non-polar aliphatic	Glycine	Gly	G
	Alanine	Ala	A
	Valine	Val	V
	Leucine	Leu	L
	Isoleucine	Ile	I
	Proline	Pro	P
Non-polar aromatic	Phenylalanine	Phe	F
	Tyrosine	Tyr	Y
	Tryptophan	Trp	W
Polar uncharged	Serine	Ser	S
	Threonine	Thr	T
	Asparagine	Asn	N
	Glutamine	Gln	Q
Polar negatively charged	Glutamic acid	Glu	E
	Aspartic acid	Asp	D
Polar positively charged	Lysine	Lys	K
	Arginine	Arg	R
	Histidine	His	H
Sulphur-containing	Cysteine	Cys	C
	Methionine	Met	M

Enantiomers

The central carbon atom is called the α -carbon. In all the amino acids except glycine, the presence of four different groups around the α -carbon means that each amino acid can exist in two **stereoisomers**. These pairs of amino acids are called **enantiomers** and are named D- (**dextro** = right) and L- (**laevo** = left). With very few exceptions, only L-amino acids are incorporated into proteins. However, D-amino acids are important constituents of the cell walls of bacteria, and as such are targets for antibiotics such as penicillin. One of the 20 amino acids, proline, is not actually an α -amino acid but an α -imino acid with an $-\text{NH}-$ group instead of the amino ($-\text{NH}_2$) group. This has important structural consequences for proteins that contain this particular amino acid (see below).

Classification of amino acids

Amino acids are classified into groups according to the chemical properties of their side chains, R. For example, amino acids with polar side chains, and particularly charged ones, will form hydrogen bonds with water and will thus tend to be found on the parts of the protein exposed to the aqueous environment. Non-polar amino acids will be **hydrophobic** and will be found in parts of the protein not exposed to water. Amino acids are classified as follows:

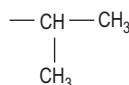
- **Non-polar aliphatic:** these have side chains which consist of carbon and hydrogen atoms with no double bonds. These saturated hydrocarbons are hydrophobic.
- **Non-polar aromatic:** these side chains all contain aromatic rings. They are generally hydrophobic except for tyrosine, which has a hydroxyl group that can be found on the hydrophilic part of proteins.
- **Polar uncharged:** these neutral amino acids have polar hydroxyl or amide groups. The hydroxyl groups on serine and threonine, like tyrosine, are often on the hydrophilic surface of proteins where they are often modified to change the activity of the protein by phosphorylation.
- **Polar negatively charged:** these acidic amino acids have carboxyl groups that are negatively charged at pH 7.

Non-polar aliphatic

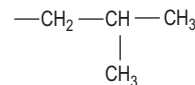
Glycine



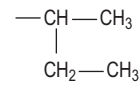
Alanine



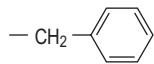
Valine



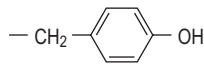
Leucine



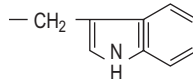
Isoleucine

Non-polar aromatic

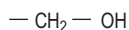
Phenylalanine



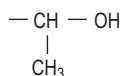
Tyrosine



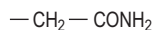
Tryptophan

Polar uncharged

Serine



Threonine



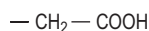
Asparagine



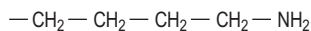
Glutamine

Polar negatively charged

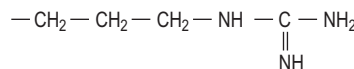
Glutamic acid



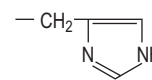
Aspartic acid

Polar positively charged

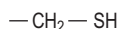
Lysine



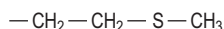
Arginine



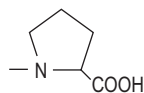
Histidine

Sulphur-containing

Cysteine



Methionine

Imino

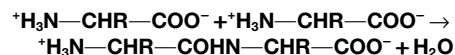
Proline

Fig. 2.3 Structures of the 20 common amino acids. This diagram shows only the R group of each amino acid.

- **Polar positively charged:** lysine and arginine are both positively charged at pH 7. However, the side chain of histidine is mainly charged at a pH of 6.8 and mainly uncharged at a pH of 7.8. Variations in pH can therefore influence whether histidine residues are charged or not.
- **Sulphur-containing:** adjacent cysteine residues on protein chains can form bonds between their sulphur atoms called disulphide bonds. These are important in maintaining the structure of the protein and in forming links between separate protein chains.

Peptide bonds

Amino acids are linked together to make proteins by the formation of bonds between the carboxyl group of one amino acid and the amino group of the other, with the elimination of a water molecule. The reaction leaves an amide bond called a **peptide bond**.



This reaction can be repeated, adding more amino acids to form a chain of amino acid residues with a carboxyl group at one end, the C-terminal, and an amino group at the other end, the N-terminal. Molecules containing two amino acid residues are called **dipeptides**, those with three, **tripeptides**. Usually peptides with more than 50 residues are called **proteins**, or **polypeptides**, and they can be made up of thousands of amino acids. Amino acid residues in a protein are numbered from the N-terminal end, which is the end from which protein synthesis starts.

The single covalent peptide bond has some of the rigidity of a double bond because electrons are unevenly shared between the COHN atoms (Fig. 2.4). This means that rotation around the C—N linkage is limited. However, rotation can still occur around the α -carbon atom and this rotation affects the way that the chains can fold.

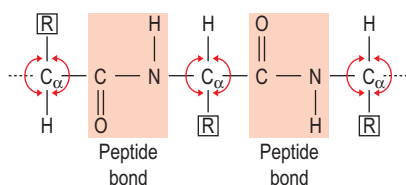


Fig. 2.4 The peptide bond. The peptide bond is relatively rigid but rotation can occur around the α -carbon.

Protein structure

Although proteins can be thought of as a linear sequence of amino acids, this does not convey the fact that, in order to function properly, each protein chain must be formed into the appropriate three-dimensional shape. There are four levels of structure that are used to describe proteins.

- 1. Primary:** the amino acid sequence of the protein linked by peptide bonds.
- 2. Secondary:** regular structures determined by hydrogen bonding between the atoms of the backbone.
- 3. Tertiary:** precise folding patterns of the polypeptide chain stabilised by a wide range of bonds between the amino acid side chains.
- 4. Quaternary:** in multimeric proteins, the number and arrangement of the subunits.

Primary structure

The primary structure of a protein is the sequence of amino acids that makes up the protein chain. A chain with a large number of hydrophobic residues will have completely different properties from one with many polar, charged side chains. Likewise, the characteristics of a particular region of a protein will depend on the average properties of the amino acids in that region. Proline residues produce a very rigid molecule with kinks in the protein chain, and glycine residues have only a single hydrogen as their side chain and so take up only a small space. Glycine and proline residues are, therefore, often found in the regions of a protein where it loops back on itself.

Secondary structure

The primary structure of a protein determines the overall folding pattern of the polypeptide chain. However, there are a number of regular patterns of folding which occur, stabilised by hydrogen bonding between the atoms of the backbone of the peptide chain. These secondary structures are not dependent on the nature of the side chains as the hydrogen bonds form between the carbonyl group (C=O) of one peptide bond and the hydrogen atom of another.

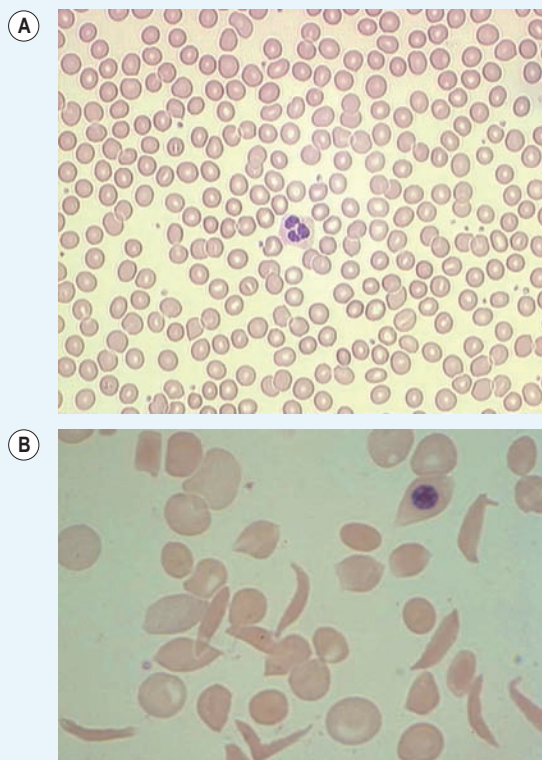
There are two common types of secondary structure (Fig. 2.5), one forms a helical structure known as an α -helix and another forms ribbons which can interact with one another to form extensive β -sheets. A third important structural motif is the **U-turn**, which allows protein chains to rapidly change direction.

The α -helix

An α -helix is formed when the polypeptide chain is twisted into a rod, with the carbonyl group of each peptide bond hydrogen bonded to the hydrogen of the amide, which is four residues further down the chain (Fig. 2.5A). This coiling is

Information box 2.1 Sickle cell anaemia

Changes in a single amino acid in a protein chain can have extreme consequences in the function of that protein. In sickle cell anaemia, the haemoglobin protein of red blood cells has a single change in one of the 146 amino acids which make up the β -globin chain. This is due to a mutation in the DNA which changes a single base from A to T, resulting in a glutamate (glutamic acid) residue at position 6 in the normal protein being replaced by a valine. This substitution of a negatively charged side chain by a non-polar hydrophobic one allows the proteins to form rigid rods under low oxygen conditions, changing the shape of the cells and making them inflexible and unable to travel through small blood vessels as readily. These rigid cells are phagocytosed more readily than normal cells, producing the characteristic anaemia.



Blood films: (A) normal; (B) with sickle cells.

always right-handed (clockwise) and has a pitch of 3.6 residues per turn. Each amino acid adds 0.15 nm to the length of the helix. The structure is rigid and all the side chains face outwards. This means that the overall hydrophobicity of these α -helical regions of the protein are determined by the side chains and not by the backbone. Proline residues are rarely found in α -helices because their rigid structure would not allow them to twist at the appropriate angle.

β -Sheets

β -sheets occur where hydrogen bonds form between peptide chains that lie alongside each other (Fig. 2.5B). Because the four bonds around the α -carbon atom are arranged in tetrahedral fashion, the chains cannot lie completely flat and the β -sheet has a pleated appearance with the side chains protruding above and below the plane of the sheet. Some proteins are formed from multiple layers of β -sheets stacked on

i Information box 2.2 α -Helical structures can cross cell membranes

The α -helix is commonly found in proteins that cross the cell membrane. These proteins have regions in which there are sequences of hydrophobic amino acid side chains that are about 25 residues long. An α -helix formed in these regions gives a rod with a hydrophobic exterior about 3.75 nm long, i.e. the length required to span the hydrophobic interior of the cell membrane.

Some of the helices have hydrophobic residues on one side of the helix and charged, hydrophilic residues on the other side. These are called amphipathic helices and they are found in proteins which form pores or channels in the cell membrane.

top of each other. An example of this is silk: its flexibility results from β -sheets sliding over one another, while its strength is due to the axis of the protein backbone lying parallel to the silk fibres.

U-turns

A third type of secondary structure formed due to hydrogen bonding is the **U-turn**. A U-turn consists of three to four residues, commonly glycine and proline, which form a short loop often found on the surface of proteins. The absence of a large side chain in glycine and the kink produced by proline allow the polypeptide chain to fold tightly back in the protein, facilitating a compact tertiary structure.

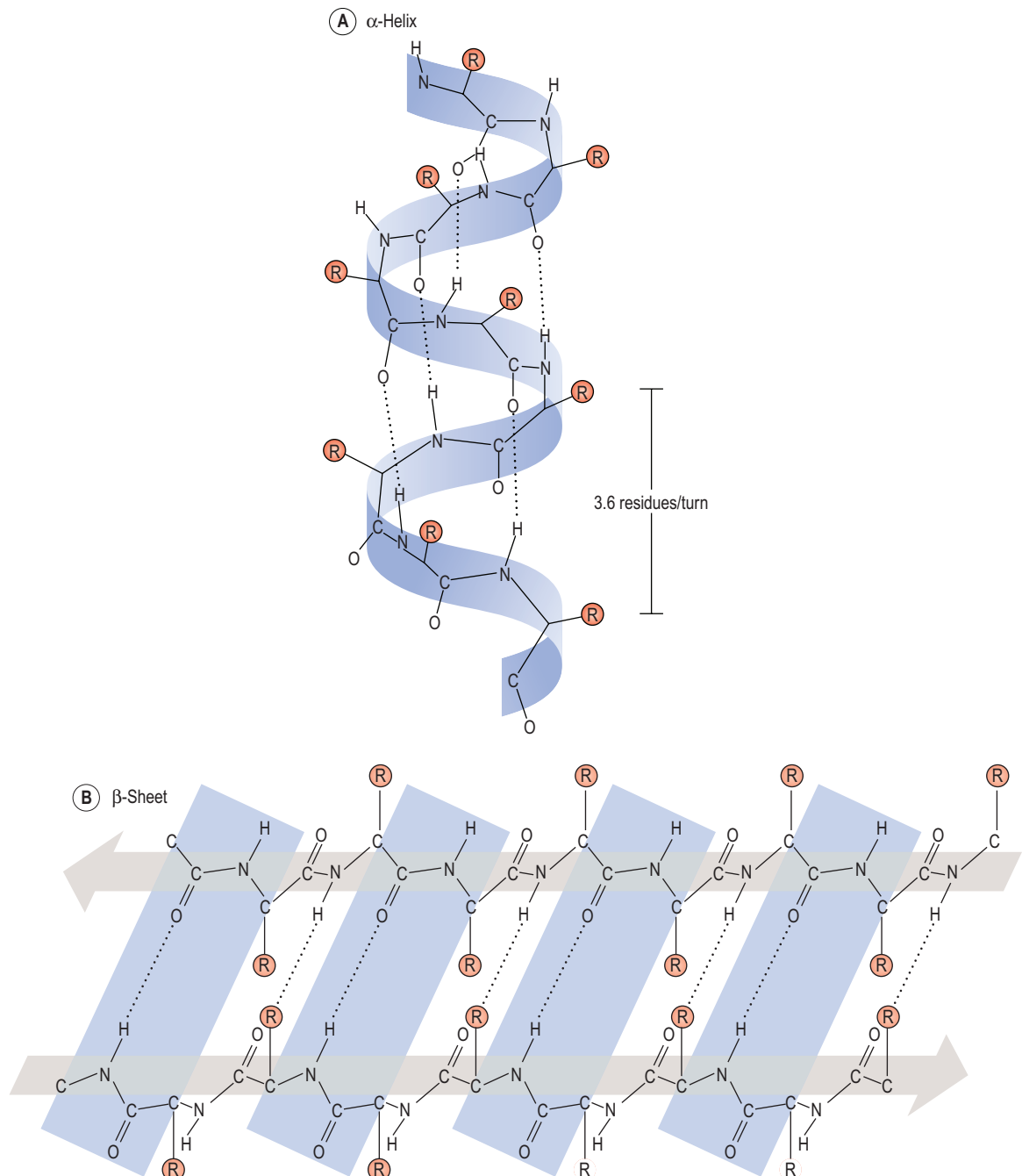


Fig. 2.5 Secondary structure of proteins: (A) α -helix and (B) β -sheet.

Tertiary structure

The next level of three-dimensional organisation of the polypeptide chain, the tertiary structure, results from a variety of interactions between the amino acid side chains. These interactions may include all the types of bonding already described. Tertiary structure specifies exactly how the entire protein is folded.

The type of interactions include:

- **Disulphide bonds**, between cysteine residues on the polypeptide chains
- **Hydrogen bonding** between side chains that have partial positive (δ^+) and negative (δ^-) charges, e.g. hydrogen bonding between serine and asparagine side chains
- **Ionic bonds** between oppositely charged side chains, e.g. between lysine and glutamate, sometimes called salt bridges
- **Hydrophobic interactions** in regions of the protein where there are large numbers of non-polar, aliphatic and aromatic side chains, which are excluded from the aqueous surroundings.

Denaturation

The tertiary structure of many proteins can be disrupted by heating, treatment with high concentrations of substances such as urea, or reducing agents which can break disulphide bonds. This is called **denaturation** and it leads to a loss of both tertiary and secondary structure, abolishing the protein's biological activity (which itself is often used as an assay of the degree of denaturation). The fact that some proteins can recover their activity when the denaturing agent is removed shows that the tertiary structure of these proteins is entirely determined by the primary structure. Examples of such proteins are lysozyme (an enzyme found in tears and saliva) and RNase (an enzyme which breaks down RNA).

Molecular chaperones

Not all denatured proteins can recover their structure and activity. The recent discovery of proteins called **molecular chaperones** or **chaperonins**, which aid the folding of the protein chain, has shown that at least some proteins require other factors to achieve the correct tertiary structure. These chaperonins may also be involved in increasing the rate at which naturally folding proteins can fold. Although many proteins are able to fold correctly in vitro this is a slow process and in vivo the chaperones make this process much more efficient.

Information box 2.3 Hair is made up of many α -helices

Hair is made up largely of a protein called keratin. Single hairs are made up of hundreds of microfibrils embedded in a protein matrix. Each microfibril is formed from a number of α -helices that are wound around each other in a superhelix. Hairs can be stretched by elongating the α -helices, which breaks the hydrogen bonds in the helices. When the hair is released it reverts to its previous length because the covalent disulphide bonds between cysteine residues in the α -helix and the protein matrix remain intact. Hair can be permanently curled or straightened by applying chemicals which first break these disulphide bonds and then re-form them in the new conformation.

Quaternary structure

Although some proteins consist of a single chain, there are many **multimeric proteins**. These consist of multiple proteins held together by non-covalent bonds. These proteins have an additional level of **quaternary** structure which specifies how many of each type of subunit is included (stoichiometry) and how they are arranged.

An example of this is the most common type of antibody molecule, immunoglobulin G (IgG) (see Ch. 6). This type of immunoglobulin is composed of four protein chains, called heavy and light chains. The Y-shaped structure consists of two heavy and two light chains, joined by disulphide bonds and hydrogen bonds. Breaking the disulphide bonds between the chains destroys the quaternary structure, and separately the proteins chains have no biological activity.

Protein modification

Almost all proteins are modified in some way after they have been synthesised. Some of these modifications are permanent while others are reversible. Some involve chemical modification of either the terminal carboxyl or amino groups and/or modification of the side chains. Other changes involve removal of regions of the protein after it has been assembled.

Acetylation

The most common modification involves the addition of an acetyl (CH_3CO) group to the N-terminal. The effect of this covalent change is to reduce the rate at which these proteins are degraded, thus increasing their lifespan.

Glycosylation

Another common modification is the addition of carbohydrate molecules to the surface of the protein, a process called glycosylation. These **glycoproteins** and **proteoglycans** have many diverse roles including cell-to-cell recognition and as components of the extracellular matrix.

Glycoproteins

Only serine, threonine or asparagine side chains can be directly glycosylated. Serine and threonine are linked to carbohydrates via their hydroxyl groups (*O*-glycosylation) and asparagine via the amide group (*N*-glycosylation) of the asparagine side chain. Another amino acid, lysine, can have carbohydrates added to it after it has been modified to hydroxylysine.

The carbohydrates found on glycoproteins are often complex, branched molecules. Many of these glycoproteins are components of the plasma membrane and have their carbohydrate residues facing outwards and are thought to be involved in cell-to-cell recognition processes.

Proteoglycans

The other types of protein which have carbohydrates linked to them are called proteoglycans. They are protein chains where the core protein is linked to large, linear carbohydrates with a repeating unit.

Lipid modifications of proteins

Some proteins have fatty acids attached, which, through their hydrophobic interactions can then be inserted into the lipid bilayer of the cell membrane. These proteins are thus concentrated at the inner face of the cell membrane

where they can interact more easily with other membrane proteins.

Phosphorylation

The activity of many proteins can be reversibly modified by **phosphorylation**. Enzymes called **kinases** add phosphate groups to the hydroxyl groups of serine, threonine and tyrosine residues found on many proteins. This change can have the effect of increasing or decreasing the activity of the protein. Other enzymes, **phosphatases**, can remove these phosphate groups. These changes in structure can act as a molecular switch that regulates the protein's activity.

Protein cleavage

Many proteins are formed from a precursor, which is then modified to make the mature protein. An example of this is the hormone, insulin, which is originally formed from a single peptide chain with 84 amino acids (pro-insulin). After the protein has been folded and disulphide bonds formed between two parts of the chain, part of the central part of the chain, called the C-peptide, is removed, leaving the two ends of the polypeptide, called the A and B chains, connected by the disulphide linkages. If the protein is denatured, for example by using mercaptoethanol which disrupts disulphide bonds, the A and B chains are separated. When the mercaptoethanol is removed the two chains do not re-associate, presumably because the most stable state for the separate chains is different from that of the original single, pro-insulin chain.

Enzymes

Almost all the biochemical reactions that take place in the body require the presence of protein catalysts called **enzymes**. Possibly as many as half of all proteins are enzymes, each with their own unique structure. They are globular proteins with irregular structures. They may contain short stretches of α -helices and β -sheets, but their overall forms vary enormously. Some enzymes fundamental to general metabolism are found in all cells, whereas others may only be found in specialised tissues. Enzymes act in the same way as inorganic catalysts in that they lower the activation energy of the reaction so that it can proceed under physiological conditions, i.e. at temperatures and pressures found in the human body. They can speed up the rate of reaction between 10^6 - and 10^{12} -fold. Enzyme activity may be increased in some diseases and reduced, or absent, in others. Enzyme analysis can be an aid to the diagnosis of many conditions.

The general mechanism of enzymic action is as follows:

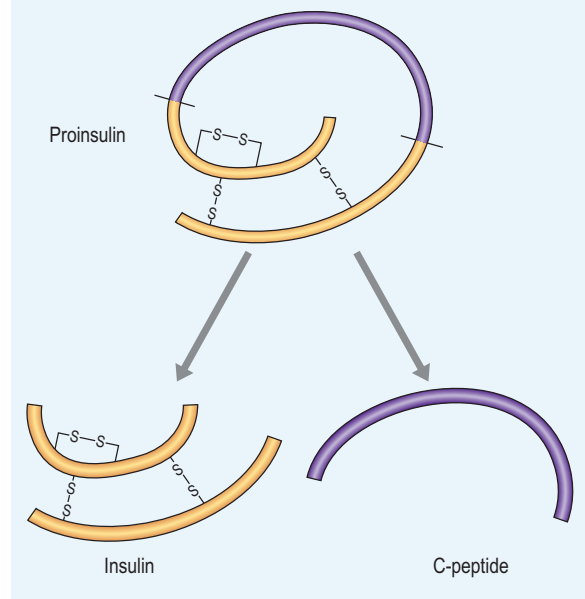
1. The reactants called **substrates** bind to the enzyme to form an enzyme–substrate complex.
2. The binding of the substrate causes changes in shape to occur within the enzyme–substrate complex. These internal rearrangements produce changes to bonds that favour the formation of the **product**.
3. The product is released from the enzyme, which reverts to its former shape. It can then go on to bind another substrate molecule.

Specificity

The substrates bind to the enzyme at the active site by a combination of ionic and hydrophobic interactions. The

Clinical box 2.1 Insulin secretion can be assessed by measuring C-peptide

In patients with diabetes who need to inject insulin, the presence of the injected insulin makes it difficult to measure endogenous insulin release, because a direct insulin assay will not be able to differentiate between the two sources of insulin. However, for every endogenous insulin molecule released into the blood, a C-peptide is also released and measuring the latter gives an estimate of endogenous insulin production. Certain conditions, however, such as a pancreatic β -cell tumour, can release large amounts of pro-insulin that will cause large errors in insulin measurement.



active site of each enzyme has a specific size and structure that allows the substrate to fit into it exactly, a characteristic that is often called a **lock and key mechanism**.

Some enzymes are very specific in their substrates, only catalysing the reaction of a single substrate, whereas others may accept a range of related substrates. Protease enzymes break down proteins by breaking peptide bonds, and different proteases act on different proteins. For example, trypsin is a proteolytic enzyme that will only hydrolyse those peptide bonds to which arginine or lysine contribute the carboxyl group. Pepsin, another gastric protease, is less specific in its substrates and will hydrolyse a wider range of peptide bonds.

Enzymes are very specific in the reaction they catalyse. The same substrate can be converted to a number of different products by the action of different reaction-specific enzymes. For example, pyruvate can be converted to lactate, alanine, oxaloacetate or acetyl CoA by different enzymes.

Enzyme nomenclature

Enzymes have specific names, depending on the reaction catalysed, the substrates and the products. They are classified into six classes, depending on the type of reaction catalysed (Table 2.7). However, many enzymes have a less formal 'common' name, which, for general use, describes their action and often their substrate.

Table 2.7 Classes of enzymes

Class	Type	Action	Examples of common names
I	Oxidoreductases	Oxidation–reduction reactions, often with coenzymes, such as NAD ⁺	Dehydrogenase, oxidase, peroxidase, reductase
II	Transferases	Transfer of amino, carboxyl, acyl, carbonyl, methyl, phosphate groups from one molecule to another	Transaminase, transcarboxylase
III	Hydrolases	Cleave bonds between carbon and another atom by inserting water	Esterase, peptidase, amylase, phosphatase, pepsin, trypsin
IV	Lyases	Break carbon–carbon, carbon–sulphur and carbon–nitrogen (but not peptide) bonds	Decarboxylase, aldolase
V	Isomerases	Racemisation of optical or geometric isomers	Epimerase, mutase
VI	Ligases	Formation of bonds between carbon and oxygen, sulphur, nitrogen, etc, often hydrolysing ATP (adenosine triphosphate) to provide the required energy	Synthetase, carboxylase

Factors affecting enzyme activity

Enzyme activity (the rate of the enzyme-catalysed reaction) is affected by a number of factors:

- Temperature
- pH
- Substrate concentration.

Temperature

Although normal body temperature is 37°C, a large number of enzymes increase their activity as the temperature is increased. This is because of the increased kinetic energy of the molecules and the increased likelihood the electrons will be excited. If the temperature falls the rate of reaction falls and sometimes the reaction will cease entirely. However, if the temperature returns to normal, the enzyme activity is regained.

If the temperature is increased above a certain level, the protein ‘melting’ point, the thermal excitation will be great enough to break essential weak bonding forces and the enzyme will become permanently disabled. At very high temperatures proteins become insoluble, as shown by the precipitation of egg white proteins during cooking.

Even though they catalyse the same reaction, some enzymes isolated from different tissues have different temperature resistance. For example, the differing denaturation temperatures of lactate dehydrogenase (LDH) have been used to distinguish heart and liver disease in assays where LDH of cardiac origin appears more heat stable than that from the liver. The difference is due to LDH being present in different forms.

Optimal pH

All enzymes have an **optimal pH** at which they function at their fastest. Below or above this the enzyme’s activity is reduced. For many enzymes the optimal pH is close to pH 7. However, there are some enzymes which have their optimal pH at very different acidities, reflecting the environment in which they work. For example, the enzyme pepsin, which digests proteins in the stomach has an optimal pH of about 2 which is the pH of the gastric juice in the stomach (see Ch. 15). Control of pH can also be a way of regulating enzyme activity. Enzymes which are inactive under normal conditions can become active when specific conditions change the pH.

Substrate concentration

The other major factor which determines the rate of reaction of an enzyme is the concentration of the substrate which is written as [S], where the square brackets represent the concentration. As the **substrate concentration** [S] is increased, the rate of the reaction increases: at first the increase is linear, then at higher [S] the rate of increase slows until the rate reaches a plateau, where any further increase in [S] does not produce any increase in the **reaction rate**.

This occurs because at low [S] the rate of reaction is limited by the limited number of substrate molecules which are bound to the active site. The initial increase in the rate is due to the increase in the binding of the substrate to the enzyme as the amount of substrate increases. As [S] increases further, more and more of the substrate binding sites are occupied until, at high [S] the enzyme is **saturated** with substrate, i.e. all the binding sites are occupied by substrate. At this point, even though there are more substrate molecules available, there are no binding sites available. The rate of the reaction is now limited by the rate at which the enzyme can catalyse the reaction and the rate at which the products leave the active site, making it available to more substrate. The maximum rate of the reaction is called the V_{\max} . V_{\max} is dependent on the quantity of enzyme present and the conditions such as temperature and pH, which affect the rate of catalysis.

Enzyme kinetics

If the initial rate of the reaction (v) at different [S] is plotted, the resulting curve is a **rectangular hyperbola**, described by the Michaelis–Menten equation. This is derived from the rate equation:



where E is the enzyme, S is the substrate and P is the product.

$$v = (V_{\max} \times [S]) / (K_m + [S])$$

V_{\max} , the maximum reaction rate, can be estimated from the plateau of the curve in Figure 2.6 where the rate of reaction is plotted against concentration. The other constant of the Michaelis–Menten equation is K_m . This is defined as the substrate concentration at which the rate is half V_{\max} . The constant K_m is informative, as it describes the ease with

which the substrate binds to the enzyme, or its **affinity**. K_m is inversely proportional to the affinity, with a low value of K_m indicating a high affinity. In other words, only a low concentration of substrate is required to saturate the enzyme and vice versa. The two constants, V_{max} and K_m , are characteristic for the enzyme–substrate combination under the conditions under which they are measured.

Curve fitting and linear transformations

Without using statistical curve-fitting programs it is very difficult to get accurate estimates of V_{max} and K_m from the Michaelis–Menten curve, especially as it is often difficult to see where the plateau stops rising, which occurs at very high levels of substrate. This leads to inaccurate estimates of V_{max} and because estimates of K_m use this inaccurate value, they themselves are inaccurate. This can be solved by using one of two possible rearrangements of the Michaelis–Menten equation that produce straight lines when plotted, where the kinetic constants can be easily identified from the slopes and the intercepts. These plots are called **Lineweaver–Burk** and **Eadie–Hofstee** plots (Fig. 2.7 and Table 2.8). Both of these linear transformations have some drawbacks because the use of reciprocals tends to underestimate errors in the data at different parts of the range. However, in the absence of curve-fitting programs, they are much better than the standard plot for estimating the kinetic constants.

The same mathematical models can be used to describe:

- **Membrane transport** (where J_{max} is the maximum rate of transport and K_m the affinity of the transport system for the substrate Carrier mediated transport)
- **Receptor binding** (where the constants are called B_{max} and K_d , see Ch. 3).

They can also be used to model drug interactions with a variety of targets.

The Michaelis–Menten equation applies only to a simple reaction with a single substrate. Reactions that involve mul-

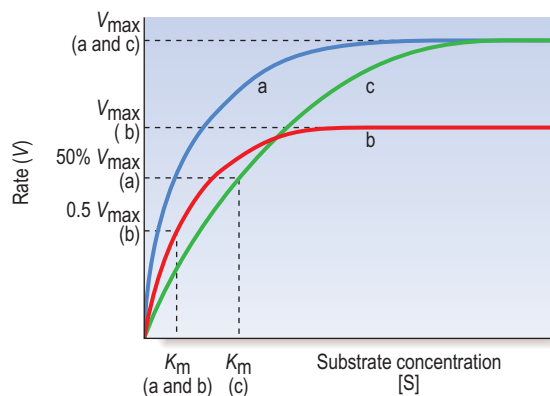


Fig. 2.6 Michaelis–Menten curves. Curves a and c have the same V_{max} , whereas curves a and b have the same K_m .

iple substrates can have different shaped curves, depending on the mechanism by which the substrates bind. In particular, reactions arising from cooperative interactions between substrate molecules, such as the binding of oxygen to haemoglobin, have sigmoidal curves,

Enzyme inhibition

Many enzymes involved in metabolic pathways are inhibited by the end-products or intermediates of those pathways. This is an important mechanism for controlling the production of metabolites in appropriate amounts. Many drugs used therapeutically act by inhibiting enzymes in order to reduce the amount of a particular reaction product. Inhibition can be reversible or irreversible.

Reversible inhibition

There are two types of reversible inhibition:

- Competitive inhibition
- Non-competitive inhibition.

They can be distinguished easily by looking at the effect of the inhibitor on the kinetic parameters of the reaction.

Competitive inhibition

Competitive inhibition occurs because the inhibitor binds reversibly to the same site on the enzyme as the substrate.

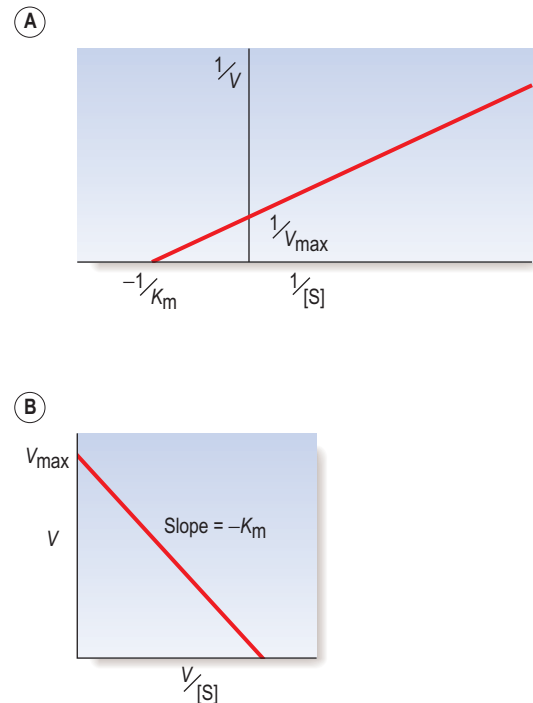


Fig. 2.7 (A) Lineweaver–Burk plot: x-intercept = $-1/K_m$, y-intercept = $1/V_{max}$. (B) Eadie–Hofstee plot: y-intercept = V_{max} , slope = $-K_m$.

Table 2.8 Kinetic constants from different plots

Type of plot	X-axis	Y-axis	V_{max}	K_m
Michaelis–Menten	[S]	v	Plateau of curve	[S] at $\frac{1}{2} V_{max}$
Lineweaver–Burk	$1/[S]$	$1/v$	$1/y$ -intercept	$-1/x$ -intercept
Eadie–Hofstee	$v/[S]$	v	Y-intercept	–slope

Clinical box 2.2 Competitive inhibition of alcohol dehydrogenase can be used to treat poisoning by methanol and ethylene glycol

Methanol is poisonous to humans and can cause blindness. It can be formed accidentally in the production of homemade spirits, or can be ingested intentionally as 'meths' by alcoholics. The metabolism of both methanol and ethanol (and the antifreeze, ethylene glycol) involves the enzyme, alcohol dehydrogenase. Part of the treatment of poisoning by both methanol and ethylene glycol is the administration of ethanol in order to block their breakdown, as it is the breakdown products of methanol and ethylene glycol that are toxic. Ethanol has a higher affinity (lower K_m) for the active site of alcohol dehydrogenase than either methanol or ethylene glycol and so blocks their binding to the enzyme, and therefore their subsequent breakdown. They can then be removed from the blood by dialysis.

At low concentrations of substrate the enzyme activity is reduced, but as $[S]$ rises the inhibitor is displaced and at very high $[S]$ the enzyme activity returns to normal. This can be seen by an increase in the K_m of the enzyme; more of the original substrate is now required to saturate the enzyme than before, with V_{max} unchanged. Competitive inhibition often occurs between alternative substrates for the same enzyme. The substrate with the lowest K_m will be preferentially metabolised.

Non-competitive inhibition

Non-competitive inhibition occurs when the inhibitor binds to a site on the enzyme other than the active site and by doing so reduces the rate at which products are formed. The binding of the substrate is unaltered so the K_m remains the same. However, at all $[S]$ the rate of reaction is reduced, so V_{max} is reduced. Several chemotherapeutic drugs act in this way.

Irreversible inhibition

Irreversible inhibition occurs when the inhibitor is so tightly bound to the enzyme that it cannot be removed and the only way that the effect can be reversed is by the removal of the enzyme and its replacement by newly synthesised enzyme. For example, penicillin covalently inactivates a key enzyme in bacterial cell wall synthesis.

Regulation of enzyme activity

Enzyme activity is regulated in several ways:

- **Feedback inhibition** by the product of the pathway. Many enzymes are inhibited by the products of their metabolic pathways. The product of the pathway can bind to a site on the enzyme and reduce the rate of reaction.
- **Allosteric regulation:** this is regulation by other small molecules which can either activate or inhibit the enzyme. The binding of the allosteric regulator to a site other than the active site causes a change in the tertiary structure of the protein, which alters the enzyme activity.
- **Gene expression:** the amount of an enzyme is altered by increased production of the enzyme in response to a metabolic signal.

- **Phosphorylation:** as described earlier, the activity of many proteins, including enzymes, can be reversibly altered by the addition (with kinases) or removal (with phosphatases) of phosphate groups.
- **Proteolysis:** enzymes can be irreversibly activated or inactivated by being broken down by other, proteolytic, enzymes. The lifespan of enzymes can be changed by altering their rate of proteolytic degradation in the cell.

Cooperativity between multiple subunits

In multimeric proteins **cooperativity** can occur between multiple subunits. This happens when the binding of a molecule, which can be a substrate or a regulator, to one subunit induces a change in the quaternary structure of the protein, which changes the affinity of the substrate-binding sites. This means that small changes in the concentration of the molecule cause a large change in the rate of catalysis.

For example, the oxygen-carrying protein, haemoglobin, can bind four oxygen molecules. The binding of a single oxygen molecule to one of the substrate-binding sites of deoxyhaemoglobin (the deoxygenated form of haemoglobin) causes a conformational change in the other subunits that increases their affinity for oxygen so that the second oxygen molecule binds more easily. This, in turn, increases the affinity of the other binding sites and so on, until all four sites are filled. The reverse also occurs in that if one oxygen molecule is unbound then the other molecules are lost more easily. This cooperativity enables haemoglobin to become fully oxygenated in the lungs and to lose this oxygen readily in the tissues (see Ch. 13).

Isoenzymes

Many enzymes exist in different tissues in slightly different forms, called **isoenzymes**. They may differ in their primary structure or their subunit composition. Isoenzymes catalyse the same reaction but may have different properties. For example, they may have different affinities for their substrates. They may also show different patterns of inhibition that help to regulate their function in different tissues.

Coenzymes

Many enzymes require the presence of an additional component in order to function. These are called **coenzymes**,

Clinical box 2.3 Lactate dehydrogenase isoenzymes can be used to diagnose heart disease

Five different types of LDH are found in different human tissues. Each type consists of a combination of four subunits of two possible types called A and B. Type 1 consists of four B subunits, type 5 of four A subunits with types 2–4 being combinations of A and B subunits. The different subunits determine how the enzyme is inhibited by its substrate, pyruvic acid, and because of this, the degree to which the particular tissue can undergo anaerobic respiration (i.e. without oxygen). Heart muscle contains virtually only type 1, whereas skeletal muscle and liver have mainly type 5. When someone has a myocardial infarction (MI) and their heart muscle is starved of oxygen, then muscle cells in the heart die and release their contents into the blood. Analysis of the LDH isoenzymes in the blood can confirm an MI as there would be more LDH-1 in the blood than normal. In contrast, in someone with liver disease, you would tend to see increased levels of LDH-5 in the plasma.

or **cofactors**. They may consist of metal ions such as copper or iron, or they may be more complex organic molecules. Some enzyme cofactors are derived from vitamins. **Vitamins** are organic molecules that are required in small amounts in the diet and without which deficiency diseases, such as scurvy (vitamin C deficiency) or beriberi (vitamin B₁ deficiency), can occur. Coenzymes are often involved in oxidation–reduction reactions, where they act as either electron donors or acceptors. For example, the coenzymes, flavine adenine dinucleotide (FAD) and nicotinamide adenine dinucleotide (NAD), which are derived from vitamin B₂ (riboflavin) and niacin, respectively, have a central role on the transduction of energy in the electron transport chain of mitochondria.

Structural proteins

Many of the proteins that form the structural elements of the body are fibrous proteins made up of long repeating units, arranged in bundles, which give strength and stability. These make up the muscles, bones, skin and connective tissues. They often have a simple repeated secondary structure of α -helices and β -sheets, but because they are often multimeric they may also have a quaternary structure.

The three major groups of structural proteins are:

- Collagens
- Muscle proteins
- Cytoskeletal proteins.

Collagens

The most abundant types of fibrous protein, which make up about 25% of the total protein of the body, are the **collagens**. They form the major component of the **extracellular matrix** (ECM) which is the material that surrounds the cells. There are a large number of different types of collagen (at least 20 have been described to date); they vary according to their protein chains but they all have a similar quaternary structure.

Collagen is made up of three polypeptide chains each of which can be up to 3000 amino acids in length. Each chain forms a left-handed helix that is tighter than the standard α -helix, with just three amino acids per turn. This tight helix can be achieved because every third amino acid is a glycine, which, with its single hydrogen, can fit into the core of the molecule. The other amino acids are very often proline and its derivative, hydroxyproline. The three helices are then wound round each other into a right-handed triple superhelix, which is stabilised by hydrogen bonding. The commonest form of collagen (type I) can be found in a variety of tissues and forms fibrils, with its molecules packed side by side that confer great strength. Other types of collagen, such as type IV, form sheets in which the molecules produce large flexible networks. The basement membrane underlying epithelial cells is partly made up of type IV collagen.

Muscle proteins

The main proteins that make up the contractile elements of muscle cells are **myosin** and **actin**. They form the thick and thin filaments of skeletal muscle, which interact to produce muscle movement (see Ch. 9, Fig. 9.16). The two types of filaments are formed in very different ways. Myosin is made up of six polypeptide chains, two heavy chains and four light



Information box 2.4

Mutations of type I collagen cause brittle bone disease (osteogenesis imperfecta)

There are many different forms of brittle bone disease, all of which are characterised by weakness in all tissues containing type I collagen. In each form there is a genetic mutation that replaces the glycine in collagen with other amino acids. These replacement amino acids are more bulky and prevent the formation of stable triple helices and collagen fibrils.

Table 2.9 Cytoskeletal proteins

Filament types	Main protein constituent	Major role
Microfilaments	Actin	Cell movement Form core of microvilli
Intermediate filaments (IFs)	Keratins*	Mechanical strength
Microtubules	Tubulins	Chromosome separation Intracellular transport

*The proteins making up IFs vary between cell types. They are made up of keratins in many cells, but there are six different classes of IFs. For example, nerve cells have IFs made from neurofilament proteins.

chains. The two heavy chains form α -helices, which then coil around each other. At their N-terminals they form two globular head regions, with the four light chains. These globular heads contain binding sites which interact with the actin and adenosine triphosphate (ATP) to power muscle movement. The strands of myosin aggregate into thick filaments made up of 300–400 myosin molecules.

The other filaments are composed of actin. Actin filaments are made by the polymerisation of molecules of globular actin (G-actin) into long fibrous strands of F-actin. These strands can be continuously extended and shortened by the addition and removal of G-actin.

Cytoskeletal proteins

All cells contain cytoskeletal proteins that make up three different types of filament, each made up of fibrous proteins (Table 2.9). Each has a different role.

Other structural proteins

Elastin is a flexible protein, associated with collagen in tissues which need to be elastic. It is found mostly in ligaments and blood vessel walls, and is also found in skin, tendons and connective tissues in much smaller amounts. Many globular elastin molecules are cross-linked covalently between lysine residues. When the tissue is stretched the globular proteins can extend, but when the tension is released they shorten back to their globular form. A mixture of collagen and elastin allows a certain degree of stretch, the degree of which is controlled by the amount of collagen.

Two other groups of glycoproteins, important in forming the basement membrane, the layer of material which lies under sheets of cells, are laminin and fibronectin. These both form part of the basal lamina, the part of the basement membrane closest to the cell surface membrane. These molecules can bind to proteins on the cell surface, as well as to collagen and other ECM proteins.

Signalling proteins and receptors

Hormones and neurotransmitters

Some, but not all, of the signalling molecules which travel from cell to cell, are proteins. These can be either **hormones**, which are released into the bloodstream or **neurotransmitters** which are released from nerve endings.

An example of a protein hormone is insulin, which regulates blood glucose levels. Insulin is secreted from cells in the pancreas into the bloodstream when blood glucose levels are high and acts on other tissues to promote the uptake and storage of the glucose. Hormonal responses are relatively slow because hormones are often released at an area distant from the site of action.

In contrast, when a rapid signal is needed, nervous impulses can produce a rapid action through the release of neurotransmitters (see Ch. 8) at the nerve terminals very close to the site of action. In fact, many nerves secrete more than one neurotransmitter. In response to low levels of stimulation, they secrete one of a number of small molecules, such as acetylcholine or noradrenaline (norepinephrine). However, in response to higher levels of stimulation, many nerves also secrete a peptide neurotransmitter. An example of a peptide neurotransmitter is the natural painkiller β -endorphin. This acts on opioid receptors, so called because they are also activated by the opiate drugs, such as morphine.

Proteins are also involved in the signalling pathways inside cells. For example, **transcription factors** are small proteins which, when activated, bind to the genetic material, DNA, to alter the type and amount of protein being produced in the cell.

Receptors

Signalling molecules produce an effect on target cells by interacting with proteins called **receptors**. The ability of a cell to respond to any chemical signal (or **ligand**) depends on the presence of receptors for that molecule. If the receptor is not present then the cell cannot respond.

Cell surface receptors

For many ligands their receptors are embedded in the cell membrane, with the binding site on the cell surface. Binding of the ligand causes a conformational change in the receptor that produces some effect inside the cell. This is called **signal transduction**.

Intracellular receptors

Some signalling molecules, such as the sex hormones, testosterone and oestrogen, can enter the cell and act on receptors present inside the cell. Some of these are present in the cytoplasm, and others are located in the nucleus. The intracytoplasmic receptors have two binding sites, one for the hormone and one for DNA. When the hormone is bound, the receptor moves to the nucleus, binding with acidic proteins of the DNA. The ligand-bound receptors, bound to the DNA, produce changes in protein production, thus effecting cell growth and metabolism.

Immune system proteins

Immunoglobulins

The body defends itself from attack by foreign bodies such as bacteria and viruses with its immune system (see Ch. 6).

Clinical box 2.4

The absence of testosterone receptors produces babies that look like girls who should have been boys

Androgen insensitivity syndrome is a condition in which the male fetus lacks the receptors for the male sex hormone, testosterone. Genetically they are XY males, but a mutation causes them to lack testosterone receptors. Despite producing testosterone, these individuals cannot respond to it and do not develop male genitalia because proteins necessary for the appropriate cell growth and metabolism are not transcribed. They do, however, produce normal amounts of other hormones, which means that they develop external female features such as breasts. Unfortunately, they do not develop internal female organs, so cannot function as normal females.

One of the key elements of the immune system is the production of vast numbers of different proteins called **immunoglobulins** or **antibodies** which can bind to the invading organism or antigen, enabling the latter to be removed and destroyed.

Antibodies have a common basic structure but include a variable region which produces the binding site. These proteins are produced with an enormous variety of variable regions to cope with the large number of possible exogenous (i.e. of foreign origin) antigens. Problems can occur, however, when the body produces antibodies that recognise its own, endogenous, molecules as antigens. This failure to recognise self results in autoimmune conditions, where the body's own cells are destroyed. A well-known example of this problem is rheumatoid arthritis (RA).

Cytokines

Another group of proteins involved in both the immune response and in the body's response to inflammation are the **cytokines**. More cytokines continue to be reported as new research techniques are developed. Cytokines mainly act locally, but some have been identified that have wide-ranging effects on metabolism. One of the better-known groups of cytokines are the **interferons**, which have actions against viruses and cancer cells. One of them, interferon beta, is currently being tested as a possible treatment for some forms of multiple sclerosis.

Transport proteins

Many substances have to be transported around the body in the extracellular fluids, and then subsequently into and out of cells.

Transport in blood

Many proteins travel around the body in the fluid phase of the blood: the plasma. **Albumin** is the major protein found in blood plasma at a concentration of 35–45 g/L, making up 50% of the total plasma protein. It has two major functions:

- To increase the osmotic pressure of the blood, thereby ensuring the correct distribution of fluid between the vascular and interstitial compartments (see Ch. 11)
- To act as a non-specific transport protein for many substances in the plasma.

Table 2.10 Some of the binding proteins in plasma

Protein	Ligand/s
Albumin	Metal cations, free fatty acids, steroids, bilirubin, haem
Transferrin	Iron
Thyroid-binding globulin	Thyroxine (T_4), triiodothyronine (T_3)
Cortisol-binding globulin	Cortisol
Sex-hormone-binding globulin	Androgens and oestrogens

Albumin has an overall negative charge at physiological pH, which means that it can easily bind metal cations such as copper and iron. Many molecules, some of which are extremely hydrophobic, can be transported in the aqueous environment of the plasma, bound to albumin. This increases the length of time they remain in the blood, due to reduced degradation and elimination (see Ch. 2).

Other plasma proteins are specific binding proteins for signalling molecules, such as sex steroids (Table 2.10).

Transport into cells

Hydrophobic substances can move into cells by simple diffusion but many substances, including glucose, cannot move into cells without the presence of specific proteins called **transporters**. These transporters are proteins embedded in the cell membranes that bind the transported molecule on one side of the membrane and move it through to the other side. Inside cells, substances can be moved into separate compartments by the action of other specific transport proteins.

CARBOHYDRATES

Carbohydrates form about 2% of body mass. They consist of carbon, oxygen and hydrogen, with a ratio of hydrogen to oxygen of about 2:1, hence the name 'hydrated carbon'. The most common organic compound on the planet is a carbohydrate, **cellulose**, which forms the principal component of plant cell walls. Unfortunately, humans cannot digest significant amounts of cellulose, although it is important to digestion as a bulk element of food. Of more importance is the carbohydrate, **glucose**, which is one of the central molecules in energy metabolism and forms the backbone for the synthesis of many other compounds. The simplest forms of carbohydrates are called **sugars**. These are either **monosaccharides**, which are single units or **disaccharides**, formed from two monosaccharides. Larger carbohydrates, called **polysaccharides** are polymers, formed from many monosaccharides.

Sugars

Monosaccharides

The basic units of all carbohydrates are called **monosaccharides**, or simple sugars. These can be obtained directly from the diet or derived from the breakdown in the gut of more complex carbohydrates. They can also be manufactured from non-carbohydrate sources when carbohydrate sources have been used up. Although food is the main source of sugars there is no dietary requirement for carbohydrates as such.

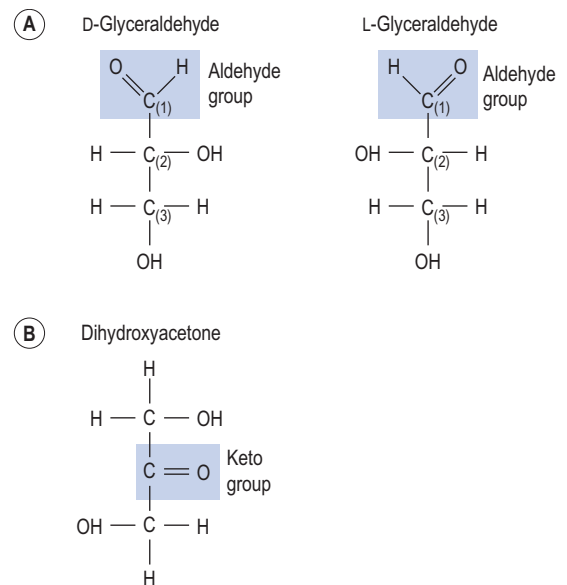


Fig. 2.8 Triose sugars: (A) aldoses and (B) ketoses.

Triose sugars

Sugars have the general formula $(CH_2O)_n$ where n is the number of carbons in the sugar. The simplest of these are the three-carbon trioses. Two types of **triose sugars** contain either an aldehyde group, aldoses, or a ketone group, ketoses (Fig. 2.8). In **aldose sugars**, the carbon atom of the terminal aldehyde group is numbered carbon-1. In **ketose sugars**, carbon-1 is the end carbon next to the carbonyl group. The central carbon of glyceraldehyde, the simplest of the aldose sugars, is chiral, and therefore it can exist as two stereoisomers, D- and L-glyceraldehyde.

D and L sugars

All sugars larger than trioses have two or more chiral groups, and so exist as a number of stereoisomers. By convention they are divided into D- and L-series, depending on the arrangement of atoms around the asymmetrical carbon furthest from the carbon-1 end. The sugar is a D-sugar if the arrangement around this carbon is the same as D-glyceraldehyde, and an L-sugar if it is like L-glyceraldehyde. Virtually all sugars found in biological systems are **D-isomers**.

Different forms of glucose

D-Glucose, also known as **dextrose**, is found in all cells and in blood plasma, where its concentration is strictly controlled. Glucose is a six-carbon sugar, or **hexose**, and has the formula, $C_6H_{12}O_6$. It can exist in four different forms, all of which are freely convertible in aqueous solution (Fig. 2.9). The straight chain form (which exists as less than 1%) has an aldehyde group at one end (where the carbon is always called carbon-1). This reacts with the hydroxyl group on carbon-5 to give a six-member ring form called a **pyranose ring**. Depending on the arrangement of the $-H$ and $-OH$ on carbon-1, the glucose molecule forms either α - or β -glucose (also known as α -D-glucopyranose and β -D-glucopyranose). These different forms are known as **anomers**. While the cyclic form of D-glucose is usually a pyranose ring, rarely it can also form a five-member **furanose ring**, D-glucofuranose.

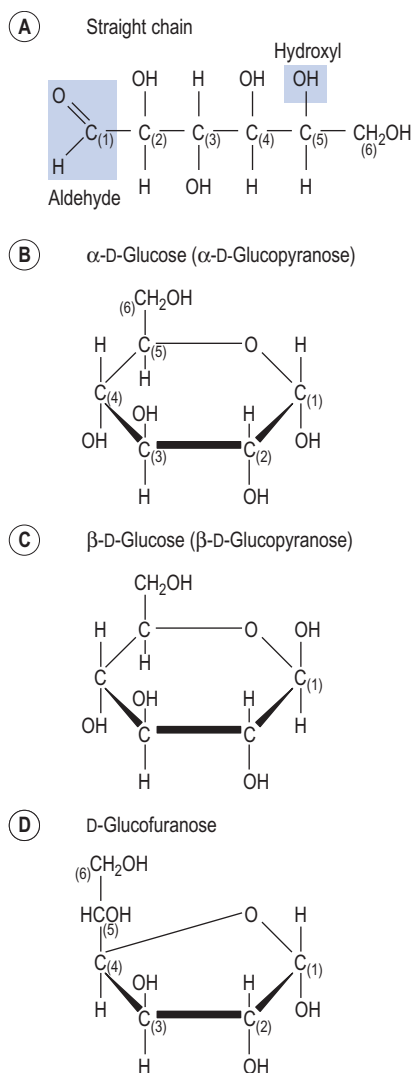


Fig. 2.9 Different forms of glucose. Each of these molecules has an identical chemical composition. A reaction between the hydroxyl group on C5 and the aldehyde on C1 produces the pyranose ring structure (B, C). If the hydroxyl group is on C4 then the furanose ring is produced (D).

Due to the tetrahedral arrangement of bonds around the carbon atoms, the pyranose and furanose rings are not planar molecules. They form two different arrangements, known as the **chair and boat conformations**. Both of these conformations are important for the packing of the molecules, although the chair conformation is the more stable.

Common hexose sugars

Other common hexoses are mannose, galactose and fructose (also known as fruit sugar) (Fig. 2.10). **Mannose** and **galactose** are both aldoses and tend to form pyranose rings. They are only different from glucose in the configuration of the other carbons, and are called **epimers**. Mannose is identical to glucose, except for the arrangement around the carbon-2. Similarly, galactose only differs from glucose by the arrangement around carbon-4. **Fructose** is a ketose sugar and forms both furanose and pyranose rings.

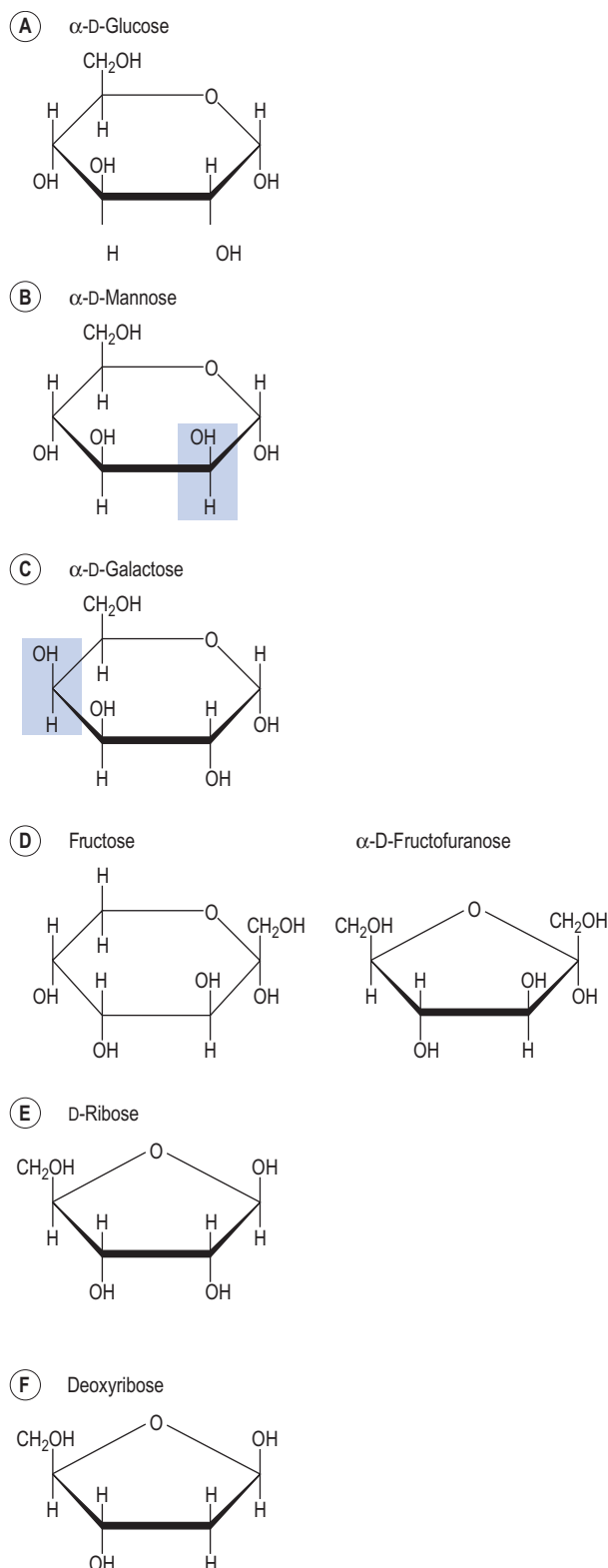


Fig. 2.10 Structures of some common sugars. In (B) and (C) the part of the molecule which differs from glucose is highlighted.

Pentose sugars

Two important **pentose** sugars are **ribose** and **deoxyribose**, which are found as components of the nucleic acids, DNA and RNA. These are both aldoses and form cyclic furanose rings.

Table 2.11 Principal disaccharides

Disaccharide	Carbon-1 sugar	Other sugar	Linkage	Digestive enzyme
Sucrose	α -Glucose	β -Fructose	1 \rightarrow 2	Sucrase-isomaltase
Lactose	β -Galactose	β -Glucose	1 \rightarrow 4	Lactase
Maltose	α -Glucose	β -Glucose	1 \rightarrow 4	Maltase
Isomaltose	α -Glucose	β -Glucose	1 \rightarrow 6	Sucrase-isomaltase

Other simple sugars

Other carbohydrates can be derived from basic carbohydrates by the inclusion of other elements. **Glucosamine** and **galactosamine** are derived from glucose and galactose, both having amino groups replacing the hydroxyl groups on carbon-2. These two sugars are major components of the proteoglycans, which make up the ECM.

Among other possible substitutions are:

- Phosphates – widespread in energy metabolism
- Sulphates and carboxylates: commonly found in proteoglycans.

Disaccharides

Monosaccharides can be linked together by **glycosidic linkages** to form disaccharides. These links are formed by a dehydration reaction between the hydrogen on carbon-1 atom of one sugar and the hydroxyl group of another, with the loss of a water molecule. This reaction is reversible. These are either α or β linkages, depending on the orientation around carbon-1 atom. This means that there are a large number of different possible bonds that could occur between two sugars, each of which gives a different product (Table 2.11). For example, glucose has five different $-\text{OH}$ groups (on carbons 1, 2, 3, 4 and 6), so could potentially form dimers in 10 different conformations. If two different sugars are combined the second sugar could be either an α -sugar or a β -sugar, increasing still further the possible combinations. However, enzymes are able to distinguish between the different conformations ensuring that the required linkage is either made or broken in order to produce the required product.

Dietary sugars

The compound known as ordinary sweet-tasting sugar is actually a **disaccharide**, **sucrose**, which is made up of α -glucose and β -fructose in an α (1 \rightarrow 2) linkage. Two other important dietary disaccharides are **maltose**, which is a dimer of two glucose molecules, and **lactose** (called milk sugar, because milk is its only significant source), which is made from galactose and glucose. Dietary disaccharides are hydrolysed by specific enzymes found in the intestine. The resulting monosaccharides are then absorbed by specific transport mechanisms (see Ch. 15).

Polysaccharides

Polysaccharides are polymers of monosaccharides and may contain thousands of units. They allow large amounts of glucose to be stored in a highly concentrated form, without the osmotic problems that would be associated with high aqueous concentrations of separate glucose molecules.

Starches

These are a group of polysaccharides which are used as sugar storage in plants. They consist of polymers of glucose

Clinical box 2.5 In certain populations adults cannot digest lactose

The enzyme, lactase, which breaks down lactose into galactose and glucose, is usually present only in children and only persists in adults of northern European descent. Lactase deficiency is common in Mediterranean countries, parts of Africa and Asia. In the absence of lactase, consuming milk products containing lactose produces watery diarrhoea, flatulence and abdominal pain. This is because the unabsorbed lactose is fermented by bacteria in the small intestine, producing gas and large numbers of metabolites that draw water into the intestine by osmosis, producing diarrhoea. This can be avoided to some extent by fermenting the milk, to produce yoghurt, for example, because the lactose is converted to lactic acid during fermentation.

linked in two different ways. The simplest starch, **amylose**, consists of long linear chains of glucose molecules, linked via α 1 \rightarrow 4 bonds. **Amylopectin**, which makes up about 80% of the starch in foods, also has 1 \rightarrow 4 linked glucose chains, but additionally every thirtieth glucose molecule is also linked via a 1 \rightarrow 6 bond, producing a branched molecule. The starches found in grains, such as wheat or rice and potatoes, are the major source of dietary carbohydrates. They can be broken down to glucose in the human gut by **amylases** secreted by the pancreas and the salivary glands. These enzymes break starches into glucose, maltose and isomaltose (from the 1 \rightarrow 6 linkages in amylopectin).

Non-starch polysaccharides

Cellulose is an unbranched glucose polymer with β 1 \rightarrow 4 linkages. Humans have no enzymes capable of hydrolysing these linkages, although a small amount of cellulose is broken down by bacteria in the colon. However, cellulose, along with other non-metabolised polysaccharides, are the major component of dietary fibre. They produce bulk in the intestine which aids defecation.

Glycogen

Glycogen is the storage form of glucose in animals, including humans. Like the starches, it is a large glucose polymer. It is similar in structure to amylopectin, except that the branch points occur every tenth glucose molecule. The high level of branching of the molecule means that it has a large number of free ends to allow glucose molecules to be added or removed rapidly. Glycogen stores are not very large in humans, with about 75 g in the liver and about 250 g in skeletal muscle. Each gram can produce 16.7 kJ (4 kcal), so glycogen stores can provide sufficient glucose for 12–18 hours in the absence of other sources of glucose. Dieting rapidly depletes glycogen, with an apparent rapid weight loss. However, this loss is mainly due to the loss of water molecules that are associated with glycogen.

Table 2.12 Major glycosaminoglycans (GAGs)

GAG	Main source
Hyaluronic acid	Joints
Chondroitin sulphate	Cartilage
Keratan sulphate	Cornea
Dermatan sulphate	Skin
Heparan sulphate	Basement membranes
Heparin	Mast cells

Functions of carbohydrates

Glucose, as mentioned above, is a major metabolic fuel, and it is the only fuel used by the brain under normal circumstances. Red blood cells are completely reliant on glucose for energy because they lack the ability to metabolise fats. Also, at the beginning of exercise, muscle preferentially uses glucose rather than other fuels. Glucose is oxidised by cells in a process called glycolysis, which liberates 16.7 kJ (4 kcal) per gram of glucose. The intermediates of glycolysis can also provide precursors for many amino acids, nucleotides and lipids.

Other functions of carbohydrates include forming structural elements surrounding cells and playing a part in the mechanisms by which cells recognise each other.

Structural carbohydrates

The proteoglycans, which form the gel part of the ECM, typically contain 95% carbohydrate. They are formed from two components: a linear polypeptide chain, to which are attached large, linear polysaccharide chains called **glycosaminoglycans (GAGs)** or **mucopolysaccharides** (Table 2.12). These GAGs consist of repeating disaccharide units. Many of these are negatively charged and thus attract cations and large amounts of water, drawn into the gel by osmosis. This results in a gel-like structure around the collagen fibres of the ECM, imparting a degree of flexibility to the tissue and acting as a shock absorber. In contrast, the GAG, heparin, although having a similar structure to the other GAGs, has a very different function. It is released from mast cells and acts as an anticoagulant (see Ch. 12).

Cell recognition by cell surface glycosylation

The carbohydrates found in glycoproteins are of two types. One type, which consists of one or two simple mono- or disaccharides, are linked to serine and threonine residues by **O-glycosylation**. The linkage is between the hydroxyl group of the amino acid side chain and the carbon-1 of *N*-acetylgalactosamine.

The other way that sugars are linked to proteins is via **N-glycosylation** to asparagine. This always occurs at specific amino acid sequences in the protein, either asparagine-X-serine or asparagine-X-threonine, where the linkage is between the amide group of asparagine and the carbon-1 of *N*-acetylglucosamine. Monosaccharide units are then added, in a specific pattern, to give highly branched carbohydrates. Many of these carbohydrates contain the sugar *N*-acetylneuraminic acid (sialic acid). This sugar is negatively charged, giving many cell surfaces a net negative charge. Lipids in the cell membrane are also glycosylated.

These distinctive patterns of glycosylation are particularly important in cell-to-cell recognition and how the immune system recognises self. An example of this specificity can be seen on red blood cells.

Blood groups

The surface of the red cell is covered with patterns of carbohydrates attached to protein and lipids, making the many different **antigens** that determine different **blood groups**. The most studied of these is the **ABO blood group system**. The O antigen is a small polysaccharide, present on all red cells of all blood groups. The A and B antigens are produced by the addition of *N*-acetylgalactosamine or galactose sugar, respectively, to this O antigen. These are added by specific transferase enzymes; individuals with the AB blood group have both enzymes and their erythrocytes have both A and B antigens (see Ch. 12).

LIPIDS

In adults who are not overweight or underweight, lipids make up about 20% of body mass, usually more in women. Unlike proteins, carbohydrates and nucleic acids, lipids are not large covalently bonded molecules. However, because of the interactions that occur between lipid molecules, they form large structures of repeating molecules. Lipids have a number of crucial roles:

- They form the membranes surrounding every cell and the compartments within the cells.
- They are also the major form of energy storage.
- They have important roles in cell signalling.

Abnormalities in lipid transport are associated with the development of atherosclerosis, where arteries are clogged with fatty deposits, and excess storage of lipids results in obesity. Diseases such as diabetes mellitus and pancreatitis have associated lipid abnormalities, and there are also rare inherited lipid storage diseases.

Triacylglycerols

The simplest type of lipids, the **triacylglycerols**, also called triglycerides or neutral fats, are formed from two components, **fatty acids** and **glycerol** (Fig. 2.11). Triacylglycerols have three ester bonds, formed by dehydration reactions between the carboxylate group of three fatty acids and each of the three hydroxyl groups of glycerol (Fig. 2.11D).

Fatty acids

Fatty acids are chains of carbon and hydrogen (Fig. 2.11A, B), with a carboxylate (COOH) group at one end (the α -carbon) and a methyl group (CH₃) at the other (the ω -carbon):

- If all the carbon atoms in the chain are single bonded then the fatty acid is **saturated**.
- If there is a single double bond it is **monounsaturated**.
- If there is more than one it is **polyunsaturated**.

Most of the double bonds in biological molecules are in the *cis*-configuration, which puts a rigid kink in the hydrocarbon chains. Many of the fatty acids found in cells have 16, 18 or 20 carbon atoms and up to three double bonds.

Glycerol

Glycerol is a sugar alcohol, a modified triose sugar formed by the reduction of the aldehyde group to hydroxyl (Fig. 2.11C).

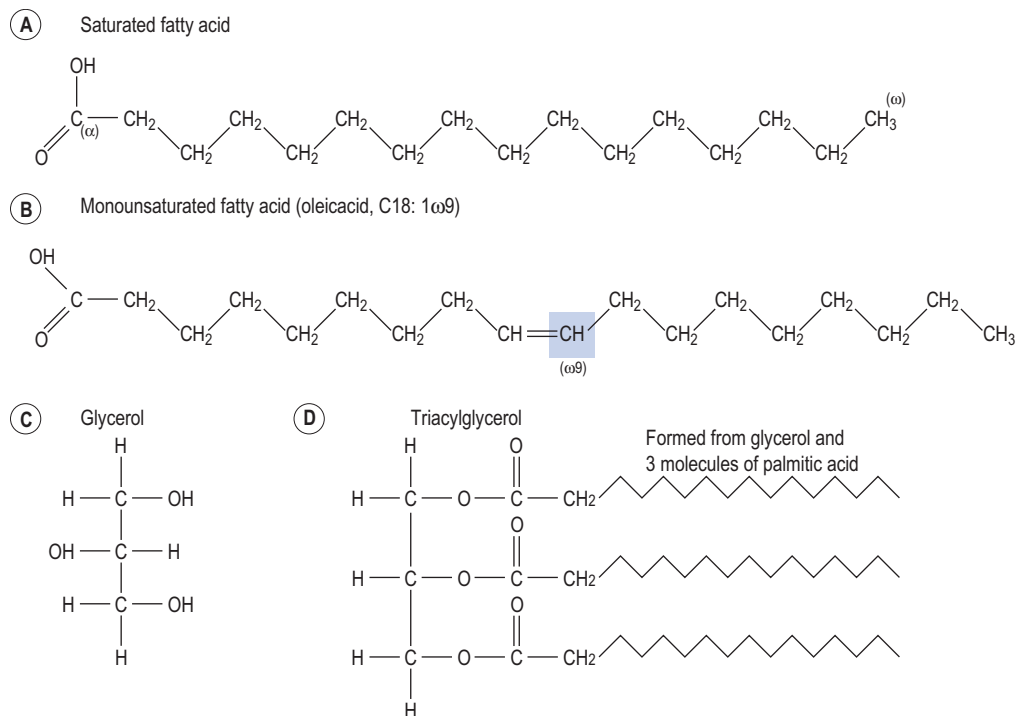


Fig. 2.11 Components of triacylglycerols.



Information box 2.5 Fatty acids can be named in several ways

Many of the different fatty acids have common names which reflect their original source, but all fatty acids can be described according to the number of carbon atoms in the chain and the number and position of the double bond/s.

A shorthand notation that is widely used gives the number of carbon atoms, the number of double bonds and the position of the first double bond, counting from the ω -carbon. For example, the saturated fatty acid found in palm oil is called palmitic acid. The formula for palmitic acid is $\text{CH}_3(\text{CH}_2)_{14}\text{COOH}$, its chemical name is n-hexadecanoic acid and the shorthand notation is C16:0, indicating that it has 16 carbon atoms and no double bonds.

Another fatty acid, arachidonic acid, is called *cis*-5,8,11,14-eicosatetraenoic acid, which is shortened to C20:4 ω -6. Another type of shorthand indicates the positions of all of the double bonds: C20:4 All *cis*- $\Delta^5, \Delta^8, \Delta^{11}, \Delta^{14}$ (see Fig. 2.14).

Dietary fats

Dietary fats, both solid and liquid (oils), are mainly composed of triacylglycerols. The fatty acids which make up the triacylglycerol molecules in each type of fat determine the physical properties of the fat. If the triacylglycerol has fatty acid chains that are predominantly short, or contains unsaturated fatty acids, it will be liquid at room temperature. Examples of these are the oils such as olive oil (which contains oleic acid) and sunflower oil, which contain polyunsaturated fatty acids. If there are more saturated fats and longer fatty acid chains then the resulting fat will be solid at room temperature, for example, butter and the fats found in meat.

Adipose tissue

Triacylglycerols tend to aggregate with the formation of hydrophobic bonds between them which exclude water. This can be seen clearly when oils are added to water. They do not mix and form droplets. Instead they form a thin layer over the surface. In the body, triacylglycerols form the major energy store. They are stored in white adipose tissue. This is found mainly under the skin, and also surrounding the abdominal organs and on the hips. Adipocytes, specialised mesenchymal cells, each consist almost entirely of a single droplet of triacylglycerols. The molecular aggregation of the lipid molecules means that the lipids do not have an osmotic effect; i.e. despite the presence of large numbers of molecules, they do not act like individual particles. Obesity is characterised by increased size and number of these cells.

Fats yield 37.7 kJ (9 kcal) per gram and the amount of fat stored in the adipose tissue in a lean average human would be enough to survive for more than two months without food. There is no specialised storage site for proteins or amino acids. In severe starvation the body will break down tissue and plasma proteins in order to produce fatty acids. Glycogen also has no specialised storage cell; it is simply accumulated in the cytoplasm, particularly in liver and muscle.

Membrane lipids

The lipids which form cell membranes are of three types:

- Phospholipids
- Sphingolipids
- Cholesterol.

Two of these have broadly similar structures. These are the **glycerolipids**, which are mainly **phospholipids**, and the **sphingolipids**.

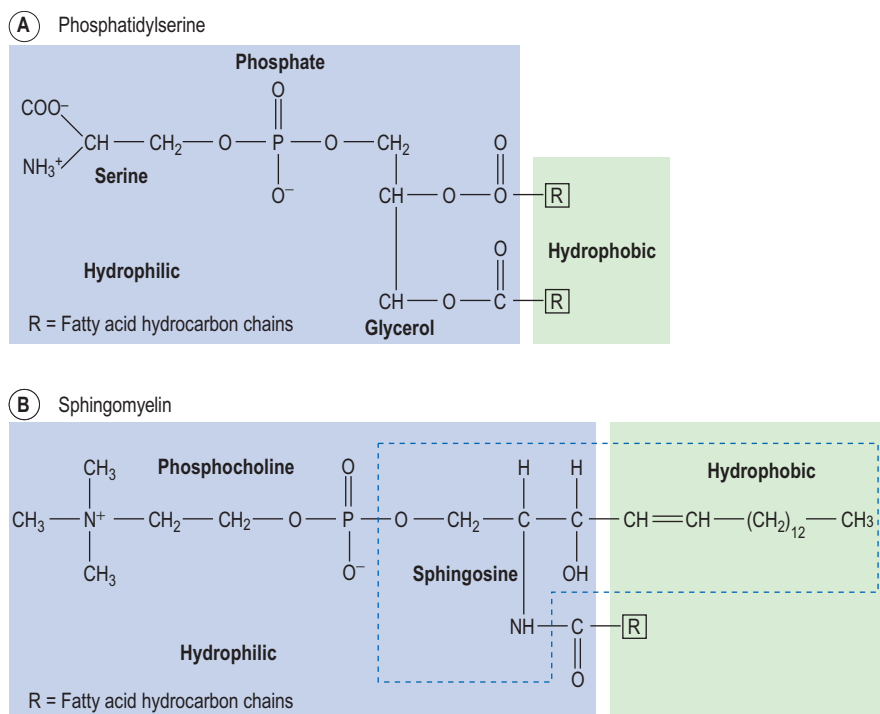


Fig. 2.12 Two membrane lipids.

Types of membrane lipids

Phospholipids

Phospholipids (Fig. 2.12A) consist of glycerol, with fatty acids attached to two adjacent OH groups via ester linkages to give two acyl groups, and with a third group attached, which is a phosphate group. The phosphate is itself linked to one of a number of possible molecules, including serine, ethanolamine (a derivative of serine), choline and inositol. The resulting molecular structure looks somewhat like a tuning fork, with hydrophobic prongs and a hydrophilic handle.

Sphingolipids

Sphingolipids (Fig. 2.12B) have a similar structure to phospholipids, but the hydrophobic portion is formed from **sphingosine**. Sphingosine contains a long hydrocarbon chain and as part of a sphingolipid it is linked to a fatty acid, giving the sphingolipid two long hydrocarbon chains. This is then linked to either a phosphate-containing group, phosphocholine, or a group of sugar residues to produce a glycolipid.

Cholesterol

A third component of cell membranes is **cholesterol**. This has a structure completely different from the other membrane lipids (Fig. 2.13), as it is formed from a four-ringed hydrocarbon called a **steroid**. Although cholesterol is almost entirely hydrocarbon, it has a hydroxyl group attached to the first ring, which can interact with water. The ring structure is quite rigid and has important effects on the fluidity of membranes.

Arrangement of membrane lipids

All the membrane lipids are **amphipathic**; that is they have a hydrophilic head group and a hydrophobic tail. They aggre-

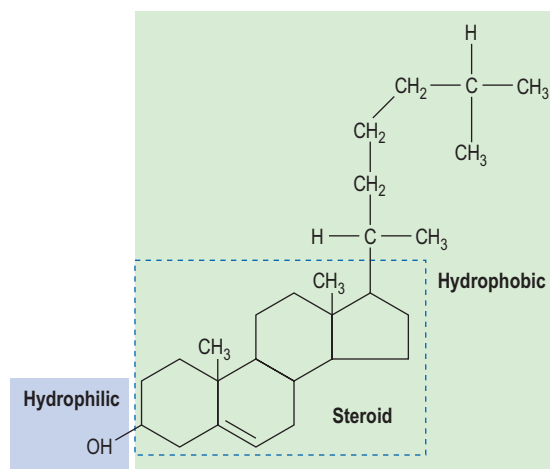


Fig. 2.13 Structure of cholesterol. The figure inside the dotted blue lines is the basic steroid structure.

gate with their head groups facing the aqueous medium and their tails excluding water. It is this arrangement which underlies the formation of all cell membranes (see Fig. 2.28). Many, although not all, of the head groups are negatively charged, giving the surface of the aggregated lipids an overall negative charge.

Different membranes vary in their relative composition. For example, the myelin sheath around nerves has a very high concentration of sphingolipids, whereas phospholipids are more prevalent in the red cell membrane.

An alternative arrangement of amphipathic lipids in aqueous solution leads to the formation of spherical micelles (see Ch. 15), structures generated, for example, by bile salts during the digestion of dietary fat.

Cholesterol and other steroids

Cholesterol is the major **sterol** (a steroid alcohol) in the human body, having both structural and transport functions. It can be synthesised in the liver or absorbed in the diet.

Studies have shown a relationship between raised plasma cholesterol (in the form of low-density lipoprotein (LDL)-cholesterol; see Ch. 16) and increased mortality from heart disease. It is therefore thought to be beneficial to reduce plasma cholesterol levels. However, reducing dietary cholesterol has less of an effect on plasma cholesterol levels than reducing the intake of saturated fats. Diets that are rich in polyunsaturated fats can also reduce serum cholesterol.

As well as forming an important part of membranes, cholesterol is the precursor of many important molecules that are formed by modifications of the basic **steroid** structure (see Fig. 2.13). These are:

- Bile salts
- Steroid hormones
- Vitamin D.

Bile salts

Bile salts act as detergents, helping to break up fats, so aiding their digestion and absorption. They are produced in the liver and are secreted from the gall bladder into the intestine (see Ch. 15).

Steroid hormones

Steroid hormones are all derived from cholesterol, which is converted to pregnenolone and then into the hormone important in maintaining pregnancy, progesterone (Table 2.13). This can then be further modified to produce all of the other steroids (see also Ch. 10). Although they are quite large molecules, because they are lipid soluble, they can cross plasma membranes and act on receptors found inside cells. They are usually carried in the blood, associated with binding proteins such as albumin (see Table 2.10).

Vitamin D

The precursor of vitamin D is synthesised in the skin from cholesterol, where it is converted by the action of ultraviolet light to **cholecalciferol**. Further processing of this molecule produces active **vitamin D** which acts to influence the absorption of calcium in the intestine, reabsorption in the kidney and bone resorption and calcification. When skin exposure to light levels are low, such as might occur in countries outside the tropics in combination with a cultural

tradition of not exposing the skin, then dietary sources of vitamin D are important (see Ch. 16).

Essential fatty acids

While most of the fatty acids used in the body are supplied in the diet (Fig. 2.14), there is almost no requirement for any fat in the diet. Saturated fatty acids can be synthesised from the breakdown products of carbohydrates and proteins in a process called lipogenesis (see Ch. 3). The requirement for mono- and polyunsaturated fatty acids is met by enzymes that convert saturated fatty acids to unsaturated fatty acids.

However, some fats **are** required:

- To aid the absorption in the gut of fat-soluble vitamins
- As precursors to the eicosanoids.

These **essential fatty acids** are **linoleic acid** and **linolenic acid**, both of which were originally identified in linseed oil. They cannot be manufactured in the body as none of the enzymes which convert fatty acids can include double bonds beyond carbon-10, and an important group of lipids, the **eicosanoids** are all made from 20 carbon fatty acids with between three and five double bonds. **Arachidonic acid**, the precursor for a large number of different eicosanoids, is synthesised from linoleic acid (see below).

Eicosanoids

These are a large group of locally acting hormones derived from fatty acids (Fig. 2.14). They are extremely potent sub-

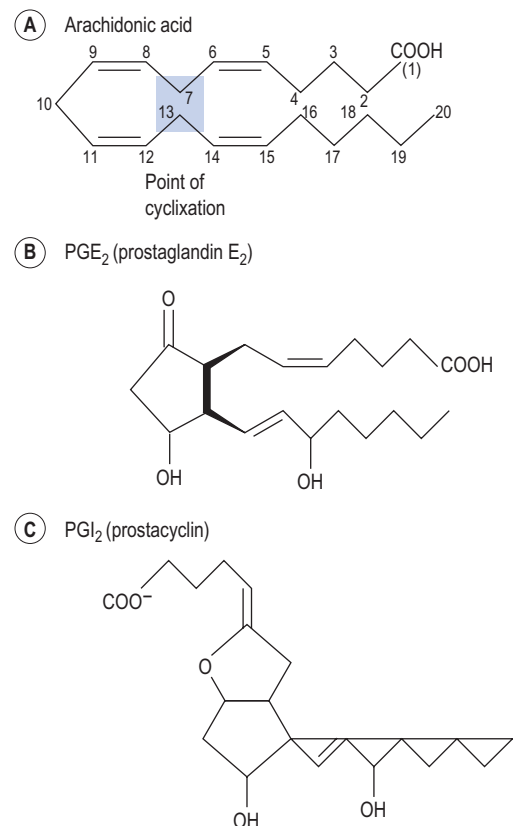


Fig. 2.14 Arachidonic acid. All *cis*, 5, 8, 11, 14 eicosatetraenoic acid (or alternatively C₂₀: 4 ω 6 C-20 Δ^5 , Δ^8 , Δ^{11} , Δ^{14}) and two of the many eicosanoids derived from it (PGE₂ and PGI₂ (prostacyclin)).

Table 2.13 Important steroid hormones

Group	Example	Major actions
Glucocorticoids	Cortisol	Formation of glucose, breakdown of proteins and lipids
Mineralocorticoids	Aldosterone	Sodium and potassium homeostasis
Progestogens	Progesterone	Pregnancy and ovarian cycles
Androgens	Testosterone	Male sexual development and characteristics
Oestrogens	Oestradiol	Female sexual development and characteristics



Information box 2.6 Prostanoids facilitate normal blood flow through vessels and aspirin acts by inhibiting the synthesis of prostaglandin

Blood is protected from clotting in blood vessels, under normal conditions, due to the release of a compound, prostacyclin, from the vessel wall. Prostacyclin acts as a vasodilator, thus encouraging flow, and inhibits the sticking together of blood platelets, which form the basis of clot formation. If the vessel wall is breached substances are released that stimulate platelets to release another prostaglandin, thromboxane A_2 . In contrast to prostacyclin, thromboxane A_2 acts as a vasoconstrictor, thus reducing blood loss through the leaking vessel wall, and stimulates platelets to aggregate, plugging the hole in the vessel.

The enzyme which acts on arachidonic acid in the first step of the production of the prostanoids is cyclo-oxygenase (COX). Aspirin (acetylsalicylic acid) is one of a group of drugs called non-steroidal anti-inflammatory drugs (NSAIDs) which inhibit COX and reduce inflammation and pain. Aspirin is also used to inhibit blood clotting by daily consumption of a low-dose, since, in the absence of COX, thromboxane A_2 cannot be produced. Because of the enzyme inhibition by NSAIDs, gastric bleeding is seen as a common side effect. COX is seen in two isoforms, COX-1 and COX-2. While COX-1 is continuously produced, the other isoform, COX-2, is only induced in response to inflammatory mediators. Although current NSAIDs inhibit both COX-1 and COX-2, it is thought that if specific COX-2 inhibitors were developed, they might be able to reduce inflammation without the side effects of current NSAIDs.

stances and are all broken down quickly after being released, so they act only over short distances. They have important roles in responses to inflammation and in the control of physiological processes, showing a wide range of functions. For example, some stimulate smooth muscle cell contraction, producing vasoconstriction, while others act as vasodilators.

The different groups of eicosanoids are:

- Prostaglandins
- Thromboxanes (together these are called the prostanoids)
- Leukotrienes.

Arachidonic acid, or **eicosatetraenoic acid**, is one of the fatty acids incorporated into membrane phospholipids. It can be released from the membrane by the action of the enzyme phospholipase A_2 (PLA $_2$) on phosphatidylcholine, one of the most common membrane phospholipids. PLA $_2$ is stimulated by a range of different signals in different tissues, including general cell damage.

Another fatty acid, **eicosapentaenoic acid**, is the precursor to a group of prostaglandins with three double bonds. One of the double bonds is at carbon-17, which is three carbons from the ω -carbon, hence the name of this type of fatty acid which is ω -3 fatty acids. It is thought to be beneficial to include this polyunsaturated fatty acid, which is found in fish oils, in the diet.

NUCLEIC ACIDS

Deoxyribonucleic acid (DNA) and the **ribonucleic acids (RNAs)** contain and transmit the genetic information in all biological organisms. Nevertheless, they are not very complex, in the sense that they are regularly arranged

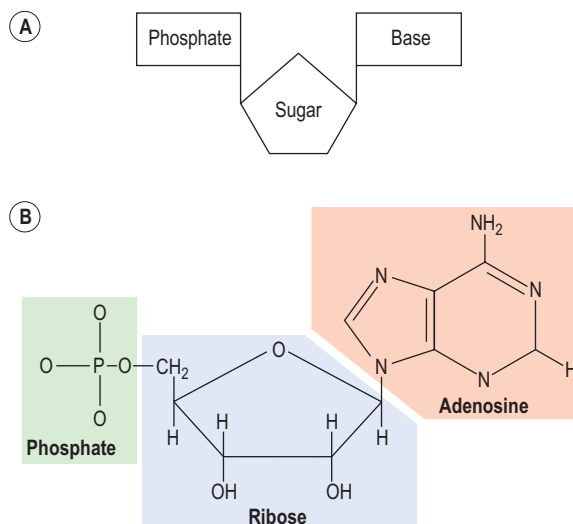


Fig. 2.15 Nucleotide structure: (A) components of a nucleotide; (B) structure of ribo-adenosine monophosphate (rAMP), usually referred to more simply as adenosine monophosphate.

molecules made from a limited number of monomers. Nucleic acids, like proteins, are unbranched chains of units that are all linked with a similar type of bond. However, instead of the 20 possible different subunits which can make up proteins, each of the nucleic acids has only four different subunits. In this way they would seem to be much simpler molecules than proteins, but the way in which the nucleic acid chains are used to store and transmit the genetic information of every cell indicates that the information contained within these macromolecules is complex (see below).

In addition to their genetic role, some of the subunits of nucleic acids also have important roles in energy transfer and as cell signalling molecules.

Nucleotides

Nucleotides form the basic units from which the larger nucleic acids are made. Nucleotides have three components (Fig. 2.15):

- A sugar
- A phosphate
- A base.

Nucleotide sugars and phosphates

Two different sugars are found in nucleotides, **deoxyribose** and **ribose**. These pentose sugars are both aldoses with a furanose ring (see Fig. 2.10E,F).

Attached to the hydroxyl on carbon-5 of these sugars by an ester bond is one of three possible phosphate-containing groups. These phosphate-containing groups can have either one, two or three phosphates attached to each other via phosphoanhydride (pyrophosphate) bonds between the phosphates.

Nucleotide purines and pyrimidines

There are two types of **bases** found in nucleotides: the **purines** and the **pyrimidines** (Fig. 2.16). Purines have two fused rings, while pyrimidines have only one. The rings contain both carbon and nitrogen atoms and form planar

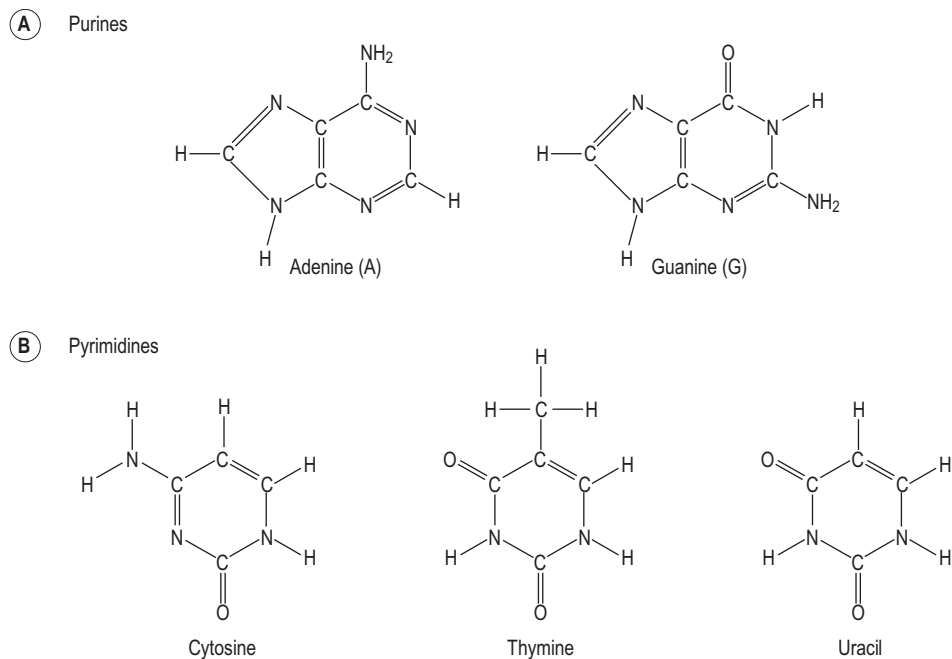


Fig. 2.16 Structure of purine and pyrimidine bases.

Table 2.14 Nucleoside phosphates*					
Base	Nucleoside	Nucleoside monophosphates	Nucleoside diphosphates	Nucleoside triphosphates	Cyclic nucleotides
Adenine	Adenosine	AMP	ADP	ATP	cAMP
Guanine	Guanosine	GMP	GDP	GTP	cGMP
Cytosine	Cytidine	CMP	CDP	CTP	
Uracil	Uridine	UMP	UDP	UTP	
Thymine	Thymidine	TMP	TDP	TTP	

*The nucleosides A, G, C and U can contain ribose or deoxyribose, but AMP infers the ribose form, which strictly is called rAMP, the deoxyribose form being called dAMP. The thymidine nucleotides are only produced as deoxynucleotides and are usually written without the d (e.g. TTP).

molecules. The nitrogen atoms are uncharged at neutral pH, which means that the bases are non-polar and hydrophobic. The major purines are adenine and guanine and the major pyrimidines are cytosine, thymine and uracil. The coenzyme NAD (nicotinamide adenine dinucleotide) contains a pyridine base, nicotinamide.

Nomenclature of nucleotides

There are two different ways of naming nucleotides, which can cause confusion. **Nucleosides** are nucleotides without the phosphate; that is, they have a pentose sugar and a base. Nucleotides with one phosphate can be called mononucleotides or nucleoside monophosphates and so on. Many of the different nucleoside phosphates can be found in cells and have a variety of functions (Table 2.14).

Nucleoside phosphates

Adenosine triphosphate

The most important of the nucleoside phosphates is **adenosine triphosphate** or ATP (Fig. 2.17). The covalent phosphoanhydride bonds which attach the second (β) and third phosphate (γ) groups to the mononucleotide, **adenosine monophosphate** (AMP), can be broken to release energy for other reactions, about 7.3 kcal/mol each. This is under standard laboratory conditions, although in the conditions

within the cell, energy released may be as high as 50 kJ (12 kcal). These bonds are usually shown as ~ indicating, that they are high energy bonds. They can be formed in two ways, either by the transfer of a phosphate group from another phosphorylated compound, called **substrate-level phosphorylation**, or by a process called **oxidative phosphorylation**, where electrons are transferred between a number of intermediates in order to power the enzyme ATP synthase.

Using these mechanisms, the metabolism of a single molecule of glucose can generate 36–38 molecules of ATP (from ADP). This is an energy transfer of about 1130 kJ (270 kcal/mol) of glucose. This energy can then be used to power other cellular processes.

Other nucleoside phosphates

Many of the nucleoside phosphates have been identified as signalling molecules, both outside and inside cells. ATP has been identified as a neurotransmitter at nerve terminals of the autonomic nervous system (see Chapters 4 and 8), as have ADP and AMP. Intracellular ATP can bind to a number of enzymes and receptors to regulate their action.

The cyclic nucleotides, cyclic AMP (**cAMP**) and cyclic guanosine monophosphate (**cGMP**), are important intracellular messengers called **second messengers**, involved in signal transduction between membrane receptors and intra-

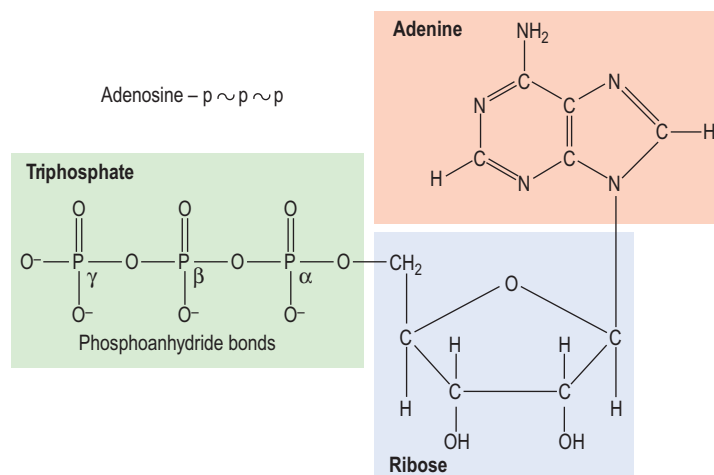


Fig. 2.17 Structure of adenosine triphosphate (ATP). The symbol ~ denotes the so-called high energy bonds.

Information box 2.7 Anti-cancer drugs may act by inhibiting nucleotide synthesis

One of the ways in which cancer is treated is to prevent the rapid proliferation of the cancer cells by blocking cell division. As the majority of normal cells will not be dividing they will not be affected by drugs which block cell division. Many of the side effects of anti-cancer drugs can be accounted for by their effect on cells that are normally rapidly dividing, such as hair and intestinal cells.

A commonly used anti-cancer drug methotrexate, acts by inhibiting an enzyme involved in the production of thymidine nucleotides. It is an analogue of dihydrofolate, which acts as a competitive inhibitor of the enzyme DHFR (dihydrofolate reductase). Other cytotoxic agents act by inhibiting other points in the production of nucleic acids.

See also Chapter 4.

cellular proteins. The coenzymes **NAD**, nicotinamide adenine dinucleotide phosphate (**NADP**), flavin adenine dinucleotide (**FAD**) and **coenzyme A** are all derived from nucleotides.

Sources of nucleotides

The nucleotides required to make DNA and RNA can be synthesised de novo by all cells. The first step involves the synthesis of ribose, as ribose 5-phosphate, which is then converted to a precursor for all the nucleosides, PRPP (5-phosphoribosyl pyrophosphate). This precursor is then used to produce either rUMP (see Table 2.14) from which the other pyrimidines are produced or rIMP (inosine monophosphate), which is the precursor for the purines. The nucleoside monophosphates are then phosphorylated by nucleoside kinases to generate the bi- and triphosphates.

Ribonucleotides are required continually by all cells but dividing cells also require deoxyribonucleotides in order to make new DNA. These are produced from the corresponding ribonucleoside diphosphates. The deoxyribonucleotide TMP is produced from dUMP by a series of reactions requiring the vitamin folic acid.

Recycling of nucleotides

The synthesis of nucleotides de novo requires a large input of energy and there are mechanisms by which free bases, either salvaged from endogenous sources or from nucleotides present in the diet, can be combined with PRPP to

produce new nucleotides. This recycling is important in some cells, particularly in T lymphocytes. Also many obligate parasites, such as viruses, do not have the enzymes required to manufacture nucleotides and are completely dependent on their hosts for a supply of nucleotides.

Deoxyribonucleic acid and ribonucleic acids

Deoxyribonucleic acid

Only a single type of nucleic acid is made from the deoxyribonucleotides, **deoxyribonucleic acid**, or **DNA**, and this is found in all cells capable of dividing.

Most of the DNA within a cell is found in the nucleus, from which derives the name nucleic acids. It is absent from red blood cells, which do not have a nucleus. It is also found in the cytoplasm of prokaryotic cells (bacteria) and some viruses (DNA viruses). A small amount of DNA is also found inside intracellular structures called mitochondria. DNA occurs as a double-stranded molecule, which has a characteristic double-helical structure. This structure was first proposed by James Watson and Francis Crick in 1953, after seeing the X-ray crystallography data obtained by Rosalind Franklin and Maurice Wilkins.

Ribonucleic acids

Nucleic acids made from **ribonucleotides** are more diverse. Three types of **ribonucleic acid** (RNA) are found in eukaryotic cells. These are:

- **Messenger RNA** (mRNA): which is copied from DNA and used as the template for the synthesis of proteins
- **Transfer RNA** (tRNA): small molecules which transfer amino acids to the site of protein synthesis, the ribosome
- **Ribosomal RNA** (rRNA): ribosomes consist of a combination of RNA and protein in two subunits. They are involved in binding mRNA and tRNA during protein synthesis.

Details of how these molecules are involved in the manufacture of proteins are covered below. RNAs are found in all cells (except red blood cells), including prokaryotes, and in some viruses (RNA viruses). All viruses have either DNA or RNA as their genetic material.

Clinical box 2.6 Excess purines can cause disease

In humans, excess purines are broken down to uric acid. If large amounts of uric acid are produced and not removed by the kidneys, uric acid crystals are deposited in joints and soft tissues, resulting in gout. One treatment of gout blocks the enzyme, xanthine oxidase, which catalyses the last steps in the production of uric acid. The intermediate breakdown products are more water soluble than uric acid and can be excreted.

Deficiencies in one of the enzymes involved in the recycling pathways leads to increased purine synthesis and a concomitant increase in uric acid, leading to a condition known as Lesch-Nyhan syndrome. One of the features of Lesch-Nyhan syndrome is gout. Patients may also go on to develop arthritis and severe mental disorders.

Constituent bases of DNA and RNA

Each nucleic acid contains two purines and two pyrimidines. The purines adenine (A) and guanine (G) are found in both DNA and RNA. However, DNA and RNA differ in one of their pyrimidines. DNA contains the pyrimidines cytosine (C) and thymine (T), but in RNA thymine is replaced by uracil (U).

Nucleotide bonds

The nucleic acid chains are formed by the sequential linkage of nucleoside triphosphates by phosphodiester bonds between the sugar and phosphate groups (Fig. 2.18), a reaction catalysed by the enzymes DNA polymerase and RNA polymerase. The free hydroxyl group of carbon-3 of the sugar is linked to the triphosphate on the carbon-5 of the next nucleotide, with the loss of a pyrophosphate (PPi). The chain grows in the direction 5'→3'. This forms a sugar-phosphate chain with bases attached to the other side of each sugar (carbon-1). At physiological pH the phosphate groups are anions, so the nucleic acids have a negative charge.

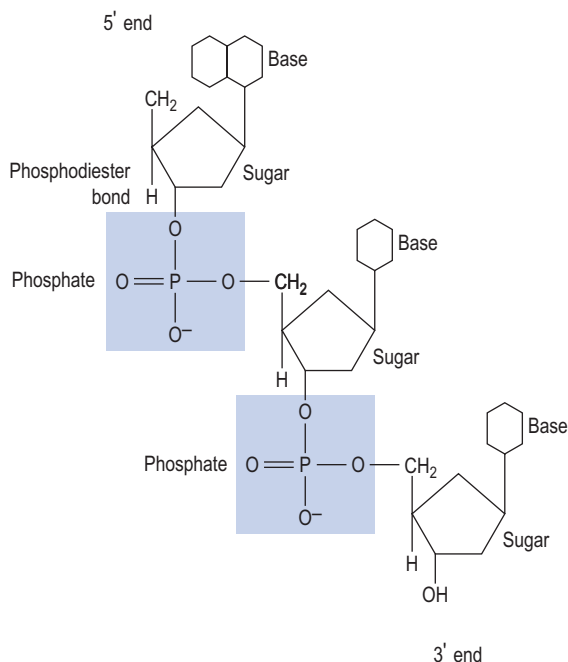


Fig. 2.18 Nucleotide bond formation.

Nucleic acid structure

While the subunits which make up DNA and RNA are fairly similar, there are great structural differences between the macromolecules themselves.

Primary structure

The primary structure of DNA and RNA consists of chains of nucleotides held together by nucleotide bonds. DNA molecules are very large. The genetic material in eukaryotes is organised in units called chromosomes, each of which contains a single DNA molecule. In humans, the largest of these is nearly 10 cm long when fully extended and has $2-3 \times 10^8$ nucleotides in each of its two strands.

RNA molecules vary enormously in size. Each mRNA molecule is used to code for a protein. Therefore the length of the mRNA is related to the size of the protein. Due to the way in which the coding occurs, the mRNA for a protein has at least three times more nucleotides than the number of amino acids in the protein and may have many more. Their sizes range from 500 to 6000 nucleotides. About 50 different tRNA molecules are produced in animal cells, each between 70 and 90 nucleotides long.

Secondary structure of DNA

The secondary structure of DNA is relatively simple, but it nevertheless gives clear indications of how the molecule can function as a source of information and also how it can be duplicated. It is dependent on interactions called **base pairing** between the bases that project from the sides of the sugar-phosphate backbone of each of the two strands. The two strands of the DNA molecule run in opposite directions from one another; they are anti-parallel, with one 3'→5' chain and one 5'→3' chain.

Base pairing

Hydrogen bonds form between pairs of bases projecting from each strand, each **base pair** consisting of a purine and a pyrimidine. The bases can only form specific pairs, depending on the number of hydrogen bonds formed between them (Fig. 2.19). Pairs consisting of **adenine** and **thymine** can form two hydrogen bonds, and pairs between **cytosine** and **guanine** have three hydrogen bonds. The two types of pairs are the same width, which means the double strands are held at the same distance apart throughout their length, with the two carbons from the sugar-phosphate backbone being 1.08 nm apart.

The planar base pairs are thus stacked one on top of another along the core of the DNA, where hydrophobic interactions occur between the non-polar base pairs. The negatively charged phosphate groups and the sugar backbone are on the outside.

The entire DNA molecule is extremely stable because of the large number of hydrogen bonds between the bases and the hydrophobic interactions between the base pairs.

Helical structure of DNA

The double-stranded DNA molecule forms a regular helical structure with 10 base pairs (bp) for each turn, advancing 3.4 nm (0.34 nm per base pair). The geometry of the backbone favours a right-handed helix, which is the normal structure. This type of DNA is called the B form. The two helical grooves along the outside of the molecule have different widths, called the major and minor grooves. The DNA mol-

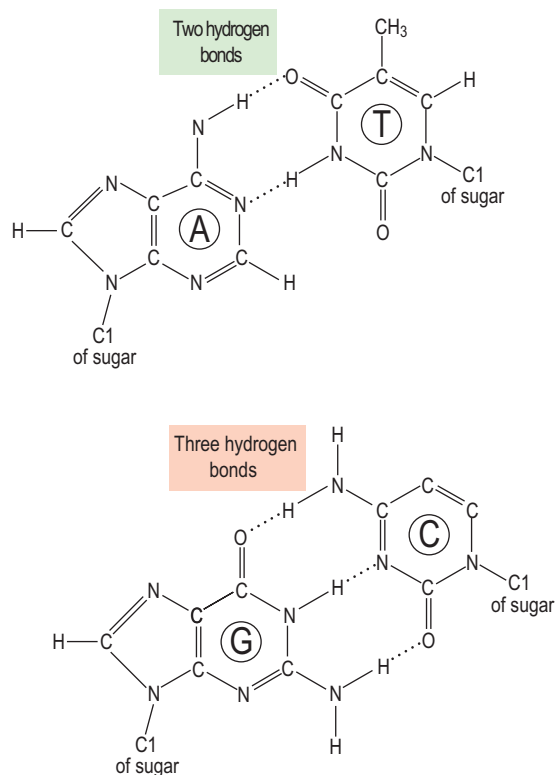


Fig. 2.19 Base pairing between A-T (or U) and C-G.

ecule is stable as well as being quite flexible as there are no hydrogen bonds between nucleotides above or below which would make the double helix more rigid. While most DNA exists in the B form, other forms of DNA are known to exist which may be involved in interactions between DNA and proteins or different regions of DNA.

Complementary strands of DNA

The two strands of the DNA are **complementary**. Because of the constraints of base pairing (A with T and G with C) a particular sequence on one strand must be reflected by a specific sequence on the other. For example the sequence TGCT in one strand would be reflected by the sequence ACGA in the other. In this complementarity lies the key to the way in which the sequence of bases can act as a template for the production of identical copies.

Secondary structure of RNA

RNA molecules are single-stranded and many have a well-defined secondary structure. Stem-loops are loops of RNA that have a base-paired sequence at the beginning and end, or hairpins (which are shorter loops that are almost all paired). These can form between regions of the single-stranded RNA by base pairing between complementary regions.

The small tRNAs all have a similar structure (Fig. 2.20). They form a cloverleaf pattern with four stem-loops, each of which forms a short double helix. rRNA forms a complex secondary structure containing many stem-loops. This pattern seems to be common among a wide variety of organisms, suggesting a common evolutionary origin.

Tertiary structure of RNA

Only some RNAs have a unique folding pattern and, as previously described for proteins, the primary structure is all that is required to reassemble the molecule if it is denatured. The

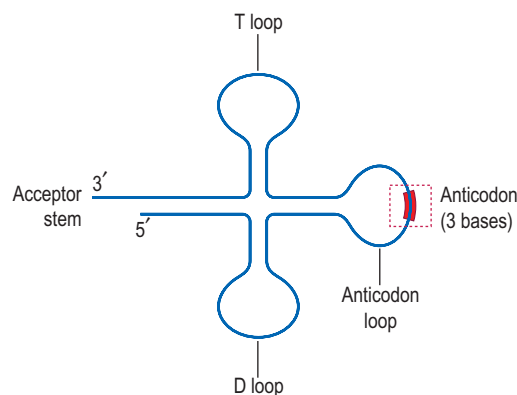


Fig. 2.20 Generalised structure of tRNA.

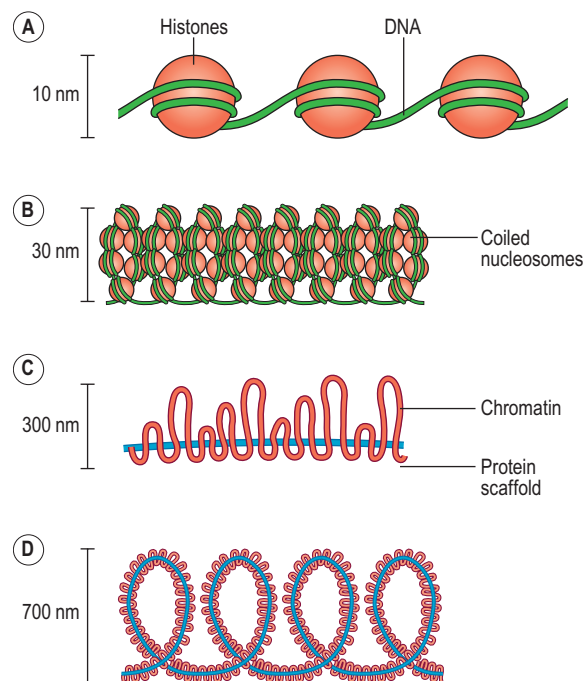


Fig. 2.21 Tertiary structure of DNA.

tertiary structure of tRNAs involves hydrogen bonding, which holds two of the loops parallel to the rest of the molecule in an L-shaped molecule, with a site for binding to mRNA at one end and a site at the other end where the amino acid can be attached.

Tertiary structure of DNA

In a human cell there are about 3×10^9 base pairs of DNA that must be packed into the nucleus. The largest **chromosome** has over 200 million base pairs, which, if extended completely, would give a DNA molecule that would be much larger than the cell. In addition, the packing must allow this DNA to be easily accessible by uncoiling, in order to copy segments.

Chromatin and nucleosomes

In the chromosome, DNA is combined with some RNA and about the same mass of protein, producing a structure called **chromatin**. Chromatin consists of a series of units called **nucleosomes** (Fig. 2.21). Each nucleosome is composed of globular proteins called histones, around which are wrapped two loops of DNA 140 bp long, with a piece of linker DNA between each nucleosome about 80 bp long (Fig. 2.21A).



Information box 2.8 The Human Genome Project has sequenced the entire human genome

A project to map and sequence the entire human genome of approximately 3 billion base pairs was started in 1990 and the first stage, described as a 'working draft', was declared complete on 26 June 2000, by two groups, the Human Genome Project led by Francis Collins, and Celera Genomics led by Craig Venter. The groups used different techniques to produce the draft, which was fully completed by 2003. Surprisingly the total number of genes in the human genome turned out to be much lower than expected (20 000–40 000) although many genes produce more than one gene product through alternative splicing. Much effort is now concerned with identifying the role of these genes and it is expected that this will lead to further advances in medicine.

This produces a 10 nm width fibre which is then coiled into a fibre 30 nm wide with 6 nucleosomes per turn (Fig. 2.21B). The coils are then formed into a supercoiled structure 300 nm in diameter (Fig. 2.21C). This multiple folding reduces the length of the DNA molecule by about 8000-fold. When cells are about to divide the chromatin is concentrated still further and the DNA molecules become visible under a light microscope as chromosomes (Fig. 2.21D). How DNA uncoils in order to be copied is an obvious topological problem.

Human cells have 23 pairs of chromosomes, each with a specific size and shape.

Mitochondrial DNA

The tertiary structure of mitochondrial DNA is very different from nuclear DNA. It is a circular molecule, which is similar to the DNA in bacteria. This is used as evidence that mitochondria originated as free-living bacteria, before forming a symbiotic relationship with what later became eukaryotic cells.

BASIC CELL STRUCTURE

The defining difference between prokaryotic and eukaryotic cells is the presence, in all eukaryotic cells, of a nucleus and other intracellular structures. In the same way that macromolecules contained within cells have similarities across a wide range of organisms, so there are common structures found in cells of different types. However, the specific structure of a particular cell varies enormously, both between cells of different organisms and between cells within a multicellular organism. While there are many multicellular organisms, such as sponges, which consist of aggregates of largely similar cells, most multicellular organisms have cells with different specialised functions. Furthermore, animals and plants have significant differences in their organelles and cell membranes.

HISTOLOGY

Cellular structures can be identified using either light microscopy or electron microscopy (EM). The study of tissue structure by microscopy is termed **histology**. Naturally, the thin layers of cells or tissues on a slide are transparent and

featureless, so both types of microscopy require that the cell or components within the cell are made opaque in some way, which usually kills the tissue. Specialised forms of light microscopy, such as phase contrast, can be used to observe living specimens, although their resolution is fairly poor. The light microscope can resolve objects which are about 0.2 μm apart with magnification up to about 1000 times.

Histochemistry

Cells can be treated by a wide range of methods so that they become visible under the light microscope. These methods rely on specific chemical reactions between the chemicals being applied and the molecules making up or contained within the cellular organelles. One of the most useful stains is a combination called haematoxylin and eosin (H and E). Haematoxylin is a basic dye that stains acidic components in the cell blue. Eosin is an acidic dye that stains basic components pink. In most cells, H and E stains the nucleus, containing the acidic DNA and RNA, blue, and the cytoplasm, most of whose proteins are basic, pink.

Other staining methods have been developed in order to label specific tissues; for example enzyme histochemistry can identify cells or parts of cells containing a particular enzyme. The recent development of methods using immunological markers, a technique referred to as immunohistochemistry, enables identification of specific proteins or other macromolecules within cells.

Electron microscopy

The detailed ultrastructure of cells can be revealed only by **electron microscopy**, where the image is formed by bombarding the specimen with electrons. This can resolve objects which are 1 nm apart, with a maximum magnification of about 100 000-fold. There are two types of EM:

- Scanning EM – used to image the surface of objects.
- Transmission EM – which uses thin sections of tissue treated with heavy metals that bind to cell components in varying amounts and reflect electrons. Those components with a large amount of bound metal reflect many electrons, are electron dense, and look dark in EM, and vice versa.

All EM images are black and white only, although false colour processing can be used to highlight differences in grey scale.

Artefacts

Care must be taken in interpreting images from both types of microscopy. Artefacts can be introduced during the many steps involved in the preparation of a specimen, and interpreting two-dimensional images as three-dimensional (3D) objects takes a lot of thought. However, computer software nowadays assists in reconstructing 3D structures from sequential sections.

ORGANELLES, STRUCTURE AND FUNCTION

Organelles are distinctive structures found within cells. Many are enclosed by one or more membranes, but there are a few structures which are not (called non-membranous organelles) (Fig. 2.22). Organelles enable cells to compartmentalise biochemical processes into different regions of the cell. Enzymes needed to carry out a particular reaction can

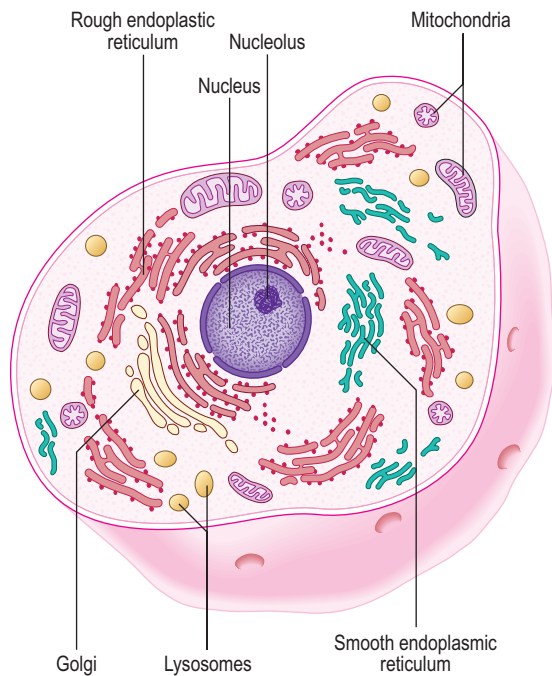


Fig. 2.22 An idealised eukaryotic cell showing the major organelles. Not all organelles are present in all cells.

be put in the same compartment as the substrates for the reaction, and the products can be passed onto the next enzyme in the chain by placing it in the same compartment. Conditions can be maintained within the restricted volume of organelles that would be harmful to the cell overall.

Cell membranes

All cells, both prokaryotic and eukaryotic, are surrounded by a **plasma membrane**, also called the **cell membrane**. It maintains the integrity of the cell and separates the fluids inside the cell from those on the outside and is typically about 9 nm thick. Under the electron microscope it is seen to consist of two outer dark layers, with a clearer inner layer. The plasma membrane regulates the movement of molecules in and out of the cell. The major component of all membranes is lipid, which on its own would make the cell impermeable to all hydrophilic molecules. However, it also contains proteins that allow the movement of hydrophilic molecules through the membrane. The outside of the plasma membrane is also studded with carbohydrates, attached to the proteins and the lipids, which have roles in cell-to-cell interactions and cell recognition. The membranes which surround the membrane-bound organelles also have a banded, double-layered appearance in electron micrographs but are thinner (about 6 nm).

Plant and bacterial cell walls

Plant cells have a rigid cell wall outside their plasma membrane, which gives the plant cell strength. Plant cell walls contain cellulose and sometimes lignin and suberin, which form wood and cork, respectively. Most bacteria also have cell walls. They often contain peptidoglycan, a polymer of amino sugars cross-linked with short peptide chains. The presence and concentration of this macromolecule is used as a way to categorise bacteria. Peptidoglycan is not

Information box 2.9 Gram staining of bacterial cell walls

Gram staining was invented by Hans Christian Gram in 1884 and is used to identify different types of bacteria. The method consists of soaking the sample of the bacterial culture firstly in a violet dye (Gentian violet), followed by treatment with iodine. The slide is then washed with alcohol and counter-stained with a red dye, safranin.

Gram-positive bacteria are coloured blue-black and Gram-negative bacteria are stained red. The different colours obtained are due to the different amounts and accessibility of the peptidoglycan in the cell walls of bacteria. Gram-positive bacteria have a thick layer of peptidoglycan surrounding their plasma membranes. This takes up the violet dye and appears blue to purple.

Gram-negative bacteria have less peptidoglycan, which is surrounded by a second outer cell membrane.

Mycobacteria, such as those causing tuberculosis, have cell walls that have a very different composition and are not stained by the Gram stain.

found in eukaryotic cells and many antibiotics, including the penicillins, inhibit bacterial cell growth by inhibiting its synthesis.

Cytoplasm

Between the plasma membrane and the nucleus is the **cytoplasm**, which contains all the other cellular elements suspended in the **cytosol**. This is a viscous aqueous solution that contains salts, proteins, sugars and the countless other molecules that are involved in the metabolic pathways. It is criss-crossed by the fibrous cytoskeleton. It has been estimated that there are about 300 mg/mL of protein and RNA in the cytosol, which makes it very crowded indeed.

Nucleus

The most prominent organelle in most eukaryotic cells is the **nucleus**, which is often at the centre of the cell, although some cells, such as skeletal muscle cells, have their nuclei displaced to one side. Some cells have more than one nucleus, whereas mammalian red blood cells lose their nucleus as they mature (see Ch. 12). The nucleus is the only organelle that can be seen under the light microscope without staining.

The nucleus contains the chromosomes in the form of chromatin, which is not normally visible as an organised structure. Actively transcribed chromatin, called **euchromatin**, is diffuse. This is where mRNA and tRNA are being transcribed. More condensed chromatin, called **heterochromatin**, is inactive.

Inside the nucleus are one or more dark-staining spherical areas called **nucleoli**. These are areas of the nucleus where ribosomes are assembled. These areas contain large amounts of RNA and proteins, and in cells which are actively secreting large amounts of protein they can be very large. Surrounding the chromatin and the nucleoli is a gel-like suspension called the **nucleoplasm**.

The nucleus is surrounded by a double layer of membrane, the **nuclear envelope**. The inner membrane is smooth, whereas the outer membrane may be continuous with the



Information box 2.10 Barr bodies represent inactive X chromosomes

In all female cells, there are two X chromosomes. However, in all somatic cells (that are not oocytes), one of these X chromosomes is inactivated at random. In about 30% of nuclei this can be seen as a condensed mass of heterochromatin called a Barr body, or X chromatin. In white blood cells called neutrophils this appears as a small drumstick. If the cell contains more than two X chromosomes, such as in triple X syndrome (XXX), then the extra chromosomes will be seen as extra Barr bodies.

endoplasmic reticulum and the outer face may be covered with ribosomes. The nuclear envelope is studded with pores. These are about 9 nm in diameter and are surrounded by an octet of protein complexes in each membrane, forming a channel through which molecules of protein and RNA can move.

Endoplasmic reticulum

Extending from the nucleus and into the cytoplasm are a series of interconnected flattened tubules and sacs called the **endoplasmic reticulum** (ER). This network of membranes extending into the cytoplasm, which is continuous with the outer nuclear membrane, encloses cavities called cisternae. There are two types of ER within the cell, one called rough ER, whose membranes are covered with dense granules, and smooth ER, which does not have these granules.

Ribosomes and rough ER

The granules which are studded all over the rough ER are called **ribosomes**. They are made from two subunits, each made up of a combination of RNA and protein which is assembled in the nucleolus. The large subunit contains a large RNA molecule of 4800 nucleotides with two smaller RNAs of 160 and 120 nucleotides, all associated with about 50 proteins. The small subunit has a single RNA of 1900 nucleotides and about 33 proteins.

Ribosomes are involved in the synthesis of proteins, which may be destined for insertion into membranes, used in lysosomes, or exported from the cell. Rough ER is very extensive in cells that have high levels of protein secretion. Many of these proteins are then processed by enzymes within the ER, which can add sugars and proteins to them. Small vesicles bud off from the ER and are seen to combine with the Golgi apparatus (see below).

Ribosomes are also found in the cytoplasm, unattached to the ER. These free ribosomes are involved in the manufacture of cytoplasmic and nuclear proteins. Mitochondria also contain ribosomes, whose rRNA is coded for by mitochondrial DNA.

Smooth ER

Smooth ER is continuous with rough ER, but is generally more tubular in nature. Smooth ER manufactures membrane phospholipids and also makes cholesterol and steroid hormones. Cells, such as those in the adrenal cortex, the testes and the ovaries, which secrete large amounts of steroid hormones, have very extensive smooth ER. In the liver, cells

called hepatocytes have rough ER, which is involved in the production of plasma proteins, such as albumin. However, a large proportion of the ER in these cells consists of smooth ER. This smooth ER is associated with granules of glycogen, which can be broken down to maintain blood glucose levels. The enzymes in the smooth ER in the liver are involved in the detoxification of a wide range of lipid-soluble molecules, including carcinogens and drugs, transforming them into polar metabolites that can then be excreted by the kidney (see Ch. 14). In two types of muscle cells, cardiac and skeletal muscle, a specialised smooth ER, called **sarcoplasmic reticulum**, has a critical role in the sequestration and release of calcium ions that are required for the activation of muscle contraction.

Golgi apparatus

The **Golgi apparatus**, named after Camillo Golgi, who first described it in 1898, appears as a number of flattened membrane-bound sacs called **cisternae**, stacked on top of each other and surrounded by large numbers of small spherical membrane-bound organelles called **vesicles**. It is involved in the concentration, modification (mainly glycosylation and phosphorylation) and packaging of proteins produced in the rough ER. The Golgi apparatus has a definite polarity. Small vesicles, called transport vesicles, derived from the ER and containing proteins and lipids, fuse onto the *cis* Golgi. After processing through the *medial* Golgi, where they are glycosylated, vesicles containing protein bud off from the *trans* Golgi, where they are sorted for a number of different destinations including secretory vesicles and lysosomes. The proteins move through the Golgi stack by being shuttled in small transport vesicles from cisterna to cisterna.

Secretory vesicles

Proteins destined to be exported from the cell are contained in vesicles derived from the *trans* Golgi. Mature vesicles are called **secretory granules**, and in some cells (for example, in the exocrine pancreas) these are some of the most prominent structures, filling a large proportion of the cytoplasm. During maturation the contents of the vesicles are concentrated and the proteins can be modified; for example modification of the insulin molecule, which removes the C-peptide, occurs in the secretory granules (see Clinical Box 2.1).

Secretion

Proteins are released into the extracellular space by a process called **exocytosis**, which involves the fusion of the vesicle's membrane with the plasma membrane. Release of the proteins may be continuous, or vesicles may be stored. Continuous secretion, called **constitutive secretion**, occurs in some immune cells called plasma cells, which continually secrete immunoglobulins.

Most secretion occurs in response to a signal, often a rise in intracellular calcium levels, and this **regulated secretion** releases stored secretory granules. Many neurons release more than one type of chemical and it is possible to identify different populations of vesicles in nerve terminals. The different chemicals, which often consist of a small non-protein and a protein, are released under different patterns of stimulation (see Ch. 8).

**Information box 2.1 Enzyme histochemistry can be used to identify lysosomes**

Despite their variable and relatively unremarkable structure, lysosomes can be easily identified using a staining technique called enzyme histochemistry, as they contain acid phosphatases. The cells are incubated with a substrate for the acid phosphatase. The reaction between enzyme and substrate releases phosphate ions. These ions react with lead ions to form the insoluble product lead phosphate, which is deposited at the location of the acid phosphatase. This is then treated with ammonium sulphide to give a black precipitate of lead sulphide, clearly visible with a light or electron microscope.

Lysosomes and proteasomes

One of the destinations for vesicles derived from the Golgi are the **lysosomes**. These are spherical or oval organelles with a variety of sizes up to 400 nm diameter. They contain about 40 different acid hydrolase enzymes which can digest most biological molecules. As well as vesicles from *trans* Golgi, other membrane-bound structures fuse with lysosomes, where their contents can be digested. These include particles such as bacteria and viruses that have been retrieved from the extracellular medium by two processes called endocytosis and phagocytosis. The lysosome is also used to destroy aged cell organelles that are first engulfed by the ER, forming a vesicle that is passed to lysosomes for destruction.

Acid hydrolases

The fluid inside lysosomes is much more acidic, at about pH 4.8, than the normal pH of about 7.0–7.3. This is the optimal pH for the activity of the **acid hydrolase** enzymes present in the lysosomes. The low pH denatures many macromolecules, which aids their degradation. However, fortunately, the acid hydrolase enzymes are relatively inactive at the pH of the cytoplasm, so if a single lysosome is ruptured little digestion of cytoplasmic contents occurs. Hydrogen ions will tend to leak out of the lysosome so the pH gradient between the cytoplasm and the interior of the lysosome is maintained by special protein pumps in the lysosomal membrane that pump hydrogen ions into the lysosome. Other pumps transport the products of the lysosomal digestion, such as amino acids and sugars, into the cytoplasm where they can be recycled.

The lysosomal membrane has an important role in compartmentalising the degradative enzymes because, if many lysosomes were ruptured, the cell could self-digest. This process is known as **autolysis** and can occur physiologically, such as, for example, during the breakdown of the uterine lining during menstruation.

Proteasomes

While membrane components and aged organelles are degraded in lysosomes, another pathway exists to degrade cytosolic proteins. **Proteasomes** are large complexes of proteins arranged in four rings around a central core. Proteins destined for destruction are first tagged with multiple copies of a protein called **ubiquitin** by specific enzymes. This acts as a signal to direct the protein into the core of the proteasome where it is broken into small peptides. These

Clinical box 2.7 Tay–Sach's disease is one of a number of lysosomal storage diseases

The absence of a particular enzyme in the breakdown pathways inside lysosomes leads to the accumulation of the substrate for that enzyme. In Tay–Sach's disease, the absence of the enzyme, hexosaminidase leads to the accumulation of a membrane glycolipid, G_{M2} , which is particularly common in nerve cells. The disease, which is inherited, causes mental retardation and central nervous system disorders, with death occurring by about the age of 5.

are then further digested in the cytoplasm by peptidases. Only proteins tagged with ubiquitin are able to enter the proteasomes, ensuring that normal cytosolic proteins are protected.

Many proteins are continually being turned over and each protein has its own rate of degradation. Some enzymes are replaced rapidly but others such as collagen have an almost imperceptible turnover rate. In starvation, proteins, particularly skeletal muscle proteins, are broken down to provide precursors for glucose synthesis.

Peroxisomes

Peroxisomes are small vesicles, also called **microbodies**, which superficially resemble lysosomes, but contain completely different enzymes that are transported into the peroxisome from the cytosol. Some of these enzymes are oxidases and catalases:

- The oxidases break down a variety of substrates, such as fatty acids, with the production of hydrogen peroxide (H_2O_2). Although potentially highly toxic to cells, H_2O_2 is produced inside some cells of the immune system in order to kill bacteria.
- The catalase enzymes use H_2O_2 to oxidise other potentially toxic compounds, or break down H_2O_2 to water.

Other peroxidase enzymes are also involved in the biosynthesis of cholesterol and other membrane components. Peroxisomes are particularly abundant in the hepatocytes of the liver.

Mitochondria

All the organelles described so far are involved in the production, processing, targeting and degradation of the proteins and other molecules made within the cell. However, most of these processes require **energy** in the form of ATP. **Mitochondria** are the organelles where ATP is generated, and so they are particularly plentiful in cells which consume large amounts of energy.

Structure of mitochondria

Mitochondria vary in size and shape but are generally sausage-shaped and about 1 μm wide and 7 μm long. They have a complex internal structure with two layers of membranes. The outer membrane is smooth and the same thickness as in other organelles. It is highly permeable to small molecules, due to the presence of a pore-forming protein called **porin**. The intermembrane space between the outer

and inner membrane is usually narrow. The inner membrane is studded with many proteins that are involved in oxidative phosphorylation, and the subsequent generation of ATP. This membrane is thinner and has multiple folds projecting inwards, called **cristae**, producing a very large surface area for ATP production. The innermost region is called the **matrix**. It is here that pyruvate, derived from the metabolism of glucose in the cytosol, and fatty acids are metabolised to provide the substrates for oxidative phosphorylation.

Mitochondrial DNA

The mitochondria are unique among the organelles, in that they contain their own DNA and the mechanism to make proteins. The matrix contains **mitochondrial DNA** (mtDNA), which codes for some of the molecules needed by the mitochondrion, including its own rRNA and tRNA. However, many of the proteins required by mitochondria are manufactured in the cytoplasm and have to be imported into the mitochondrial matrix. Receptors on the surface of mitochondria act to transfer these proteins required in the matrix through the outer and inner membranes.

Evolutionary origin of mitochondria

Bacterial cells do not contain mitochondria but mitochondria themselves have many similarities with a group of bacteria called purple bacteria. This has led to the hypothesis called the **endosymbiotic theory**: that mitochondria and other intracellular organelles of eukaryotes could have arisen when one type of bacteria was engulfed by another, without digesting it. The two bacteria could have become dependent on one another until they could no longer exist separately. Another piece of evidence for the bacterial origin of mitochondria is that the genetic code which translates the bases of the mitochondrial DNA molecule into amino acids is slightly different from that used by the rest of the cell, and can vary both between different mammals and other organisms.

Inheritance of mitochondrial DNA

Mitochondria are passed from parent to child during reproduction, but because the volume of the cytoplasm contributed by the sperm is very small compared to that of the ovum, virtually all the mitochondria (estimated at 99.99%) are inherited from the mother. mtDNA is therefore passed, virtually unchanged, down the maternal line and geneticists interested in the evolution of humans have utilised mtDNA lineages to examine population movements.

Plant 'mitochondria'

In plants, the equivalent organelle is the chloroplast, which generates glucose from light, water and CO₂ by photosyn-

thesis. This reaction not only provides all the food we eat, but also produces oxygen as a 'waste product', without which we could not live.

Non-membranous organelles

Non-membranous organelles are cellular structures not bound by membranes. These include ribosomes and lipid droplets, which are composed of triacylglycerols, seen in fat-storing cells called adipocytes. In liver hepatocytes, large numbers of glycogen granules can be seen in the cytoplasm either as single particles or as aggregates called rosettes.

Cytoskeleton

All cells have an internal skeleton of protein fibres called the **cytoskeleton** (Table 2.9). These are made up of:

- Microfilaments
- Intermediate filaments
- Microtubules.

Microfilaments

Microfilaments are composed of twisted double strands of actin (Fig. 2.23A) and are the thinnest of the cytoskeletal fibres at 7 nm in diameter. They enable cells to move in an amoeboid fashion and change their shape. They form a mesh of fibres below the plasma membrane attached to proteins embedded in the membrane, which themselves connect with

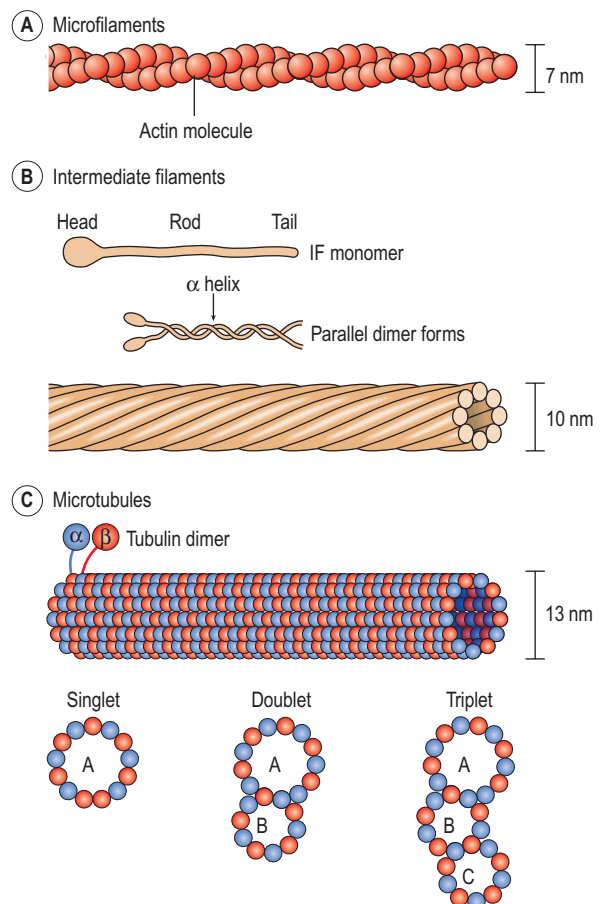


Fig. 2.23 Structure of cytoskeletal elements. IF, intermediate filament.

Information box 2.12 Mutations in mitochondrial DNA are inherited maternally

Any mutations that occur in mtDNA are passed from a mother to all her children. Because many of these mutations occur in the enzymes responsible for energy metabolism and impair ATP production they often result in an increase in anaerobic metabolism, with the production of lactic acid. This may cause cell death in tissues such as skeletal and cardiac muscle which are heavily dependent on ATP, causing myopathies.

molecules of the ECM. This enables a cell to fix its shape with respect to the external environment and, if necessary, move around in it. Cells move by extending either thin finger-like structures, called **filopodia**, or thin sheets, called **lamellipodia**, which contain large amounts of F-actin that is continuously added to in the direction of movement. The fibres are thought to be moved forward by the movement of myosin pushing the actin fibres against the plasma membrane. They are attached to long actin fibres, called stress fibres, which extend backwards. Microfilaments also form the core of the specialised plasma membrane structures, microvilli and stereocilium.

Intermediate filaments

Intermediate filaments (IFs) are made from a variety of proteins. However, they all form filaments 10 nm wide which are extremely stable. They all form α -helical dimers, which twist around one another forming rope-like structures (Fig. 2.23B), which give mechanical support to structures such as the nucleus and the plasma membrane. IFs do not change dynamically like microfilaments and microtubules, and so they form the most stable element of the cytoskeleton.

There are five classes of IFs, which are characteristic of each cell type. In epithelial cells, IFs made from keratin are anchored to structures called desmosomes, which link cells together, and hemidesmosomes, which link cells to the ECM. In nerve cells, neurofilaments form the core of axons which can stretch for long distances.

Microtubules

Microtubules are made up of two globular proteins, α - and β -tubulin, which form a dimer. These proteins are polymerised into a protofilament of alternating α - and β -tubulin. These are then arranged to form a tube from 13 protofilaments, which is 25 nm wide, called a singlet (Fig. 2.23C). Other arrangements involve adding extra protofilaments around the singlet to form doublets and triplets. The microtubule can be extended by the addition of further dimers to the '+' end of each protofilament, the end furthest away from the nucleus. Microtubules can be shortened from both ends by disassembly, which, using electron microscopy, looks as if the tube is being frayed. The growth of microtubules requires GTP, which is hydrolysed to GDP when the dimer of tubulin is incorporated into the growing protofilaments.

A number of different proteins are attached to microtubules, called **microtubule-associated proteins** (MAPs). One type of MAP cross-links microtubules in the cytoplasm, or binds to IFs. One of the MAPs found in nerve cells is called **tau**. A model of Alzheimer's disease suggests that accumulation and cross-linking of tau proteins can lead to the formation of the neurofibrillary tangles which are found in the brains of many people with Alzheimer's.

Transport along microtubules

Inside the cell, organelles and proteins are moved to specific locations. Microtubules are used as tracks along which organelles can be moved. For example, in nerve cells, the distance between the nucleus and the synapse, the point at the end of the axon where chemicals are released, can be several metres. The synapse does not contain the ribosomes required for protein synthesis, and in any case the DNA is in the distant nucleus. Proteins and membrane components

required in the synapse are manufactured in and near the nucleus in the cell body and must be transported along microtubules to the synapse. This process is called **fast axonal transport**.

Transport occurs in both directions along the same microtubule and is rapid (1–2 $\mu\text{m/s}$):

- **Anterograde transport** carries new material from the cell body to the synapse.
- **Retrograde transport** carries unwanted materials back to the cell body for destruction by lysosomes.

Anterograde transport

The anterograde transport, from the – end to the + end of each microtubule, is carried out by proteins called **kinesins**. These consist of four peptide chains: two heavy chains that bind to tubulin at their head end, and which are then coiled in an α -helix until they reach the tail end, where there are two light chains which bind to the transported substance. The force required to move material is provided by the heavy chains which, like the muscle protein myosin, change their shape on the hydrolysis of ATP.

Retrograde transport

Another type of motor protein, **dynein**, is responsible for retrograde transport. Dyneins have a different structure from kinesins but they are similar in that a globular head region binds to the microtubules and moves with the consumption of ATP. In nerve cells, microtubules do not extend all the way to the synapse; they stop at the end of the axon, where they are replaced by microfilaments. Materials transported along the microtubules are taken the rest of the way along these microfilaments, the movement being powered by a type of myosin and ATP. Other, slower transport may involve IFs.

Microtubule structures

Three distinct structures are formed from microtubules:

- Centrosomes
- Cilia
- Flagella.

Centrosomes

Most cells contain a **centrosome** which is found near the nucleus. This is the organising centre for microtubules, which radiate from here to the plasma membrane. The centrosome contains two **centrioles**, each of which is composed of a cylinder made up of nine microtubule triplets. This arrangement is called a 9 + 0 array. Centrioles are essential to the movement of the chromosomes to the opposite ends of a cell during cell division.

Cilia and flagella

Cilia and **flagella** (singular: cilium and flagellum) are mobile projections from the plasma membrane that are also composed of microtubules. The arrangement is slightly different from that found in centrioles, as the **core** or **axoneme** is formed from nine microtubule doublets, arranged around a doublet running down the central axis, in what is called a 9 + 2 array. Just underneath the plasma membrane, cilia and flagella are anchored to the **basal body**, which is identical in structure to the centrioles. The doublets of the cilia and flagella merge into the triplets of the basal body. The entire structure is covered in extensions of the plasma membrane. Inner and outer arms attached to the nine doublets are

Clinical box 2.8 Cigarette smoke stops the mucociliary escalator

The cilia which move mucus and trapped particles out of the lungs are inhibited and eventually destroyed by the inhalation of cigarette smoke. When this occurs, the only way that mucus can be removed from the lungs is by coughing. Irritants in the smoke cause chronic inflammation of the airways leading to excess production of thick mucus. This narrows the airways and traps pathogens, which can cause infection. Chronic bronchitis, along with emphysema, are two pathologies found in chronic obstructive pulmonary disease (COPD), a serious respiratory disease which occurs almost exclusively in smokers. See also Chapter 5.

Clinical box 2.9 Cardiac muscle is very sensitive to changes in K^+ concentration

Changes in the extracellular K^+ concentration alter the resting membrane potential of cardiac myocytes, which can change their excitability. Normal extracellular K^+ levels are 3.5–5.0 mM but if levels fall below this then the myocytes will become hyperpolarised and cardiac excitation is reduced. If the K^+ concentration rises above 5.5 mM, at first the cardiac excitation increases, which can cause arrhythmias and may lead to heart attacks (ventricular fibrillation). However, if K^+ levels increase further, then heart block may occur. Levels of K^+ above 7 mM can cause complete heart block, where the heart stops completely (see Ch. 11).

dyneins, which move the cilia and flagella using energy derived from ATP.

Cilia are hair-like projections up to 10 μm long, whereas flagella are much longer and occur individually or in pairs. Many cells are covered with cilia, which serve to move fluids or particles across their surface. There may be up to 300 cilia on each cell and they move in a coordinated manner. In the lungs, the cells which cover the larger airways are covered in cilia, which move mucus and particulates out of the lungs – the **mucociliary escalator**.

In single-celled organisms, cilia and flagella are the means by which the organism propels itself through the medium in which it lives. The only human cells that have flagella are sperm. Each sperm moves by beating the flagellum that forms its tail. Many bacteria also have flagella but these are completely different in structure, consisting of a protein strand that is rotated in its basal body, using energy derived from proton gradients.

INTRACELLULAR AND EXTRACELLULAR FLUIDS

All cells contain fluid and are bathed in fluid, but the composition of the intracellular and extracellular fluids is very different (Table 2.15). The **extracellular fluid** resembles a dilute version of seawater, possibly reflecting our evolutionary origins as single-celled marine organisms. It is rich in sodium and chloride ions, with quite high levels of calcium but very little protein compared with intracellular fluids. The **intracellular fluid** has a high concentration of potassium and a large number of intracellular anions such as phosphates and protein side chains, which balance the positively charged potassium ions. It also has a relatively high concentration of

protein. Free calcium levels inside cells are very low, reflecting its role as an intracellular messenger.

Ion gradients

While a pure lipid bilayer is virtually impermeable to ions, there are a number of transport processes acting across cell membranes which not only allow the dissipation of the ion gradient but also use the energy contained within it. The concentration gradients of Na^+ and K^+ across the plasma membrane are maintained by the activity of a special transport protein called the Na^+/K^+ -ATPase, which transports sodium ions out of the cell and potassium into it. Without the activity of this pump the ion gradients would disappear and many cellular processes would cease to function.

Resting membrane potential

The differential distribution of sodium and potassium cations, and the fact that cells are more permeable (at rest) to potassium than sodium, leads to the establishment of an electrical potential. This is called the **resting membrane potential** which acts across the cell membrane. In most cells it is of the order of -60 mV (inside negative), but is higher in **excitable cells**, such as neurons and muscle (see Ch. 8). Variations in this electrical potential underlie the electrical signalling carried out by excitable cells, which is entirely dependent on the ion gradients of sodium and potassium. The resting membrane potential is very sensitive to the external potassium concentration, which must always be closely controlled otherwise the electrical signalling will be disrupted (Clinical box 2.9).

Osmotic pressure**Osmolarity**

Osmosis is the movement of water by diffusion through a semi-permeable membrane from a region of high water concentration to a lower concentration. If two solutions of differing solute concentration are separated by a semi-permeable membrane water will tend to move from the lower solute concentration to the higher one until the solutions have equal concentrations of both water and solute. The solution with the higher solute concentration is said to have a higher **osmotic potential**.

Water is a polar molecule and pure phospholipid bilayers, which are non-polar, only allow limited water movement across them. However, in many cells, particularly in erythrocytes and water-absorbing cells of the collecting ducts of the kidney, there are large numbers of specific water channels, formed by proteins called **aquaporins**. Using these channels

Table 2.15 Composition of intracellular and extracellular fluids

Ion	Extracellular concentration (mM)	Intracellular concentration (mM)
Na^+	145	12
K^+	4	139
Cl^-	116	4
Bicarbonate, HCO_3^-	29	12
Phosphates	2	13
Protein anions	9	138
Ca^{2+}	1.8	<0.0002
Mg^{2+}	1.5	0.8

Clinical box 2.10 Intravenous fluid must have the right osmolarity

When administering intravenous fluids it is important to match not only the correct ionic concentrations, but also the correct osmotic potential. The osmolarity of body fluids is about 285 mOsm/L. If an isotonic sodium chloride solution is added to the extracellular fluid, because the sodium and chloride remain outside the cells, there is no osmotic movement of water into the cells. However, if a hypertonic sodium chloride solution is administered, water is drawn out of the cells, further increasing the extracellular fluid volume, with an increase in the osmolarity of both compartments. Severe hypotension is often treated by a rapid infusion of hypertonic solutions, increasing the circulating plasma volume. The inverse occurs with a hypotonic solution. There is an increase in the intracellular volume as water diffuses into the more concentrated intracellular compartment and both compartments are enlarged, with a reduction in osmolarity.

water moves freely across these plasma membranes. Therefore the internal and external media must have the same osmotic potential (they must be **isotonic**), otherwise cells would shrink or swell and burst. If cells, for example erythrocytes, are put in a dilute solution which has a lower osmotic potential (it is hypotonic) than normal plasma, then water will move into the cells and they will swell and eventually break open (lysis). Equally, if they are put into a more concentrated hypertonic solution then water would move out of the cells and they would shrink. The osmotic pressure of the intracellular and extracellular fluids is normally the same at about 285 milliosmoles per litre (mOsm/L (see Ch. 14)).

Vascular compartment

The **intravascular** compartment, which contains the blood, is separated from the **extracellular** fluid surrounding most tissues by a barrier, which in the smallest blood vessels, the capillaries, consists of endothelial cells. Although in some tissues, such as the liver, this barrier is very permeable to both water and solutes, in most tissues there is a significant difference between the intravascular and extracellular fluid composition. Although they both contain sodium as their predominant cation, the blood also contains significant amounts of protein, mainly in the form of albumin. This means that blood has a higher osmotic potential than extracellular fluid, which tends to draw water out of the interstitial spaces between the cells. This does not mean that all the fluid around cells moves into the blood. The blood has a significant hydrostatic pressure and this tends to push water out of the blood and into the interstitium. While this varies around the body, the net effect is that the movement of water into and out of the vasculature is balanced (see Clinical box 2.11 and Ch. 11).

PROTEIN SYNTHESIS AND PROCESSING

The central dogma of molecular biology is that, briefly stated, 'DNA makes RNA makes protein' and not the other way round. That is, you cannot make protein from protein or DNA from RNA. This holds true for all cellular organisms. Only in some viruses, which are obligate parasites, is this rule broken (see Information box 2.14).

Clinical box 2.11 Low albumin levels in the blood lead to oedema

In liver disease, the levels of albumin in the blood may be lowered as the liver loses its ability to make enough protein. Equally, in the kidney, if there is an increased permeability that allows proteins to be lost in the urine (they are normally retained), then the concentration of albumin in the blood will fall. If this happens then the osmotic potential of the blood will be reduced and water will tend to accumulate in the extravascular compartment. This causes the tissues to become puffy and they will indent when pressed. This is more evident in the lower limbs as water accumulates there under the force of gravity. Called 'pitting oedema', this is a diagnostic sign of right heart failure, as well as of some liver and kidney disorders.

The synthesis of proteins involves two processes which convert the base sequence of the appropriate DNA into a polypeptide chain.

1. Transcription: the conversion of a specific section of DNA into a complementary strand of mRNA. This transcription occurs in the nucleus after which the mRNA is processed and translocated to the cytoplasm where it is attached to a ribosome.
2. Translation: the strand of mRNA is used as a template to assemble the sequence of amino acids which make up the protein. This involves converting the code within the bases of the mRNA into amino acid residues and each of the 20 possible amino acids is coded for by three bases in what is called the genetic code. Both tRNA and rRNA are involved in this process.

GENES AND THE GENOME

The entire complement of double-stranded DNA within a cell is called the **genome**, but only about 10% of the DNA in mammals is thought to carry information in the form of sequences that are expressed. The sequences of bases within the long DNA molecules are of two different types:

- Those which only occur in single or a few copies, which are the **genes**
- Regions of repeated sequences, which may be repeated millions of times within the genome.

Some of these repeated sequences code for proteins which are needed in large amounts, such as histones. But the function of much of this DNA is unknown, hence the term '**junk**' DNA. Some of this DNA may be 'old' genes which have undergone mutations and are no longer transcribed. However, it has also been suggested that this DNA may act as a reservoir for the evolution of new genes.

Genome size

The size of the genome varies enormously between organisms; the well-studied bacterium *Escherichia coli* has about 5 million base pairs, compared to the 3 trillion base pairs of the human genome. In humans there are estimated to be about 30 000 genes, each of which contains the information for a gene product. Most gene products are proteins; the only other gene products are the RNA molecules that form rRNA and tRNA.

Table 2.16 The genetic code (RNA to amino acids)

First position	Second position				Third position
	U	C	A	G	
U	Phe	Ser	Tyr	Cys	U
	Phe	Ser	Tyr	Cys	C
	Leu	Ser	Stop	Stop	A
	Leu	Ser	Stop	Trp	G
C	Leu	Pro	His	Arg	U
	Leu	Pro	His	Arg	C
	Leu	Pro	Gln	Arg	A
	Leu	Pro	Gln	Arg	G
A	Ile	Thr	Asn	Ser	U
	Ile	Thr	Asn	Ser	C
	Ile	Thr	Lys	Arg	A
	Met (start)	Thr	Lys	Arg	G
G	Val	Ala	Asp	Gly	U
	Val	Ala	Asp	Gly	C
	Val	Ala	Glu	Gly	A
	Val (start)	Ala	Glu	Gly	G

The code given here translates mRNA codons into amino acids; the original sequence of DNA would be complementary to the mRNA and would contain thymine (T) instead of uracil (U).

THE GENETIC CODE

DNA is only made up of a sequence of four different nucleotides (A, G, C and T) which are used to code for 20 amino acids. Obviously this could not be achieved by one-to-one coding as, with only four different nucleotides, it would only be possible to specify four different amino acids. If each amino acid were coded by two nucleotides, there would still only be 16 (4^2) different possible combinations. The way in which DNA codes for different amino acids uses sets of three nucleotides to code for each amino acid giving 64 (4^3) possible combinations. Each group of three nucleotides is called a **codon**. The genetic code is usually shown as a table showing the sequence of three nucleotides in the mRNA codons and the corresponding amino acid (Table 2.16). It can be seen from the table that not all of the codons represent an amino acid. There are three codons that cause the protein chain to terminate, called **stop codons**, and one codon (AUG) that codes for methionine but also acts as a **start codon** to initiate the protein synthesis. The codon GUG usually codes for valine, but occasionally also acts as a start codon.

All 64 possible codons are used, many amino acids being coded for by more than one possible codon, but this is not evenly spread. Some amino acids are represented by six different codons, whereas others only have a single one. The code is said to be 'degenerate'. The code is virtually the same in all organisms and this universality is used as an argument that life on Earth evolved once only. However, there are minor differences in the genetic code in mitochondria, in some protozoans and in single-celled plants. These minor changes are thought to be later mutations.

Reading frame

In most cases the code has to be read, in threes, starting at the AUG codon, to one of the stop codons. This is known as the **reading frame**, and most mRNAs are read within a single frame, although in some cases it has been found that by starting at a different AUG codon the same mRNA can produce a different protein. Also some unusual RNAs contain sites where either four or two bases are read and then the remaining RNA is read with a **frame shift**.

Information box 2.13 Deletion of a single base from DNA will frame shift the protein

If a single nucleotide is removed from a gene then the resulting protein will be very different. The protein chain will be normal up to the point of the deletion. After this all the resulting codons will be read out of phase. This results in different amino acids being substituted and the frame shift will continue adding the wrong amino acids until one of the 'new' codons corresponds to a stop codon. If the deletion is near the end of the coding sequence, then the protein may be nearly normal, but if it is close to the beginning then changes can be catastrophic. Normal haemoglobin is composed of two polypeptide chains, α - and β -globin, folded to hold the haem molecule. β -thalassaemia is an inherited anaemia characterised by a point mutation in the β -globin chain. The resulting frame shift produces a protein that contains a series of unexpected amino acids, rendering the chain unstable and the molecule unbalanced. Precipitation of the globin chains results in red cells that are vulnerable to haemolysis, producing a severe anaemia in those individuals who have inherited the abnormality from both parents.

TRANSCRIPTION

Transcription is started by the binding of the enzyme **RNA polymerase** (RP) to the appropriate section of the appropriate strand of the DNA (Fig. 2.24). RNA polymerases can start a new RNA strand (initiation) and extend it (elongation). In mammals, there are three types of RNA polymerase (RPI, RPII, RPIII), which make pre-rRNA, pre-mRNA and pre-tRNA, respectively.

As stated previously, the two strands of DNA are anti-parallel and complementary. Starting at the same end of each strand would yield two different sequences of bases, so the RP must bind to the correct strand. The strand serving as a template is called the **antisense** strand. The other strand, the coding strand, is not usually transcribed and is called the **sense** strand.

The sequence of bases represented by the sequence of the mRNA is the same as the original coding strand of the

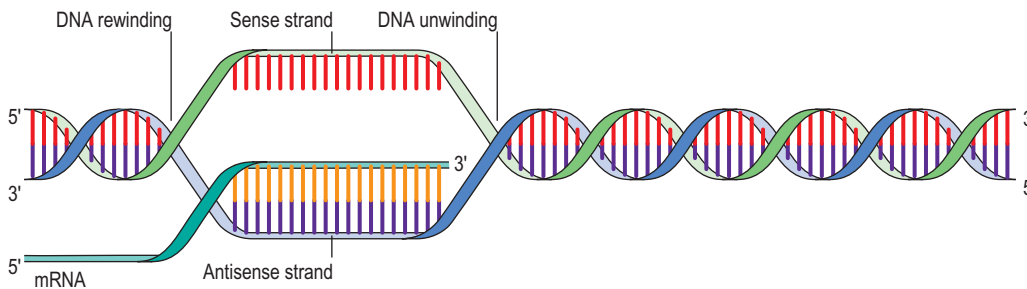


Fig. 2.24 Transcription of DNA into RNA.

Table 2.17 Complementary DNA and mRNA sequences

Sequence	Direction	Codons or amino acids
DNA coding strand (sense)	5'→3'	ATG AGA CTA TTC AGC TAA
Complementary DNA (antisense)	3'→5'	TAC TCT GAT AAG TCG ATT
mRNA	5'→3'	AUG AGA CUA UUC AGC UAA
Amino acids	N → C	Met Arg Leu Phe Ser (start)

Spaces are shown between the codons for clarity.

DNA, except that mRNA contains uracil instead of thymine. This is because a complementary copy of a complementary copy produces the original sequence. This can be followed through step by step as shown in Table 2.17. It used to be thought that only the antisense strand was used; however, in very limited sections of DNA, there are proteins encoded on both strands.

Starting transcription

The antisense strand is read from its 3' end to the 5' end, and the RNA chain grows from the 5' end to the 3' end. Each gene has a region called the initiation site, where the RP must bind to start transcription. In many genes, upstream (25–35 bp) from the transcriptional start site is a sequence called the **TATA box**, which contains the sequence TATA(A or T)A, that binds transcription factors that initiate transcription. Other regions contain areas called **promoters**, which are sequences that control the rate at which the gene will be transcribed. They are located both upstream and downstream from the gene to be transcribed and bind molecules called **trans-acting factors**, which can either enhance gene transcription or silence it. Many slow-acting hormones, such as the steroid hormones and thyroid hormones, change the rate of gene expression, causing changes in the rate of transcription of specific mRNAs by forming a ligand–receptor complex that binds to promoter regions of DNA.

Binding of RNA polymerase

The binding of RP produces separation of the two strands of DNA in the region of the gene. This separation occurs more easily between T-A pairs than G-C pairs, because T-A pairs are joined by two hydrogen bonds whereas G-C pairs have three. The weaker links of the TATA box may help the DNA to unwind more easily. The first exposed base then binds the complementary free nucleotide from the cytoplasm. The first nucleotide of the RNA chain at the 5' end is different from all

the others because, being at the beginning of the chain, it remains a triphosphate. The second base of the DNA then binds the corresponding free nucleotide, which is then linked to the first by the RP enzyme. As each nucleotide is added to the chain, a pyrophosphate is lost, leaving only the single α -phosphate in the backbone of the growing chain. Travelling along the DNA in the 3' to 5' direction the RP adds each successive complementary base to the growing RNA chain. As the RP proceeds, the DNA unwinds, while areas that have already transcribed will rewind.

Transcription can proceed quite rapidly at about 40 bases/s. At this rate a very large gene can take many hours to transcribe. The gene which codes for dystrophin, the protein that is absent in Duchenne muscular dystrophy, is more than 2×10^6 bp and would take more than 13 hours to transcribe. However, most gene transcripts are considerably smaller. The error rate in copying RNA is about 1 error every 10^4 – 10^5 nucleotides. This is high compared to the error rate in DNA replication but, because RNA is a short-lived molecule and is not repeatedly copied, this rate of error is not significant.

Termination of transcription

After the gene has been transcribed, elongation terminates and the RP is dislodged from the DNA strand. How this occurs in eukaryotes is not yet understood, although it may involve the formation of hairpin-like secondary structures in the mRNA that destabilise the RP. Termination often occurs up to 2000 bp past the end of the last exon (see below) of the gene. The excess mRNA will be ignored, because nothing after the stop codon is translated. The mRNA has a number of nuclear proteins associated with it, which are necessary for the transport out of the nucleus.

Post-transcriptional processing

All eukaryotic mRNAs are modified after they have been transcribed and before they are translated into protein in the cytosol. While the rest of the gene is being transcribed, and after only about 35 nucleotides have been added, the **primary transcript** of mRNA is capped at the 5' end by the addition of 7-methylguanosine to the 5' triphosphate group via an unusual 5'-5' linkage. The two 5' end nucleotides may also be 2'-methylated. This capping is thought to have a number of functions that aid further processing and protect the mRNA from destruction by nuclease enzymes.

Poly-A tails

Next, the 3' end is cleaved at a region called the poly-A site at the sequence AAUAAA (or rarely AUUAAA). This 3' end of

the mRNA then has a sequence of up to 300 adenosine (A) nucleotides added, hence the term poly-A tail. These adenosine nucleotides bind proteins to help prevent degradation of the mRNA. This polyadenylation does not occur in all RNAs; for example it is absent in histones. However, the poly-A tails make it easier to isolate mRNAs for experimental use by 'fishing' for them with molecular probes that consist of poly-T nucleotides.

RNA splicing: introns and exons

Most eukaryotic genes are much longer than would be predicted by the number of amino acids in the resulting protein. This is because these genes contain regions, called **exons** (short for expressed sequences), which code for the protein, interspersed with non-coding regions, called **introns**. Many genes have multiple introns, which are removed from the initial primary transcript of mRNA to produce the mature mRNA.

The final modification to the pre-mRNA involves the removal of the **introns** by a process called **splicing**. This is done by a protein–RNA complex called a **spliceosome**. The positions of the sections to be removed are indicated by special sequences in the introns which allow the correct positioning of the different proteins involved in removing the intron.

Splicing of the same pre-mRNA does not always yield the same mature mRNA. Exons may also be removed. This **alternative splicing** can produce mRNAs with different exons, which will give different proteins. This can be controlled in a tissue-specific manner. For example, the extracellular protein fibronectin, which is secreted by different cell types, has two exons that code for parts of the protein which interact strongly with cell surface receptors. These are included in the mRNA produced in fibroblasts but are removed in the mRNA from hepatocytes. This change results in a functional difference such that the hepatocyte fibronectin does not attach to cells very readily and circulates easily in the plasma.

RNA editing

This is a recently discovered process, which changes the sequence of an mRNA. This process is widespread among protozoans and plants, particularly in the chloroplast. In mammals, a few examples have been shown. For example, in a particular protein, by changing a single base in a specific position from A to G this changes the codon from CAG to CGG. This changes the amino acid inserted at that position from glutamine to arginine, with a subsequent change in the behaviour of the protein.

Processing rRNA and tRNA

The other gene products, rRNA and tRNA, are also processed before they leave the nucleus. In their mature form, tRNA molecules contain a number of modified bases. These modifications are important in determining the secondary structure of the tRNAs and the way in which they recognise the enzymes involved in protein synthesis. Ribose sugars are also modified by methylation, which makes them more stable.

Assembly of ribosomes

Ribosomal subunits are assembled in the nucleolus, where the rRNA is produced. Proteins are imported from the cytoplasm to be combined with the rRNA to make the small and large ribosomal subunits. The pre-rRNA is split into two

subunits from a single original transcript. This is a way of ensuring that the separate subunits are produced, as needed, in equal amounts.

TRANSLATION

After the mRNA has been transcribed and modified in the nucleus it is transported to the cytoplasm where the production of the proteins occurs by translation of the mRNA transcript.

Transport from the nucleus

It has been estimated that more than one million molecules pass in or out of the nucleus every minute through approximately 4000 pores. These include the mRNAs, as well as other molecules exported from the nucleus such as ribosomal subunits and tRNAs. Moving in the opposite direction are nuclear and ribosomal proteins.

Nuclear pores

After it has been processed in the nucleus, the mature mRNA is transported to the cytoplasm through one of the **nuclear pores**. The normal diameter of the pores is 9 nm, but they can open to a maximum of 25 nm. It is thought that mRNAs that are ready for export include a sequence designating them for transport to the cytoplasm. Additionally, nuclear proteins are associated with their 5' end, and these possibly act as a signal sequence to guide the RNA. The 5' end leaves the nucleus first, then the nuclear proteins are removed and returned to the nucleus for reuse.

tRNAs

The function of tRNA is to link a particular amino acid onto the acceptor stem (see Fig. 2.20) and to bind to mRNA, through base pairing of three complementary base pairs, the **anticodon**, which corresponds to the linked amino acid on the opposite loop, called the anticodon loop. Because of the way that tRNA is folded into a clover leaf, the acceptor stem contains both the 3' and 5' ends of the molecule. The amino acid is attached to the 5' end of the tRNA in a sequence of reactions involving the enzyme, aminoacyl-tRNA synthetase. Using ATP, the enzyme forms an AMP-amino acid, which is then transferred to the tRNA, releasing AMP. This reaction is referred to as **charging**, because the energy from the ATP bond is passed to the tRNA bond, and is later used to form the peptide bond, when the amino acid is incorporated into the protein chain. There are different synthetase enzymes for each of the amino acids; this ensures that each specific amino acid is charged with the correct tRNA.

Codon wobble

While there are 61 different codons that code for amino acids – remember that three of the 64 codons are stop codons – there are only about 50 different types of tRNA found in animal cells. This is also more than the number that would be required if only one tRNA was used for each amino acid. The reason for this anomaly lies in the fact that each tRNA is able to recognise more than one codon, a phenomenon called **wobble** (Table 2.18). This happens because the first base in the anticodon can recognise different bases in the third position on the mRNA. These non-standard base pairings are particularly common between G-U bases, which

Table 2.18 Codon/anticodon base pairings

Third base of the codon	First base of the anticodon
Normal base pairing	
A	U
G	C
U	A (rare)
C	G
Unusual base pairing	
G	U
U	G
A or C or U	I
I, inosine.	

are formed almost as easily as the normal G-C pairing. Thus the mRNA codons for phenylalanine, UUU and UUC, are both recognised by the tRNA anticodon GAA.

Ribosomes and protein synthesis

Ribosomes direct protein synthesis in the cytoplasm in a cycle that consists of three phases:

1. The **initiation phase**, when the ribosomal subunits combine with the mRNA and move to the start codon.
2. The **elongation phase**, when the mRNA is translated into protein.
3. The **termination phase**, when the protein is released and the ribosome dissociates.

Initiation phase

The first part of the **initiation phase** is when the small ribosomal subunit binds to the cap at the 5' end of the mRNA. It then travels along the mRNA until it reaches the start codon, AUG. This then binds Met-tRNA and the large ribosomal subunit then binds to form the complete ribosome.

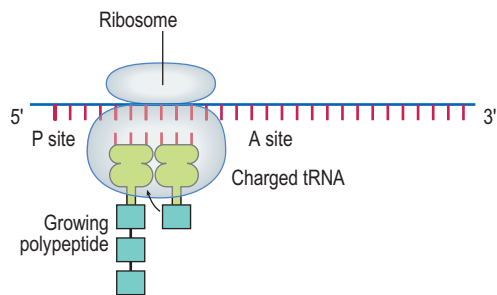
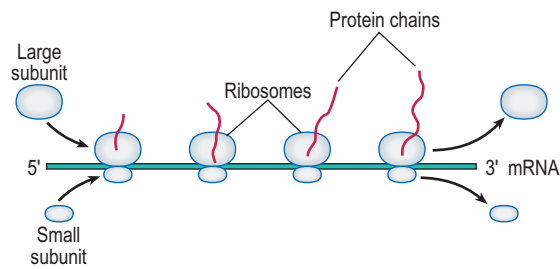
There are two tRNA binding sites on the ribosome, A and P (Fig. 2.25). The A site receives the tRNA carrying the next amino acid to be added to the chain. The P site holds the tRNA attached to the growing peptide chain.

Elongation phase

Elongation starts with the Met-tRNA bound to the P site and the ribosome assembled. The next codon of the mRNA is in the A site of the ribosome and is then base paired with the appropriate tRNA. A peptide bond is formed between the carboxyl group of the first amino acid, which becomes the N-terminal amino acid, and the amino group of the second. The ribosome moves one codon down the mRNA and the first tRNA is released to be recharged in the cytoplasm. The second tRNA is shifted to the P site of the ribosome and a new tRNA occupies the A site. The energy required to form the peptide bond comes from the tRNA-amino acid bond, but the movement of the ribosome and the movement of the tRNA from the A site to the P site requires the hydrolysis of GTP to GDP.

Termination phase

The process of elongation continues, with each step adding another amino acid to the peptide. When the A site is occupied by a stop codon the elongation ceases. The last step, which also requires GTP, removes the peptide chain from its attachment to the last tRNA and dissociates the ribosome into its subunits.

**Fig. 2.25** Translation of mRNA.**Fig. 2.26** Polysomes produce multiple copies of a protein from a single mRNA.

Polysomes

A given mRNA can be translated simultaneously by more than one ribosome. Once the first ribosome has moved far enough down the mRNA, the initiation site can be occupied by another ribosome. This gives rise to a structure called a **polysome**, which consists of a single mRNA molecule with many ribosomes attached to it, each with a lengthening peptide chain attached (Fig. 2.26).

Lifetime of mRNAs

A given mRNA can be translated many times until it is degraded. Enzymes in the cytoplasm gradually remove the A residues from the poly-A tail. When this becomes too short then the 5' cap is also removed. The mRNA is then degraded by exonucleases, which remove nucleotides from each end of the mRNA. Most mRNAs last a few hours, but some only last a few minutes and others last a few days. Many of the short-lived mRNAs code for proteins that are involved in the control of gene transcription. The half-life of these mRNAs – the time required for the concentration to halve – is determined partly by sequences in their genetic code, which allow them to be targeted for degradation, and partly by regulation depending on cellular requirements. An example of this is the mRNA coding for the milk protein, casein. In the presence of the hormone prolactin, which stimulates milk production, the rate of transcription of the casein gene rises about threefold. However, the amount of casein mRNA is increased 100-fold. This is due to a large increase in the stability of the mRNA, which in some cases is due to the binding of specific proteins to sequences in the 3' region after the stop codon.

Rate of protein synthesis

Protein synthesis is a major activity of cells; it is estimated that, in an average mammalian cell, about one million amino acids are being added to growing polypeptide chains each second. A single polypeptide grows at the rate of three to five amino acids per second. This implies that at any given time a cell is busy making some 200 000 proteins. Clearly,


Information box 2.14 Some viruses break the rules

In some viruses, the rule observed in cellular organisms, 'DNA makes RNA makes protein' is not followed. Viruses can have either DNA or RNA as their genetic material and they rely on the synthetic machinery of the host cell to make more copies. In many of the RNA viruses, DNA is not generally involved in that process as the virus directs the manufacture of more RNA particles, which also make the proteins necessary to make more viruses. There are, however, a group of RNA viruses, called retroviruses, which use a specific viral enzyme, reverse transcriptase, which is packaged in the virus along with the RNA, to make a DNA copy of the viral RNA. This DNA is then incorporated into the host cell genome, from where it can make more copies of the viral RNA and proteins.

The most well-known retrovirus is the human immunodeficiency virus (HIV). One of the consequences of this packaging of DNA is that the virus can 'hide' in the genome and remain undetected for years until activated and new virus particles are produced.

Knowledge of reverse transcriptase enzymes has revolutionised the study of molecular biology, enabling researchers to reverse engineer DNA from mRNAs, which are more easily accessible, allowing the identification of products.

See also Chapter 5.

small proteins of 100 to 200 amino acids can be made in less than a minute. However, many proteins are larger. The largest protein known, titin, which is found in muscles, has about 30 000 amino acids and, at the rate of five amino acids per second, it would take over 1.5 hours to synthesise each molecule.

PROCESSING PATHWAYS FOR PROTEINS

As described previously, almost all proteins are modified after they are translated. Many modifications are not permanent but serve to help the targeting of proteins to the appropriate destinations. As they are produced, proteins are folded into their final shapes. This is partly determined by their primary structure and partly by the presence of molecular chaperones, which aid folding.

Protein can be produced in three different sites:

1. Some proteins are synthesised entirely on 'free' ribosomes within the cytosol. These are the proteins that are either destined to remain in the cytosol or those which are used in the nucleus, mitochondria or peroxisomes.
2. Proteins destined for the other organelles (ER, Golgi, lysosomes), insertion into membranes (organelle or plasma), or for secretion, are synthesised on ribosomes on the rough ER.
3. Some mitochondrial proteins are synthesised on ribosomes within mitochondria themselves using mitochondrial DNA.

Regardless of their final destination, all proteins start their synthesis on cytoplasmic ribosomes. Their subsequent fate is determined by the presence (or not) of a signal sequence of 16–30 amino acids at the N-terminal end of the protein.

PROTEIN PRODUCTION IN THE CYTOSOL

Proteins for use in the nucleus are transported through the nuclear pores. These pores are large enough for the protein to be transported without being unfolded. Nuclear localisation sequences, often rich in positively charged lysine and arginine residues, interact with the nuclear pores to facilitate uptake.

Those mitochondrial proteins not made in mitochondria are manufactured on cytosolic ribosomes. The proteins are targeted by N-terminal signal sequences, which not only determine that the protein will be transported into the mitochondria but also specify whether it will end up in the outer or inner membrane, or in the matrix or intermembrane space. Mitochondrial proteins have to be transported through the membrane unfolded. They are unfolded by molecular chaperones, transported through the membrane/s, and then refolded using other chaperones inside the mitochondria.

The enzymes within peroxisomes are made in the cytosol and imported into the organelles. They are thought to bind to receptors on the surface of the peroxisomes before entering through a transport protein. Rare disorders such as Zellweger's syndrome, which is characterised by central nervous system, liver and renal abnormalities, seems to involve a defect in protein import into peroxisomes.

Protein production in the ER

For proteins destined for insertion into the rough ER, as the signal sequence emerges from the free ribosome it is bound to a complex of protein and RNA, called a **signal recognition particle** (SRP). This inhibits further elongation of the protein until the ribosome reaches the ER. The cytoplasmic surface of the ER has receptors that bind the SRP and stimulate the formation of a transmembrane channel, through which the signal sequence can pass. The SRP then dissociates and is free to bind another signal sequence. The peptide continues to grow, but with the N-terminal end being extruded into the lumen of the ER. The signal sequence is then cleaved by an enzyme in the ER and degraded. As the protein grows, it is modified by the addition of carbohydrate and other residues, folded using chaperones, and when the C-terminal is reached it is released into the lumen of the ER. The ribosome is released from the rough ER and dissociates, as does the transmembrane channel. Many of the varying signal sequences contain a series of hydrophobic amino acids, particularly leucine, which help the signal sequence to cross the membrane. When these are cleaved, the resulting protein is more hydrophilic and will be retained within the ER.

Transmembrane proteins

Proteins that are destined to be part of a membrane are inserted into the membrane as it is being made. For a simple protein with only one transmembrane segment (one end of the protein inside the cell and one end outside, with the middle section spanning the membrane), this is achieved due to the presence of a second signal sequence. The first part of the protein, the N-terminal end, is made in the same way as secreted proteins. This will eventually become the external part of the protein, and it has the signal sequence removed and is glycosylated as normal. However, when this part reaches its appropriate size, there is a second signal sequence that consists of about 22 hydrophobic amino



Information box 2.15 One type of cystic fibrosis is due to a protein not being processed correctly

The most common form of cystic fibrosis is caused by a deletion of a single amino acid (a phenylalanine at position 508) in the gene that codes for the membrane protein, CFTR (cystic fibrosis transmembrane regulator). This results in the ER being unable to export the CFTR protein which is needed to produce the proper ion channels needed for cAMP-regulated chloride transport across the plasma membrane. At normal body temperature, there is an error in folding of the protein which prevents its export from the ER. However, curiously, it was found that at lower temperatures (25°C) the proteins fold normally.

These channels are found particularly in the specialised linings of the bowel and lung. Defective ion transport results in salt imbalance, drawing water from the airways and producing the thick lung secretions typical of this disease. Defective ion transport explains the high concentration of salt ions in sweat secretions in patients with CF (the 'sweat' test).

See also Chapter 14.

acids. This forms an α -helix, which spans the membrane and is anchored in it. This signal is called the stop-transfer membrane-anchor sequence. The transmembrane channel is disassembled leaving the ribosome to finish the C-terminal end of the protein on the cytosolic side of the ER.

Protein processing in the ER

In the ER, a number of important steps take place:

- Multimeric proteins are formed into oligomers.
- Mutated proteins which misfold are identified and retained.
- Unwanted excess protein subunits are degraded.
- Asparagine residues undergo *N*-linked glycosylation.
- Precursors of the lysosomal enzymes, prohydrolases, are glycosylated.

Processing in the Golgi

Proteins move from the ER to the Golgi (Fig. 2.27) where they are processed further. In the Golgi, as in the ER, there is a wide range of modifications which occur before proteins can be sent to their final destinations.

Glycosylation

Many of the *N*-linked sugars that are added in the ER are modified, and other sugars are added by *O*-linked glycosylation to serine and threonine residues, and then further modified. Many proteins, especially peptide hormones, are converted from pro-hormones to their active forms by cleavage of peptides by proteolytic enzymes. Sphingolipids are manufactured in the Golgi and glycosylated.

Phosphorylation

The phosphorylation of glycosylated lysosomal prohydrolase enzymes generates terminal mannose-6-phosphate groups, which are then used to target them to **lysosomes**. These mannose-6-phosphate groups bind to receptors on the internal face of the *trans*-Golgi. These are clustered into an area of membrane that is coated on the cytoplasmic side by the protein, clathrin. The areas of

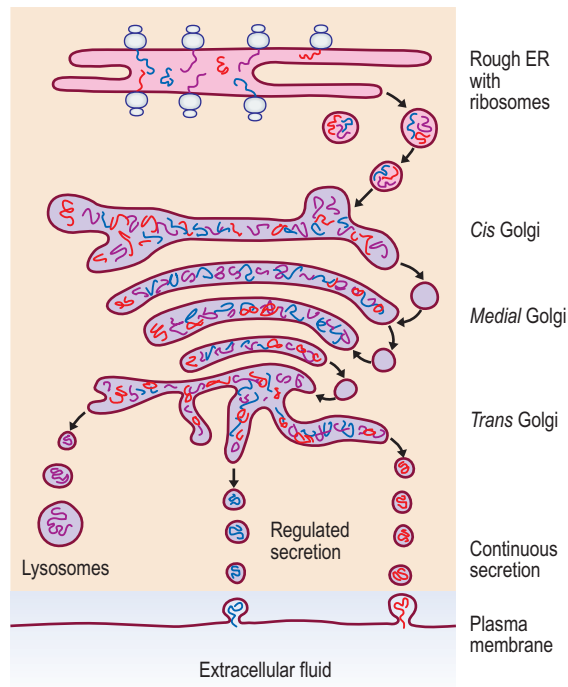


Fig. 2.27 Destinations of proteins.

membrane coated with clathrin bud off from the Golgi, after which the clathrin is removed. After a number of complicated steps, involving other vesicles called **late endosomes**, a vesicle containing the dephosphorylated enzyme fuses with the lysosome, and another vesicle containing the empty mannose receptors returns to the Golgi.

Secretory proteins

Proteins in the Golgi that are destined for secretion do not have mannose-6-phosphate and are not targeted to lysosomes. However, they are separated according to whether or not they are continuously secreted, or their secretion is regulated (Fig. 2.27). Regulated secretion requires that vesicles are stored in the cytoplasm and attached to the cytoskeleton before they are secreted. A large number of proteins have been identified that carry out the necessary docking, priming and fusion. Research carried out in the secretion of neurotransmitters has helped determine these mechanisms (see Ch. 8).

MEMBRANE STRUCTURE AND FUNCTIONS

FLUID MOSAIC MODEL OF MEMBRANES

A typical cell membrane has hundreds of different types of protein, which may be pictured as floating in the phospholipid sea. This model of proteins embedded in a lipid bilayer, some of them free to diffuse and some fixed, was originally proposed by Singer and Nicholson in 1972 and is called the **fluid mosaic model** (Fig. 2.28).

It is largely the proteins in a cell membrane which determine its biological function. So, for example, a given cell responds to a particular chemical signal because its plasma membrane contains a specific protein. Any cell lacking the particular protein in its membrane would be completely

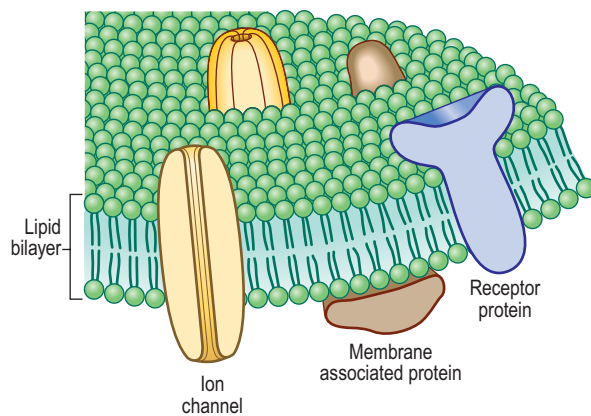


Fig. 2.28 Fluid mosaic model of membrane structure, showing integral and membrane-associated proteins.

unresponsive to the signal. The functions of cell membrane proteins include:

- Enzymes
- Transporters
- Transmembrane channels
- Receptors
- Cell markers
- Linking the cytoskeleton and the ECM.

Lipid bilayer

All cell membranes, including the plasma membrane, consist in part of a **bilayer** of phospholipids, sphingolipids and cholesterol. Phospholipids are all **amphipathic** molecules because they have two acyl groups that are hydrophobic, and a substituted phosphate, which is hydrophilic because it is polar. Similarly, sphingolipids and cholesterol both have hydrophobic and hydrophilic regions. Because of this property, in an aqueous environment phospholipids aggregate so that the acyl tails are as far away from the water as possible. One configuration that permits this is a bilayer.

Movement of lipids in the bilayer

Of itself, a phospholipid bilayer is remarkably dynamic. Individual phospholipid molecules are free to rotate rapidly around their long axis and migrate rapidly within their side of the bilayer. However, transfer of a phospholipid between the sides of the bilayer (flipping) is very rare. Overall, a phospholipid bilayer has the fluidity of olive oil, and it is a highly impenetrable barrier to most materials. It will, however, allow the passage of small molecules, such as water, oxygen and carbon dioxide, and anything that is fat soluble, but it will exclude almost all polar molecules or charged solutes.

Differential distribution of lipids in the bilayer

Different membranes have different lipid compositions, including the two leaflets of the same bilayer. An example of this is the red cell plasma membrane, which has sphingomyelin and phosphatidylcholine as the major components of the outer face of the membrane (called the **exoplasmic leaflet**), and phosphatidylethanolamine and phosphatidylserine as the main lipids in the inner side (the **cytoplasmic leaflet**). The difficulty with which lipids flip between the two leaflets helps to maintain this differential distribution. Phospholipids can be moved between organelles by exchange proteins. Proteins called **flippases** are found in some membranes,

particularly in the ER and in erythrocytes. These can move phospholipids from one side of the bilayer to the other.

Cholesterol in membranes

At normal human body temperature the presence of cholesterol reduces the fluidity of the membrane. It inserts itself between the phospholipids with its hydrophilic hydroxyl group facing outwards where the planar structure of cholesterol reduces the movement of phospholipids. The plasma membrane of cells has more cholesterol than the intracellular membranes, which makes it more rigid, less permeable to small molecules and thicker than intracellular membranes. Some intracellular membranes such as the ER contain very little cholesterol.

Membrane proteins

The second major component of membranes is protein. Membrane proteins are of two main types:

- Integral (intrinsic)
- Peripheral (extrinsic).

Integral membrane proteins can be removed only by disrupting the membrane structure. Many of these proteins span the bilayer and have hydrophobic amino acid residues on the part of the protein in contact with the membrane, while the parts inside and outside the cell are hydrophilic and may also carry charged groups. These charged groups are often glycosylations in the extracellular face, and phosphate groups on the intracellular face.

Peripheral membrane proteins, in contrast, may be removed without disrupting the membrane. They may be associated with the cytoskeleton, or involved in signal transduction across the membrane, or they may be attached to the ECM.

Protein composition of different membranes

The amount of protein varies enormously between different membranes. The erythrocyte plasma membrane is about half lipid and half protein with about 8% carbohydrate. In contrast, the myelin membrane, which surrounds nerve cell axons and acts as an insulator, is almost 80% lipid. The inner mitochondrial membrane, which contains all the enzymes associated with metabolism, is over 75% protein, with no carbohydrate.

SPECIALISED PLASMA MEMBRANE STRUCTURES

Many cells, particularly those that make up the solid tissues of the body, are polarised. That is, different functions are observed in different parts of the cell. Given the fluidity of the plasma membrane, and the ability of membrane lipids and proteins to move laterally within the membrane, in order for cells to have polarity there needs to be a way of separating the different areas of membrane. Cells adhere to one another through interactions of cell adhesion molecules (see below), but cells that form sheets, separating different body compartments, must have structures which link them together with sufficient strength to maintain the integrity of the layer.

Sheets of epithelial cells

Epithelial cells form layers covering most body surfaces. A typical polarised epithelial cell has an apical surface in

contact with the lumen (e.g. the intestinal lumen). This may have one of a number of membrane specialisations (Table 2.19). The basolateral surface is in contact with the basement membrane, which consists of the basal lamina and the reticular lamina. The proteins in the plasma membrane are restricted to the apical and basolateral domains by junctions between the cells (Fig. 2.29).

Junctions between cells

There are a number of different types of junctions between cells, and between cells and the ECM (Fig. 2.29). Many sheets of cells sit on a fibrous network of collagens, proteoglycans, laminin and fibronectin called the basal lamina.

These junctions can be classified as:

- **Tight junctions** – ones that prevent movement of substances
- **Adhering junctions** – ones that maintain cellular positions
- **Gap junctions** – ones that allow movement of substances.

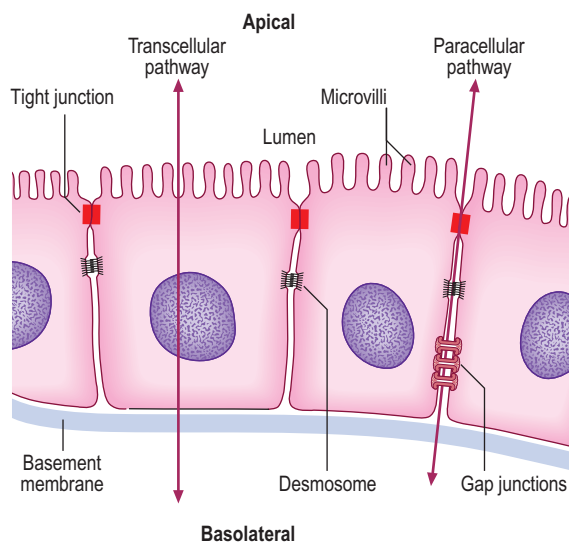


Fig. 2.29 A sheet of epithelial cells.

Specialisation	Function	Example
Cilia	Movement across surface	Bronchial epithelium
Microvilli	Increased area for absorption	Intestinal epithelium
Stereocilium	Sperm maturation	Epididymal epithelium

Table 2.20 Different types of adhering junctions

Adhering junction	Alternative name	Interaction	Linker
Belt desmosome	Zonula adherens	Actin–actin	Cadherins
Spot desmosome	Macula adherens	IFs–IFs	Cadherins
Hemidesmosome	–	IFs–basal lamina	Integrin and laminin

IF, intermediate filament.

Tight junctions

Tight junctions, also known as **occluding junctions**, form a belt around the cell attached to the neighbouring cells known as the zonula occludens. The zonula occludens is formed from a complex of proteins (ZO1-3, AF-6 and occludin) that connect the two membranes tightly together in a series of studs appearing as interconnected ridges with no visible extracellular space. Tight junctions are not connected to the cytoskeleton and prevent the movement of membrane components between the separated regions of the cell. This mechanism allows each part to have its own set of proteins and lipids. Tight junctions also prevent the passage of molecules across the sheet of cells through the route passing between the cells, the paracellular pathway. This barrier is produced by a family of proteins called claudins.

Adhering junctions

Three types of **adhering junctions**, also known as **anchoring junctions**, are all different types of **desmosomes**. These are formed from plaques just below the plasma membrane, bound to cytoskeletal elements. The plaque is attached to the plaque of another cell, or to the ECM, by proteins which span the intercellular space (Table 2.20). Zonula adherens form a belt around the cell, whereas the macula adherens only contact the other cell at spots around the cell.

Gap junctions

Gap junctions, also known as **communicating junctions**, are formed from integral membrane proteins called **connexins**. A pore in the membrane, called a **connexon**, is formed from six connexins, and this is aligned with a connexon in the opposing cell. This forms a channel between the cells which allows the free passage between the cytoplasms of the cells of water, small molecules (up to 1.2 nm diameter), including some signalling molecules, and ions. It is this electrical coupling of cells via gap junctions that allows the movement of excitation across sheets of muscle. This is particularly important in smooth muscle and cardiac cells. The connexons tend to cluster in patches on the membrane. The pore is closed by high concentrations of calcium, which may be important during apoptosis.

Microvilli and stereocilia

In cells that are required to absorb substances from the extracellular medium, such as those found in the intestine and kidney, the surface area for absorption can be increased up to 30-fold by the presence of **microvilli**. There may be up to 3000 microvilli per cell, forming what is known as the **brush border**. Microvilli are small (<1 μm) projections of the plasma membrane, with a core formed of actin microfilaments. They are attached to the terminal web, a network of actin filaments running under the apical surface of the cell. **Stereocilia** are longer and branched, with a similar actin core

to microvilli. They are rare, mainly being found in the epididymis.

CELLULAR TRANSPORT PROCESSES

In order for chemical reactions to occur, the reagents must be placed appropriately along with their enzyme catalysts. The products and reaction waste products may need to be moved elsewhere. All cells have a plasma membrane, which is a barrier to the movement of substances that are not membrane soluble. The movement of compounds across this and other intracellular membranes depends on the chemistry of the molecule and often requires specialised transport mechanisms.

Passive diffusion

Small molecules such as water, oxygen, carbon dioxide, and fat-soluble molecules are able to move through cell membranes readily by **simple** or **passive diffusion**. Net diffusion occurs from high concentration to low concentration, that is, down the concentration gradient. The shorter the distance a molecule has to diffuse, the quicker the journey will be. Electron microscopy shows that cell membranes are only some 9 nm across, so the diffusion distance is extremely short.

This is important for respiration in the lungs, in which oxygen and carbon dioxide must diffuse through five cell membranes. To enter the red blood cell, oxygen in the lung must cross the alveolar cell lining the lungs (two cell membranes), the endothelial cell lining the capillary through which blood flows (two cell membranes) and must then diffuse across the red cell membrane; carbon dioxide takes the reverse journey. The total diffusion distance is about 1 μm and only 3 s is required for this **gas exchange** to be completed.

Fick's first law of diffusion

Passive diffusion occurs through a series of random steps, which result in the molecule being equilibrated between the compartments. Molecules can move randomly in all directions but the net movement occurs down the concentration gradient, and the relationship between the rate of transport and the concentration gradient is linear (Fig. 2.30). Diffusion across a plasma membrane can be described by Fick's first law of diffusion:

$$J = -DA\Delta c/\Delta x$$

where J is the net rate of diffusion; D is the diffusion coefficient; A is the membrane area; Δc is the concentration difference across the membrane and Δx is the thickness of the membrane.

The diffusion coefficient (D) is defined as:

$$D = kT/(6\pi r\eta)$$

where k is Boltzmann's constant; T is the absolute temperature; r is the molecular radius of the diffusing compound and η is the viscosity of the medium.

Diffusion coefficient

From the equation above it can be seen that the diffusion coefficient (D) is dependent on the viscosity of the diffusing medium and the molecular radius of the diffusing compound. The larger the compound and the thicker the medium, the

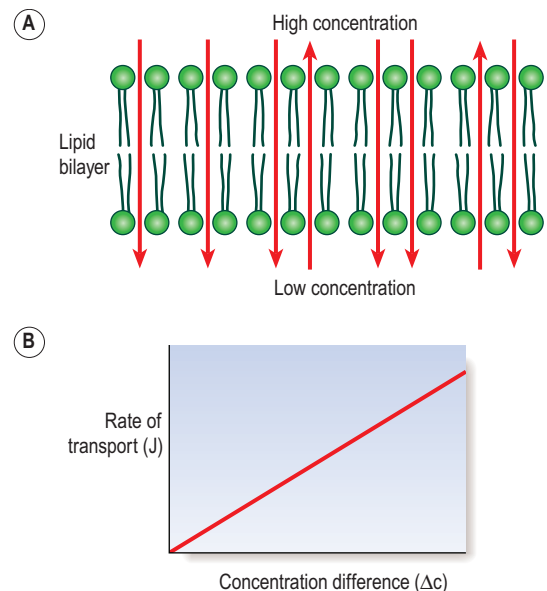


Fig. 2.30 Passive diffusion: (A) transport and (B) relationship between rate and concentration difference.

slower the rate of diffusion. An 'average' small water-soluble molecule, with a diffusion coefficient of 1×10^{-5} cm²/s, will only take 0.5 ms to diffuse 1 μm (the thickness of a capillary wall), but it will take 5 s to move 100 μm (the length of many cells), and 14 hours to diffuse 1 cm.

In prokaryotic cells there are no membrane-bound structures within the cytoplasm, so all compounds can diffuse within the cell to their target. However, over large distances, this type of transport is very slow, which is possibly one of the size-limiting factors for both prokaryotic organisms and within single cells.

Carrier-mediated transport

Many materials that cells need to import or export are either polar or charged, and so will not diffuse through cell membranes. These charged molecules go via special proteins within the membrane, called **transporters**. Any transport of materials that requires transporters is said to be **carrier-mediated transport**. Transporters are specific, in the same way that enzymes are specific; they have preferred substrates and the kinetics of transport show the same characteristics as enzymes. Transporters are both specific and saturable, so the maximum rate of transport and the affinity of the transporter for the substrate can be described using the Michaelis–Menten equation.

There are three types of carrier-mediated transport:

- Facilitated diffusion
- Active transport
- Secondary active transport.

Facilitated diffusion

In **facilitated diffusion** the substrate diffuses from high to low concentration – just as in passive diffusion – except that a transporter is essential for diffusion to occur (Fig. 2.31). These transporters are protein channels – pores, also called **permeases** – that span the entire width of the membrane. It is the size and chemistry of the channel lining that allows specific substrates to pass through it.